ANALYSIS OF VARIANCE AND DESIGN OF EXPERIMENTS

The Pennsylvania State University

LibreTexts™

STAT 502: Analysis of Variance and Design of Experiments

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the hundreds of other texts available within this powerful platform, it is freely available for reading, printing and "consuming." Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects.

Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.

The LibreTexts mission is to unite students, faculty and scholars in a cooperative effort to develop an easy-to-use online platform for the construction, customization, and dissemination of OER content to reduce the burdens of unreasonable textbook costs to our students and society. The LibreTexts project is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning by developing an Open Access Resource environment. The project currently consists of 14 independently operating and interconnected libraries that are constantly being optimized by students, faculty, and outside experts to supplant conventional paper-based books. These free textbook alternatives are organized within a central environment that is both vertically (from advance to basic level) and horizontally (across different fields) integrated.

The LibreTexts libraries are Powered by NICE CXOne and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org. More information on our activities can be found via Facebook (https://facebook.com/Libretexts), Twitter (https://twitter.com/libretexts), or our blog (http://Blog.Libretexts.org).

This text was compiled on 03/18/2025

TABLE OF CONTENTS

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Overview of ANOVA

Objectives

Upon completion of this lesson, you should be able to:

- Become familiar with the standard ANOVA basics.
- Apply the Exploratory Data Analysis (EDA) basics for ANOVA appropriate data.

In previous statistics courses analysis of variance (ANOVA) has been applied in very simple settings, mainly involving one group or factor as the explanatory variable. In this course, ANOVA models are extended to more complex situations involving several explanatory variables. The experimental design aspects are discussed as well. Even though the ANOVA methodology developed in the course is for data obtained from designed experimental settings, the same methods may be used to analyze data from observational studies as well. However, let us keep in mind that the conclusions made may not be as sound because observational studies do not satisfy the rigorous conditions that the designed experiments are subjected to.

Note!

If you aren't familiar with the difference between observational and experimental studies, you should be reviewing introductory statistical concepts which are **essential** for success in this course!

"Classic" analysis of variance (ANOVA) is a method to compare average (mean) responses to experimental manipulations in controlled environments. For example, if people who want to lose weight are randomly selected to participate in a weight-loss study, each person might be randomly assigned to a dieting group, an exercise group, and a "control" group (for which there is no intervention). The **mean** weight loss for each group is compared to every other group.

Recall that a fundamental tenet of the scientific method is that results should be reproducible. A designed experiment provides this through replication and generates data that requires the calculation of mean (average) responses.

This page titled 1: Overview of ANOVA is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

1.1: The Working Hypothesis

Using the scientific method, before any statistical analysis can be conducted, a researcher must generate a guess, or hypothesis about what is going on. The process begins with a **Working Hypothesis**. This is a direct statement of the research idea. For example, a plant biologist may think that plant height may be affected by applying different fertilizers. So they might say: "*Plants with different fertilizers will grow to different heights*".

But according to the Popperian Principle of Falsification, we can't conclusively affirm a hypothesis, but we can conclusively negate a hypothesis. So we need to translate the working hypothesis into a framework wherein we state a null hypothesis that the average height (or mean height) for plants with the different fertilizers will all be the same. The alternative hypothesis (which the biologist hopes to show) is that they are not all equal, but rather some of the fertilizer treatments have produced plants with different mean heights. The strength of the data will determine whether the null hypothesis can be rejected with a specified level of confidence.

Pictured in the graph below, we can imagine testing three kinds of fertilizer and also one group of plants that are untreated (the control). The plant biologist kept all the plants under controlled conditions in the greenhouse, to focus on the effect of the fertilizer, the only thing we know to differ among the plants. At the end of the experiment, the biologist measured the height of each plant. Plant height is the dependent or response variable and is plotted on the vertical (y) axis. The biologist used a simple boxplot to plot the difference in the heights.

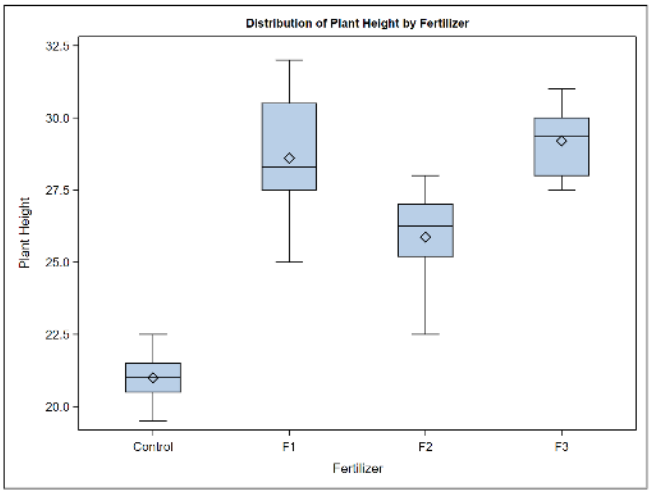


Figure 1.1.1: Boxplot of plant height distribution by fertilizer.

This boxplot is a customary way to show treatment (or factor) level differences. In this case, there was only one treatment: fertilizer. The fertilizer treatment had four levels that included the control, which received no fertilizer. Using this language convention is important because later on we will be using ANOVA to handle multi-factor studies (for example if the biologist manipulated the amount of water AND the type of fertilizer) and we will need to be able to refer to different treatments, each with their own set of levels.

Another alternative for viewing the differences in the heights is with a **means plot** (a scatter or interval plot):

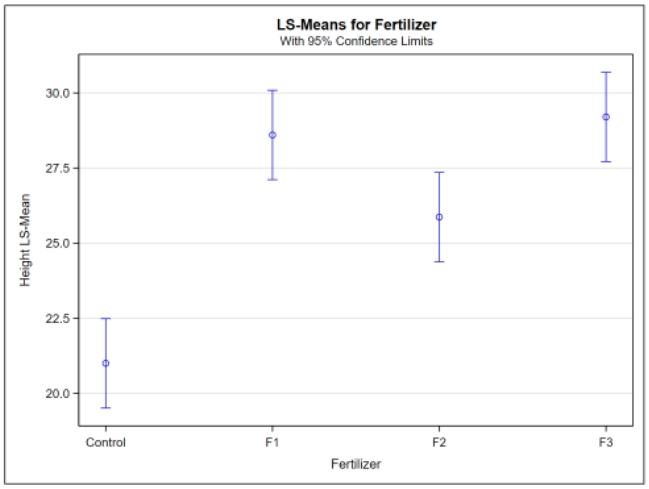


Figure 1.1.2: Means plot for fertilizer with 95% confidence limits.

This second method to plot the difference in the means of the treatments provides essentially the same information. However, this plot illustrates the variability in the data with 'error bars' that are the 95% confidence interval limits around the means.

In between the statement of a Working Hypothesis and the creation of the 95% confidence intervals used to create this means plot is a 7-step process of statistical hypothesis testing, presented in the following section.

This page titled 1.1: The Working Hypothesis is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

1.2: The 7-Step Process of Statistical Hypothesis Testing

We will cover the seven steps one by one.

Step 1: State the Null Hypothesis

The null hypothesis can be thought of as the opposite of the "guess" the researchers made: in this example, the biologist thinks the plant height will be different for the fertilizers. So the null would be that there will be no difference among the groups of plants. Specifically, in more statistical language the null for an ANOVA is that the means are the same. We state the null hypothesis as:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_T \tag{1.2.1}$$

for T levels of an experimental treatment.

> **Note**
>
> Why do we do this? Why not simply test the working hypothesis directly? The answer lies in the Popperian Principle of Falsification. Karl Popper (a philosopher) discovered that we can't conclusively confirm a hypothesis, but we can conclusively negate one. So we set up a null hypothesis which is effectively the opposite of the working hypothesis. The hope is that based on the strength of the data, we will be able to negate or reject the null hypothesis and accept an alternative hypothesis. In other words, we usually see the working hypothesis in H_A.

Step 2: State the Alternative Hypothesis

$$H_A: \text{ treatment level means not all equal} \tag{1.2.2}$$

The reason we state the alternative hypothesis this way is that if the null is rejected, there are many possibilities.

For example, $\mu_1 \neq \mu_2 = \ldots = \mu_T$ is one possibility, as is $\mu_1 = \mu_2 \neq \mu_3 = \ldots = \mu_T$. Many people make the mistake of stating the alternative hypothesis as $mu_1 \neq mu_2 \neq \ldots \neq \mu_T$, which says that every mean differs from every other mean. This is a possibility, but only one of many possibilities. To cover all alternative outcomes, we resort to a verbal statement of "not all equal" and then follow up with mean comparisons to find out where differences among means exist. In our example, this means that fertilizer 1 may result in plants that are really tall, but fertilizers 2, 3, and the plants with no fertilizers don't differ from one another. A simpler way of thinking about this is that at least one mean is different from all others.

Step 3: Set α

If we look at what can happen in a hypothesis test, we can construct the following contingency table:

	In Reality	
Decision	H_0 is TRUE	H_0 is FALSE
Accept H_0	correct	Type II Error β = probability of Type II Error
Reject H_0	Type I Error α = probability of Type I Error	correct

You should be familiar with type I and type II errors from your introductory course. It is important to note that we want to set α before the experiment (*a priori*) because the Type I error is the more grievous error to make. The typical value of α is 0.05, establishing a 95% confidence level. **For this course, we will assume α=0.05, unless stated otherwise.**

Step 4: Collect Data

Remember the importance of recognizing whether data is collected through an experimental design or observational study.

Step 5: Calculate a test statistic

For categorical treatment level means, we use an F statistic, named after R.A. Fisher. We will explore the mechanics of computing the F statistic beginning in Chapter 2. The F value we get from the data is labeled $F_{\text{calculated}}$.

Step 6: Construct Acceptance / Rejection regions

As with all other test statistics, a threshold (critical) value of F is established. This F value can be obtained from statistical tables or software and is referred to as F_{critical} or F_{α}. As a reminder, this critical value is the minimum value for the test statistic (in this case the *F* test) for us to be able to reject the null.

The F distribution, F_{α}, and the location of acceptance and rejection regions are shown in the graph below:

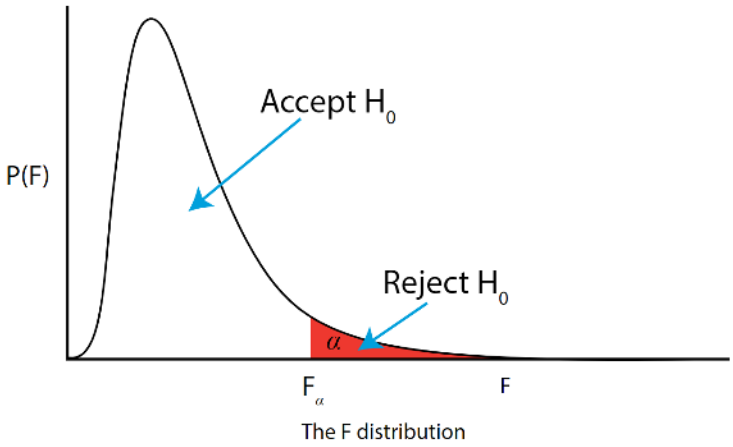


Figure 1.2.1: The F distribution, with F_{α} and acceptance and rejection regions.

Step 7: Based on steps 5 and 6, draw a conclusion about H0

If the $F_{\backslash\text{calculated}}$ from the data is larger than the F_{α}, then you are in the rejection region and you can reject the null hypothesis with $(1-\alpha)$ level of confidence.

Note that modern statistical software condenses steps 6 and 7 by providing a p-value. The p-value here is the probability of getting an $F_{\text{calculated}}$ even greater than what you observe assuming the null hypothesis is true. If by chance, the $F_{\text{calculated}} = F_{\alpha}$, then the p-value would exactly equal α. With larger $F_{\text{calculated}}$ values, we move further into the rejection region and the p-value becomes less than α. So the decision rule is as follows:

If the p-value obtained from the ANOVA is less than α, then reject H_0 and accept H_A.

Note

If you are not familiar with this material, we suggest that you review course materials from your basic statistics course.

This page titled 1.2: The 7-Step Process of Statistical Hypothesis Testing is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

1.3: Chapter 1 Summary

The emphasis of this lesson was to reinforce the basics of ANOVA, which perhaps you may have seen in other courses. Using the greenhouse example, the seven important steps of hypothesis testing in a single factor ANOVA setting were explored. Step 2 highlighted the correct way to state and also interpret the alternative hypothesis (H_A), while Step 3 discusses the Truth Table that includes possible errors in hypothesis testing. Step 6 discusses in detail the rejection region of the null hypothesis (H_0).

The lesson also introduced us to some basics in ANOVA-related explanatory data analysis (EDA). The graphics such as side-by-side boxplots and mean plots are useful tools in producing a visual summary of the raw data and ANOVA results. These will serve as stepping stones to more elaborate graphical techniques we will learn throughout the course.

The concepts and methodology learned in this lesson, though seem straight forward will help us navigate more complex topics addressed in future lessons. The keywords and phrases learned in this lesson are:

- null and alternative hypotheses (H_0 and H_A)
- Type 1 and Type II errors
- significance level (α)
- rejection region
- F statistic and its critical and calculated values.

This page titled 1.3: Chapter 1 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

CHAPTER OVERVIEW

2: ANOVA Foundations

Objectives

Upon completion of this chapter, you should be able to:

- Perform basic computations for Single Factor ANOVA and interpret the results.
- Carry out the Tukey pairwise mean comparison method.
- Learn about other pairwise mean comparison methods.
- Conduct a contrast analysis that accommodates the comparison of group means.

In this chapter, we will begin to learn the notation and the formulas to compute the fundamental quantities necessary for ANOVA-related hypothesis testing as well as mean comparison procedures. The application of these statistical procedures will be illustrated using the Greenhouse example from **Chapter 1**.

This page titled 2: ANOVA Foundations is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Building the ANOVA Table - Notation

The idea of ANOVA is to compare different sources of variability: between sample variability and within sample variability.

As a point of review, the alternative hypothesis is what we think is going on (or what we need to conclude). Typically we are looking to find differences among at least one pair of our treatment means. Because of this, the null hypothesis (the opposite of the alternative) states that there are no differences among the group means (or that they are all equal).

To test the Null hypothesis (which is traditionally written as $H_0: \mu_1 = \mu_2 = \ldots = \mu_T$, we need to compute the **test (F) statistic** that compares the between sample variability to within sample variability.

To see how we compute this statistic it is helpful to look at the ANOVA table. The table below is an ANOVA table (here presented blank, with no entries yet):

Source	df	SS	MS	F

Figure 2.1.1: Blank ANCOVA table.

To define the elements of the table and fill in these quantities, let's return to our example data (Lesson 1 Data) for the hypothetical greenhouse experiment:

Control	F1	F2	F3
21	32	22.5	28
19.5	30.5	26	27.5
22.5	25	28	31
21.5	27.5	27	29.5
20.5	28	26.5	30
21	28.6	25.2	29.2

Notation

Each observation in the dataset can be referenced by two indicator subscripts, i and j, as Y_{ij}.

For those of you not familiar with this notation, we use Y to indicate that it is a response variable. The subscript i refers to the i^{th} level of the treatment; our example has 4 treatments, so i will take on the values $1, 2, 3,$ and 4.) The subscript j refers to the j^{th} observation (again, our example has 6 observations for each treatment so j takes the values $1, 2, 3, 4, 5,$ and 6). It is important to note that the j^{th} observation is occurring **within** the i^{th} treatment level.

subscripts	$i = 1$	$i = 2$	$i = 3$	$i = 4$
	Control	**F1**	**F2**	**F3**
$j = 1$	21	32	22.5	28

$j = 2$	19.5	30.5	26	27.5
$j = 3$	22.5	25	28	31
$j = 4$	21.5	27.5	27	29.5
$j = 5$	20.5	28	26.5	30
$j = 6$	21	28.6	25.2	29.2
For example, $Y_{4,2} = 27.5$.				

We now can define the various means explicitly using these subscripts. The overall or Grand Mean is given by

$$\text{Grand Mean} = \bar{Y}_{..} \tag{2.1.1}$$

where the dots indicate that the quantity has been averaged over that subscript. For the Grand Mean, we have averaged over all j observations in all i treatment levels. The treatment means are given by

$$\text{Treatment Mean} = \bar{Y}_{i} \tag{2.1.2}$$

indicating that we have averaged over the j observations in each of the i treatment levels.

We can find these in the output from the summary procedure that can be generated in SAS and the coding details are discussed in Chapter 3:

Summary Output for Lesson 1 Data

Fert	_Type_	_FREQ_	mean
	0	24	26.1667
Control	1	6	21.0000
F1	1	6	28.6000
F2	1	6	25.8667
F3	1	6	29.2000

In the output we see the column heading _TYPE_. The summary procedure in SAS calculates all possible means when specified, and so the _TYPE_ indicates what mean is being computed. _TYPE_ = 0 is the Grand Mean, and we can see this from the number of observations (given by _FREQ_) of 24. Each of the treatment level means is listed as _TYPE_ = 1 and we confirm that 6 replications were made for each treatment level (remember that *j* took on values 1 through 6).

Note that SAS automatically has ordered the treatment levels alphabetically.

The grand mean and treatment means are all we need in this example to compute the quantities for the ANOVA table.

This page titled 2.1: Building the ANOVA Table - Notation is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Computing Quantities for the ANOVA Table

When working with ANOVA, we start with the total variability in the response variable and divide or "partition" it into different parts: the between sample variability (i.e. variability due to our treatment) and the within sample variability (i.e. residual variability). The variability that is due to our treatment we of course hope is significantly large and variability in the response that is leftover can be thought of as the nuisance, "error", or "residual" variability.

To help you imagine this a bit more, think about the data storage capacity of a computer. If you have 8GB of storage total, you can ask your computer to show the types of files that are occupying the storage. The ANOVA model is (in a very elementary fashion) going to compare the variability due to the treatment to the variability left over.

From elementary statistics, when we think of computing a variance of a random variable (say X), we use the expression:

$$\text{variance} = \frac{\sum \left(X_i - \bar{X}\right)^2}{N-1} = \frac{SS}{df} \tag{2.2.1}$$

The numerator of this expression is referred to as the Sum of Squares, or Sum of Squared deviations from the mean, or simply SS. (If you don't recognize this, then we suggest you sharpen your introductory statistics skills!) The denominator is the degrees of freedom, $(N-1)$, or df.

ANOVA Table Rules

1. Total SS = sum of the SS of all Sources (i.e., Total SS = Treatment SS + Error SS)
2. Total df = sum of df of all Sources
3. MS = SS/df
4. $F_{\text{calculated}} = \dfrac{\text{Treatment MS}}{\text{Error MS}}$ with numerator df = number of treatments - 1 and denominator df = error df

The ANOVA table is set up to generate quantities analogous to the simple variance calculation above. In our greenhouse experiment example:

1. We start by considering the TOTAL variability in the response variable. This is done by calculating the SS_{Total}

$$\begin{aligned} \text{Total SS} &= \textstyle\sum_i \sum_j \left(Y_{ij} - \bar{Y}_{..}\right)^2 \\ &= \mathbf{312.47} \end{aligned} \tag{2.2.2}$$

   The degrees of freedom for the Total SS is $N-1 = 24-1 = 23$, where N is the total sample size.
2. Our next step determines how much of the variability in Y is accounted for by our treatment. We now calculate $SS_{\text{Treatment}}$ or SS_{Trt}:

$$\text{Treatment SS} = \sum_i n_i \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2 \tag{2.2.3}$$

   Note!

   The sum of squares for the treatment is the deviation of the group mean from the grand mean. So in some sense, we are "aggregating" all of the responses from that group and representing the "group effect" as the group mean.

   and for our example:

$$\begin{aligned} &\text{Treatment SS} = 6 \times (21.0 - 26.1667)^2 + 6 \times (28.6 - 26.1667)^2 + \\ &\ldots + 6 \times (25.8667 - 26.1667)^2 + 6 \times (29.2 - 26.1667)^2 = 251.44 \end{aligned} \tag{2.2.4}$$

   Note that in this case we have equal numbers of observations (6) per treatment level, and it is, therefore, a balanced ANOVA.
3. Finally, we need to determine how much variability is "left over". This is the Error or Residual sums of squares by subtraction:

$$\begin{aligned} \text{Error SS} &= \textstyle\sum_i \sum_j \left(Y_{ij} - \bar{Y}_{i.}\right)^2 = \text{Total SS} - \text{Treatment SS} \\ &= 312.47 - 251.44 = \mathbf{61.033} \end{aligned} \tag{2.2.5}$$

Note here that the "leftover" is really the deviation of any score from its group mean.

We can now fill in the following columns of the table:

ANOVA

Source	df	SS	MS	F
Treatment	T - 1 = 3	251.44		
Error	23-3=20	61.033		
Total	N - 1 =23	312.47		

We have T treatment levels and so we use $T-1$ for the df for the treatment. In our example, there are 4 treatment levels (the control and the 3 fertilizers) so $T=4$ and $T-1=4-1=3$. Finally, we obtain the error df by subtraction as we did with the SS.

The Mean Squares (MS) can now be calculated as:

$$MS_{Trt}=\frac{SS_{Trt}}{df_{Trt}}=\frac{251.44}{3}=83.813 \tag{2.2.6}$$

and

$$MS_{Error}=\frac{SS_{Error}}{df_{Error}}=\frac{61.033}{20}=3.052 \tag{2.2.7}$$

NOTE: MS_{Error} will sometimes be referred as MSE and we don't need to calculate the MS_{Total}.

ANOVA

Source	df	SS	MS	F
Treatment	3	251.44	83.813	
Error	20	61.033	3.052	
Total	23	312.47		

Finally, we can compute the F statistic for our ANOVA. Conceptually we are comparing the ratio of the variability due to our treatment (remember we expect this to be relatively large) to the variability leftover, or due to error (and of course, since this is an error we want this to be small). Following this logic, we expect our F to be a large number. If we go back and think about the computer storage space we can picture most of the storage space taken up by our treatment, and less of it taken up by error. In our example, the F is calculated as:

$$F=\frac{MS_{Trt}}{MS_{Error}}=\frac{83.813}{3.052}=27.46 \tag{2.2.8}$$

Source	df	SS	MS	F
Treatment	3	251.44	83.813	**27.46**
Error	20	61.033	3.052	
Total	23	312.47		

So how do we know if the F is large enough to conclude we have a significant amount of variability due to our treatment? We look up the critical value of F and compare it to the value we calculated. Specifically, the critical F is $F_{\alpha}=F_{(0.05,3,20)}=3.10$. The critical value can be found using tables or technology.

Finding a Critical Value of F

Using a Table:

Appendix Table B4

Using SAS:

```
data Fvalue;
    q=finv(0.95, 3, 20);
    put q=;
run;

proc print data=work.Fvalue;
    run;
```

The Print Procedure

Data Set WORK.FVALUE

Obs	q
1	3.09839

Most F tables actually index this value as $1-\alpha = .95$

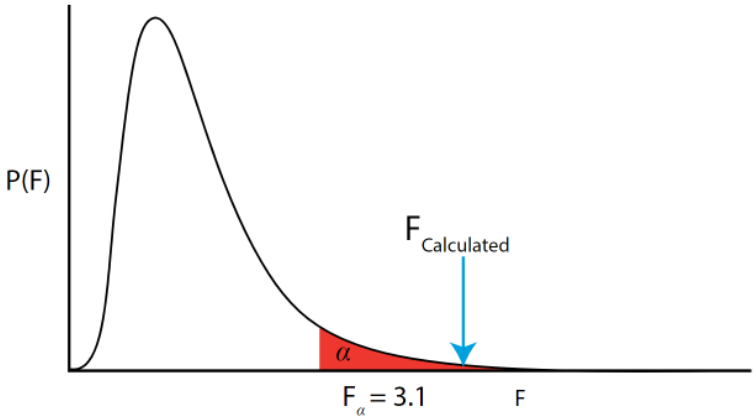


Figure 2.2.1: The F distribution.

The $F_{calculated} > F_{\alpha}$ so we **reject H_0** and accept the alternative H_A. The p-value (which we don't typically calculate by hand) is the area under the curve to the right of the $F_{calculated}$ and is the way the process is reported in statistical software. Note that in the unlikely event that the $F_{calculated}$ is exactly equal to the F_{α} then the p-value $= \alpha$. As the calculated F statistic increases beyond the F_{α} and we go further into the rejection region, the area under the curve (hence the p-value) gets smaller and smaller. This leads us to the decisions rule: If the p-value is $< \alpha$ then we reject H_0.

This page titled 2.2: Computing Quantities for the ANOVA Table is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Tukey Test for Pairwise Mean Comparisons

If (and only if) we reject the null hypothesis, we then conclude at least one group is different from one other (importantly, we do *not* conclude that all the groups are different).

If it is the case that we reject the null, then we will want to know *which* group or groups are different. In our example we are not satisfied knowing at least one treatment level is different, we want to know where the difference is and the nature of the difference. To answer this question, we can follow up the ANOVA with a mean comparison procedure to find out which means differ from each other and which ones don't.

You might think we could not bother with the ANOVA and proceed with a series of t-tests to compare the groups. While that is intuitively simple, it creates inflation of the type I error. How does this inflation of type I error happen? For a single test,

$$\alpha = 1 - (.95) \tag{2.3.1}$$

The probability of committing a type I error (by random chance) for two simultaneous tests follows from the Multiplication Rule for independent events in probability. Recall that, for two independent events A and B the probability of A and B both occurring is $P(\text{A and B}) = P(\text{A}) * P(\text{B})$. So for two tests, we have

$$\alpha = 1 - ((.95) * (.95)) = 0.0975 \tag{2.3.2}$$

which is now larger than the α that we originally set. For our example, we have 6 comparisons, so $\alpha = 1 - (.95^6) = 0.2649$ which is a much larger (inflated) probability of committing a type I error than we originally set.

The multiple comparison procedures compensate for the type I error inflation (although each does so in a slightly different way).

There are several comparison procedures that can be employed, but we will start with the one most commonly used, the Tukey procedure. In the Tukey procedure, we compute a "yardstick" value based on the MS_{Error} and the number of means being compared. If any two means differ by more than the Tukey w value, then they are significantly different.

Step 1: Compute Tukey's w value

$$w = q_{\alpha(p,df_{Error})} \cdot s_{\bar{Y}} \tag{2.3.3}$$

where q_α is obtained from a table of Tukey q values
p = the number of treatment levels
$s_{\bar{Y}}$ = standard error of a treatment mean = $\sqrt{MS_{Error}/r}$
r = number of replications

Show Tukey q Values Table

df for Error Term	α	p = Number of Treatments								
		2	3	4	5	6	7	3	9	10
5	0.05 0.01	3.64 5.70	4.6 6.98	5.22 7.80	5.67 8.42	6.03 8.91	6.33 9.32	6.58 9.67	6.80 9.97	6.99 10.24
6	0.05 0.01	3.46 5.24	4.34 6.33	4.90 7.03	5.30 7.56	5.63 7.97	5.90 8.32	6.12 8.61	6.32 8.87	6.49 9.10
7	0.05 0.01	3.34 4.95	4.16 5.92	4.68 6.54	5.06 7.01	5.36 7.37	5.61 7.68	5.82 7.94	6.00 8.17	6.16 8.37
8	0.05 0.01	3.26 4.75	4.04 5.64	4.53 6.20	4.89 6.62	5.17 6.96	5.40 7.24	5.60 7.47	5.77 7.68	5.92 7.86
9	0.05 0.01	3.20 4.60	3.95 5.43	4.41 5.96	4.76 6.35	5.02 6.66	5.24 6.91	5.43 7.13	5.59 7.33	5.74 7.49

df for Error Term	α	p = Number of Treatments								
		2	3	4	5	6	7	3	9	10
10	0.05 0.01	3.15 4.48	3.88 5.27	4.33 5.77	4.65 6.14	4.91 6.43	5.12 6.67	5.30 6.87	5.46 7.05	5.60 7.21
11	0.05 0.01	3.11 4.39	3.82 5.15	4.26 5.62	4.57 5.97	4.82 6.25	5.03 6.48	5.20 6.67	5.35 6.84	5.49 6.99
12	0.05 0.01	3.08 4.32	3.77 5.05	4.20 5.50	4.51 5.84	4.75 6.10	4.95 6.32	5.12 6.51	5.27 6.67	5.39 6.81
13	0.05 0.01	3.06 4.26	3.73 4.96	4.15 5.40	4.45 5.73	4.69 5.98	4.88 6.19	5.05 6.37	5.19 6.53	5.32 6.67
14	0.05 0.01	3.03 4.21	3.70 4.89	4.11 5.32	4.41 5.63	4.64 5.88	4.83 6.08	4.99 6.26	5.13 6.41	5.25 6.54
15	0.05 0.01	3.01 4.17	3.67 4.84	4.08 5.25	4.37 5.56	4.59 5.80	4.78 5.99	4.94 6.16	5.08 6.31	5.20 6.44
16	0.05 0.01	3.00 4.13	3.65 4.79	4.05 5.19	4.33 5.49	4.56 5.72	4.74 5.92	4.90 6.08	5.03 6.22	5.15 6.35
17	0.05 0.01	2.98 4.10	3.63 4.74	4.02 5.14	4.30 5.43	4.52 5.66	4.70 5.85	4.86 6.01	4.99 6.15	5.11 6.27
18	0.05 0.01	2.97 4.07	3.61 4.70	4.00 5.09	4.28 5.38	4.49 5.60	4.67 5.79	4.82 5.94	4.96 6.08	5.07 6.20
19	0.05 0.01	2.96 4.05	3.59 4.67	3.98 5.05	4.25 5.33	4.47 5.55	4.65 5.73	4.79 5.89	4.92 6.02	5.04 6.14
20	0.05 0.01	2.95 4.02	3.58 4.64	3.96 5.02	4.23 5.29	4.45 5.51	4.62 5.69	4.77 5.84	4.90 5.97	5.01 6.09
24	0.05 0.01	2.92 3.96	3.53 4.55	3.90 4.91	4.17 5.17	4.37 5.37	4.54 5.54	4.68 5.69	4.81 5.81	4.92 5.92
30	0.05 0.01	2.89 3.89	3.49 4.45	3.84 4.80	4.10 5.05	4.30 5.24	4.46 5.40	4.60 5.54	4.72 5.65	4.83 5.76
40	0.05 0.01	2.86 3.82	3.44 4.37	3.79 4.70	4.04 4.93	4.23 5.11	4.39 5.27	4.52 5.39	4.63 5.50	4.74 5.60

For our greenhouse example we get: $w = q_{.05(4,20)}\sqrt{(3.052/6)} = 3.96(0.7132) = 2.824$

Step 2: Rank the means, calculate differences

For the greenhouse example, we rank the means as:

29.20	28.6	25.87	21.00

Start with the largest and second-largest means and calculate the difference, $29.20 - 28.60 = 0.60$, which is *less* than our w of 2.824, so we indicate there is no significant difference between these two means by placing the letter "a" under each:

29.20	28.6	25.87	21.00
a	a		

Then calculate the difference between the largest and third-largest means, $29.20 - 25.87 = 3.33$, which exceeds the critical w of 2.824, so we can label these with a "b" to show this difference is significant:

29.20	28.6	25.87	21.00
a	a	b	

Now we have to consider whether or not the second-largest and third-largest differ significantly. This is a step that sets up a back-and-forth process. Here $28.6 - 25.87 = 2.73$, less than the critical w of 2.824, so these two means do not differ significantly. We need to add a factor of "b" to show this:

29.20	28.6	25.87	21.00
a	ab	b	

Continuing down the line, we now calculate the next difference: $28.60 - 21.00 = 7.60$, exceeding the critical w, so we now add a "c":

29.20	28.6	25.87	21.00
a	ab	b	c

Again, we need to go back and check to see if the third-largest also differs from the smallest: $25.87 - 21.00 = 4.87$, which it does. So we are done.

These letters can be added to figures summarizing the results of the ANOVA.

The Tukey procedure explained above is valid only with equal sample sizes for each treatment level. In the presence of unequal sample sizes, more appropriate is the Tukey-Cramer Method, which calculates the standard deviation for each pairwise comparison separately. This method is available in SAS, R, and most other statistical softwares.

This page titled 2.3: Tukey Test for Pairwise Mean Comparisons is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

2.4: Other Pairwise Mean Comparison Methods

Although the Tukey procedure is the most widely used multiple comparison procedure, there are many other multiple comparison techniques.

An older approach, no longer offered in many statistical computing packages, is Fisher's Protected Least Significant Difference (LSD). This is a method to compare all possible means, two at a time, as t-tests. Unlike an ordinary two-sample t-test, however, the method does rely on the experiment-wide error (the MSE). The LSD is calculated as:

$$LSD(\alpha) = t_{\alpha,df} s_{\bar{d}} \tag{2.4.1}$$

where t_α is based on α and $df =$ error degrees of freedom from the ANOVA table. The standard error for the difference between two treatment means ($s_{\bar{d}}$ or SE) is calculated as:

$$s_{\bar{d}} = \sqrt{\frac{2s^2}{r}} \tag{2.4.2}$$

where r is the number of observations per treatment mean (replications) and s^2 is the MSE from the ANOVA. As in the Tukey method, any pair of means that differ by more than the LSD value differ significantly. The major drawback of this method is that it does not control α over for an entire set of pair-wise comparisons (the experiment-wise error rate) and hence is associated with Type 1 inflation.

The following multiple comparison procedures are much more assertive in dealing with Type 1 inflation. In theory, while we can set α for a single test, the fact that we have T treatment levels means there are $T(T-1)/2$ tests (the number of pairs of possible comparisons), and so we need to adjust α to have the desired confidence level for the set of tests. The Tukey, Bonferroni, and Scheffé methods control the experiment-wise error, but in different ways. All three use a "multiplier" $* SE$ form, but differ in the form of the multiplier.

Contrasts are comparisons involving two or more factor level means (discussed more in the following section). Mean comparisons can be thought of as a subset of possible contrasts among the means. If only pairwise comparisons are made, the Tukey method will produce the narrowest confidence intervals and is the recommended method. The Bonferroni and Scheffé methods are used for general tests of possible contrasts. The Bonferroni method is better when the number of contrasts being tested is about the same as the number of factor levels. The Scheffé method covers all possible contrasts, and as a result, is the most conservative of all the methods. The drawback for such a highly conservative test, however, is that it becomes more difficult to resolve differences among means, even though the ANOVA would indicate that they exist.

When treatment levels include a control and mean comparisons are restricted to only comparing treatment levels against a control level, Dunnett's mean comparison method is appropriate. Because there are fewer comparisons made in this case, the test provides more power compared to a test (see Section 3.7) using the full set of all pairwise comparisons.

To illustrate these methods, the following output was obtained (as we will see later on in the course) for the hypothetical greenhouse data of our example. We will be running these types of analyses later.

Fisher's Least Significant Difference (LSD)

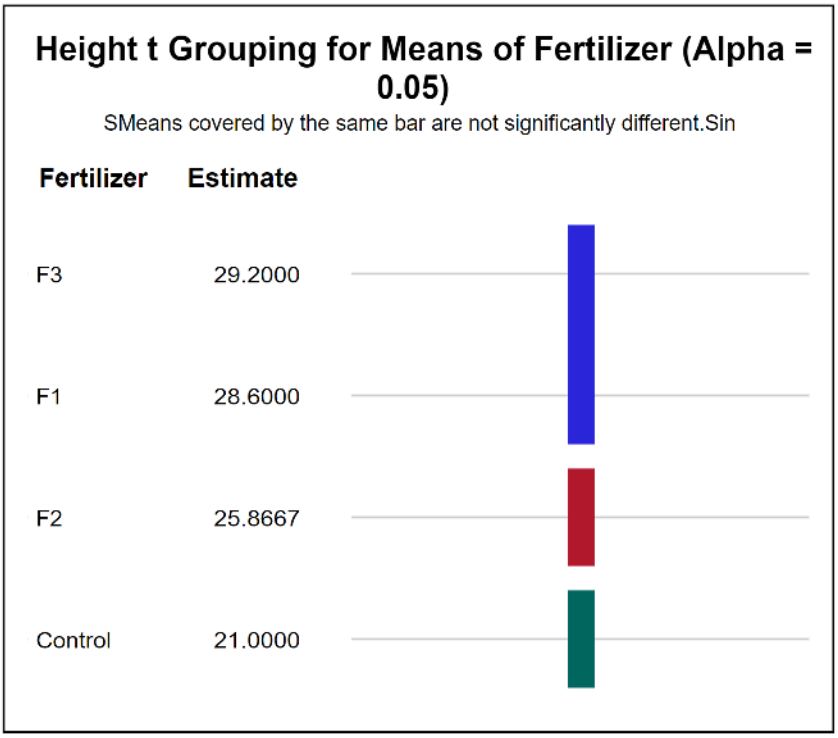


Figure 2.4.1: LSD height groupings for fertilizer treatments.

Since the estimated means for F1 and F3 are covered by the same colored bar, they are not significantly different using the LSD approach.

Tukey

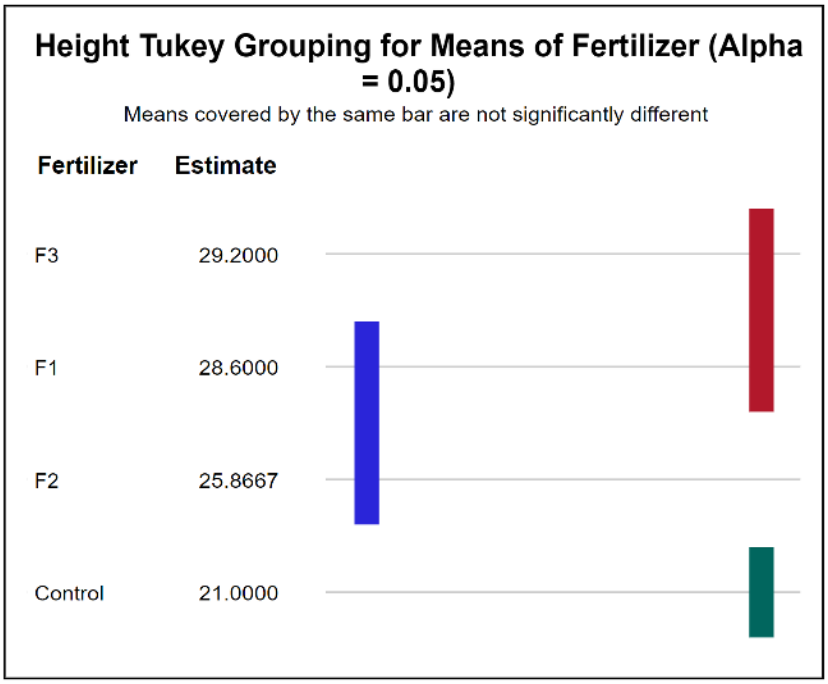


Figure 2.4.2: Tukey height groupings for fertilizer treatments.

Since the estimated means for F1 and F3 are covered by the same colored bar (red bar), they are not significantly different using Tukey's approach. Similarly, since F1 and F2 are covered by the same colored bar (blue bar) they are not significantly different

using Tukey's approach.

Bonferroni

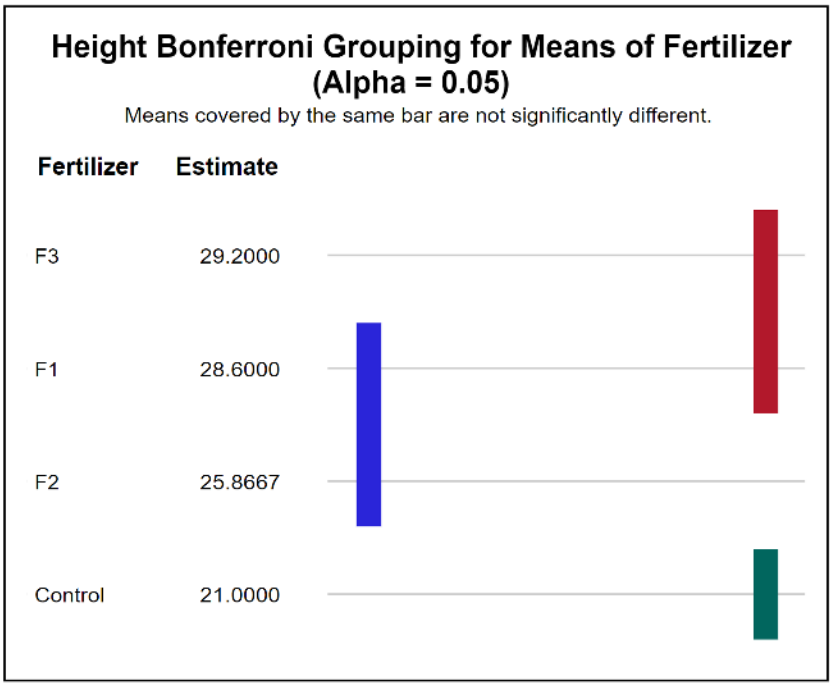


Figure 2.4.3: Bonferroni height groupings for fertilizer treatments.

Observations from the Bonferroni approach are similar to the ones from Tukey's approach.

Scheffé

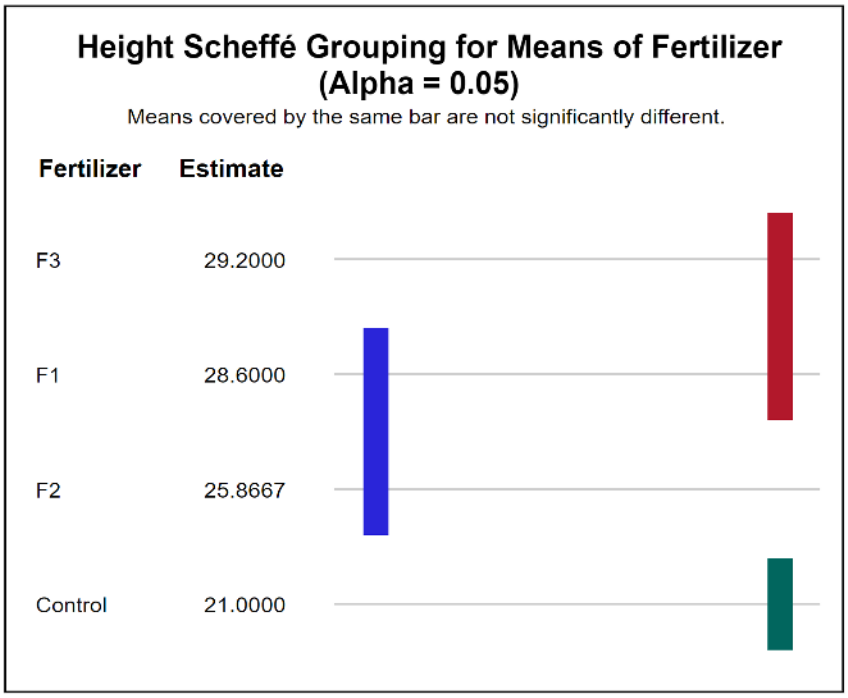


Figure 2.4.4: Scheffé height groupings for fertilizer treatments.

Observations from the Scheffé approach are similar to the ones from Tukey's and Bonferroni's approaches.

Dunnett

				Comparisons significant at the 0.05 level are indicated by ***.
Fertilizer Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		*******
F3 - Control	8.200	5.638	10.762	***
F1 - Control	7.600	5.038	10.162	***
F2 - Control	4.867	2.305	7.429	***

We can see that the LSD method was the most liberal, that is, it indicated the largest number of significant differences between means. In this example, Tukey, Bonferroni, and Scheffé produced the same results. The Dunnett test was consistent with the other 4 methods, and this is not surprising given the small value of the control mean compared to the other treatment levels.

To get a closer look at the results of employing the different methods, we can focus on the differences between the means for each possible pair:

Comparison		**Difference between means**
Control	F1	7.6000
Control	F2	4.8667
Control	F3	8.2000
F1	F2	2.7333
F1	F3	0.6000
F2	F3	3.3333

and compare the 95% confidence intervals produced:

Type		**LSD**		**Tukey**		**Bonferroni**		**Scheffé**		**Dunnett**	
Comparison		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Control	**F1**	5.496	9.704	4.777	10.423	4.648	10.552	4.525	10.675	5.038	10.162
Control	**F2**	2.763	6.971	2.044	7.690	1.914	7.819	1.792	7.942	2.305	7.429
Control	**F3**	6.096	10.304	5.377	11.023	5.248	11.152	5.125	11.275	5.638	10.762
F1	**F2**	0.629	4.837	-0.090	5.556	-0.2189	5.686	-0.342	5.808	X	X
F1	**F3**	-1.504	2.704	-2.223	3.423	-2.352	3.552	-2.475	3.675	X	X
F2	**F3**	1.229	5.437	0.510	6.156	0.3811	6.286	0.258	6.408	X	X

You can see that the LSD produced the narrowest confidence intervals for the differences between means. Dunnett's test had the next most narrow intervals, but only compares treatment levels to the control. The Tukey method produced intervals that were similar to those obtained for the LSD, and the Scheffé method produced the broadest confidence intervals.

What does this mean? When we need to be REALLY sure about our results, we should use conservative tests. If you are working in life-and-death situations, such as in most clinical trials or bridge building, you might want to be surer. If the consequences are less severe you can use a more liberal test, understanding there is more of a chance you might be incorrect (but still able to detect differences). In reality, you need to be consistent with the rigor used in your discipline. While we can't tell you which comparison to use, we can tell you the differences among the tests and the trade-offs for each one.

This page titled 2.4: Other Pairwise Mean Comparison Methods is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

2.5: Contrast Analysis

The paired comparisons discussed in sections 2.2 and 2.3 have the limitation that the comparisons are made only between treatment mean pairs. The contrast analysis procedure can be used to carry out comparisons of a much wider context such as comparisons of treatment level groups or even testing of trends prompting regression modeling to express the response vs. treatment relationship with treatment as a numerical predictor. In the context of a single factor ANOVA model, a linear contrast can be defined as a linear combination of the treatment means such that their numerical coefficients add to zero. Mathematically, a contrast can be represented by...

$$A = \sum_{i=1}^{T} a_i \bar{y}_i \tag{2.5.1}$$

where $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_T$ represent the sample treatment means and $\sum_{i=1}^{T} a_i = 0$. The quantity A is a sample statistic and serves as an estimate for the population contrast $\sum_{i=1}^{T} a_i \mu_i$. By choosing the numerical coefficients appropriately, linear contrasts can be used to make different comparisons among groups of treatment means but not limited to only mean pairs. The table below gives 4 linear contrasts defined in terms of the 3 fertilizer levels F1, F2, F3, and the Control in the greenhouse example.

Table: Greenhouse example contrasts

Ex	a_1	a_2	a_3	a_4	**Contrast**
1	1	-1	0	0	F1-F2
2	1	1	1	-3	F1+F2+F3-3C
3	1	1	-2	0	F1+F2-2F3
4	0	1	-1	0	F2-F3

Notice that values of each list of a_i $(i = 1, 2, 3, 4)$ add to zero. The first contrast compares the first two fertilizer types in terms of their means, and the second compares the means of the 3 fertilizer types with the Control mean. The third is a comparison between the combined effect of fertilizer types 1 and 2 with fertilizer type 3, while the last contrast compares the second and third fertilizer types.

A pair of contrasts $A = \sum_{i=1}^{T} a_i \bar{y}_i$ and $B = \sum_{i=1}^{T} b_i \bar{y}_i$ is orthogonal if the products of their numerical coefficients add to zero. This can be expressed mathematically as

$$\sum_{i=1}^{T} a_i b_i = 0 \tag{2.5.2}$$

A set of contrasts is said to be orthogonal if every pair of contrasts in the set is orthogonal. Two orthogonal contrasts are not correlated which means that if A and B are orthogonal, then $\text{Covariance}(A, B) = 0$. Furthermore, the sum of squares of the treatment usually displayed in the ANOVA table can be partitioned into a set of $(T-1)$ orthogonal contrasts each with 1 degree of freedom. Note that the maximal number of orthogonal contrasts associated with a treatment of T levels is $(T-1)$ and each of them would be associated with one specific comparison independent of each other. In the table above, contrasts 1, 2, and 3 form an orthogonal set of contrasts and contrast 4 cannot be admitted into this set.

The statistical significance of a linear contrast, which can be equated to testing for the zero contrast value can be formulated using the null and alternative hypotheses:

$$H_0: \sum_{i=1}^{T} a_i \mu_i = 0 \text{ vs. } H_A: \sum_{i=1}^{T} a_i \mu_i \neq 0 \tag{2.5.3}$$

and can be tested using either,

$$t = \frac{\sum_{i=1}^{T} a_i \bar{y}_i}{\sqrt{\text{MSE} \sum_{i=1}^{T} \frac{a_i^2}{n_i}}} \text{ with } (N-T) \text{ degrees of freedom} \tag{2.5.4}$$

or

$$F = \frac{\left(\sum_{i=1}^{T} a_i \bar{y}_i\right)^2}{\text{MSE} \sum_{i=1}^{T} \frac{a_i^2}{n_i}} \tag{2.5.5}$$

with numerator and denominator degrees of freedom equal to 1 and $(N - T)$ respectively.

Note that MSE can be obtained from the ANOVA table. Applying the above formula, the t-statistic for testing contrast 2 above is...

$$t = \frac{\sum_{i=1}^{T} a_i \bar{y}_i}{\sqrt{\text{MSE} \sum_{i=1}^{T} \frac{a_i^2}{n_i}}} = \frac{28.6 + 25.867 + 29.2 + (3 * 21)}{3.052 \times \sqrt{\frac{(1+1+1+9)}{6}}} = 8.365 \tag{2.5.6}$$

with $df = 20$ and has a p-value of .0028, indicating that the average plant height due to the combined treatment of the 3 fertilizer types differs significantly from the average plant height yielded by the control.

The above testing procedure is applicable to non-orthogonal contrasts as well. But, as non-orthogonal contrasts are not guaranteed to be uncorrelated, the conclusions arrived at may be "overlapping" and lead to redundancies. In Chapter 3, examples are provided to illustrate how software can be used to conduct contrast testing. The hypothesis testing for trends using contrasts will be discussed in Chapter 10: ANCOVA II.

This page titled 2.5: Contrast Analysis is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

2.6: Try It!

Exercise 2.6.1: Teaching Effectiveness

To compare the teaching effectiveness of 3 teaching methods, the semester average based on 4 midterm exams from five randomly selected students enrolled in each teaching method were used.

1. What is the response in this study?
2. How many replicates are there?
3. Write the appropriate null and alternative hypotheses.
4. Complete the partially filled ANOVA table given below. Round your answers to 4 decimal places.

Source	df	SS	MS	F	p-value
teach_mtd		245			
error					
total		345.1			

5. Find the critical value at $\alpha = .01$
6. Make your conclusion.
7. From the ANOVA analysis, you performed, can you detect the teaching method which yields the highest semester average? If not, suggest a technique that will.

Solution

1. Average of 4 mid-terms
2. 5
3. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$, where μ_1, μ_2, μ_3 are the actual semester average of a student enrolled in teaching method 1, method 2, and method 3 respectively. Ha: Not all semester averages are equal. (This means that there are at least two teaching methods that differ in their actual semester averages)
4.

Source	df	SS	MS	F	p-value
teach_mtd	**2**	245	**122.5000**	**14.6853**	**0.0006**
error	**12**	**100.1**	**8.3417**		
total	**14**	345.1			

5. 6.925
6. As the calculated F-statistic value = 14.6853 is more than the critical value of 6.925, H_0 should be rejected. Therefore, we can conclude that all 3 teaching methods do not have the same semester average, indicating that at least 2 teaching methods differ in their actual semester average.
7. The ANOVA conclusion indicated that not all 3 teaching methods are equally effective, but did not indicate which one yields the highest mean score. The Tukey comparison method is one procedure that shows the teaching method that yields the significantly highest average semester score.

Exercise 2.6.2: Commuter Times

In a local commuter bus service, the number of daily passengers for 50 weeks was recorded. The purpose was to determine if the passenger volume is significantly less during weekends compared to workdays. Below are summary statistics for each day of the week. The partially filled ANOVA table, along with a Tukey plot, is shown below.

Statistics

Day	N	Mean	SE Mean	Std Dev

Day	N	Mean	SE Mean	Std Dev
Sun	50	486.500	9.003	63.661
Mon	50	514.600	6.891	48.724
Tue	50	501.340	7.922	56.018
Wed	50	520.640	7.055	49.886
Thu	50	512.880	10.258	72.532
Fri	50	512.600	8.086	57.174
Sat	50	469.860	8.988	63.555

a) State the appropriate null and alternative hypotheses for this test.

Solution

$H_0: \mu_{Sun} = \mu_{Mon} = \mu_{Tues} = \mu_{Wed} = \mu_{Thurs} = \mu_{Fri} = \mu_{Sat}$

\(H_{A}: \ \text{At least one } \mu_{day \ i} \neq \mu_{day \ j}, \text{ for some\) i, j = 1, 2, \ldots, 7 \text{ OR not all means are equal}\)

b) Complete the partially filled ANOVA table given below. Use two decimal places in the F statistic.

Source	df	SS	MS	F	p-value
Groups		100391			
Error					
Total		1306887			

Solution

Source	df	SS	MS	F	p-value
Day	**6**	100391	**16731.8**	**4.76**	**0.0001**
Error	**343**	**1206496**	**3517.5**		
Total	**349**	1306887			

c) Use the appropriate F-distribution cumulative probabilities to verify that the p-value for the test is approximately zero.

Solution

p-value ≈ 0 (from the F-distribution with 6 and 343 degrees of freedom)

d) Use $\alpha = 0.05$ to test if the mean passenger volume differs significantly by day of the week.

Solution

Since the p-value $\leq \alpha = 0.05$, we reject H_0. There is strong evidence to indicate that the mean passenger volume differs significantly by day of the week (i.e., for some days of the week, the average number of commuters is more than others, but this test does not indicate which days have a higher passenger volume).

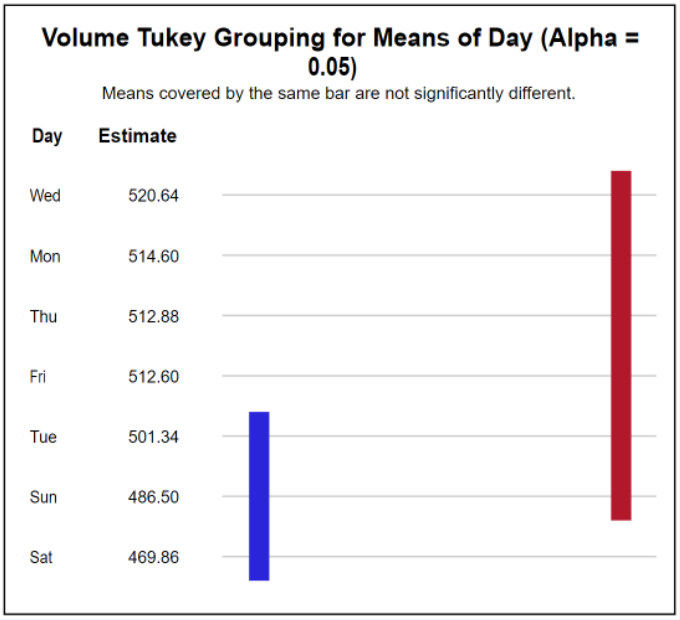


Figure 2.6.1: Grouping information using the tukey method and 95% confidence.

e) Use the output to make a statement about how the mean daily passenger volume differs significantly by day of the week.

Solution

The passenger volume on Sundays is not statistically different from Saturdays and also from Tuesdays. The mean passenger volume on Saturdays is significantly lower than on workdays other than Tuesdays.

f) The management would like to know if the overall number of commuters is significantly more during workdays than during weekends. An appropriate comparison to respond to their query would be to compare the average number of commuters between workdays (Monday through Friday) and the weekend. Write the weight (coefficients) for a linear contrast to make this comparison. Test the hypothesis that the average commuter volume during the weekends is less.

Solution

The weights (coefficients) for the appropriate contrast are given below.

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
weight	1	1	1	1	1	-2.5	-2.5

$$t = \frac{\sum_{i=1}^{T} a_i \bar{y}_i}{\sqrt{MSE \sum_{i=1}^{T} \frac{a_i^2}{n_i}}} = \frac{171.16}{\sqrt{3517.5 * \frac{17.5}{50}}} = 4.878$$

Under the null hypothesis, this test statistic has a t-distribution with 343 degrees of freedom. You can obtain the p-value using statistical software. Recall this is a one-tailed test.

Student's t distribution with 343 DF

x	$P(X \leq x)$
4.878	$8.216815 \cdot 10^{-7} \approx 0$

This p-value indicates that the difference in the average number of passengers is statistically significant between workdays and weekends.

See the table below for computations:

Factor	N	Mean	weights	product	weight^2
Mon	50	514.6	1.0	514.6	1.00
Tue	50	501.34	1.0	501.34	1.00
Wed	50	520.64	1.0	520.64	1.00
Thu	50	512.88	1.0	512.88	1.00
Fri	50	512.6	1.0	512.6	1.00
Sat	50	469.86	-2.5	-1174.65	6.25
Sun	50	486.5	-2.5	-1216.25	6.25

Recall that the MSE (error mean squares) is 3517.5 with $df_{error} = 343$.

This page titled 2.6: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

2.7: Chapter 2 Summary

In this lesson, we became familiar with the ANOVA methodology to test for equality among treatment means. As follow-up procedures, we were also exposed to the Tukey method for paired mean comparisons which helped to identify significantly different treatment (factor) levels. The contrast analysis was also discussed as a means to compare differences among group means.

This page titled 2.7: Chapter 2 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

3: ANOVA Models Part I

In Chapter 2 we learned that ANOVA is based on testing the effect of the treatment relative to the amount of random error. In statistics, we call this the partitioning of variability (due to treatment and due to random variability in the measurements). This partitioning of the deviations can be written mathematically as:

$$\underbrace{Y_{ij} - \bar{Y}_{..}}_{(1)} = \underbrace{\bar{Y}_{i.} - \bar{Y}_{..}}_{(2)} + \underbrace{Y_{ij} - \bar{Y}_{i}}_{(3)} \tag{3.1}$$

Thus, the total deviation $Y_{ij} - \bar{Y}_{..}$ in (1) can be viewed as the sum of two components:

(2) Deviation of estimated factor level mean around overall mean, and

(3) Deviation of the j^{th} response of the i^{th} factor around the estimated factor level mean.

These two deviations are also called variability between groups, a reflection of differences between treatment levels and the variability within groups that serves as a proxy for the error variability among individual observations. A practitioner would however be more interested in the variability between groups as it is the indicator of treatment level differences and may have little interest in the within-group variability, expecting it to be in fact small. However, it will be seen that both these variability measures will play an important role in statistical procedures.

There are several mathematically equivalent forms of ANOVA models describing the relationship between the response and the treatment. In this chapter we will focus on the **effects model,** and in the next chapter three other alternative models will be introduced.

This lesson will also cover the topic of model assumptions needed to employ the ANOVA. Model diagnostics, which deal with verifying the validity of model assumptions, are also discussed, along with power analysis techniques to assess the power associated with a statistical study. How software can be used to analyze data using the statistical techniques discussed will also be presented.

This page titled 3: ANOVA Models Part I is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

3.1: The Model

The effects model for one way ANOVA is a linear additive statistical model which relates the response to the treatment and can be expressed as

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{3.1.1}$$

where μ is the grand mean, τ_i $(i = 1, 2, \ldots, T)$ are the deviations from the grand mean due to the treatment levels and ϵ_{ij} are the error terms. The quantities τ_i $(i = 1, 2, \ldots, T)$ which add to zero, are also referred to as the treatment level effects and the errors show the amount "left over" after considering the grand mean and the effect of being in a particular treatment level.

Here's the analogy in terms of the greenhouse experiment. Think of someone who is not aware that different fertilizers have been used walking into the greenhouse to simply inquire about plant heights in general. The overall sample mean, an estimate of the grand mean, will be a suitable response to this inquiry. On the other hand, the overall mean would not be satisfactory to the experimenter of the study, who obviously suspects that there will be height differences among different fertilizer types. Instead, what is more acceptable to the experimenter are the plant height estimates after including the effect of the treatment τ_i.

Note

The actual plant height can never be known because there is an unknown measurement error associated with any observation. This unknown error is associated with the ith treatment level, and the jth observation is denoted ϵ_{ij} $(i = 1, 2, \ldots, T,\ j = 1, 2, \ldots, n_i)$ is a random component (noise) that reflects the unexplained variability among plants within treatment levels.

Under the null hypothesis where the treatment effect is zero, ***the reduced model*** can be written $Y_{ij} = \mu + \epsilon_{ij}$.

Under the alternative hypothesis, where the treatment effects are not zero, ***the full model*** for at least one treatment level can be written $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$.

If $SSE(R)$ denotes the error sums of squares associated with the reduced model and $SSE(F)$ denotes the error sums of squares associated with the full model, we can utilize the General Linear Test approach to test the null hypothesis by using the test statistic:

$$F = \frac{\left(\dfrac{SSE(R) - SSE(F)}{df_R - df_F} \right)}{\left(\dfrac{SSE(F)}{df_F} \right)} \tag{3.1.2}$$

which under the null hypothesis has an F distribution with the numerator and denominator degrees of freedom equal to $df_R - df_F$ and df_F respectively, where df_R is the degrees of freedom associated with $SSE(R)$ and df_F is the degree of freedom associated with $SSE(F)$. It is easy to see that $df_R = N - 1$ and $df_F = N - T$ where $N = \sum_{i=1}^{N} n_i$. Also,

$$SSE(R) = \sum_i \sum_j \left(Y_{ij} - \bar{Y}_{..} \right)^2 = SS_{Total} \quad \text{See Section 2.2} \tag{3.1.3}$$

Therefore,

$$F = \frac{\left(\dfrac{SS_{Total} - SSE}{T - 1} \right)}{\left(\dfrac{SSE}{df_{Error}} \right)} \tag{3.1.4}$$

$$= \frac{\left(\dfrac{SS_{Treatment}}{df_{Treatment}} \right)}{\left(\dfrac{SSE}{df_{Error}} \right)} \tag{3.1.5}$$

$$= \frac{MS_{Trt}}{MSE} \tag{3.1.6}$$

Note that this is the same test statistic derived in Section 2.2 for testing the treatment significance. If the null hypothesis is true, then the treatment effect is not significant. If we reject the null hypothesis, then we conclude that the treatment effect is significant, which leads to the conclusion that at least one treatment level is better than the others!

This page titled 3.1: The Model is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.2: Assumptions and Diagnostics

Before we draw any conclusions about the significance of the model, we need to make sure we have a "valid" model. Like any other statistical procedure, the ANOVA has assumptions that must be met. Failure to meet these assumptions means any conclusions drawn from the model are not to be trusted.

Assumptions

So what are these assumptions being made to employ the ANOVA model? The ***errors*** are assumed to be independent and identically distributed (*iid*) with a normal distribution having a mean of 0 and unknown equal variance.

As the model residuals serve as estimates of the unknown error, diagnostic tests to check for validity of model assumptions are based on **residual plots**, and thus, the implementation of diagnostic tests is also called **Residual Analysis**.

Diagnostic Tests

Most useful is the residual vs. predicted value plot, which identifies the violations of zero mean and equal variance. Residuals are also plotted against the treatment levels to examine if the residual behavior differs among treatments.

The normality assumption is checked by using a normal probability plot.

Residual plots can help identify potential outliers, and the pattern of residuals vs. fitted values or treatments may suggest a transformation of the response variable.

Lesson 4: SLR Model Assumptions of STAT 501 online notes discuss various diagnostic procedures in more detail.

There are various statistical tests to check the validity of these assumptions, but some may not be that useful. For example, Bartlett's test for homogeneity is too sensitive and indicates that problems exist when they really don't. It turns out that the ANOVA is very robust and is not badly affected by minor violations of these assumptions. In practice, a good deal of common sense and the visual inspection of the residual plots are sufficient to determine if serious problems exist.

We will employ statistical software such as SAS to conduct the residual analysis. Here are common patterns that you may encounter in the residual analysis (i.e. plotting residuals, e, against the predicted values, $\hat{y}$).

Figure $3.2.1a$ shows the prototype plot when the ANOVA model is appropriate for data. The residuals are scattered randomly around mean zero and variability is constant (i.e. within the horizontal bands) for all groups.

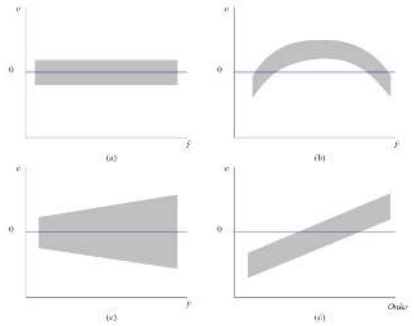


Figure 3.2.1: Common patterns in residual analysis.

Figure $3.2.1b$ suggests that although the variance is constant, there are some trends in the response that is not explained by a linear model. Using Figure $3.2.1c$, we can depict that the linear model is appropriate as the central trend in data is a line. However, the megaphone patterns in Figure $3.2.1c$ suggest that variance is not constant.

Alert!

A common problem encountered in ANOVA is when the variance of treatment levels is not equal (heterogeneity of variance). If the variance is increasing in proportion to the mean (panel (c) in Figure $3.2.1$), a logarithmic transformation of Y can "stabilize" the variances. If the residuals vs. predicted values instead show a curvilinear trend (panel (b) in Figure $3.2.1$), then a quadratic or other transformation may help. Since finding the correct transformation can be challenging, the Box-Cox method is often used to identify the appropriate transformation, given in terms of λ as shown below.

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda}-1}{\lambda}, \text{ if } \lambda \neq 0, \\ \ln y_i, \text{ if } \lambda = 0 \end{cases} \tag{3.2.1}$$

Some λ values result some common transformations.

transformations.

λ	Y^{λ}	Transformation
2	Y^2	Square
1	Y^1	Original (No transform)
1/2	$\sqrt{Y}$	Square Root
0	$\log(Y)$	Logarithm
-1/2	$\frac{1}{\sqrt{Y}}$	Reciprocal Square Root
-1	$\frac{1}{Y}$	Reciprocal

Using Technology

Using Minitab

To run the Box-Cox procedure in Minitab, set up the data (Simulated Data), as a stacked format (a column with treatment (or trt combination) levels, and the second column with the response variable.

Treatment	Response Variable
A	12
A	23
A	34
B	45
B	56
B	67
C	14
C	25
C	36

Steps in Minitab

1. On the Minitab toolbar, choose **`Stat`** > **`Control Charts`** > **`Box-Cox Transformation`**

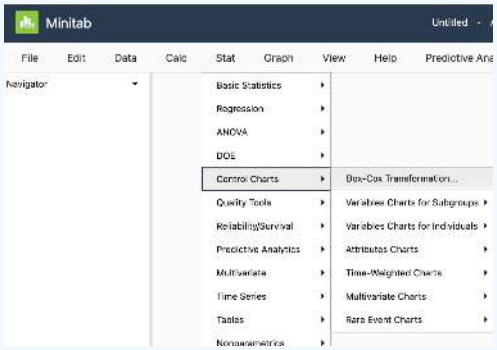


Figure 3.2.2: Selecting Box-Cox Transformation stat option.

2. Place "Response Variable" and "Treatment" in the boxes as shown below.

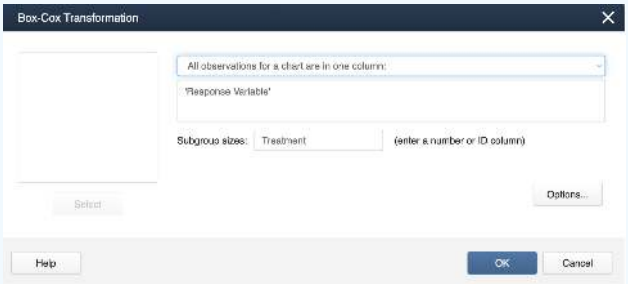


Figure 3.2.3: Inputting "Response Variable" and "Treatment" in pop-up window.

3. Click **`OK`** to finish. You will get an output like this:

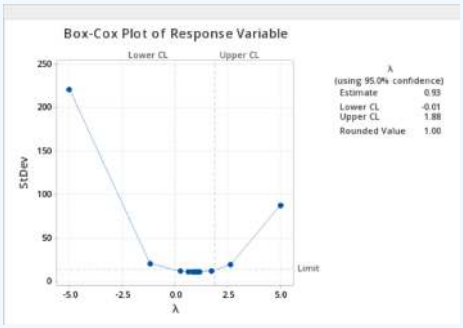


Figure 3.2.4: Minitab Box-Cox plot output.

In the upper right-hand box, the rounded value for λ is given from which the appropriate transformation of the response variable can be found using the chart above. Note, with a λ of 1, no transformation is recommended.

Using SAS

The Box-Cox procedure in SAS is more complicated in a general setting. It is done through the Transreg procedure, by obtaining the ANOVA solution with regression which first requires coding the treatment levels with effect coding discussed in Chapter 4.

However, for one-way ANOVA (ANOVA with only one factor) we can use the SAS *Transreg* procedure without much hassle.

Steps in SAS

Suppose we have SAS data as follows.

Obs	Treatment	ResponseVariable
1	A	12
2	A	23
3	A	34
4	B	45
5	B	56
6	B	67
7	C	14
8	C	25
9	C	36

We can use the following SAS commands to run the Box-Cox analysis.

```
proc transreg data=boxcoxSimData;
model boxcox(ResponseVariable)=class(Treatment);
run;
```

This would generate an output as follows, which suggests a transformation using $\lambda = 1$ (i.e. no transformation).

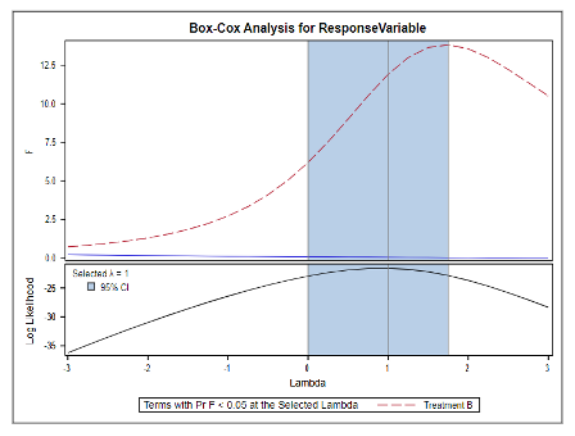


Figure 3.2.5: SAS Box-cox plot output.

Using R

Steps in R

Load the simulated data and perform the Box-Cox transformation. Note that simulated data are in the stacked format (a column with treatment levels and a column with the response variable)

```
setwd("~/path-to-folder/)
simulated_data<-read.table("simulated_data.txt",header=T)
attach(simulated_data)
library(AID)#Load package AID so that we can use the Box-Cox Procedure
boxcoxfr(Response_Variable,Treatment)#Box-Cox command for One-Way ANOVA
```

Output

Box-Cox power transformation	
data:	Response_Variable and Treatment
lambda.hat:	0.93

Shapiro-Wilk normality test for transformed data (alpha = 0.05)				
	Level	statistic	p.value	Normality
1	A	0.9998983	0.9807382	YES
2	B	0.9999840	0.9923681	YES
3	C	0.9999151	0.9824033	YES

Bartlett's homogeneity test for transformed data (alpha = 0.05)				
	Level	statistic	p.value	Homogeneity
1	All	0.008271728	0.9958727	YES

From the output, we can see that the lambda value for the transformation is 0.93 (the same value as Minitab suggested). Since this value is very close to 1 we can use $\lambda = 1$ (no transformation).

In addition, from the output, we can see that normality exists in all 3 levels (Shapiro-Wilk test) and we have the same variance (Bartlett's test).

Alternative:

We can use the command `boxcox` from package MASS

```
library(MASS)
Box_Cox_Plot<-boxcox(aov(Response_Variable~Treatment),lambda=seq(-3,3,0.01))
```

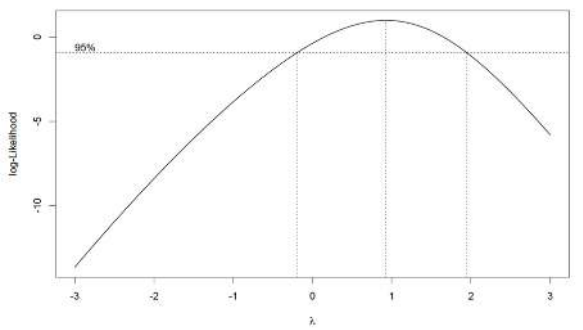


Figure 3.2.6: R-generated plot of log-likelihood vs λ.

From the plot, we can see the 95% CL. Since $\lambda = 1$ is within the interval there is no need for transformation.

```
#Exact lambda
lambda<-Box_Cox_Plot$x[which.max(Box_Cox_Plot$y)] #0.93
detach(simulated_data)
```

This page titled 3.2: Assumptions and Diagnostics is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.3: Anatomy of SAS Programming for ANOVA

The statistical software SAS is widely used in this course, and in previous sections we came across outputs generated through SAS programs. In this section, we begin to delve further into SAS programming with a special focus on ANOVA-related statistical procedures. The STAT 480-course series is also a useful resource for additional help.

Here is the program used to generate the summary output in Section 2.1:

```
data greenhouse;
input Fert $ Height;
```

The first line begins with the word `data` and invokes the data step. Notice that the end of each SAS statement has a semicolon. This is essential. In the dataset, the data to be used and its variables are named. Note that SAS assumes variables are numeric in the input statement, so if we are going to use a variable with alpha-numeric values (e.g. F1 or Control), then we have to follow the name of the variable in the input statement with a `$` sign.

A simple way to input small datasets is shown in this code, wherein we embed the data in the program. This is done with the word `datalines` .

```
datalines;
Control     21
Control     19.5
Control     22.5
Control     21.5
Control     20.5
Control     21
F1     32
F1     30.5
F1     25
F1     27.5
F1     28
F1     28.6
F2     22.5
F2     26
F2     28
F2     27
F2     26.5
F2     25.2
F3     28
F3     27.5
F3     31
F3     29.5
F3     30
F3     29.2
;
```

The semicolon here ends the dataset.

SAS then produces an output of interest using `proc` statements, short for “procedure”. You only need to use the first four letters, so SAS code is full of `proc` statements to do various tasks. Here we just wanted to print the data to be sure it read it in OK.

```
proc print data= greenhouse;
title 'Raw Data for Greenhouse Data'; run;
```

Notice that the data set to be printed is specified in the proc print command. This is an important habit to develop because if not specified, SAS will use the last created data set, out of both input data sets, and output datasets that may have been generated as a result of any SAS procedures run up to that point.

The summary procedure which was then run can be very useful in both EDA (exploratory data analysis) and obtaining descriptive statistics such as mean, variance, minimum, maximum, etc. SAS procedures including the summary procedure categorical variables are specified in the class statement. Any variable NOT listed in the class statement is treated as a continuous variable. The target variable for which the summary will be made is specified by the `var` (for variable) statement.

The `output` statement creates an output dataset and the `out=` part assigns a name of your choice to the output. Descriptive statistics also can be named. For example, in the `output` statement below, `mean=mean` and `stderr=se` have named the mean of the variable `fert` as `mean` and standard error as `se` . The output data sets of any SAS procedure will not be automatically printed. As illustrated in the code below, the print procedure would then have to be used to print the generated output. In the proc print command a title can be included as a means of identifying and describing the output contents.

```
proc summary data= greenhouse;
class fert;
var height;
output out=output1 mean=mean stderr=se;
run;
proc print data=output1;
title 'Summary Output for Greenhouse Data';
run;
```

The two commands `title ; run;` right after will erase the title assignment. This prevents the same title to be used in every output generated thereafter, which is a default feature in SAS.

```
title; run;
```

Summary Output for Greenhouse Data

Obs	Fert	TYPE	FREQ	mean	se
1		0	24	26.1667	0.75238
2	Control	1	6	21.0000	0.40825
3	F1	1	6	28.6000	0.99499
4	F2	1	6	25.8667	0.77531
5	F3	1	6	29.2000	0.52599

This page titled 3.3: Anatomy of SAS Programming for ANOVA is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.4: Greenhouse Example in SAS

In this section we will modify our previous program for greenhouse data to run the ANOVA model. The two SAS procedures that are commonly used are: `proc glm` and `proc mixed` .

```
data greenhouse;
input fert $ Height;
datalines;
Control     21
Control     19.5
Control     22.5
Control     21.5
Control     20.5
Control     21
F1     32
F1     30.5
F1     25
F1     27.5
F1     28
F1     28.6
F2     22.5
F2     26
F2     28
F2     27
F2     26.5
F2     25.2
F3     28
F3     27.5
F3     31
F3     29.5
F3     30
F3     29.2
;

/*
Any lines enclosed between starting with "/*" & ending with "*/" will be ignored by S
*/

/* Recall how to print the data and obtain summary statistics. See section 3.3*/

/*To run the ANOVA model, use proc mixed procedure*/

proc mixed data=greenhouse method=type3 plots=all;
class fert;
model height=fert;
store myresults; /*myresults is an user defined object that stores results*/
title 'ANOVA of Greenhouse Data';
```

```
run;

/*To conduct the pairwise comparisons using Tukey adjustment*/
/*lsmeans statement below outputs the estimates means,
performs the Tukey paired comparisons, plots the data. */
/*Use proc plm procedure for post estimation analysis*/
proc plm restore=myresults;
lsmeans fert / adjust=tukey plot=meanplot cl lines;
run;

/* Testing for contrasts of interest with Bonferroni adjustment*/
proc plm restore=myresults;
lsmeans fert / adjust=tukey plot=meanplot cl lines;
estimate 'Compare control + F3 with F1 and F2 ' fert 1 -1 -1 1,
                'Compare control + F2 with F1' fert 1 -2 1 0/ adjust=bon;
run;
```

This page titled 3.4: Greenhouse Example in SAS is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.5: SAS Output for ANOVA

The first output of the ANOVA procedure as shown below, gives useful details about the model.

ANOVA of Greenhouse Data: The Mixed Procedure

Model Information	
Data Set	WORK.GREENHOUSE
Dependent Variable	Height
Covariance Structure	Diagonal
Estimation Method	Type 3
Residual Variance Method	Factor
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
fert	4	Control F1 F2 F3

Dimensions	
Covariance Parameters	1
Columns in X	5
Columns in Z	0
Subjects	0
Max Obs Per Subject	24

The output below titled '**Type 3 Analysis of Variance**' is similar to the ANOVA table we are already familiar with. Note that it does not include the Total SS, however it can be computed as the sum of all SS values in the table.

Type 3 Analysis of Variance								
Sources	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
fert	3	251.440000	83.813333	Var(Residual)+Q(fert)	MS(Residual)	20	27.46	<.0001
Residual	20	61.033333	3.051667	Var(Residual)				

Covariance Parameter Estimates	
Cov Parm	Estimate
Residual	3.0517

Fit Statistics	
-2 Res Log Likelihood	86.2
AIC (smaller is better)	88.2
AICC (smaller is better)	88.5
BIC (smaller is better)	89.2

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
fert	3	20	27.46	<.0001

The output above titled "**Type 3 Tests of Fixed Effects**" will display the $F_{calculated}$ and p-value for the test of any variables that are specified in the model statement. Additional information can also be requested. For example, the `method = type 3` option will include the Expected Mean Squares for each source, which will prove to be useful and will be seen in Chapter 6.

The Mixed Procedure also produces the following diagnostic plots:

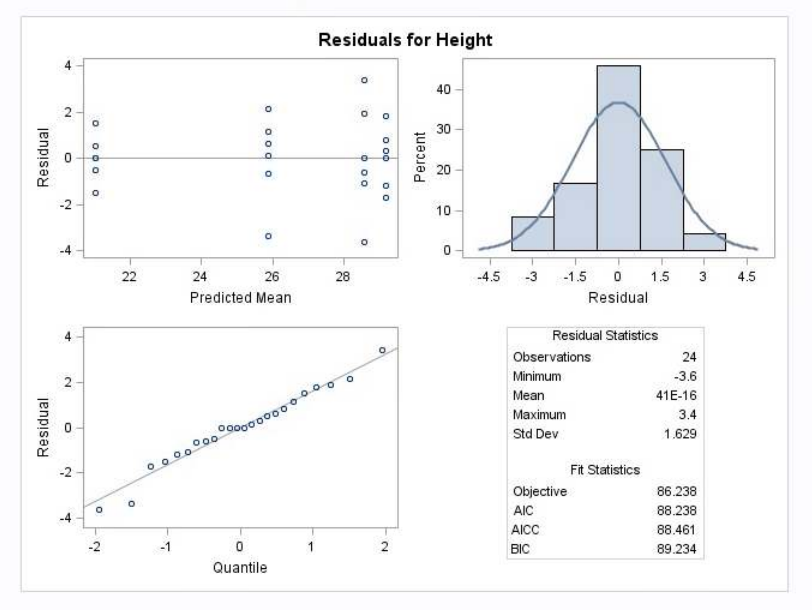


Figure 3.5.1: Diagnostic plots for residuals for height.

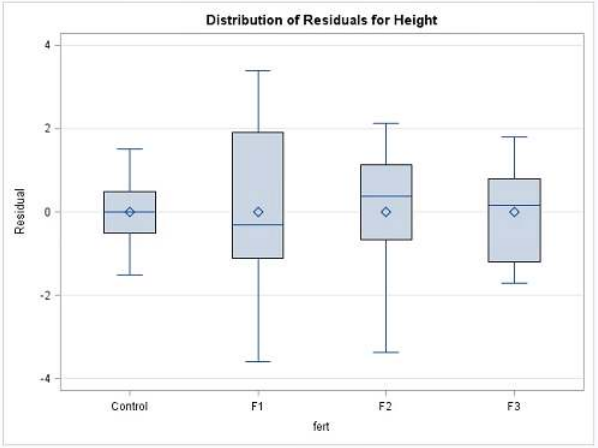


Figure 3.5.2: Box plots for distribution of residuals for height.

The following display is a result of the LSmeans statement in the PLM procedure which was included in the programming code.

Differences of fert Least Squares Means								
fert	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Control	21.0000	0.7132	20	29.45	<.0001	0.05	19.5124	22.4876
F1	28.6000	0.7132	20	40.10	<.0001	0.05	27.1124	30.0876
F2	25.8667	0.7132	20	36.27	<.0001	0.05	24.3790	27.3543
F3	29.2000	0.7132	20	40.94	<.001	0.05	27.7124	30.6876

In the "Least Squares Means" table above, note that the t-value and $Pr > |t|$ are testing null hypotheses that each group mean= 0. (These tests usually do not provide any useful information). The Lower and Upper values are the 95% confidence limits for the group means. Note also that the least square means are the same as the original arithmetic means that were generated in the Summary procedure in Section 3.3 because all 4 groups have the same sample sizes. With unequal sample sizes or if there is a covariate present, the least square means can differ from the original sample means.

Next, the Plot= mean plot option in the LSmeans statement yields a mean plot and also a diffogram, shown below. The confidence intervals in the mean plot are commonly used to identify the significantly different treatment levels or groups. If two confidence intervals do not overlap, then the difference between the two associated means is statistically significant, which is a valid conclusion. However, if they overlap, it may be the case that the difference might still be significant. Consequently, conclusions made based on the visual inspection of the mean plot may not match with those arrived at using the table of "Difference of Least Square Means", another output of the Tukey procedure, and is displayed below.

Notice that this is different from the previous table because it displays the results of each pairwise comparison. For example, the first row shows the comparison between the control and F1. The interpretation of these results is similar to any other confidence interval for the difference in two means—if the confidence interval does not contain zero, then the difference between the two associated means is statistically significant.

Differences of fert Least Squares Means Adjustment for Multiple Comparisons: Tukey												
fert	_fert	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
Control	F1	-7.6000	1.0086	20	-7.54	<.0001	<.0001	0.05	-9.7038	-5.4962	-10.4229	-4.7771
Control	F2	-4.8667	1.0086	20	-4.83	0.0001	0.0006	0.05	-6.9705	-2.7628	-7.6896	-2.0438
Control	F3	-8.2000	1.0086	20	-8.13	<.0001	<.0001	0.05	-10.3038	-6.0962	-11.0229	-5.3771
F1	F2	2.7333	1.0086	20	2.71	0.0135	0.0599	0.05	0.6295	4.8372	-0.08957	5.5562
F1	F3	-0.6000	1.0086	20	-0.59	0.5586	0.9324	0.05	-2.7038	1.5038	-3.4229	2.2229
F2	F3	-3.3333	1.0086	20	-3.30	0.0035	.0171	0.05	-5.4372	-1.2295	-6.1562	-0.5104

This discrepancy between the mean plot and the "Difference of Least Square Means" results occurs because the testing is done in terms of the difference of two means, using the standard error of the difference of the two-sample means, but the confidence intervals of the mean plot are computed for the individual means which are in terms of the standard error of individual sample means. Consistent results can be achieved by using the diffogram as discussed below or the confidence intervals displayed in the "difference in mean plot" available in SAS 14, but not included here.

The diffogram has two useful features. It allows one to identify the significant mean pairs and also gives estimates of the individual means. The diagonal line shown in the diffogram is used as a reference line. Each group (or factor level) is marked on the horizontal and vertical axes and has vertical and horizontal reference lines with their intersection point falling on the diagonal reference line. The x or the y coordinates of this intersection point which are equal is the sample mean of that group. For example, the sample mean for the Control group is about 21, which matches with the estimate provided in the "Least Squares Means" table displayed above. Furthermore, each slanted line represents a mean pair. Start with any group label from the horizontal axis and run your cursor up, along the associated vertical line until it meets a slanted line, and then go across the intersecting horizontal line to identify the other group (or factor level). For example, the lowermost solid line (colored blue) represents the Control and F2. As stated at the bottom of the chart, the solid (or blue) lines indicate significant pairs, and the broken (or red) lines correspond to the non-significant pairs. Furthermore, a line corresponding to a nonsignificant pair will cross the diagonal reference line.

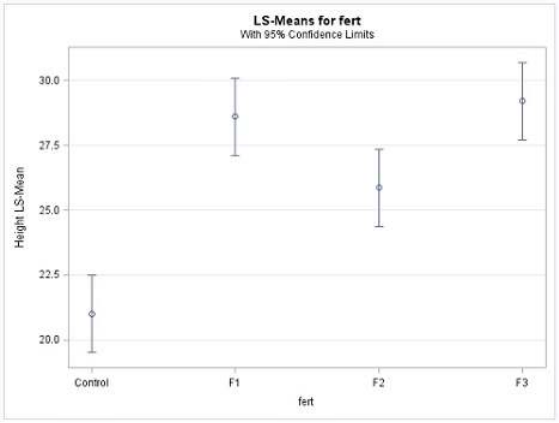


Figure 3.5.3: LS-Means plot.

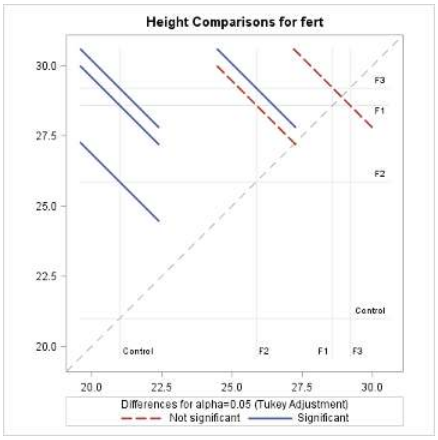


Figure 3.5.4: Diffogram.

The non-overlapping confidence intervals in the mean plot above indicate that the average plant height due to control is significantly different from those of the other 3 fertilizer levels and that the F2 fertilizer type yields a statistically different average plant height from F3. The diffogram also delivers the same conclusions and so, in this example, conclusions are not contradictory. In general, the diffogram always provides the same conclusions as derived from the confidence intervals of difference of least-square means shown in the "Difference of Least Square Means" table, but the conclusions based on the mean plot may differ.

There are two contrasts of interest: contrast to compare the control and F3 with F1 (i.e. $\mu_{control} - \mu_{F1} - \mu_{F2} + \mu_{F3}$) and the contrast to compare control and F2 with F1 (i.e., $\mu_{control} - 2\mu_{F1} + \mu_{F2}$). Since we are testing for two contrasts, we should adjust for multiple comparisons. We use Bonferroni adjustment. In SAS, we can use the `estimate` command under `proc plm` to make these computations.

In general, the `estimate` command estimates linear combinations of model parameters and performs t-tests on them. Contrasts are linear combinations that satisfy a special condition. We will discuss the model parameters in Chapter 4.

Estimates Adjustment for Multiplicity: Bonferroni						
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P
Compare control + F3 with F1 and F2	-4.2667	1.4263	20	-2.99	0.0072	0.0144
Compare control + F2 with F1	-10.3333	1.7469	20	-5.92	<.0001	<.0001

SAS returns both unadjusted and adjusted p-values. Suppose we wanted to make the comparisons at 1% level. If we ignored the multiple comparisons (i.e. using unadjusted p-values), the both comparisons are statistically significant. However, if we consider the adjusted p-values, we will fail to reject the hypothesis corresponding to the first contrast at the 1% level.

This page titled 3.5: SAS Output for ANOVA is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.6: One-Way ANOVA Greenhouse Example in Minitab

Step 1: Import the data

The data (Lesson 1 Data) can be copied and pasted from a word processor into a worksheet in Minitab:

	C1	C2	C3	C4
	Control	F1	F2	F3
1	21.0	32.0	22.5	28.0
2	19.5	30.5	26.0	27.5
3	22.5	25.0	28.0	31.0
4	21.5	27.5	27.0	29.5
5	20.5	28.0	26.5	30.0

Figure 3.6.1: Worksheet in Minitab of Lesson 1 data.

Step 2: Run the ANOVA

To run the ANOVA, we use the sequence of tool-bar tabs: **`Stat`** > **`ANOVA`** > **`One-way…`**

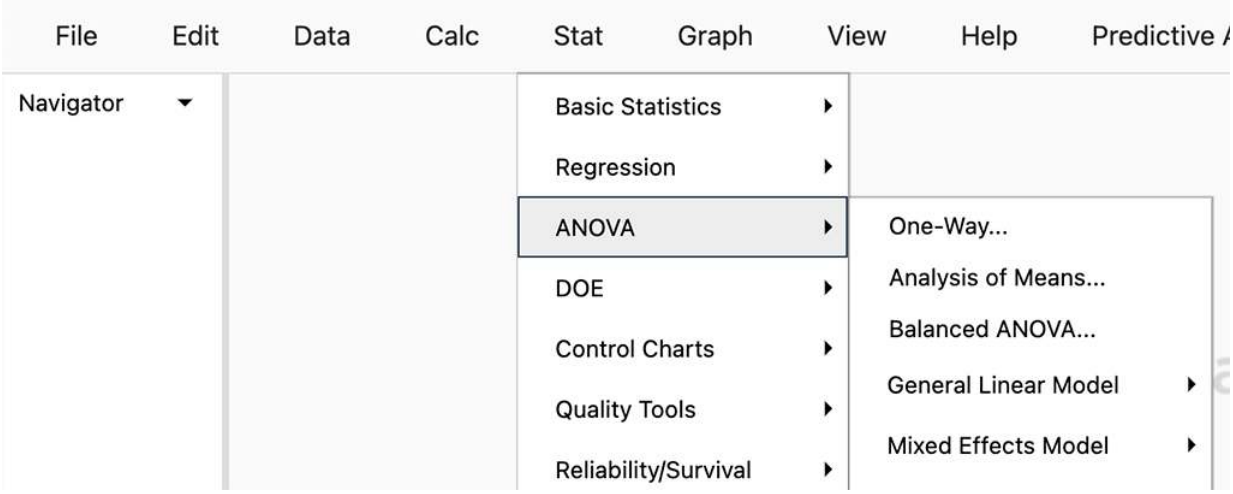


Figure 3.6.2: Selecting toolbar tabs in Minitab.

You then get the pop-up box seen below. Be sure to select from the drop-down in the upper right, "Response data are in a separate column for each factor level":

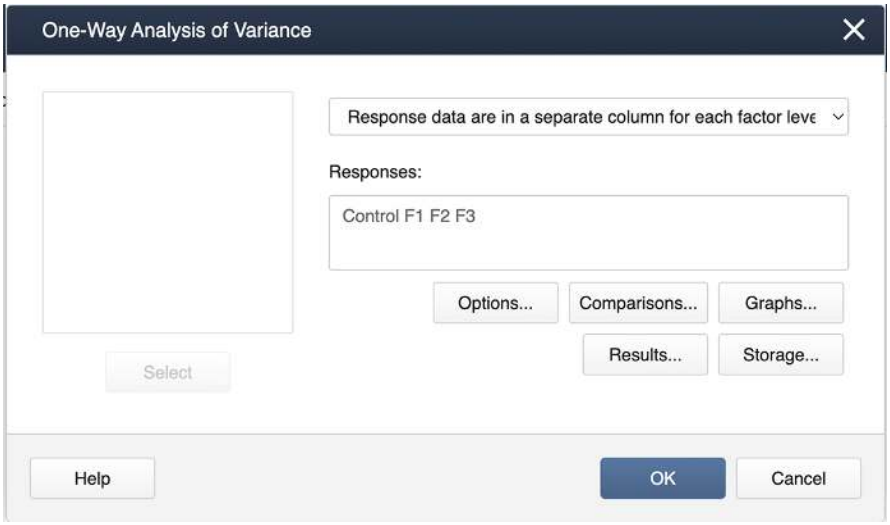


Figure 3.6.3: ANOVA pop-up window in Minitab.

Then we double-click from the left-hand list of factor levels to the input box labeled "Responses", and then click on the box labeled **Comparisons**.

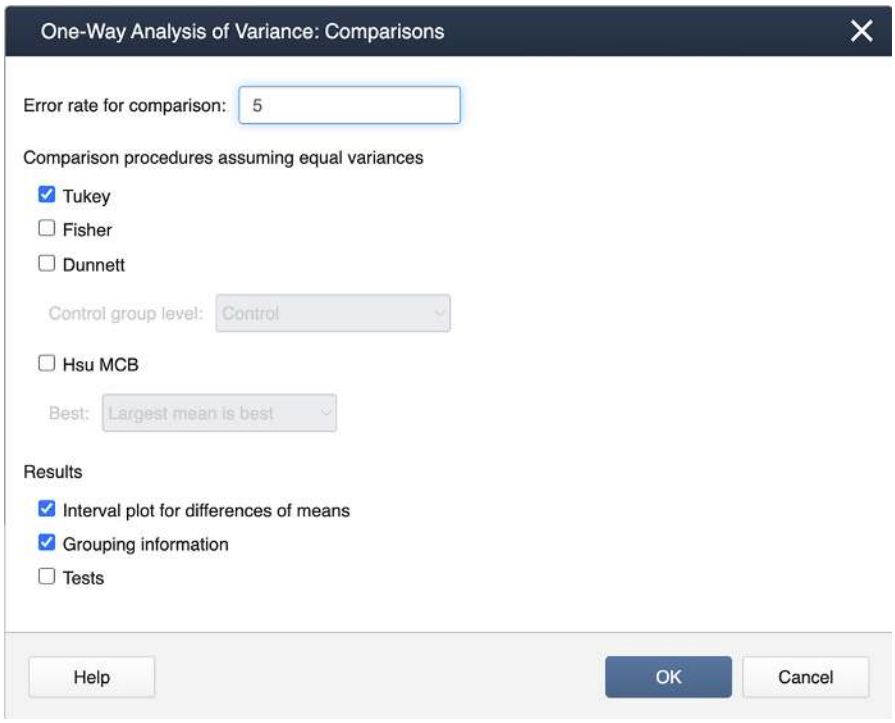


Figure 3.6.4: ANOVA: Comparisons pop-up window in Minitab.

We check the box for Tukey and then exit by clicking on **OK**. To generate the Diagnostics, we then click on the box for **Graphs** and select the "Three in one" option:

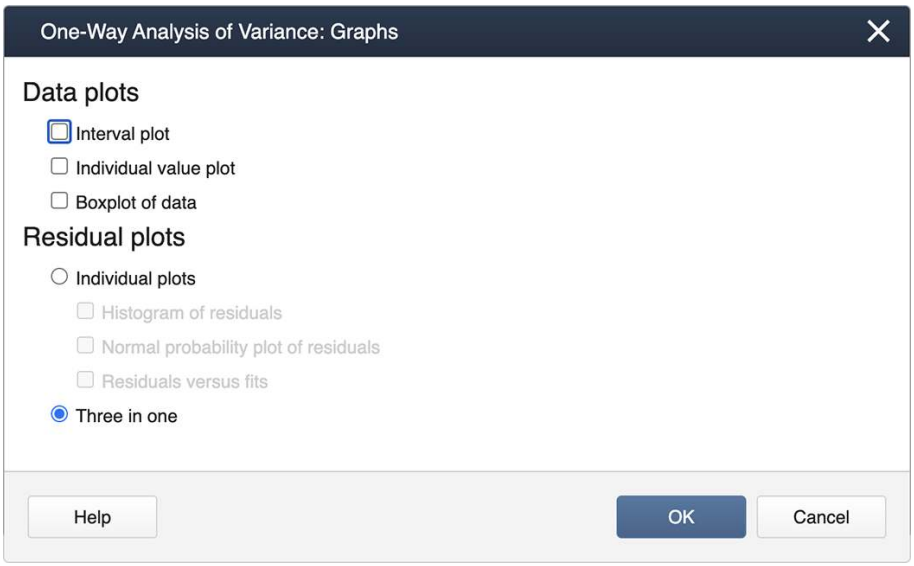


Figure 3.6.5: ANOVA: Graphs pop-up window in Minitab.

You can now "back out" by clicking on **OK** in each nested panel.

Step 3: Results

Now in the Session Window, we see the ANOVA table along with the results of the Tukey Mean Comparison:

One-Way ANOVA: Control, F1, F2, F3

Method

Null Hypothesis: All means are equal

Alternative Hypothesis: Not all means are equal

Significance Level: $\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	4	Control, F1, F2, F3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	251.44	83.813	27.46	0.000
Error	20	61.03	3.052		
Total	23	312.47			

(Extracted from the output that follows from above.)

Grouping Information Using Tukey Method

	N	Mean	Grouping		
F3	6	29.200	A		
F1	6	28.600	A	B	
F2	6	25.867		B	
Control	6	21.000			C

Means that do not share a letter are significantly different.

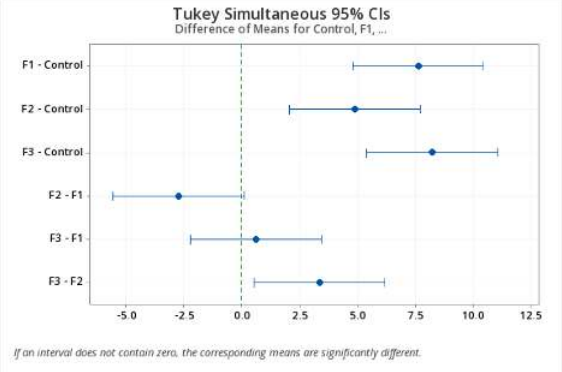


Figure 3.6.6: Minitab difference in means plot.

As can be seen, Minitab provides a difference in means plot, which can be conveniently used to identify the significantly different means by following the rule: if the confidence interval does not cross the vertical zero line, then the difference between the two associated means is statistically significant.

The diagnostic (residual) plots, as we asked for them, are in one figure:

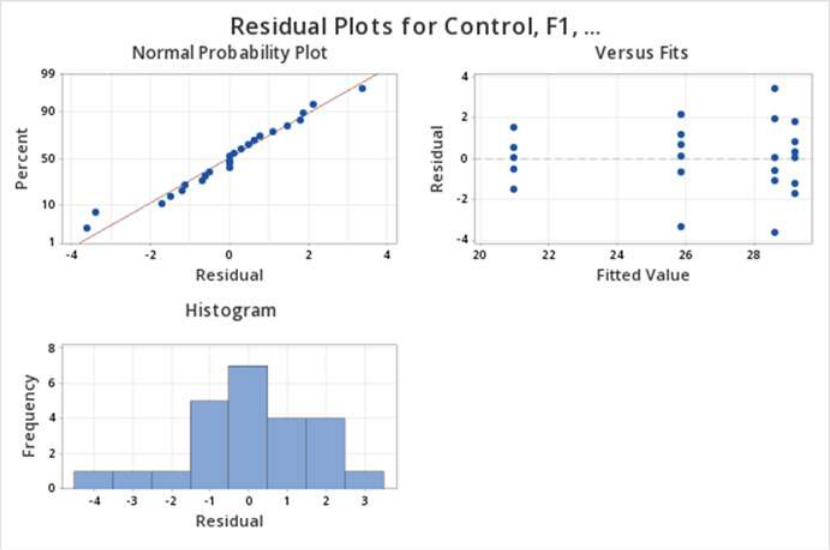


Figure 3.6.7: Residual plots generated by Minitab.

Note that the Normal Probability plot is reversed (i.e, the axes are switched) compared to the SAS output. Assessing straight line adherence is the same, and the residual analysis provided is comparable to SAS output.

This page titled 3.6: One-Way ANOVA Greenhouse Example in Minitab is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.7: One-Way ANOVA Greenhouse Example in R

R Instructions: Code for the Greenhouse Data

- Load the greenhouse data.
- Calculate the overall mean, standard deviation, and standard error.
- Calculate the mean, standard deviation, and standard error for each group.
- Produce a boxplot to plot the differences in heights for each fertilizer.
- Produce a "means plot" (interval plot) to view the differences in heights for each fertilizer.
- Obtain the ANOVA table.
- Obtain Tukey's multiple comparisons CIs and difference in means plot.
- Produce diagnostic (residuals) plots.
- Power analysis.

```
setwd("~/path-to-folder/")
greenhouse_data<-read.table("greenhouse_data.txt",header=T)
```

Note that greenhouse data are in separate columns.

```
attach(greenhouse_data)
my_data<-stack(greenhouse_data)
```

With this command, we put our data in a stacked format (the first column has the response variable (values) and the second column has the treatment levels (ind).

To calculate the overall mean, standard deviation, and standard error we can use the following commands:

```
overall_mean<-mean(my_data$values) #26.16667
overall_sd<-sd(my_data$values) #3.685892
overall_standard_error<-overall_sd/sqrt(length(my_data$values)) #0.7523795
```

To calculate the group means we can use the following command:

```
group_means<-aggregate(my_data[, 1],list(my_data$ind),mean)
#  group_means
#  Group.1        x
# 1 Control 21.00000
# 2      F1 28.60000
# 3      F2 25.86667
# 4      F3 29.20000
```

To calculate the group standard deviations and standard errors we can use the following commands:

```
group_sd<-aggregate(my_data[, 1],list(my_data$ind),sd)
# group_sd  Group.1        x
# 1 Control 1.000000
# 2      F1 2.437212
# 3      F2 1.899123
# 4      F3 1.288410
```

```
group_standard_error<-group_sd$x/sqrt(length(my_data$ind)/4)
# group_standard_error
# 0.4082483 0.9949874 0.7753136 0.5259911
```

To produce the Boxplot we can use the following commands:

```
library("ggpubr")
boxplot(values~ind,data=my_data,
xlab="Fertilizer",ylab="Plant Height",
main="Distribution of Plant Heights by Fertilizer",
frame=TRUE)
```

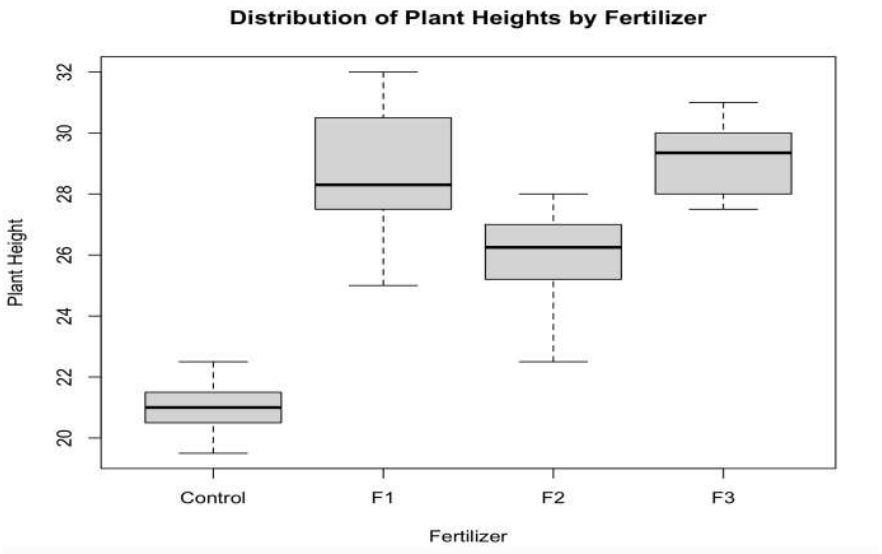


Figure 3.7.1: Box plot for distribution of plant heights by fertilizer.

To produce the means plot (interval plot) we can use the following commands:

```
library("gplots")
plotmeans(values~ind,data=my_data,connect=FALSE,
xlab="Fertilizer",ylab="Plant Height",
main="Means Plot with 95% CI")
```

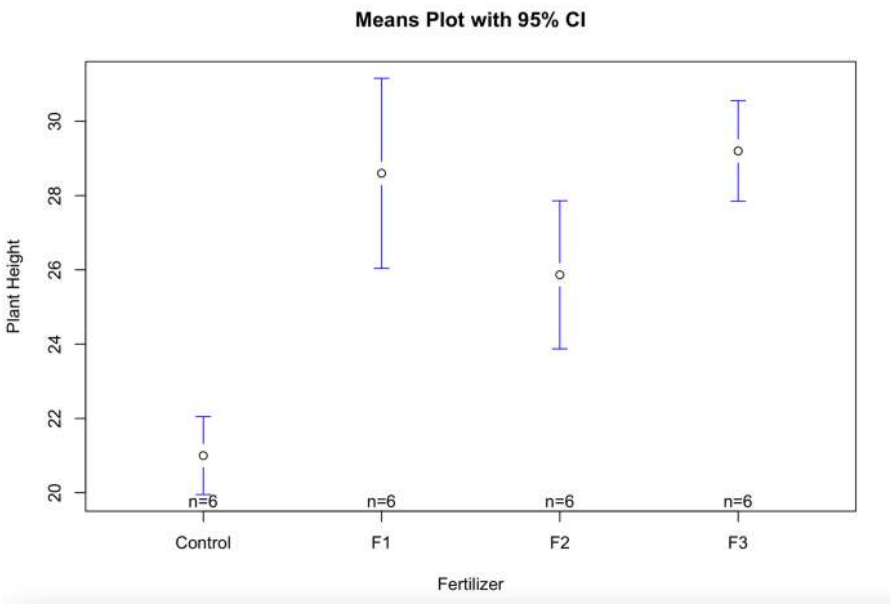


Figure 3.7.2: Means plot with 95% confidence interval.

To obtain the ANOVA table we can use the following commands:

```
anova<-aov(values~ind,my_data)
summary(anova)
```

The command summary (anova) will give you the following output:

```
summary(anova)
Df Sum Sq Mean Sq F value   Pr(>F)
ind          3 251.44   83.81   27.46 2.71e-07 ***
Residuals   20  61.03    3.05
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can see the degrees of freedom in the first column, the sum of squares in the second column, the mean sum of squares in the third column, the F-test statistic in the fourth column, and finally, we can see the p-value.

Note that the output doesn't give the $SSTO$. To find it, use the identity $SSTO = SSR + SSE$. Similarly, for the df associated with $SSTO$, add the df of SSR and SSE.

For our example, $SSTO = 251.44 + 61.03 = 312.47$

To obtain Tukey multiple comparisons of means with a 95% family-wise confidence level we use the following command:

```
library(multcomp)
library(multcompView)
tukey_multiple_comparisons<-TukeyHSD(anova,conf.level=0.95)
plot(tukey_multiple_comparisons)
tukey_multiple_comparisons
Tukey multiple comparisons of means
95% family-wise confidence level
Fit: aov(formula = values ~ ind, data = my_data)
```

```
$ind
diff        lwr         upr     p adj
F1-Control  7.600000  4.7770648 10.42293521 0.0000016
F2-Control  4.866667  2.0437315  7.68960188 0.0005509
F3-Control  8.200000  5.3770648 11.02293521 0.0000005
F2-F1      -2.733333 -5.5562685  0.08960188 0.0598655
F3-F1       0.600000 -2.2229352  3.42293521 0.9324380
F3-F2       3.333333  0.5103981  6.15626854 0.0171033
```

Based on this output, the Control group is significantly different from the 3 treatment groups and F3 is significantly different from F2.

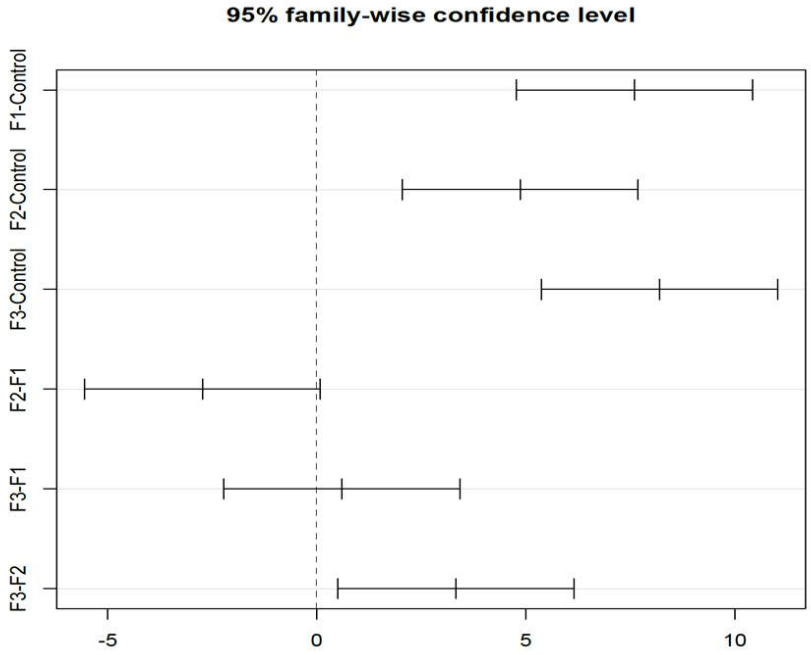


Figure 3.7.3: 95% famil-wise confidence level plot.

To produce diagnostic (residuals) plots we use the following commands:

```
#Residuals vs Fits plot
plot(anova,1)
```

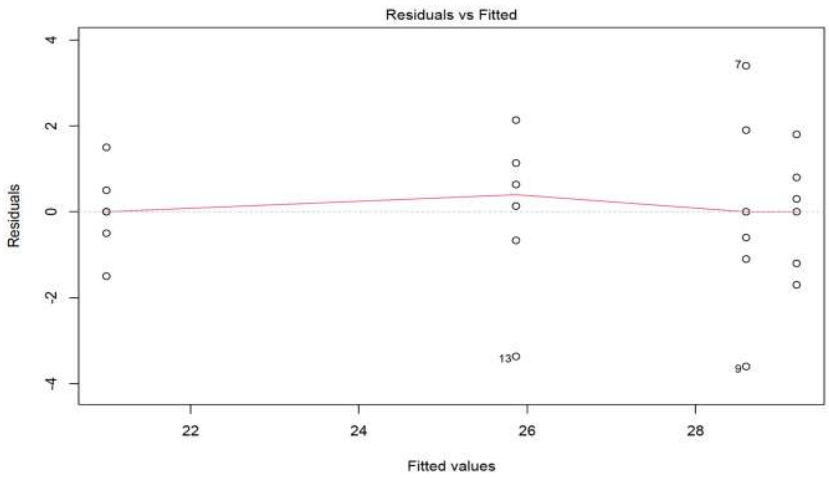


Figure 3.7.6: Residuals vs fitted values plot.

```
#QQ plot
plot(anova,2)
```

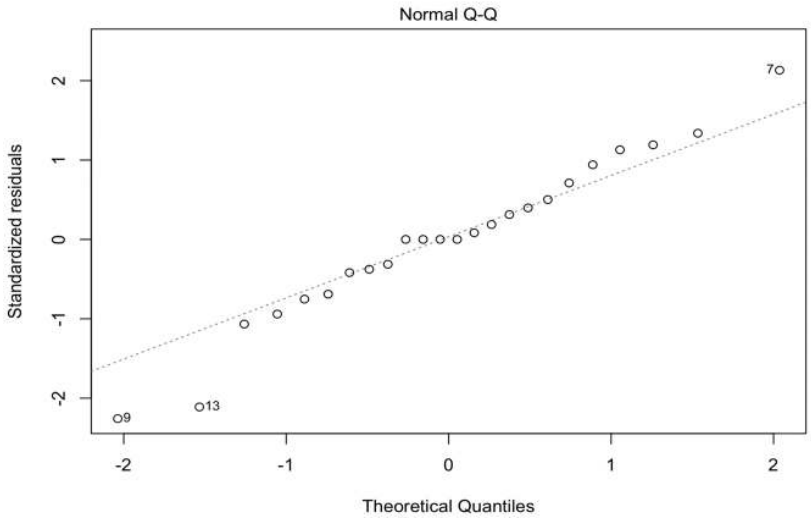


Figure 3.7.7: Normal Q-Q plot.

```
#Histogram of residuals
residuals<-anova$res #with this command we get the residuals from ANOVA
hist(residuals)
```

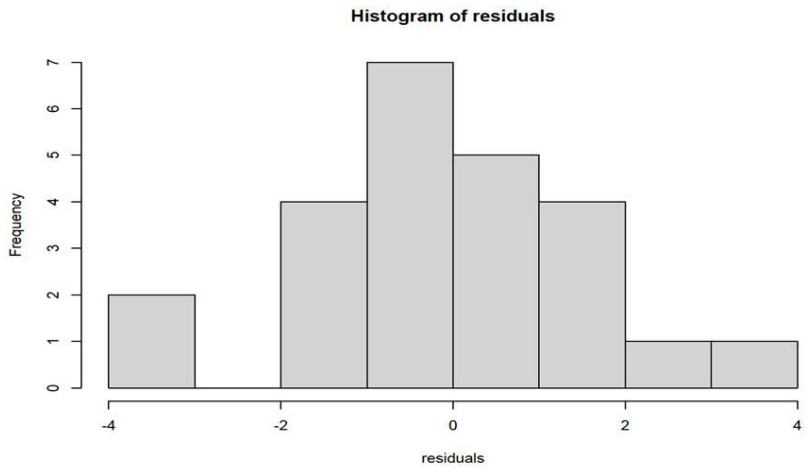


Figure 3.7.8: Histogram of residuals.

This page titled 3.7: One-Way ANOVA Greenhouse Example in R is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.8: Power Analysis

After completing a statistical test, conclusions are drawn about the null hypothesis. In cases where the null hypothesis is not rejected, a researcher may still feel that the treatment did have an effect. Let's say that three weight loss treatments are conducted. At the end of the study, the researcher analyzes the data and finds there are no differences among the treatments. The researcher believes that there really are differences. While you might think this is just wishful thinking on the part of the researcher, there MAY be a statistical reason for the lack of significant findings.

At this point, the researcher can run a **power analysis**. Recall from your introductory text or course that **power** is the ability to reject the null when the null is really false. The factors that impact power are sample size (larger samples lead to more power), the effect size (treatments that result in larger differences between groups will have differences that are more readily found), the variability of the experiment, and the significance of the type 1 error.

As a note, the most common type of power analysis are those that calculate needed sample sizes for experimental designs. These analyses take advantage of pilot data or previous research. When power analysis is done ahead of time, it is a PROSPECTIVE power analysis. This example is a retrospective power analysis, as it is done after the experiment is completed.

So back to our greenhouse example. Typically we want power to be at 80%. Again, power represents our ability to reject the null when it is false, so a power of 80% means that 80% of the time our test identifies a difference in at least one of the means correctly. The converse of this is that 20% of the time we risk not rejecting the null when we really should be rejecting the null.

Using our greenhouse example, we can run a retrospective power analysis (just a reminder, we typically don't do this unless we have some reason to suspect the power of our test was very low).

This is one analysis where Minitab is much easier and still just as accurate as SAS, so we will use Minitab to illustrate this simple power analysis in detail and follow up the analysis with SAS.

Power Analysis Techniques

? Using SAS

Steps in SAS

Let us now consider running the power analysis in SAS. In our greenhouse example with 4 treatments (control, F1, F2, and F3), the estimated means were $21, 28.6, 25.877, 29.2$respectively. Using ANOVA, the estimated standard deviation of errors was 1.747 (which is obtained by $\sqrt{MSE} = \sqrt{3.0517}$. There are 6 replicates for each treatment. Using these values, we could employ SAS `POWER` procedure to compute the power of our study retrospectively.

```
proc power;
onewayanova alpha=.05 test=overall
groupmeans=(21 28.6 25.87, 29.2)
npergroup=6 stddev=1.747
power=.;
run;
```

Fixed Scenario Elements	
Method	Exact
Alpha	0.05
Group Means	21 28.6 25.87 29.2
Standard Deviation	1.747
Sample Size per Group	6

Computed Power	
	Power
	>.999

As with MINITAB, we see that the retrospective power analysis for our greenhouse example yields a power of 1. If we re-do the analysis ignoring the *CONTROL* treatment group, then we only have 3 treatment groups: F1, F2, and F3. The ANOVA with only these three treatments yields an MSE of 3.735556 Therefore the estimated standard deviation of errors would be 1.933 We will have a power of 0.731 in this modified scenario, as shown in the below output.

Fixed Scenario Elements	
Method	Exact
Alpha	0.05
Group Means	28.6 25.87 29.2
Standard Deviation	1.933
Sample Size per Group	6

	Computed Power
	Power
	0.731

Suppose, we ask the question of how many replicates we would need to obtain at least 80% power to detect a difference in the means of our greenhouse example with the same group means but with different variability in data (i.e. standard deviations should be different). We can use SAS `POWER` to answer this question.

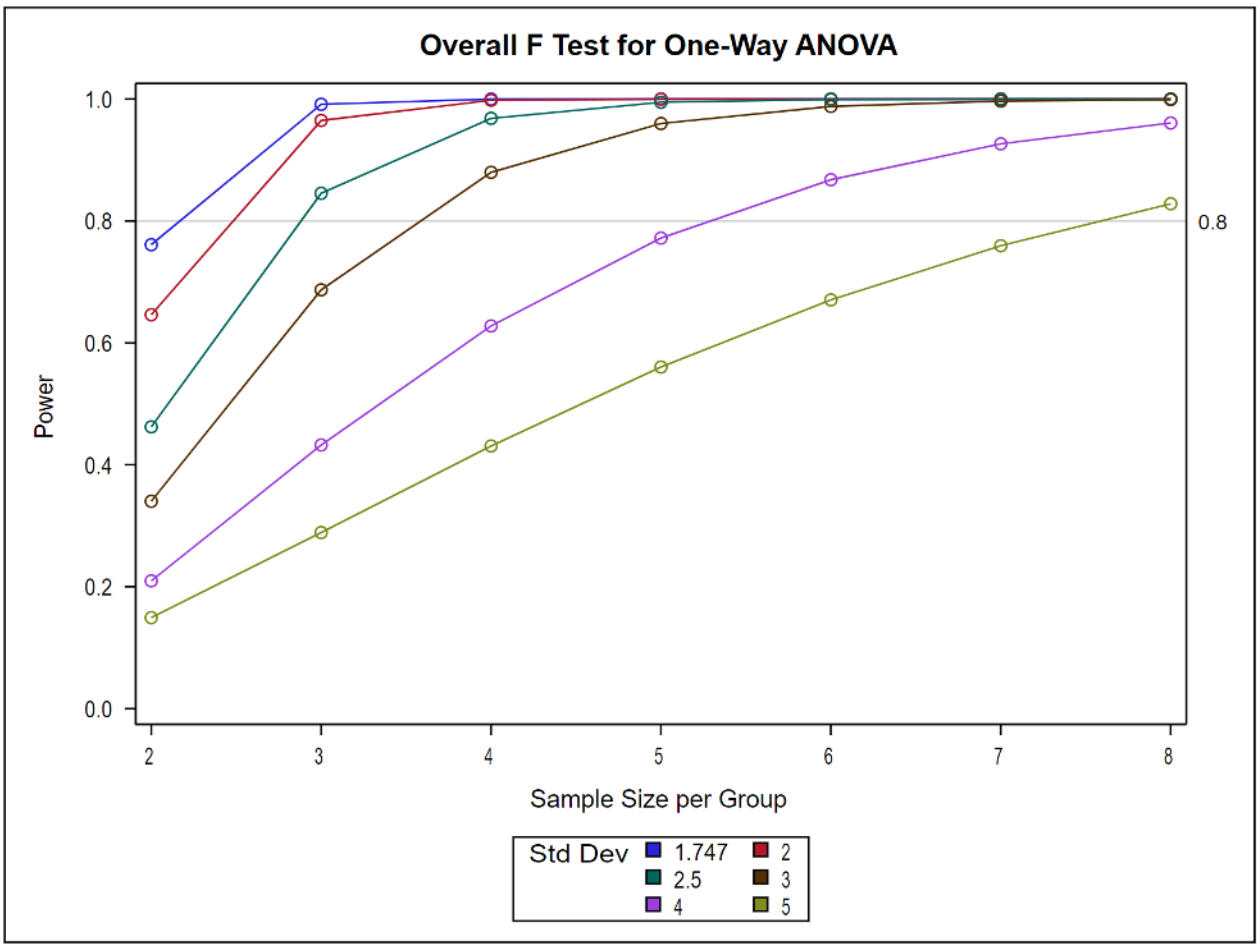


Figure 3.8.a1: Plot for overall F-test.

We can see that with a standard deviation of 1.747, if we have only 2 replicates in each of the four treatments we can detect the differences in greenhouse example means with more than 80% power. However, as the data get noisier (i.e. as standard deviation increases) we need more replicates to achieve 80% power in the same example.

Using Minitab

Steps in Minitab

In Minitab select **STAT** > **Power and Sample Size** > **One-Way ANOVA**

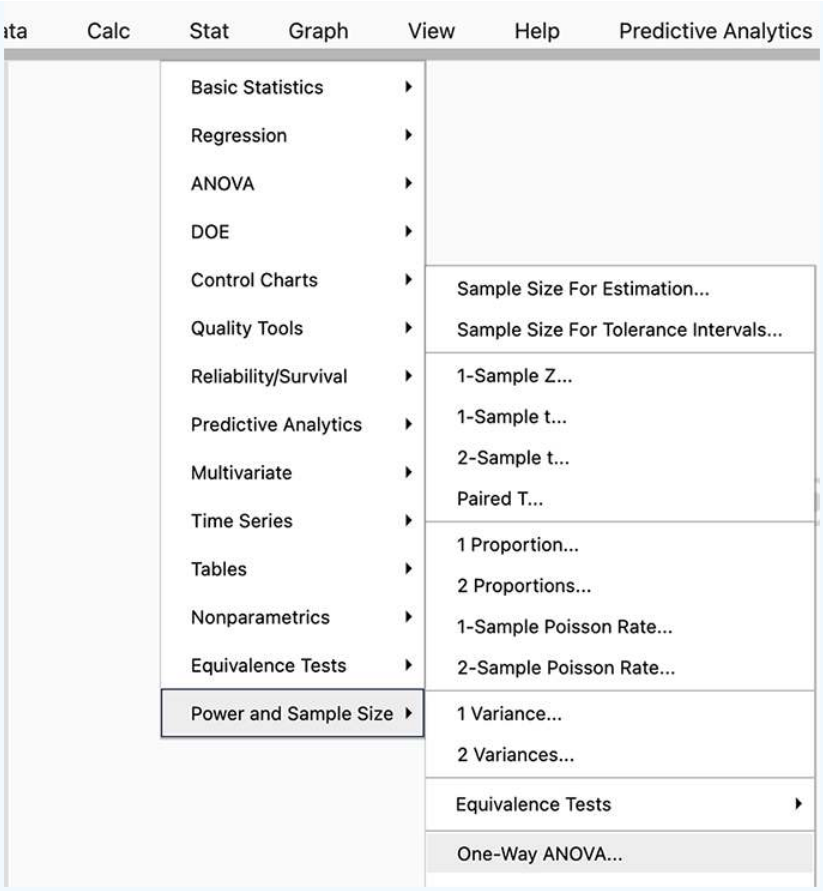


Figure 3.8.*b*1: Selecting the One-Way ANOVA tab in Minitab.

Since we have a one-way ANOVA we select this test (you can see there are power analyses for many different tests, and SAS will allow even more complicated options).

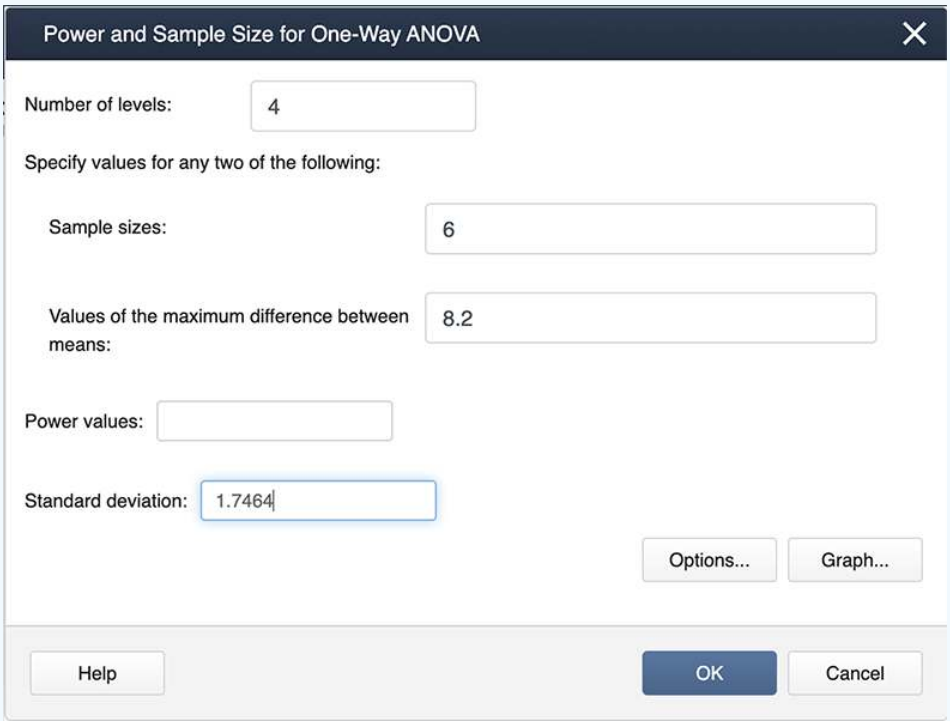


Figure 3.8.*b*2: Entering values in the Power and Sample Size pop-up window.

When you look at our filled-in dialogue box, you will notice we have not entered a value for power. This is because Minitab will calculate whichever box you leave blank (so if we needed sample size we would leave sample size blank and fill in a value for power). From our example, we know the number of levels is 4 because we have four treatments. We have six observations for each treatment so the sample size is 6. The value for the maximum difference in the means is 8.2 (we simply subtracted the smallest mean from the largest mean, and the standard deviation is 1.747. Where did this come from? The MSE, available from the ANOVA table, is about 3, and hence the standard deviation is $\sqrt{3} = 1.747$).

After we click **OK** we get the following output:

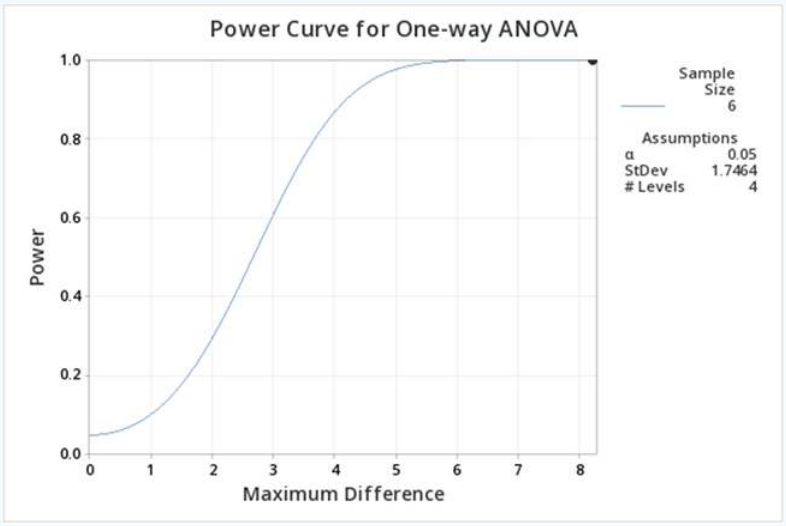


Figure 3.8.*b*3: Power curve for greenhouse data one-way ANOVA, with 4 treatment levels.

If you follow this graph you see that power is on the y-axis and the power for the specific setting is indicated by a red dot. It is hard to find, but if you look carefully the red dot corresponds to a power of 1. In practice, this is very unusual, but can be easily explained given that the greenhouse data was constructed to show differences.

We can ask the question, what about differences among the treatment groups, not considering the control? All we need to do is modify some of the input in Minitab.

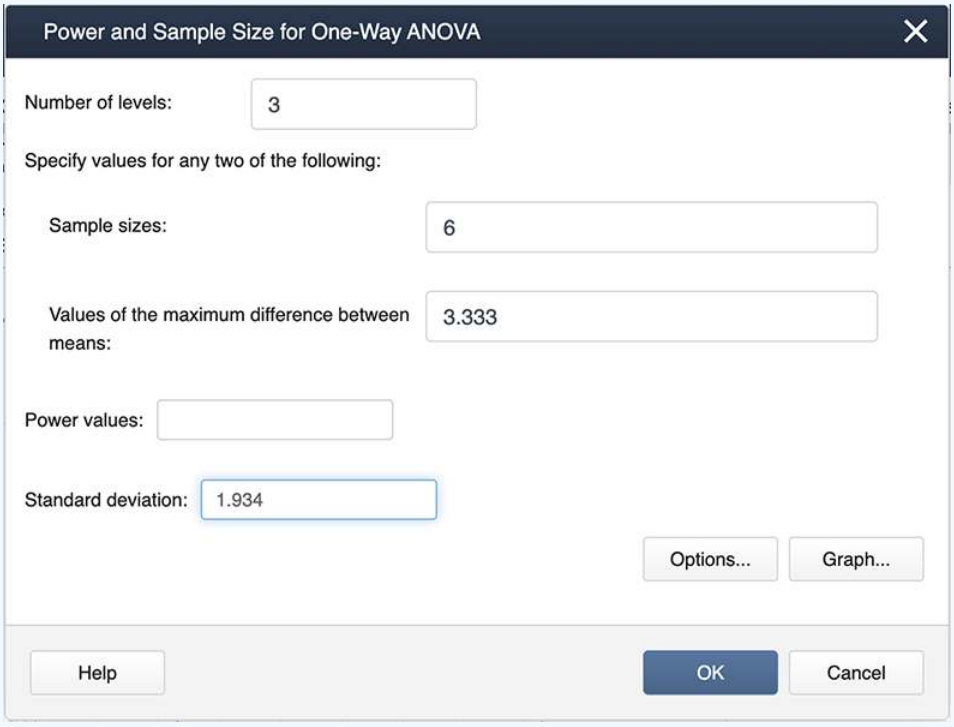


Figure 3.8.*b*4: Entering modified values in the Power and Sample Size window.

Note the differences here as in the previous screenshot. We now have 3 levels because we are only considering the three treatments. The maximum differences among the means and also the standard deviation are also different.

The output now is much easier to see:

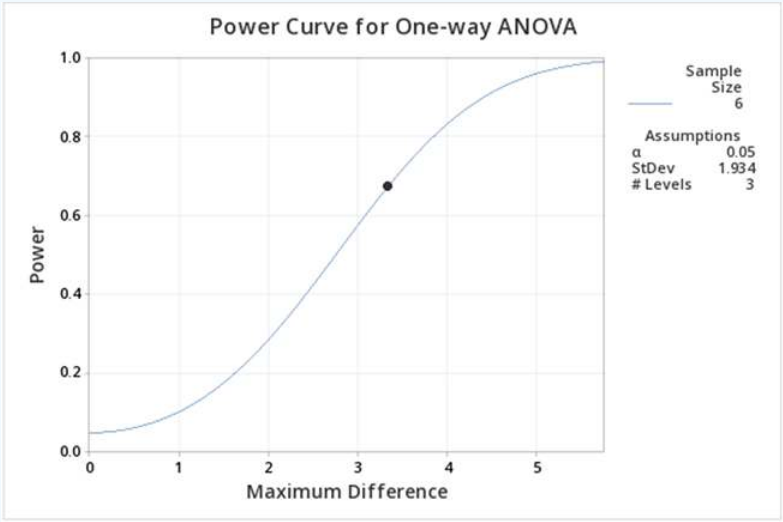


Figure 3.8.*b5*: Power curve for greenhouse data one-way ANOVA, with 3 treatment levels (control omitted).

Here we can see the power is lower than when including the control. The main reason for this decrease is that the difference between the means is smaller.

You can experiment with the power function in Minitab to provide you with sample sizes, etc. for various powers. Below is some sample output when we ask for various power curves for various sample sizes, a kind of "what if" scenario.

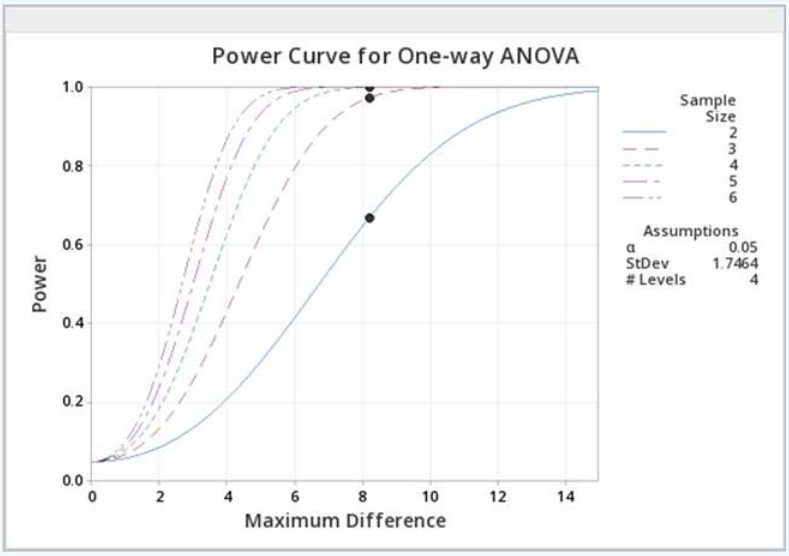


Figure 3.8.*b6*: Power curves for greenhouse data, with varying sample sizes.

Just as a reminder, power analyses are most often performed BEFORE an experiment is conducted, but occasionally, a power analysis can provide some evidence as to why significant differences were not found.

? Using R

Steps in R

With the following commands we will get the power analysis for the greenhouse example:

```
groupmeans<-c(21,28.6,25.87,29.2)
power.anova.test(groups=4,n=6,between.var=var(groupmeans),within.var=3.05,sig.le
Balanced one-way analysis of variance power calculation
groups = 4
n = 6
between.var = 13.96823
within.var = 3.05
sig.level = 0.05
power = 1
```

NOTE: n is the number in each group.

If we want to produce a power plot by increasing the sample size and the variance (like the one produced by SAS) we can use the following commands:

```
groupmeans<-c(21,28.6,25.87,29.2)
n<-c(seq(2,8,by=1))
p<power.anova.test(groups=4,n=n,between.var=var(groupmeans),within.var=3.05,sig.
p1<power.anova.test(groups=4,n=n,between.var=var(groupmeans),within.var=4,sig.le
p2<power.anova.test(groups=4,n=n,between.var=var(groupmeans),within.var=6.25,sig
p3<power.anova.test(groups=4,n=n,between.var=var(groupmeans),within.var=9,sig.le
p4<power.anova.test(groups=4,n=n,between.var=var(groupmeans),within.var=16.05,si
p5<power.anova.test(groups=4,n=n,between.var=var(groupmeans),within.var=25,sig.l
plot(n,p$power,ylab="Power",xlab="Sample size per group",main="Overall F test fo
lines(n,p$power, col = "blue")
abline(h=0.80)
par(new=TRUE)
plot(n,p1$power,ylab="Power",xlab="Sample size per group",main="Overall F test f
lines(n,p1$power, col = "red")
par(new=TRUE)
plot(n,p2$power,ylab="Power",xlab="Sample size per group",main="Overall F test f
lines(n,p2$power, col = "green")
par(new=TRUE)
plot(n,p3$power,ylab="Power",xlab="Sample size per group",main="Overall F test f
lines(n,p3$power, col = "brown")
par(new=TRUE)
plot(n,p4$power,ylab="Power",xlab="Sample size per group",main="Overall F test f
lines(n,p4$power, col = "purple")
par(new=TRUE)
plot(n,p5$power,ylab="Power",xlab="Sample size per group",main="Overall F test f
lines(n,p5$power, col = "gray")
text(locator(1),"var=3.05",col="blue")
text(locator(1),"var=4",col="red")
text(locator(1),"var=6.25",col="green")
text(locator(1),"var=9",col="brown")
text(locator(1),"var=16",col="purple")
text(locator(1),"var=25",col="gray")
```

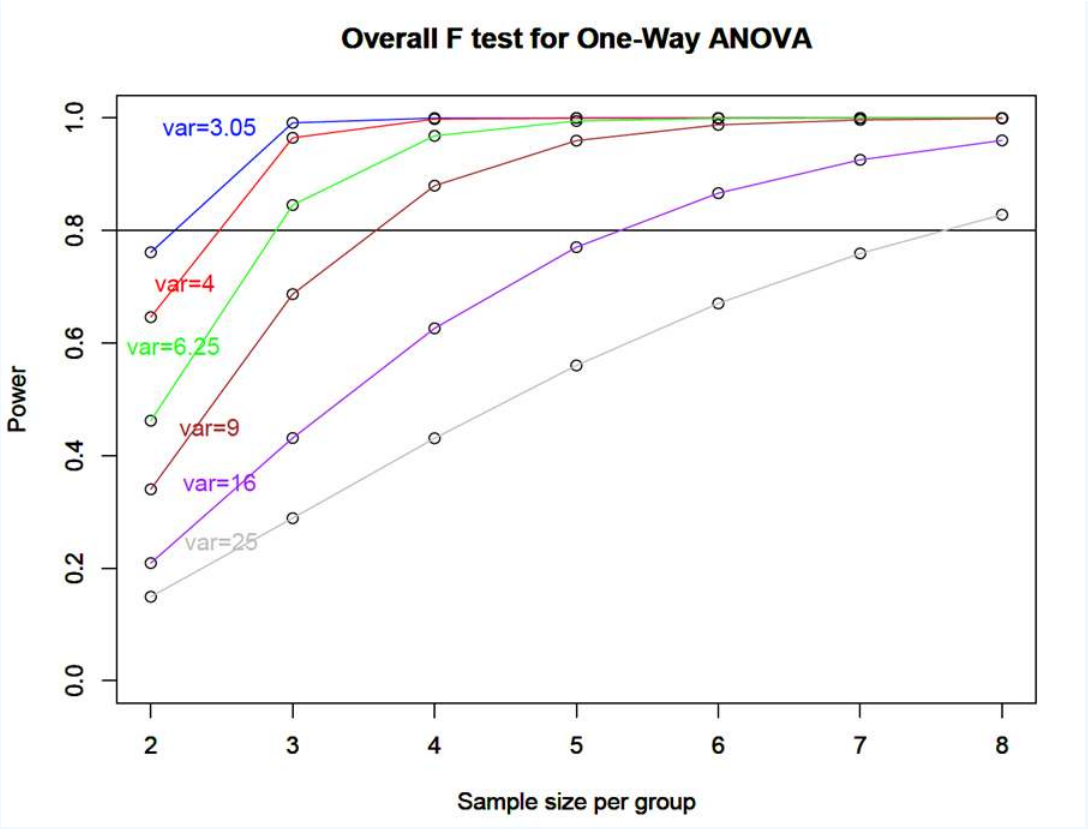


Figure 3.8.c1: Plot of overall F-test for one-way ANOVA of greenhouse data.

This page titled 3.8: Power Analysis is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.9: Try It!

Exercise 3.9.1: Diet Study

The weight gain due to 4 different diets given to 24 calves is shown below.

diet1	diet2	diet3	diet4
12	18	10	19
10	19	12	20
13	18	13	18
11	18	16	19
12	19	14	18
09	19	13	19

a) Write the appropriate null and alternative hypotheses to test if the weight gain differs significantly among the 4 diets.

Solution

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. $H_a: \mu_i \neq \mu_j for some \backslash (i, j = 1, 2, 3, 4$ OR "Not all means are equal"

Note: Here, $\mu_i,\ i = 1, 2, 3, 4$ are the actual mean weight gains due to diet1, diet2, diet3, and diet4, respectively.

b) Analyze the data and write your conclusion.

Solution

Using SAS...

```
data Lesson3_ex1;
input diet $ wt_gain;
datalines;
diet1 12
diet1 10
diet1 13
diet1 11
diet1 12
diet1 09
diet2 18
diet2 19
diet2 18
diet2 18
diet2 19
diet2 19
diet3 10
diet3 12
diet3 13
diet3 16
diet3 14
diet3 13
diet4 19
diet4 20
diet4 18
diet4 19
diet4 18
diet4 19
;
ods graphics on;
proc mixed data= Lesson3_ex1 plots=all; class diet;
model wt_gain = diet;
contrast 'Compare diet1 with diets 2,3,4 combined ' diet 3 -1 -1 -1;
store result1;
title 'ANOVA of Weight Gain Data';
run;
ods html style=statistical sge=on;
proc plm restore=result1;
lsmeans diet/ adjust=tukey plot=meanplot cl lines;
run;
```

The ANOVA results shown below indicate that the diet effect is significant with an F-value of 51.27 (p-value <.0001). This means that not all diets provide the same mean weight gain. The diffogram below indicates the significant different pairs of diets identified by solid blue lines. The estimated mean weight gains from diets 1, 3, 2, and 4 are 11, 13, 18.1, and 19 units respectively. The diet pairs that have significantly different mean weight gains are (1,2), (1,4), (3,2), and (3,4).

Partial Output:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
diet	3	20	51.27	<.0001

diet Least Squares Means

diet	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
diet1	11.1667	0.5413	20	20.63	<.0001	0.05	10.0374	12.2959
diet2	18.5000	0.5413	20	34.17	<.0001	0.05	17.3708	19.6292
diet3	13.0000	0.5413	20	24.01	<.0001	0.05	11.8708	14.1292
diet4	18.8333	0.5413	20	34.79	<.0001	0.05	17.7041	19.9626

Differences of diet Least Squares Means
Adjustment for Multiple Comparisons: Tukey

diet	_diet	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
diet1	diet2	-7.3333	0.7656	20	-9.58	<.0001	<.0001	0.05	-8.9303	-5.7364	-9.4761	-5.1906
diet1	diet3	-1.8333	0.7656	20	-2.39	0.0265	0.1105	0.05	-3.4303	-0.2364	-3.9761	0.3094
diet1	diet4	-7.6667	0.7656	20	-10.01	<.0001	<.0001	0.05	-9.2636	-6.0697	-9.8094	-5.5239
diet2	diet3	5.5000	0.7656	20	7.18	<.0001	<.0001	0.05	3.9030	7.0970	3.3572	7.6428
diet2	diet4	-0.3333	0.7656	20	-0.44	0.6679	0.9716	0.05	-1.9303	1.2636	-2.4761	1.8094
diet3	diet4	-5.8333	0.7656	20	-7.62	<.0001	<.0001	0.05	-7.4303	-4.2364	-7.9761	-3.6906

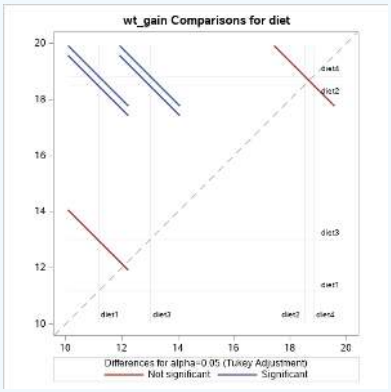


Figure 3.9.a1: SAS-generated diffogram for weight gain comparisons by diet.

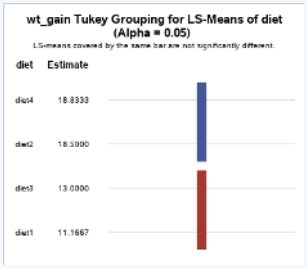


Figure 3.9.a2: SAS-generated Tukey grouping of weight gains for diet LS-means.

? Exercise 3.9.2: Commuter Times

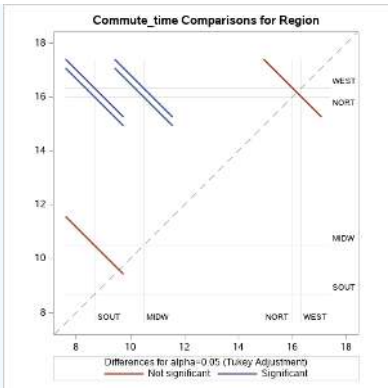


Figure 3.9.*b*1: Commute time comparisons in hours by region.

Above is a diffogram depicting the differences in daily commuter time (in hours) among regions of a metropolitan city. Answer the following.

a) Name the regions included in the study.

Solution

SOUT, MIDW, NORT, and WEST

b) How many red or blue lines are to be expected?

Solution

4 choose 2 = 6 red or blue lines

c) Which pairs of regions have significantly different average commuter times?

Solution

(SOUT and NORT), (SOUT and WEST), (MIDW and NORT), and (MIDW and WEST) have significantly different mean commuter times.

d) Write down the estimated mean daily commuter time for each region.

Solution

Region	SOUT	MIDW	NORT
Estimated mean commuter time in hours	8.7	10.5	16

This page titled 3.9: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

3.10: Chapter 3 Summary

The primary focus in this chapter was to establish the foundation for developing mathematical models for a one-way ANOVA setting. The effects model was then discussed along with the ANOVA model assumptions and diagnostics. The other focus was to illustrate, using the greenhouse example, how SAS and Minitab can be utilized to run an ANOVA model. Sections 3.3-3.6 were devoted to this purpose and include details on SAS and Minitab ANOVA basics, together with guidance in the interpretation of the outputs. Software-based diagnostics tests to detect the validity of model assumptions were also discussed, along with the power analysis procedure which computes any one of the four quantities of sample size, power, effect size, and the significance level, given the other three.

The next chapter will be a continuation of this lesson. Three more different versions of ANOVA model equations that represent a single factor experiment will be discussed. These are known as **Overall Mean**, **Cell Means**, and **Dummy Variable Regression** models.

This page titled 3.10: Chapter 3 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

CHAPTER OVERVIEW

4: ANOVA Models Part II

Objectives

By the end of this chapter, students will be able to:

- Apply the overall mean, cell means, and dummy variable regression models for a one-way ANOVA and interpret the results.
- Identify the design matrix and the parameter vector for each ANOVA model studied.
- Recognize aspects of ANOVA programming computations.

This is a continuation of the previous lesson, and in this lesson, three more alternative ANOVA models are introduced. ANOVA models are derived under the assumption of linearity of model parameters and additivity of model terms so that every model will follow the general linear model (GLM): $Y = X\beta + \mathcal{E}$. In later sections of this lesson, we will see that the appropriate choice of X, the design matrix, will result in a different ANOVA model. This lesson will also shed insight into the similarities of how ANOVA calculations are done by most software, regardless of which model is being used. Finally, the concept of a study diagram is also discussed, demonstrating its usefulness when building a statistical model and designing an experiment.

This page titled 4: ANOVA Models Part II is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

4.1: How is ANOVA Calculated?

In the past lessons, we carried out the ANOVA computations conceptually in terms of deviations from means. For the calculation of total variance, we used the deviations of the individual observations from the overall mean, while the treatment SS was calculated using the deviations of treatment level means from the overall mean, and the residual or error SS was calculated using the deviations of individual observations from treatment level means. In practice, however, to achieve higher computational efficiency, SS for ANOVA is computed utilizing the following mathematical identity:

$$SS = \sum \left(Y_i - \bar{Y}\right)^2 = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{N} \tag{4.1.1}$$

This identity is commonly called the **working formula** or **machine formula**. The second term on the right-hand side is often referred to as the **correction factor (CF)**.

For computing the SS for the total variance of the responses, the formula above can be used as it is, but modifications need be made for others. For example, to compute the treatment SS, the above equation has to be modified as:

$$SS_{treatment} = \sum_{i=1}^{T} \frac{\left(\sum_{j=1}^{n_i} Y_{ij}\right)^2}{n_i} - \frac{\left(\sum Y_i\right)^2}{N} \tag{4.1.2}$$

We will examine three new ANOVA models (Models 1, 2, 3), as well as the effects model (Model 4) from the previous lesson, defined as follows:

Model 1 - The Overall Mean Model

$$Y_{ij} = \mu + \epsilon_{ij} \tag{4.1.3}$$

which simply fits an overall or "grand" mean'. This model reflects the situation where H_0 is true, implying that $\mu_1 = \mu_2 = \ldots = \mu_T$.

Model 2 - The Cell Means Model

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{4.1.4}$$

where $\mu_i,\ i = 1, 2, \ldots, T$ are the factor level means. Note that in this model, there is no overall mean being fitted.

Model 3 - Dummy Variable Regression

$$Y_{ij} = \mu + \mu_i + \epsilon_{ij},\ \text{fitted as}\ Y_{ij} = \beta_0 + \beta_{Level\,1} + \beta_{Level\,2} + \ldots + \beta_{Level\,r-1} + \epsilon_{ij} \tag{4.1.5}$$

where $\beta_{Level\,1}, \beta_{Level\,2}, \ldots, \beta_{Level\,T-1}$ are regression coefficients for $T-1$ indicator-coded regression "dummy" variables that are correspond to the $T-1$ categorical factor levels. The T^{th} factor level mean is given by the regression intercept β_0.

Model 4 - The Effects Model

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{4.1.6}$$

where τ_i are the the deviations of each factor level mean from the overall mean so that $\sum_{i=1}^{T} \tau_i = 0$.

Each of these four models can be written as a **general linear model (GLM)**: $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\mathcal{E}}$ simply by changing the design matrix $\mathbf{X}$. Thus to perform the data analysis, in terms of the computer coding instructions, the appropriate numerical values for the $\mathbf{X}$ matrix elements will need to be inputted.

This page titled 4.1: How is ANOVA Calculated? is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

4.2: The Overall Mean Model

Model 1 - The Overall Mean Model

$$Y_{ij} = \mu + \epsilon_{ij} \tag{4.2.1}$$

which simply fits an overall or "grand" mean. This model reflects the situation where H_0 is true, implying that $\mu_1 = \mu_2 = \ldots = \mu_T$.

To understand how various facades of the model relate to each other, let us look at a toy example with 3 treatments (or factor levels) and 2 replicates of each treatment.

We have 6 observations, which means that $\mathbf{Y}$ is a column vector of dimension 6 and so is the error vector $\boldsymbol{\mathcal{E}}$ where its elements are the random error values associated with the 6 observations. In the GLM model of $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, the design matrix $\mathbf{X}$ for the overall mean model turns out to be a 6-dimensional column vector of ones. The parameter vector, $\boldsymbol{\beta}$, is a scalar equal to μ, the overall population mean.

That is,

$$\mathbf{Y} = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{\beta} = [\mu], \text{ and } \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix} \tag{4.2.2}$$

Using the method of least squares, the estimates of the parameters in $\boldsymbol{\beta}$ are obtained as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{4.2.3}$$

Using the estimate $\hat{\boldsymbol{\beta}}$, the i^{th} predicted response $\hat{\mathbf{y}}_\mathbf{i}$ can be computed as $\hat{\mathbf{y}}_\mathbf{i} = \mathbf{x_i}'$, where $\mathbf{x_i}'$ denotes the i^{th} row vector of the design matrix.

In this simplest of cases, we can see how the matrix algebra works. The term $\mathbf{X}'\mathbf{X}$ would be:

$$[1\ 1\ 1\ 1\ 1\ 1] * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 1+1+1+1+1+1 = 6 = n \tag{4.2.4}$$

The term $\mathbf{X}'\mathbf{Y}$ would be:

$$[1\ 1\ 1\ 1\ 1\ 1] * \begin{bmatrix} 2 \\ 1 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} = 2+1+3+4+5+6 = 21 = \sum Y_i \tag{4.2.5}$$

So in this case, the estimate $\mathbf{b}$ as expected is simply the overall mean $= \hat{\mu} = \bar{y}_{..} = 21/6 = 3.5$

Note that the exponent of $\mathbf{X}'\mathbf{X}$ in the formula above indicates arithmetic division as $\mathbf{X}'\mathbf{X}$ is a scalar increase in this case. In the more general setting, the superscript of '-1 ' indicates the inverse operation in matrix algebra.

To perform these matrix operations in SAS IML, we will open a regular SAS editor window, and then copy and paste three components from the file (IML Design Matrices) as shown below.

SAS: Overall Mean Model

Steps in SAS

Step 1

Procedure initiation, and specification of the dependent variable vector, $\mathbf{Y}$.

For our example we have:

```
/* Initiate IML, define response variable */
proc iml;
y={
2,
1,
3,
4,
6,
5};
```

Step 2

We then enter a design matrix $\mathbf{X}$. For the Overall Mean model and our example data, we have:

```
x={
1,
1,
1,
1,
1,
1};
```

Step 3

We can now copy and paste a program for the matrix computations to generate results (regression coefficients and ANOVA output):

```
beta=inv(x`*x)*x`*y;
beta_label={"Beta_0","Beta_1","Beta_2","Beta_3"};
print beta [label="Regression Coefficients"
                        rowname=beta_label];
n=nrow(y);
p=ncol(x);
j=j(n,n,1);
ss_tot = (y`*y) - (1/n)*(y`*j)*y;
ss_trt = (beta`*(x`*y)) - (1/n)*(y`*j)*y;
ss_error = ss_tot - ss_trt;
total_df=n-1;
trt_df=p-1;
error_df=n-p;
ms_trt = ss_trt/(p-1);
```

```
ms_error = ss_error / error_df;
F=ms_trt/ms_error;

empty={.};
row_label= {"Treatment", "Error", "Total"};
col_label={"df" "SS" "MS" "F"};
trt_row= trt_df || ss_trt || ms_trt || F;
error_row= error_df || ss_error || ms_error || empty;
tot_row=total_df || ss_tot || empty || empty;
aov = trt_row // error_row // tot_row;
print aov [label="ANOVA"
                        rowname=row_label
                        colname=col_label];
```

Here is a quick video walk-through to show you the process for how you can do this in SAS. (Right-click and select "Show All" if your browser does not display the entire screencast window.)

Video 4.2.1: Walkthrough for ANOVA using the SAS overall mean model.

The program can then be run to produce the following output:

Regression Coefficients	
Beta_0	3.5

ANOVA				
Treatment	DF	SS	MS	F
	0	0		
Error	5	17.5	3.5	
Total	5	17.5		

We see the estimate of the regression coefficient for β_0 equals 3.5, which indeed is the overall mean of the response variable, and is also the same value we obtained above using "by-hand" calculations. In this simple case, where the treatment factor has not entered the model, the only item of interest from the ANOVA table would be the SS_{Error} for later use in the General Linear F-test.

If you like to see the internal calculations further, you may optionally add the following few lines, to the end of the calculation code.

```
/* (Optional) Intermediates in the matrix computations */
xprimex=x`*x; print xprimex;
xprimey=x`
*y; print xprimey;
xprimexinv=inv(x`*x); print xprimexinv;
check=xprimexinv*xprimex; print check;
SumY2=beta`
*(x`*y); print SumY2;
CF=(1/n)*(y`
*j)*y; print CF;
```

This additional code produces the following output:

xprimex	xprimey	xprimeinv
6	21	0.1666667

check	SumY2	CF
1	73.5	73.5

From this we can verify the computations for the $SS_{treatment} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} = \sum Y_2 - CF = 0$.

The "check" calculation confirms that $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = 1$, which in fact defines the matrix division operation. In this simple case, it amounts to simple division by n, but in other models that we will work with, the matrix division process is more complicated and is explained here. In general, the inverse of a matrix $\mathbf{A}$, denoted $\mathbf{A}^{-1}$, is defined by the matrix identity $\mathbf{A}^{-1}\mathbf{A} = I$, where I is the identity matrix (a diagonal matrix of 1's). In this example, $\mathbf{A}$ is replaced by $\mathbf{X}'\mathbf{X}$, which is a scalar and equals 6.

R: Overall Mean Model

Steps in R

1. Define response variable and design matrix

```
y<-matrix(c(2,1,3,4,6,5), ncol=1)
x<-matrix(c(1,1,1,1,1,1), ncol=1)
```

2. Regression coefficients

```
beta<-solve(t(x)%*%x)%*%(t(x)%*%y) #3.5
```

3. Calculate the entries of the ANOVA Table

```
n<-nrow(y)
p<-ncol(x)
J<-matrix(1,n,n)
ss_tot = (t(y)%*%y) - (1/n)*(t(y)%*%J)%*%y #17.5
```

```
ss_trt = t(beta)%*%(t(x)%*%y) - (1/n)*(t(y)%*%J)%*%y #0
ss_error = ss_tot - ss_trt #17.5
total_df=n-1 #5
trt_df=p-1 #0
error_df=n-p #5
MS_trt = ss_trt/(p-1)
MS_error = ss_error / error_df #3.5
F=MS_trt/MS_error
```

4. Creating the ANOVA table

```
ANOVA <- data.frame(
c ("","Treatment","Error", "Total"),
c("DF", trt_df,error_df,total_df),
c("SS", ss_trt, ss_error, ss_tot),
c("MS", "", MS_error, ""),
c("F","","",""),
stringsAsFactors = FALSE)
names(ANOVA) <- c(" ", "  ", " ","","")
```

5. Print the ANOVA table

```
print(ANOVA)
# 1           DF   SS  MS F
# 2 Treatment  0    0
# 3     Error  5 17.5 3.5
# 4     Total  5 17.5
```

6. Intermediates in the matrix computations

```
xprimex<-t(x)%*%x # 6
xprimey<-t(x)%*%y # 21
xprimexinv<-solve(t(x)%*%x) # 0.1666667
check<-xprimexinv*xprimex # 1
SumY2<-t(beta)%*%(t(x)%*%y) # 73.5
CF<-(1/n)*(t(y)%*%J)%*%y # 73.5
```

This page titled 4.2: The Overall Mean Model is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

4.3: Cell Means Model

Model 2 - The Cell Means Model

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{4.3.1}$$

where $\mu_i,\ i = 1, 2, \ldots, T$ are the factor level means. Note that in this model, there is no overall mean being fitted.

The cell means model does not fit an overall mean, but instead fits an individual mean for each of the treatment levels. Let us run this model for the same data assuming that each pair of observations arise from one treatment level, so that T, the number of treatment levels equals 3. We then have to replace the design matrix in the IML code with:

```
/* The Cell Means Model */
x={
1    0    0,
1    0    0,
0    1    0,
0    1    0,
0    0    1,
0    0    1};
```

Notice that each column represents a specific treatment level and is using indicator coding: 1 for the rows corresponding to the observations receiving the specified treatment level, and 0 for the other rows. It can be seen that $r = 2$ is the number of replicates for each treatment level. Observe that column 1 generates the mean for treatment level 1, column 2 for treatment level 2, and column 3 for treatment level 3.

To write the cell means model as a GLM, let

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{x_1}' \\ \mathbf{x_2}' \\ \mathbf{x_3}' \\ \mathbf{x_4}' \\ \mathbf{x_5}' \\ \mathbf{x_6}' \end{bmatrix} \tag{4.3.2}$$

where $\mathbf{x_i}'$ is the i^{th} row vector of the design matrix.

The parameter vector $\boldsymbol{\beta}$ is a 3-dimensional column vector and is defined by

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \tag{4.3.3}$$

The parameter estimates $\hat{\boldsymbol{\beta}}$ can again be found using the least squares method. One can verify that $\mu_i = \bar{y}_i$, the i^{th} treatment mean, for $i = 1, 2, 3$. Using this estimate, the resulting estimated regression equation for the cell means model is,

$$\hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \tag{4.3.4}$$

which produces $\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{x}_{\boldsymbol{i}}' \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix}$.

We then re-run the program with the new design matrix to get the following output:

Regression Coefficients

Regression Coefficients	1.5
Beta_0	1.5
Beta_1	3.5
Beta_2	5.5
Beta_1	3.5

ANOVA				
Treatment	dF	SS	MS	F
	2	16	8	16
Error	3	1.5	0.5	
Total	5	17.5		

The regression coefficients β_0, β_1, and β_2 are now the means for each treatment level, and in the ANOVA table, we see that the SS_{Error} is 1.5. This reduction in the SS_{Error} is the $SS_{treatment}$. Notice that the error SS of the Overall Mean model is the sum of the SS values for Treatment and Error term in this model, which means that by not including the treatment effect in that model, its error SS has been unduly inflated.

Adding the optional code given in Section 4.2 to compute additional Internal computations, we can obtain:

xprimex			check			xprimey
2	0	0	1	0	0	3
0	2	0	0	1	0	7
0	0	2	0	0	1	11

SumY2	CF
89.5	73.5

xprimexinv		
0.5	0	0
0	0.5	0
0	0	0.5

Here we can see that $\mathbf{X}'\mathbf{X}$ now contains diagonal elements that are the n_i = number of observations for each treatment level mean being computed. In addition, we can verify that $CF = \sum Y^2 - CF = 16$, or the working formula equals the treatment SS.

We can now test for the significance of the treatment by using the General Linear F test:

$$F = \frac{SSE_{reduced} - SSE_{full}/dfE_{reduced} - dfE_{full}}{SSE_{full}/dfE_{full}} \tag{4.3.5}$$

The Overall Mean model is the "Reduced" model, and the Cell Means model is the "Full" model. From the ANOVA tables, we get:

$$F = \frac{17.5 - 1.5/5 - 3}{1.5/3} = 16 \tag{4.3.6}$$

which can be compared to $F_{.05,2,3} = 9.55$.

Using R

Steps in R - Cell Means Model

1. Define response variable and design matrix

https://stats.libretexts.org/@go/page/33491

```
y<-matrix(c(2,1,3,4,6,5), ncol=1)
x<matrix(c(1,0,0,1,0,0,0,1,0,0,1,0,0,0,1,0,0,1),ncol=3,nrow=6,byrow=TRUE)
```

2. Regression coefficients

```
beta<-solve(t(x)%*%x)%*%(t(x)%*%y)
#  beta
#      [,1]
# [1,]  1.5
# [2,]  3.5
# [3,]  5.5
```

3. Calculate the entries of the ANOVA Table

```
n<-nrow(y)
p<-ncol(x)
J<-matrix(1,n,n)
ss_tot = (t(y)%*%y) - (1/n)*(t(y)%*%J)%*%y #17.5
ss_trt = t(beta)%*%(t(x)%*%y) - (1/n)*(t(y)%*%J)%*%y #16
ss_error = ss_tot - ss_trt #1.5
total_df=n-1 #5
trt_df=p-1 #2
error_df=n-p #3
MS_trt = ss_trt/(p-1) #8
MS_error = ss_error / error_df #0.5
F=MS_trt/MS_error #16
```

4. Creating the ANOVA table

```
ANOVA <- data.frame(
c ("","Treatment","Error", "Total"),
c("DF", trt_df,error_df,total_df),
c("SS", ss_trt, ss_error, ss_tot),
c("MS", MS_trt, MS_error, ""),
c("F",F,"",""),
stringsAsFactors = FALSE)
names(ANOVA) <- c(" ", "  ", " ","","")
```

5. Print the ANOVA table

```
print(ANOVA)
# 1           DF   SS  MS  F
# 2 Treatment  2   16   8 16
# 3     Error  3  1.5 0.5
# 4     Total  5 17.5
```

6. Intermediates in the matrix computations

```
xprimex<-t(x)%*%x
#  xprimex
#      [,1] [,2] [,3]
# [1,]    2    0    0
# [2,]    0    2    0
# [3,]    0    0    2
xprimey<-t(x)%*%y
#  xprimey
#      [,1]
# [1,]    3
# [2,]    7
# [3,]   11
xprimexinv<-solve(t(x)%*%x)
#  xprimexinv
#      [,1] [,2] [,3]
# [1,]  0.5  0.0  0.0
# [2,]  0.0  0.5  0.0
# [3,]  0.0  0.0  0.5
check<-xprimexinv%*%xprimex
#  check
#      [,1] [,2] [,3]
# [1,]    1    0    0
# [2,]    0    1    0
# [3,]    0    0    1
SumY2<-t(beta)%*%(t(x)%*%y) #89.5
CF<-(1/n)*(t(y)%*%J)%*%y #73.5
```

This page titled 4.3: Cell Means Model is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

4.4: Dummy Variable Regression

Model 3 - Dummy Variable Regression

$$Y_{ij} = \mu + \mu_i + \epsilon_{ij}, \text{ fitted as } Y_{ij} = \beta_0 + \beta_{Level\,1} + \beta_{Level\,2} + \ldots + \beta_{Level\,T-1} + \epsilon_{ij} \tag{4.4.1}$$

where $\beta_{Level\,1}, \beta_{Level\,2}, \ldots, \beta_{Level\,T-1}$ are regression coefficients for $T-1$ indicator-coded regression "dummy" variables that are correspond to the $T-1$ categorical factor levels. The T^{th} factor level mean is given by the regression intercept β_0.

The General Linear Model (GLM) applied to data with categorical predictors can be viewed from a regression modeling perspective as an ordinary multiple linear regression (MLR) with "dummy" coding, also known as indicator coding, for the categorical treatment levels. Typically, software performing the MLR will automatically include an intercept, which corresponds to the first column of the design matrix and is a column of 1's. This automatic inclusion of the intercept can lead to complications when interpreting the regression coefficients.

The SAS Mixed procedure, and also the GLM procedure which we may encounter later, use the "Dummy Variable Regression" model. For the Y data used in sections 4.2 and 4.3, the design matrix for this model can be entered into IML as:

```
/* Dummy Variable Regression Model */
x = {
1    1    0,
1    1    0,
1    0    1,
1    0    1,
1    0    0,
1    0    0};
```

Notice that in the above design matrix, there are only two indicator columns even though there are three treatment levels in the study. It is because, similar to the matrix below, if we were to have a design matrix with another indicator column representing the third treatment level, the resulting 4 columns would form a set of linearly dependent columns, a mathematical condition that will hinder the computation process any further as explained below.

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \tag{4.4.2}$$

The above matrix containing all 4 columns has the property that the sum of columns 2-4 will equal the first column representing the intercept. As a result, a mathematical condition called singularity is created and the matrix computations will not run. So one of the treatment levels is omitted from the coding in the design matrix above for IML and the eliminated level is called the ‘reference’ level. In SAS, typically, the treatment level with the highest label is defined as the reference level and so, in this study, it is treatment level 3.

Note that the parameter vector for the dummy variable regression model is

$$\boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \tag{4.4.3}$$

.

Running IML, with the design matrix for the dummy variable regression model, we get the following output;

Regression Coefficients	
Beta_0	5.5
Beta_1	-4
Beta_2	-2

The coefficient β_0 is the mean for treatment level 3. The mean for treatment level 1 is then calculated from $\hat{\beta}_0 + \hat{\beta}_1 = 1.5$. Likewise, the mean for treatment level 2 is calculated as $\hat{\beta}_0 + \hat{\beta}_2 = 3.5$.

Notice that the F statistic calculated from this model is the same as that produced from the Cell Means model.

ANOVA				
Treatment	df	SS	MS	F
	2	16	8	16
Error	3	1.5	0.5	
Total	5	17.5		

Using Technology

? Minitab Example

We can confirm our ANOVA table now by running the analysis in software such as Minitab.

Steps in Minitab

First input the data:

	C1	C2-T
	y	trt
1	2	A
2	1	A
3	3	B
4	4	B
5	6	C
6	5	C
7		

Figure 4.4.a1: Inputting data.

In Minitab, different coding options allow the choice of the design matrix which can be done as follows:

`Stat` > **`ANOVA`** > **`General Linear Model`** > **`Fit General Linear Model`** and place the variables in the appropriate boxes:

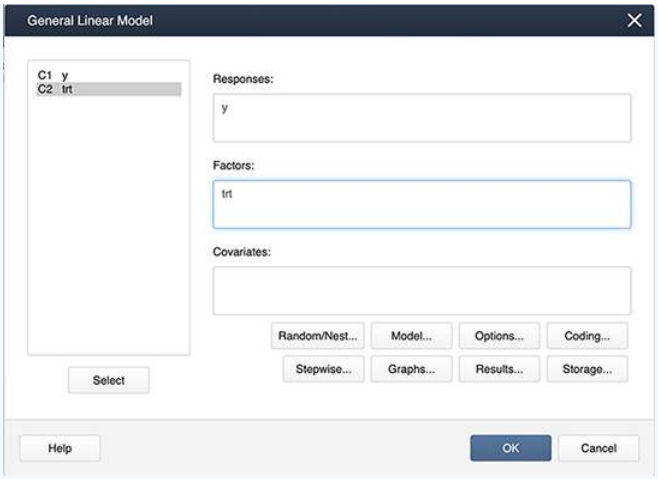


Figure 4.4.a2: Placing variables in the General Linear Model pop-up window.

Then select **Coding**… and choose the (1,0) coding as shown below:

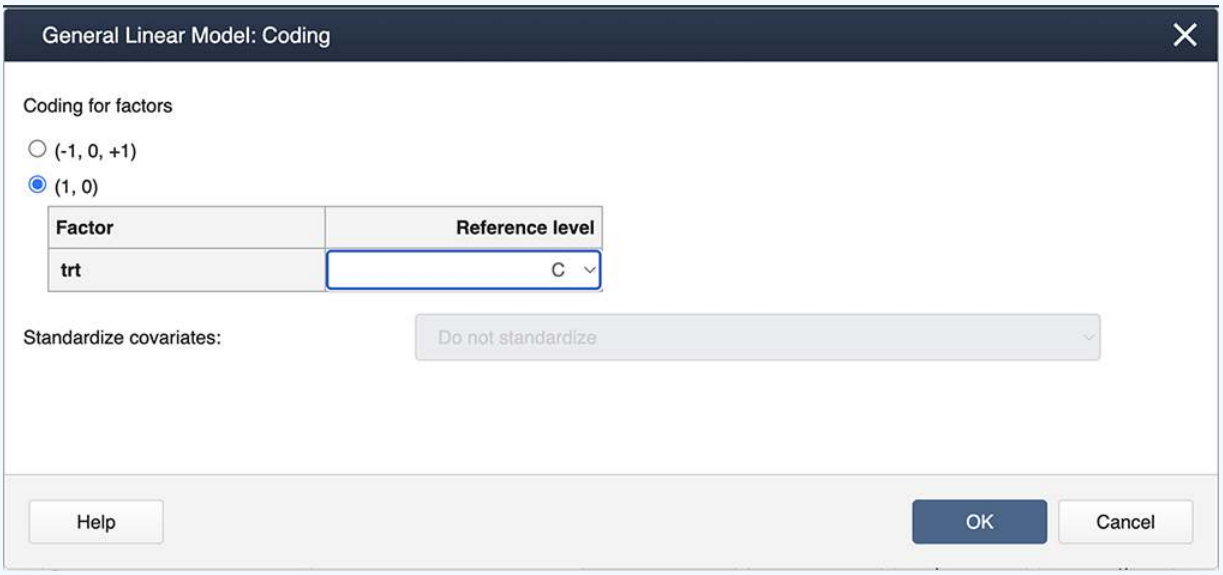


Figure 4.4.a3: Selecting options in the General Linear Model: Coding window.

Select **OK** to exit the nested windows. This produces the regular ANOVA output:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
trt	2	16.000	8.0000	16.00	0.025
Error	3	1.500	0.5000		
Total	5	17.500			

And also the Regression Equation:

Regression Equation

y = 5.500 - 4.000 trt_level1 - 2.000 trt_level2 + 0.0 trt_level3

SAS Example

Steps in SAS

In SAS, the default coding is indicator coding, so when you specify the option

```
model y=trt / solution;
```

Copy code

you get the regression coefficients:

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		5.5000	0.5000	3	11.00	0.0016
trt	**level1**	-4.0000	0.7071	3	-5.66	0.0109
trt	**level2**	-2.0000	0.7071	3	-2.83	0.0663
trt	**level3**	0				

And the same ANOVA table:

Type 3 Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
trt	2	16.000000	8.000000	Var(Residual)+Q(trt)	MS(Residual)	3	16.00	0.0251
Residual	3	1.500000	0.500000	Var(Residual)				

The Intermediate calculations for this model are:

xprimex			check		xprimey
6	2	2	1	-2.22E-16	0 21
2	2	0	3.331E-16	1	0 3
2	0	2	0	0	1 7

SumY2	CF
89.5	73.5

xprimexinv		
0.5	-0.5	-0.5
-0.5	1	0.5
-0.5	0.5	1

R Example

Steps in R

1. Define response variable and design matrix

```
y<-matrix(c(2,1,3,4,6,5), ncol=1)
x = matrix(c(1,1,0,1,1,0,1,0,1,1,0,1,1,0,0,1,0,0),ncol=3,nrow=6,byrow=TRUE)
```

2. Regression coefficients

```
beta<-solve(t(x)%*%x)%*%(t(x)%*%y)
#  beta
#      [,1]
# [1,]  5.5
# [2,] -4.0
# [3,] -2.0
```

3. Calculate the entries of the ANOVA Table

```
n<-nrow(y)
p<-ncol(x)
J<-matrix(1,n,n)
ss_tot = (t(y)%*%y) - (1/n)*(t(y)%*%J)%*%y #17.5
ss_trt = t(beta)%*%(t(x)%*%y) - (1/n)*(t(y)%*%J)%*%y #16
ss_error = ss_tot - ss_trt #1.5
total_df=n-1 #5
trt_df=p-1 #2
error_df=n-p #3
MS_trt = ss_trt/(p-1) #8
MS_error = ss_error / error_df #0.5
F=MS_trt/MS_error #16
```

4. Creating the ANOVA table

```
ANOVA <- data.frame(
c ("","Treatment","Error", "Total"),
c("DF", trt_df,error_df,total_df),
c("SS", ss_trt, ss_error, ss_tot),
c("MS", MS_trt, MS_error, ""),
c("F",F,"",""),
stringsAsFactors = FALSE)
names(ANOVA) <- c(" ", "  ", " ","","")
```

5. Print the ANOVA table

```
print(ANOVA)
# 1           DF   SS  MS  F
# 2 Treatment  2   16   8 16
```

```
# 3     Error  3  1.5 0.5
# 4     Total  5 17.5
```

6. Intermediates in the matrix computations

```
xprimex<-t(x)%*%x
#  xprimex
#      [,1] [,2] [,3]
# [1,]    6    2    2
# [2,]    2    2    0
# [3,]    2    0    2
xprimey<-t(x)%*%y
#   xprimey
#      [,1]
# [1,]   21
# [2,]    3
# [3,]    7
xprimexinv<-solve(t(x)%*%x)
#  xprimexinv
#      [,1] [,2] [,3]
# [1,]  0.5 -0.5 -0.5
# [2,] -0.5  1.0  0.5
# [3,] -0.5  0.5  1.0
check<-xprimexinv%*%xprimex
#  check
#               [,1]          [,2] [,3]
# [1,]  1.000000e+00  0.000000e+00    0
# [2,] -1.110223e-16  1.000000e+00    0
# [3,]  0.000000e+00 -1.110223e-16    1
SumY2<-t(beta)%*%(t(x)%*%y) # 89.5
CF<-(1/n)*(t(y)%*%J)%*%y # 73.5
```

7. Regression Equation and ANOVA table

```
trt_level1<-x[,2]
trt_level2<-x[,3]
model<-lm(y~trt_level1+trt_level2)
```

8. With the command summary(model) we can get the following output:

```
Call:
lm(formula = y ~ trt_level1 + trt_level2)
Residuals:
1    2    3    4    5    6
0.5 -0.5 -0.5  0.5  0.5 -0.5
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.5000     0.5000  11.000  0.00161 **
```

```
trt_level1   -4.0000     0.7071  -5.657  0.01094 *
trt_level2   -2.0000     0.7071  -2.828  0.06628 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.7071 on 3 degrees of freedom
Multiple R-squared:  0.9143,    Adjusted R-squared:  0.8571
F-statistic:    16 on 2 and 3 DF,  p-value: 0.02509
```

From the output, we can see the estimates for the coefficients are b0=5.5, b1=-4, b2=-2 and the F-statistic is 16 with a p-value of 0.02509.

By using the estimates we can write the regression equation:

> y=5.5-4 trt_level1-2 trt_level2+0 trt_level3

9. With the command anova(model) we can get the following output

```
Analysis of Variance Table
Response: y
Df Sum Sq Mean Sq F value  Pr(>F)
trt_level1  1   12.0    12.0      24 0.01628 *
trt_level2  1    4.0     4.0       8 0.06628 .
Residuals   3    1.5     0.5                 ---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Note: R is giving the sequential sum of squares in the ANOVA table.

This page titled 4.4: Dummy Variable Regression is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

4.5: Computational Aspects of the Effects Model

Model 4 - The Effects Model

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{4.5.1}$$

where τ_i are the the deviations of each factor level mean from the overall mean so that $\sum_{i=1}^{T} \tau_i = 0$.

In the effects model that we discussed in chapter 3, the treatment means were not estimated but instead, the τ_i's, or the deviations of treatment means from the overall mean, were estimated. The model must include the overall mean, which is estimated by the intercept, and hence the design matrix to be inputted for IML is:

```
/* The Effects Model */
x={
1    1    0,
1    1    0,
1    0    1,
1    0    1,
1    -1    -1,
1    -1    -1};
```

Here we have another omission of a treatment level, but for a different reason. In the effects model, we have the constraint $\sum \tau_i = 0$. As a result, coding for one treatment level can be omitted.

Running the IML program with this design matrix yields:

Regression Coefficients	
Beta_0	3.5
Beta_1	-2
Beta_2	0

ANOVA				
Treatment	dF	SS	MS	F
	2	16	8	16
Error	3	1.5	0.5	
Total	5	17.5		

The regression coefficient Beta_0 is the overall mean and the coefficients β_1 and β_2 are τ_1 and τ_2, respectively. The estimate for τ_3 is obtained as $-(\tau_1) - (\tau_2) = 2.0$.

In Minitab, if we change the coding now to be Effect coding (-1,0,+1), which is the default setting, we get the following:

Regression Equation

y = 3.500 - 2.000 trt_A - 0.000 trt_B + 2.000 trt_C

The ANOVA table is the same as for the dummy-variable regression model above. We can also observe that the factor level means and General Linear F Statistics values obtained for all 3 representations (cell means, dummy coded regression and effects coded regression) are identical, confirming that the 3 representations are identical.

The intermediates were:

xprimex			check			xprimey
6	0	0	1	0	0	21
0	4	2	0	1	0	-8
0	2	4	0	0	1	-4

xprimexinv		
0.1666667	0	0
0	0.3333333	-0.166667
0	-0.166667	0.3333333

SumY2	CF
89.5	73.5

By coding treatment or factor levels into numerical terms, we can use regression methods to perform the ANOVA.

To state the effects model

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{4.5.2}$$

as a regression model, we need to include $\mu, \tau_i, \ldots, \tau_T$ as elements in the parameter vector $\boldsymbol{\beta}$ of the GLM model. Note that, in the case of equal replication at each factor level, the deviations satisfy the following constraint:

$$\sum_{i=1}^{T} \tau_i = 0 \tag{4.5.3}$$

This implies one of the τ_i parameters is not needed since it can be expressed in terms of the other $T-1$ parameters and need not be included in the $\boldsymbol{\beta}$ parameter vector. We shall drop the parameter τ_T from the regression equation, as it can be expressed in terms of the other $T-1$ parameters τ_i as follows:

$$\tau_T = -\tau_1 - \tau_2 - \ldots - \tau_{T-1} \tag{4.5.4}$$

Thus, the $\boldsymbol{\beta}$ vector of the GLM is a $T \times 1$ vector containing only the parameters $\mu, \tau_1, \ldots, \tau_{T-1}$ for the linear model.

To illustrate how a linear model is developed with this approach, consider a single-factor study with $T = 3$ factor levels when $n_1 = n_2 = n_3 = 2$. The $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ matrices for this case are as follows:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \ \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix} \tag{4.5.5}$$

where β_0, β_1, and β_2 correspond to μ, τ_1, and τ_2 respectively.

Note that the vector of expected values $\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$ yields the following:

$$\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\beta \tag{4.5.6}$$

$$\begin{bmatrix} E\{Y_{11}\} \\ E\{Y_{12}\} \\ E\{Y_{21}\} \\ E\{Y_{22}\} \\ E\{Y_{31}\} \\ E\{Y_{32}\} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \tag{4.5.7}$$

$$= \begin{bmatrix} \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu-\tau_1-\tau_2 \\ \mu-\tau_1-\tau_2 \end{bmatrix} \tag{4.5.8}$$

Since $\tau_3 = -\tau_1 - \tau_2$, as shown above, we see that $E\{Y_{31}\} = E\{Y_{32}\} = \mu + \tau_3$. Thus, the above $\mathbf{X}$ matrix and β vector representation provides the appropriate expected values for all factor levels as expressed below:

$$E\{Y_{ij}\} = \mu + \tau_i \tag{4.5.9}$$

? Using R: Effects Model

Steps in R

1. Define response variable and design matrix

```
y<-matrix(c(2,1,3,4,6,5), ncol=1)
x = matrix(c(1,1,0,1,1,0,1,0,1,1,0,1,1,-1,-1,1,-1,-1),ncol=3,nrow=6,byrow=TRUE)
```

2. Regression coefficients

```
beta<-solve(t(x)%*%x)%*%(t(x)%*%y)
#  beta
#     [,1]
# [1,]  3.5
# [2,] -2.0
# [3,]  0.0
```

3. Calculate the entries of the ANOVA Table

```
n<-nrow(y)
p<-ncol(x)
J<-matrix(1,n,n)
ss_tot = (t(y)%*%y) - (1/n)*(t(y)%*%J)%*%y #17.5
ss_trt = t(beta)%*%(t(x)%*%y) - (1/n)*(t(y)%*%J)%*%y #16
ss_error = ss_tot - ss_trt #1.5
total_df=n-1 #5
trt_df=p-1 #2
error_df=n-p #3
MS_trt = ss_trt/(p-1) #8
MS_error = ss_error / error_df #0.5
F=MS_trt/MS_error #16
```

4. Creating the ANOVA table

```
ANOVA <- data.frame(
c ("","Treatment","Error", "Total"),
c("DF", trt_df,error_df,total_df),
c("SS", ss_trt, ss_error, ss_tot),
c("MS", MS_trt, MS_error, ""),
c("F",F,"",""),
stringsAsFactors = FALSE)
names(ANOVA) <- c(" ", "  ", " ","","")
```

5. Print the ANOVA table

```
print(ANOVA)
# 1          DF   SS  MS  F
# 2 Treatment  2   16   8 16
# 3     Error  3  1.5 0.5
# 4     Total  5 17.5
```

Copy code

6. Intermediates in the matrix computations

```
xprimex<-t(x)%*%x
#  xprimex
#      [,1] [,2] [,3]
# [1,]    6    0    0
# [2,]    0    4    2
# [3,]    0    2    4
xprimey<-t(x)%*%y
#  xprimey
#      [,1]
# [1,]   21
# [2,]   -8
# [3,]   -4
xprimexinv<-solve(t(x)%*%x)
#  xprimexinv
#           [,1]       [,2]       [,3]
# [1,] 0.1666667  0.0000000  0.0000000
# [2,] 0.0000000  0.3333333 -0.1666667
# [3,] 0.0000000 -0.1666667  0.3333333
check<-xprimexinv%*%xprimex
# check
#      [,1] [,2] [,3]
# [1,]    1    0    0
# [2,]    0    1    0
# [3,]    0    0    1
```

```
SumY2<-t(beta)%*%(t(x)%*%y) #89.5
CF<-(1/n)*(t(y)%*%J)%*%y # 73.5
```

7. Regression Equation and ANOVA table

```
trt_level1<-x[,2]
trt_level2<-x[,3]
model<-lm(y~trt_level1+trt_level2)
```

8. With the command summary(model) we can get the following output:

```
Call:
lm(formula = y ~ trt_level1 + trt_level2)
Residuals:
1    2    3    4    5    6
0.5 -0.5 -0.5  0.5  0.5 -0.5
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.500e+00  2.887e-01  12.124  0.00121 **
trt_level1  -2.000e+00  4.082e-01  -4.899  0.01628 *
trt_level2  -1.282e-16  4.082e-01   0.000  1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.7071 on 3 degrees of freedom
Multiple R-squared:  0.9143,    Adjusted R-squared:  0.8571
F-statistic:    16 on 2 and 3 DF,  p-value: 0.02509
```

From the output we can see the estimates for the coefficients are b0=3.5, b1=-2, b2=0 and the F-statistic is 16 with a p-value of 0.02509.

By using the estimates we can write the regression equation:

> y=3.5-2 trt_level1-0 trt_level2+2 trt_level3

The estimator τ_3 is obtained as $-\tau_1 - \tau_2 = 2$

9. With the command anova(model) we can get the following output:

```
Analysis of Variance Table
Response: y
Df Sum Sq Mean Sq F value  Pr(>F)
trt_level1  1   16.0    16.0      32 0.01094 *
trt_level2  1    0.0     0.0       0 1.00000
Residuals   3    1.5     0.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Note that R is giving the sequential sum of squares in the ANOVA table.

This page titled 4.5: Computational Aspects of the Effects Model is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

4.6: The Study Diagram

In Section 1.1 we encountered a brief description of an experiment. The description of an experiment provides a *context* for understanding how to build an appropriate statistical model. All too often, mistakes are made in statistical analyses because of a lack of understanding of the setting and procedures in which a designed experiment is conducted. Creating a study diagram is one of the best ways to address this, in addition to being intuitive. A study diagram is a schematic diagram that captures the essential features of the experimental design. Here, as we explore the computations for a single factor ANOVA in a simple experimental setting, the study diagram may seem trivial. However, in practice and in lessons to follow in this course, the ability to create accurate study diagrams usually makes a substantial difference in getting the model right.

In our example, as described in Section 1.1, a plant biologist thinks that plant height may be affected by the fertilizer type and three types of fertilizer were chosen to investigate this claim. Next, 24 plants were randomly chosen and 4 batches, with 6 plants in each, were assigned individually to the 3 fertilizer types; the last batch was left untreated, constituting the Control group. The researchers kept all the plants under controlled conditions in the greenhouse. The individual containerized plants were randomly assigned the fertilizer treatment levels to produce 6 replications of each of the fertilizer applications.

Here is the data from the example that we were using in this lesson:

Control	F1	F2	F3
21	32	22.5	28
19.5	30.5	26	27.5
22.5	25	28	31
21.5	27.5	27	29.5
20.5	28	26.5	30
21	28.6	25.2	29.2

So we have a description of the treatment levels and how they were assigned to individual experimental units (the potted plant), and we see the data organized in a table. But what are we missing? A key question is: how was the experiment conducted? This question is a practical one and is answered with a study diagram. These are usually hand-drawn depictions of a real setting, indicating the treatments, levels of treatments, and how the experiment was laid out. They are not typically works of art and no one should ever feel embarrassed by a lack of artistic ability to draw one. For this example, we need to draw a greenhouse bench, capable of holding the $4 \times 6 = 24$ experimental units:

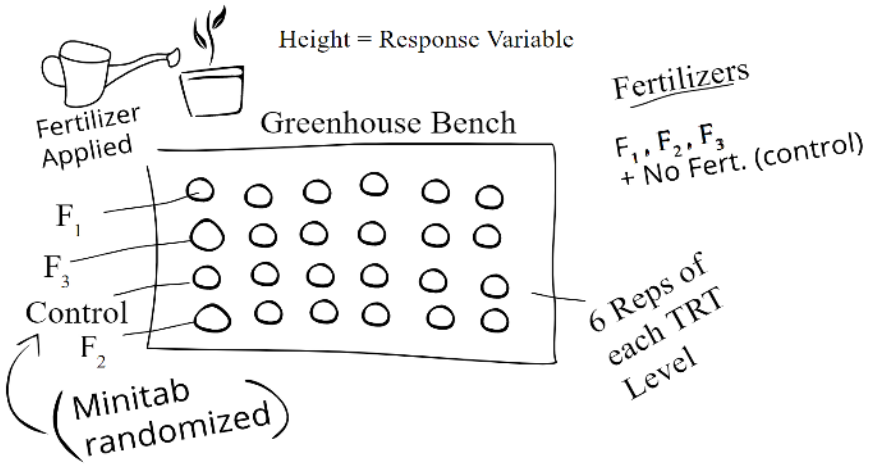


Figure 4.6.1: Study diagram for response variable of height, showing 4 treatment levels with 6 units in each.

The diagram identified the response variable, listed the treatment levels, and indicated the random assignment of treatment levels to these 24 experimental units on the greenhouse bench.

This randomization and the subsequent experimental layout we would identify as a Completely Randomized Design (CRD). We know from this schematic diagram that we need a statistical model that is appropriate for a one-way ANOVA in a Completely Randomized Design (CRD).

Furthermore, once the plant heights are recorded at the end of the study, the experimenter may observe that the variability in the growth may possibly be influenced by a second factor besides the fertilizer level. A careful examination of the layout of the plants in the study diagram may perhaps reveal this additional factor. For example, if the growth is higher in the plants placed on the row nearer to the windows, it is reasonable to assume that sunlight also plays a role and to redesign the experiment as a randomized completely block design (RCBD) with rows as a blocking factor. Note that design aspects of experiments are covered in Chapters 7 and 8.

Being able to draw and reproduce a study diagram is very useful in identifying the components of the ANOVA models.

This page titled 4.6: The Study Diagram is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

4.7: Try It!

Exercise 4.7.1: Design Matrix

Below is a design matrix for a data set of a recent study.

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

a) Identify the number of treatment levels and replicates.

Solution

4 treatment levels and 3 replicates

b) Name the model and write its equation.

Solution

This design matrix corresponds to the effects model, and the model equation is $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$, where $i = 1, 2, 3, 4$, $j = 1, 2, 3$, and $\sum_{i=1}^{4} \tau_i = 0$.

c) Write the equation and the design matrix that corresponds to the cell means model.

Solution

The equation for the cell means model is: $Y_{ij} = \mu + \epsilon_{ij}$, where $i = 1, 2, 3, 4$ and $j = 1, 2, 3$. The design matrix corresponding to the cell means model is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

d) Write the equation and the design matrix that corresponds to the dummy variable regressions model.

Solution

The equation for the 'dummy variable regression' model is: $Y_{ij} = \mu + \mu_i + \epsilon_{ij}$ for $i = 1, 2, 3$ and $j = 1, 2, 3$. $Y_{4j} = \mu + \epsilon_{4j}$

The design matrix is given below. Note that the last 3 rows correspond to the 4th treatment level which is the reference category and its effect is estimated by the model intercept.

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

This page titled 4.7: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

4.8: Chapter 4 Summary

This chapter, together with Chapter 3, covered four different versions of single-factor ANOVA models. They are: Overall Mean, Cell Means, Dummy Variable Regression, and Effects Coded Regression models. This lesson also provided the coding compatible with the SAS IML procedure, which facilitates the ANOVA computations using Matrix Algebra in a GLM setting. The method of least squares was used to estimate model parameters yielding a prediction equation for the response in terms of the treatment level. This prediction tool will show to be more useful in ANCOVA settings where model predictors are both categorical and numerical (more details on ANCOVA in Chapters 9 and 10). The prediction process can be utilized effectively only with a sound knowledge of the parameterization process for each ANOVA model, which we have been able to acquire as the design matrix was an input resource for running the IML code and the knowledge of the parameter vector was useful in interpreting the prediction (regression) equations.

Finally, using the greenhouse example, the concept of a study diagram was discussed. Though a simple visual tool, a study diagram may play an important role in identifying new predictors so that perhaps a pre-determined ANOVA model can be extended to include additional factors to create a multi-factor model discussed in Chapters 5 and 6. In addition to identifying the treatment design, the study diagram also helps in choosing an appropriate randomization design, a topic discussed in Chapters 7 and 8.

This page titled 4.8: Chapter 4 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

CHAPTER OVERVIEW

5: Multi-Factor ANOVA

Learning Objectives

Upon completion of this lesson, you should be able to:

1. Identify factorial, nested, and cross-nested treatment designs.
2. Use main effects and interaction effects in factorial designs.
3. Create nested designs and identify the nesting effects.
4. Use statistical software to analyze data from different treatment designs via ANOVA and mean comparison procedures.

Researchers often identify more than one experimental factor of interest. One alternative is to set up separate, independent experiments in which a single treatment (or factor) is used in each experiment, and data from each experiment to be analyzed as we have done using a one-way ANOVA. This approach might have the advantage of a concentrated focus on the single treatment of interest and the simplicity of computations. However, there are several disadvantages as well.

- First, environmental factors or experimental material conditions may change during the process. This could distort the assessment of the relative importance of different treatments on the response variable.
- Second, it is inefficient. Setting up and running multiple separate experiments usually will involve more work and resources.
- Last, and probably the most important, this one-at-a-time approach does not allow the examination of how several treatments jointly impact the response.

ANOVA methodology can be extended to accommodate this multi-factor setting. Here are Dr. Rosenberger and Dr. Shumway talking about some of the things to look out for as you work your way through this lesson.

Video 5.1: Experimental design drives analysis.

To put it into perspective, let’s take a look at the phrase "Experimental Design" a term that you often hear. We are going to take this colloquial phrase and divide it into two formal components:

A. The Treatment Design
B. The Randomization Design

We will use the treatment design component to address the nature of the experimental factors under study and the randomization design component to address how treatments are assigned to experimental units. An experimental unit is defined to be that which

receives a specific treatment level and in a multi-factor setting, a specific treatment or factor combination. In the single-factor greenhouse example, which is an experiment, the experimental unit is a single plant receiving one specific fertilizer level. Note that the ANOVA model pertaining to a given study depends on both the treatment design and the randomization process.

The following figure illustrates the conceptual division between the treatment design and the randomization design. The terms that are in boldface type will be addressed in detail in this or future lessons.

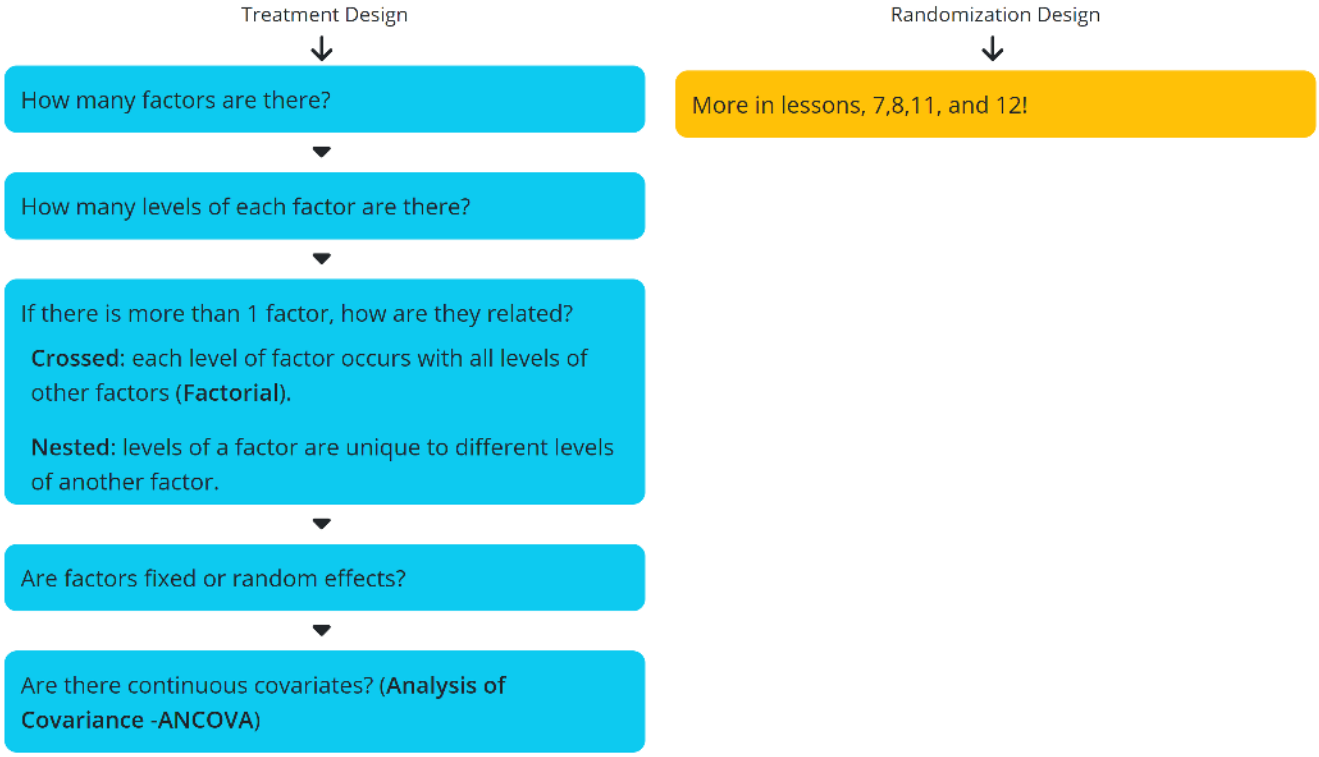


Figure 5.1: Concepts to consider in experimental design, for treatment design and randomization design.

This page titled 5: Multi-Factor ANOVA is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Factorial or Crossed Treatment Designs

In multi-factor experiments, combinations of factor levels are applied to experimental units. The single-factor greenhouse experiment discussed in previous lessons can be extended to a multi-factor study by including plant species as an additional factor along with fertilizer type. This addition of another factor may prove to be useful, as one fertilizer type may be most effective on one specific plant species! In other words, the optimal height growth is perhaps attainable by a unique combination of fertilizer type and plant species. A treatment design that provides the opportunity to determine this best combination is a factorial design, where responses are observed at each level of a given factor combined with each level of all other factors. In this setting, factors are said to be crossed.

A factorial design with t factors is identified using the $l_1 l_2 \ldots l_t$ notation, where l_i is the number of levels of factor i $(i = 1, 2, \ldots, t)$. For example, a factorial design with 2 factors A and B, where A has 4 levels and B has 3 levels, will have the 4×3 notation.

One complete replication of a factorial design with t factors requires $(l_1 \times l_2 \times \ldots \times l_t)$ experimental units, and this quantity is called the replicate size. If r is the number of complete replicates, then N, the total number of observations, equals $r \times (l_1 \times l_2 \times \ldots \times l_t)$.

It is easy to see that with the addition of more and more crossed factors, the replicate size will increase rapidly and design modifications have to be made to make the experiment more manageable.

In a factorial experiment, as combinations of different factor levels play an important role, it is important to differentiate between the lone (or main) effects of a factor on the response and the combined effects of a group of factors on the response.

The main effect of factor A is the effect of A on the response ignoring the effect of all other factors. The main effect of a given factor is equivalent to the factor effect associated with the single-factor experiment using only that particular factor.

The combined effect of a specific combination of l different factors is called the interaction effect (more details later). The interaction effect of most interest is the two-way interaction effect and is denoted by the product of the two letters assigned to the two factors. For example, the two-way interaction effects of a factorial design with 3 factors A, B, C are denoted AB, AC, and BC. Likewise, the three-way interaction effect of these 3 factors is denoted by ABC.

Let us now examine how the degrees of freedom (df) values of a single-factor ANOVA can be extended to the ANOVA of a two-factor factorial design. Note that the interaction effects are additional terms that need to be included in a multi-factor ANOVA, but the ANOVA rules studied in Chapter 2 for single-factor situations still apply for the main effect of each factor. If the two factors of the design are denoted by A and B with a and b as their number of levels respectively, then the df values of the two main effects are $(a-1)$ and $(b-1)$. The df value for the two-way interaction effect is $(a-1)(b-1)$, the product of df values for A and B. The ANOVA table below gives the layout of the df values for a 2×2 factorial design with 5 complete replications. Note that in this experiment, r equals 5, and N is equal to 20.

Source	d.f.
Factor A	$(a-1) = 1$
Factor B	$(b-1) = 1$
Factor A × Factor B	$(a-1)(b-1) = 1$
Error	$19 - 3 = 16$
Total	$N - 1 = (nab) - 1 = 19$

If in the single-factor model of

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{5.1.1}$$

τ_i is effectively replaced with $\alpha_i + \beta_i + (\alpha\beta)_{ij}$, then the resulting equation shown below will represent the model equation of a two-factor factorial design.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{5.1.2}$$

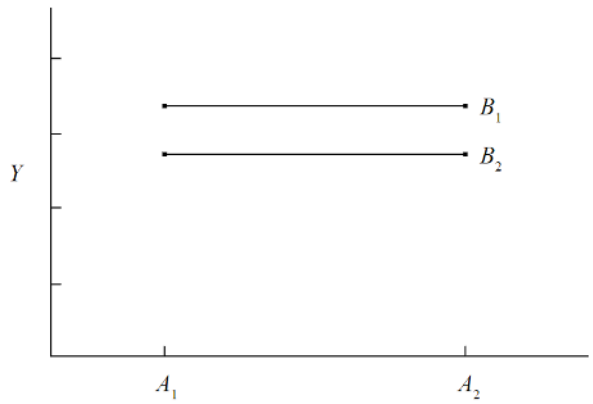
where α_i is the main effect of factor A, β_j is the main effect of factor B, and $(\alpha\beta)_{ij}$ is the interaction effect $(i = 1, 2, \ldots, a,\ j = 1, 2, \ldots, b,\ k = 1, 2, \ldots, r)$.

This reflects the following partitioning of treatment deviations from the grand mean:

$$\underbrace{\bar{Y}_{ij.} - \bar{Y}_{...}}_{\substack{\text{Deviation of estimated treatment mean} \\ \text{around overall mean}}} = \underbrace{\bar{Y}_{i..} - \bar{Y}_{...}}_{A \text{ main effect}} + \underbrace{\bar{Y}_{.j.} - \bar{Y}_{...}}_{B \text{ main effect}} + \underbrace{\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}}_{AB \text{ interaction effect}} \tag{5.1.3}$$

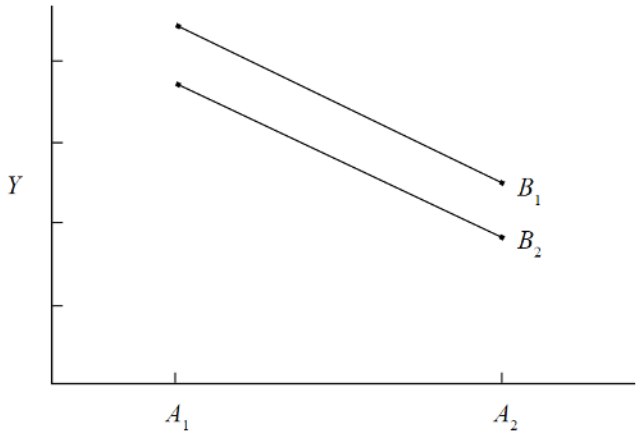
The main effects for Factor A and Factor B are straightforward to interpret, but what is an interaction? Delving more, an interaction can be defined as the failure of the response to one factor to be the same at different levels of another factor. Notice that $(\alpha\beta)_{ij}$, the interaction term in the model, is multiplicative, and as a result may have a large and important impact on the response variable. Interactions go by different names in various fields. In medicine, for example, physicians most times ask what medication you are on before prescribing a new medication. They do this out of a concern for interaction effects of either interference (a canceling effect) or synergism (a compounding effect).

Graphically, in a two-factor factorial with each factor having 2 levels, the interaction can be represented by two non-parallel lines connecting means (adapted from Zar, H. *Biostatistical Analysis*, 5th Ed., 1999). It is because the interaction reflects the failure of the difference in response between the two different levels of one factor to be the same, for both levels of the other factor. So, if there is no interaction, then this difference in response will be the same, which will graphically result in two parallel lines. In the interaction plots below, parallel lines are a consistent feature in all settings with no interaction. In plots depicting interaction, the lines do cross (or would cross if the lines kept going).



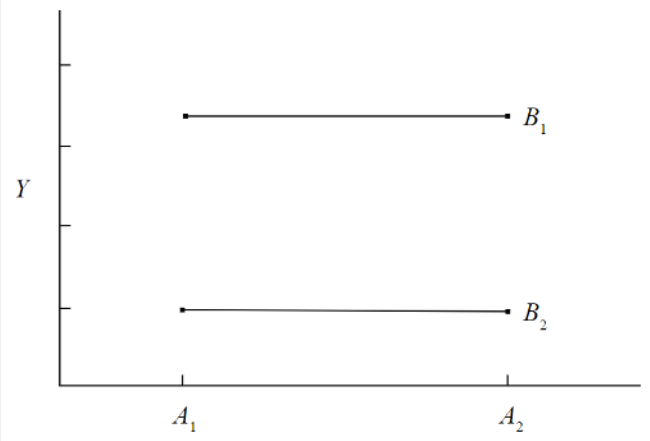
Graph 1

In graph 1 there is no effect of Factor A, a small effect of Factor B (and if there were no effect of Factor B the two lines would coincide), and no interaction between Factor A and Factor B.



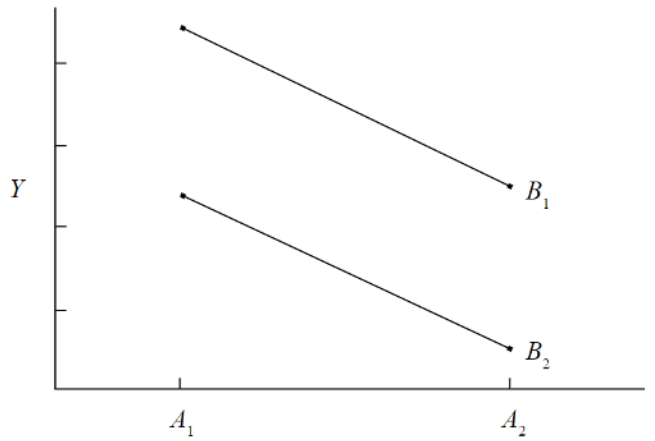
Graph 2

Graph 2 shows a large effect of Factor A, small effect of Factor B, and no interaction.



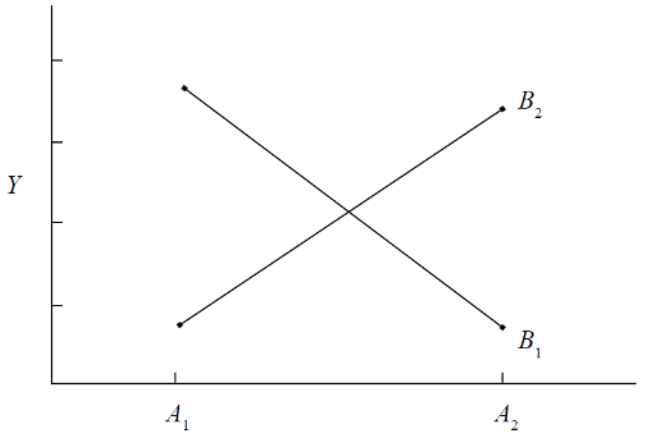
Graph 3

Graph 3 shows no effect of Factor A, larger effect of Factor B, and no interaction.



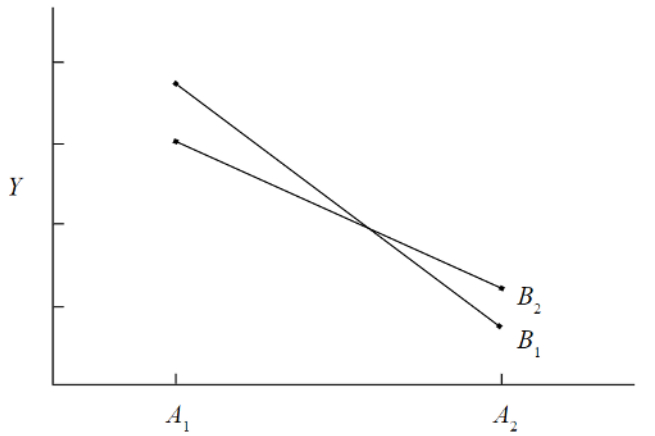
Graph 4

In graph 4 there is a large effect of Factor A, a large effect of Factor B , and no interaction.



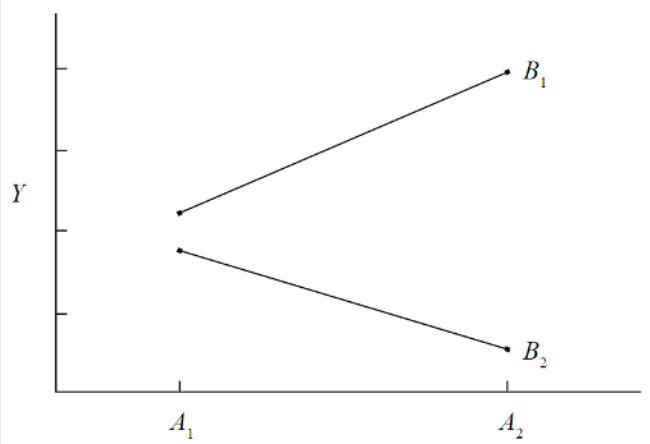
Graph 5

In graph 5 there is no effect of Factor A and no effect of Factor B, but an interaction between A and B.



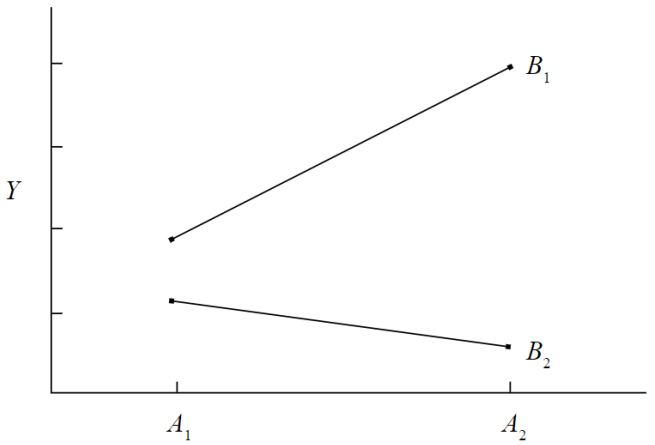
Graph 6

In graph 6 there is a large effect of Factor A and no effect of Factor B, with a slight interaction between A and B.



Graph 7

In graph 7 there is no effect of Factor A and a large effect of Factor B, with a very large interaction.



Graph 8

In graph 8 there is a small effect of Factor A and a large effect of Factor B, with a large interaction.

In the presence of multiple factors with their interactions, multiple hypotheses can be tested and for a two-factor factorial design. They are:

Main Effect of Factor A:

$$\begin{aligned} H_0 &: \ \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0 \\ H_A &: \ \text{not all } \alpha_i \text{ are equal to } 0 \end{aligned} \tag{5.1.4}$$

Main Effect of Factor B:

$$\begin{aligned} H_0 &: \ \beta_1 = \beta_2 = \ldots = \beta_b = 0 \\ H_A &: \ \text{not all } \beta_j \text{ are equal to } 0 \end{aligned} \tag{5.1.5}$$

A × B Interaction:

$$\begin{aligned} H_0 &\ \text{there is no interaction} \\ H_A &: \ \text{an interaction exists} \end{aligned} \tag{5.1.6}$$

When testing these hypotheses, it is important to test for the significance of the interaction effect first. If the interaction is significant, the main effects are of no consequence; rather, the differences among different factor level combinations should be looked into. The greenhouse example, extended to include a second (crossed) factor, will illustrate the steps.

This page titled 5.1: Factorial or Crossed Treatment Designs is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.1.1: Two-Factor Factorial - Greenhouse Example (SAS)

Let's return to the greenhouse example with *plant species* also as a predictive factor, in addition to *fertilizer type*. The study then becomes a 2×4 factorial as 2 types of plant species and 4 types of fertilizers are investigated. The total number of experimental units (plants) that are needed now is 48, as r=6 and there are 8 plant species and fertilizer type combinations.

The data might look like this:

		Fertilizer Treatment			
		Control	F1	F2	F3
Species	A	21.0	32.0	22.5	28.0
		19.5	30.5	26.0	27.5
		22.5	25.0	28.0	31.0
		21.5	27.5	27.0	29.5
		20.5	28.0	26.5	30.0
		21.0	28.6	25.2	29.2
	B	23.7	30.1	30.6	36.1
		23.8	28.9	31.1	36.6
		23.7	34.4	34.9	37.1
		22.8	32.7	30.1	36.8
		22.8	32.7	30.1	36.8
		24.4	32.7	25.5	37.1

The ANOVA table would now be constructed as follows:

Source	df	SS	MS	F
Fertilizer	$(4-1)=3$			
Species	$(2-1)=1$			
Fertilizer × Species	$(2-1)(4-1)=3$			
Error	$47-7=40$			
Total	$N-1=47$			

The data presented in the table above are in unstacked format. One needs to convert this into a stacked format when attempting to use statistical software. The SAS code is as follows.

The data presented in the table above are in unstacked format. One needs to convert this into a stacked format when attempting to use statistical software. The SAS code is as follows.

```
data greenhouse_2way;
input fert $ species $ height;
datalines;
```

```
control SppA     21.0
control SppA     19.5
control SppA     22.5
control SppA     21.5
control SppA     20.5
control SppA     21.0
control SppB     23.7
control SppB     23.8
control SppB     23.8
control SppB     23.7
control SppB     22.8
control SppB     24.4
f1      SppA     32.0
f1      SppA     30.5
f1      SppA     25.0
f1      SppA     27.5
f1      SppA     28.0
f1      SppA     28.6
f1      SppB     30.1
f1      SppB     28.9
f1      SppB     30.9
f1      SppB     34.4
f1      SppB     32.7
f1      SppB     32.7
f2      SppA     22.5
f2      SppA     26.0
f2      SppA     28.0
f2      SppA     27.0
f2      SppA     26.5
f2      SppA     25.2
f2      SppB     30.6
f2      SppB     31.1
f2      SppB     28.1
f2      SppB     34.9
f2      SppB     30.1
f2      SppB     25.5
f3      SppA     28.0
f3      SppA     27.5
f3      SppA     31.0
f3      SppA     29.5
f3      SppA     30.0
f3      SppA     29.2
f3      SppB     36.1
f3      SppB     36.6
f3      SppB     38.7
f3      SppB     37.1
f3      SppB     36.8
```

```
f3      SppB     37.1
;
run;
/*The code to generate the boxplot
for distribution of height by species organized by fertilizer
in Figure 5.1*/

proc sort data=greenhouse_2way; by fert species;
proc boxplot data=greenhouse_2way;
plot height*species (fert);
run;
```

As a preliminary step in Exploratory Data Analysis (EDA), a side-by-side boxplot display of height vs. species organized by fertilizer type would be an ideal graphic. As the plot shows, the height differences between species are variable among fertilizer types (see for example the difference in height between *SppA* and *SppB* for *Control* is much less than that for *F3*). This indicates that *fert*species* could be a significant interaction prompting a factorial model with interaction.

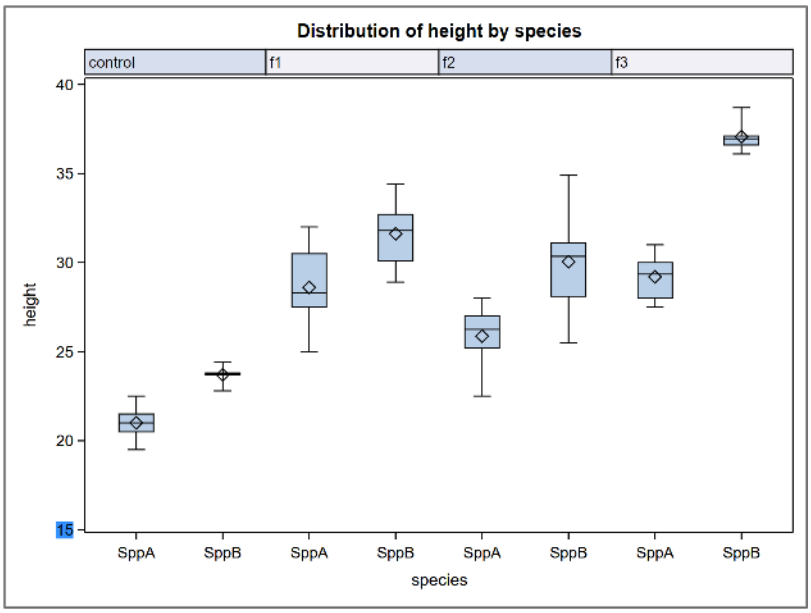


Figure 5.1.1.1: Boxplot for distribution of height by species organized by fertilizer.

To run the two-factor factorial model with interaction in SAS `proc mixed`, we can use:

```
/*Runs the two-factor factorial model with interaction*/
proc mixed data=greenhouse_2way method=type3;
class fert species;
model height = fert species fert*species;
store out2way;
run;
```

In the `proc mixed` procedure, similar to when running the single factor ANOVA. The name of the data set is specified in the `proc mixed` statement and so is the `method=type 3` option that specifies the way the F test is calculated. The `fert` and `species` factors that are both categorical are included in the class statement. The terms (or effects) in the model statement are consistent with the *source effects* in the layout of the "theoretical" ANOVA table illustrated in 5.1. Finally, the `store` command stores the elements necessary for the generation of the LS-Means interval plot.

Recall the two ANOVA rules, applicable to any model: (*a*). the df values add up to total df and (*b*). the sums of squares add up to total sums of squares. As seen by the output below, the df values and also the sums of squares follow these rules. (It is easy to confirm that the total sum of squares = 1168.732500, by the 2nd ANOVA rule.)

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
fert	3	745.437500	248.479167	Var(Residual)+Q(fert,fert*species)	MS(Residual)	40	73.10	<.0001
species	1	236.740833	236.740833	Var(Residual)+Q(species,fert*species)	MS(Residual)	40	69.65	<.0001
fert*species	3	50.584167	16.861389	Var(Residual)+Q(fert*species)	MS(Residual)	40	4.96	0.0051
Residual	40	135.970000	3.399250	Var(Residual)				

Rule

In a model with the interaction effect, the interaction term should be interpreted first. If the interaction effect is significant, then do NOT interpret the main effects individually. Instead, compare the mean response differences among the different factor level combinations.

In general, a significant interaction effect indicates that the impact of the levels of Factor A on the response depends upon the level of Factor B and vice versa. In other words, in the presence of a significant interaction, a stand-alone main effect is of no consequence. In the case where an interaction is not significant, the interaction term can be dropped and a model without the interaction should be run. See Section 5.1.1a: The Additive Model (No Interaction)).

Now applying the above rule for this example, the small p-value of 0.0051 displayed in the table above indicates that the interaction effect is significant, which means that the main effects of either *fert* or *species* should not be considered individually. It is the average response differences among the *fert* and *species* combinations that matter. In order to determine the statistically significant *fert* and *species* combinations, a suitable multiple comparison procedure, such as Tukey and Kramer procedure can be performed on the LS-Means of the interaction effect (i.e.: the treatment combinations).

The necessary follow-up SAS code to perform this procedure is given below.

```
ods graphics on;
proc plm restore=out2way;
lsmeans fert*species / adjust=tukey plot=(diffplot(center) meanplot(cl ascending)) cl
/* Because the 2-factor interaction is significant, we work with
   the means for treatment combination*/
run;
```

SAS Output for the LSmeans:

fert*species Least Squares Means

fert*species Least Squares Means									
fert	species	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
control	SppA	21.0000	0.7527	40	27.90	<.0001	0.05	19.4788	22.5212
control	SppB	32.7000	0.7527	40	31.49	<.0001	0.05	22.1788	25.2212
f1	SppA	28.6000	0.7527	40	38.00	<.0001	0.05	27.0788	30.1212
f1	SppB	31.6167	0.7527	40	42.00	<.0001	0.05	30.0954	33.1379
f2	SppA	25.8667	0.7527	40	34.37	<.0001	0.05	24.3454	27.3879
f2	SppB	30.0500	0.7527	40	39.92	<.0001	0.05	28.5288	31.5712
f3	SppA	29.2000	0.7527	40	38.79	<.0001	0.05	27.6788	30.7212
f3	SppB	37.0667	0.7527	40	49.25	<.0001	0.05	35.5454	38.5879

Note that the *p*-values here (*Pr* > *t*) are testing the hypotheses that the fert and species combination means = 0. This may be of very little interest. However, a comparison of mean response values for different species and fertilizer combinations may prove to be more beneficial and can be derived from the diffogram shown in Figure 5.1.1.2 Again recall that, if the confidence interval does not contain zero, then the difference between the two associated means is statistically significant.

Notice also that we see a single value for the standard error based on the MSE from the ANOVA, rather than a separate standard error for each mean (as we would get from Proc Summary for the sample means). Again in this example, with equal sample sizes and no covariates, the *lsmeans* will be identical to the ordinary means displayed in the Summary Procedure.

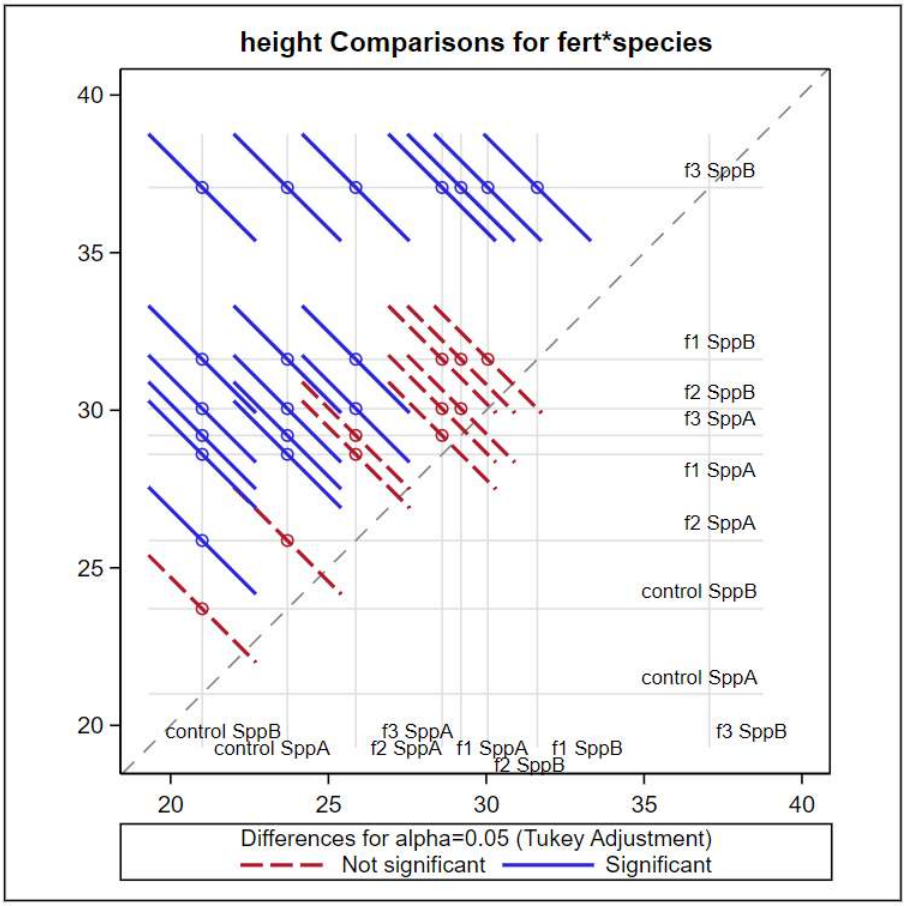


Figure 5.1.1.2: Diffogram for species and fertilizer combinations.

https://stats.libretexts.org/@go/page/33634

There are total of 8 *fert*species* combinations resulting a total of $\binom{8}{2} = 28$ pairwise comparisons. From the diffogram for differences in *fert*species* combinations, we see that 10 of them are not significant and 18 of them are significant at a 5% level after Tukey adjustment (more about diffograms). The information used to generate the diffogram is presented in the table for *differences of fert*species least squares means* in the SAS output (this table is not displayed here).

We can save the *differences* estimated in SAS `proc mixed` and utilize `proc sgplot` to create the plot of differences in mean response for the *fert*species* combinations as shown in Figure 5.1.1.3 The CIs shown are the Tukey adjusted CIs. SAS code to produce Figure 5.1.1.3 is not given in these notes. The interpretations of the plot are similar to what we observed from the diffogram in Figure 5.1.1.2

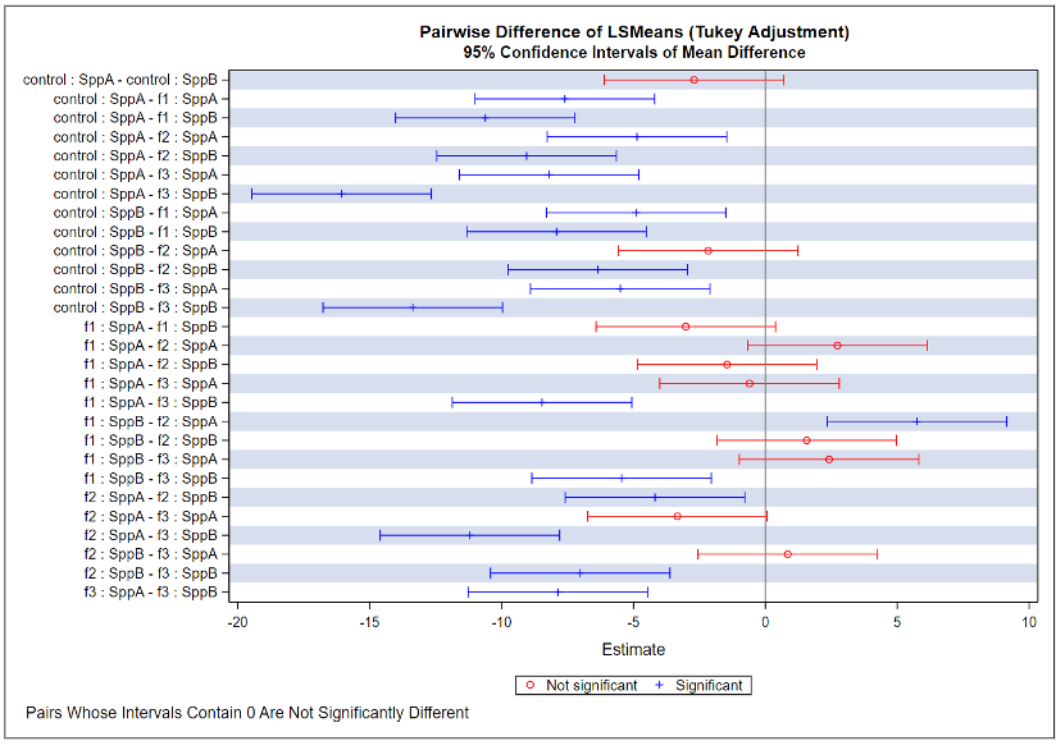


Figure 5.1.1.3: Plot of differences in mean response for the *fert*species* combinations.

In addition to comparing differences in mean responses for the *fert*species* combinations, the SAS code shared above will also produce the line plot for multiple comparisons of means for *fert*species* combinations (shown in Figure 5.1.1.4) and the plot of means responses organized in the ascending order with 95% CIs for *fert*species* combinations (shown in Figure 5.1.1.5).

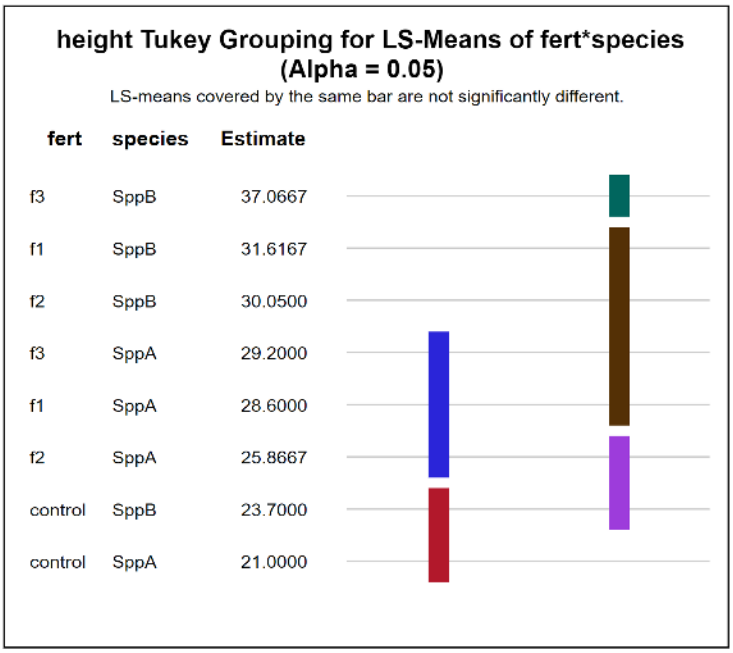


Figure 5.1.1.4: The line plot for multiple comparisons of means for *fert*species* combinations.

The line plot in Figure 5.1.1.4 connects groups in which the LS-means are not statistically different and displays a summary of which groups have *similar* means. The plot of means with 95% CIs in Figure 5.1.1.5 illustrates the same result, although it uses unadjusted CIs. We have organized the plot in the ascending order of estimated means to make it easy to draw conclusions.

Copy and Paste
Image here.
Delete this
placeholder image

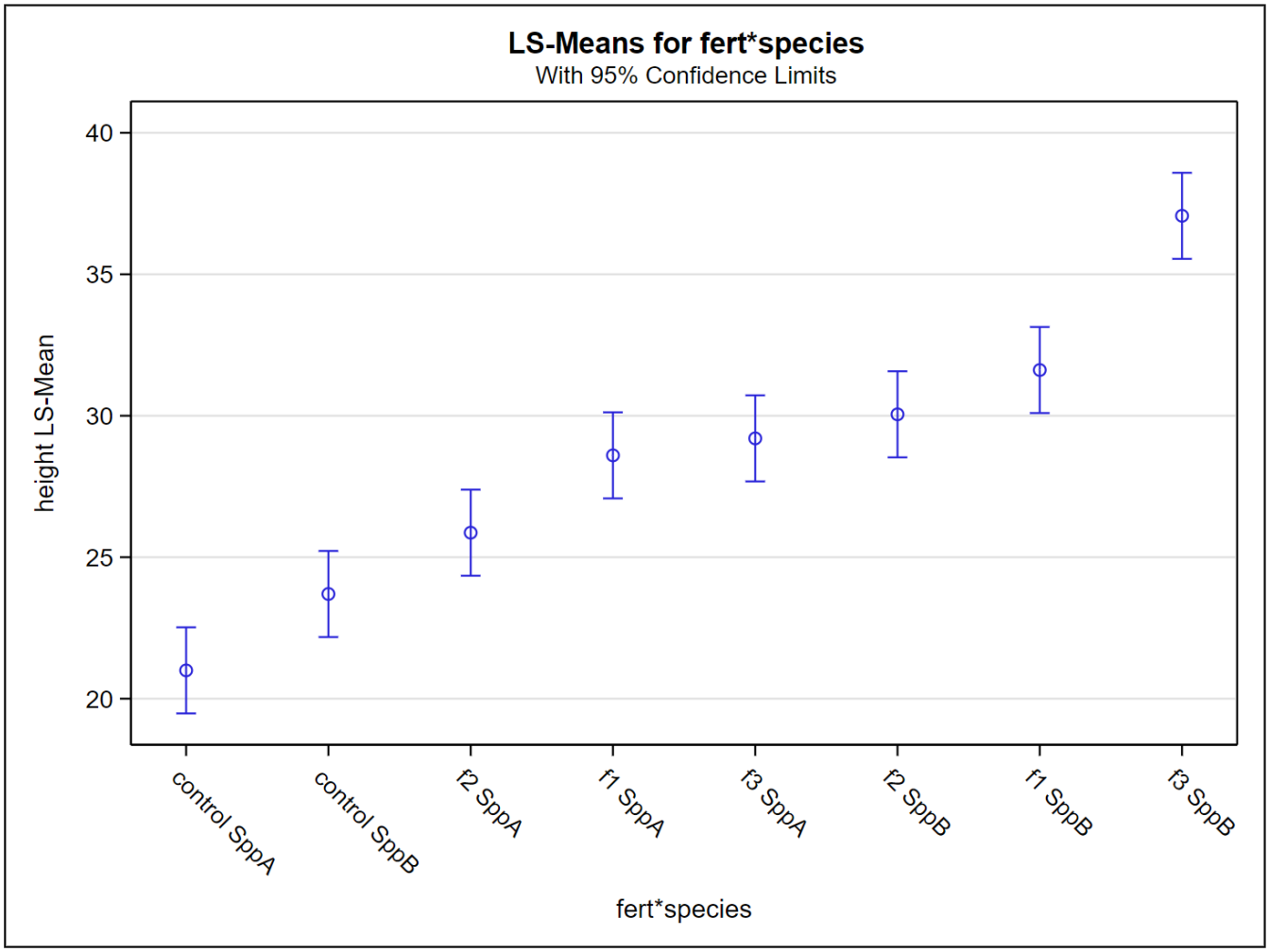


Figure 5.1.1.5: The plot of means with 95% CIs for *fert*species* combinations.

Using LSMEANS, subsequent to performing an ANOVA will help to identify the significantly different treatment level combinations. In other words, the ANOVA doesn't end with a p-value for an F-test. A small p-value signals the need for a mean comparison procedure.

This page titled 5.1.1: Two-Factor Factorial - Greenhouse Example (SAS) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.1.1a: The Additive Model (No Interaction)

In a factorial design, we first look at the interactions for significance. In the case where interaction is not significant, then we can drop the interaction term from our model, and we end up with an additive model.

For a two-factor factorial, the model we initially consider (as we have discussed in Section 5.1) is:

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{5.1.1a.1}$$

Note that the interaction term, $(\alpha\beta)_{ij}$, is a multiplicative term.

If the interaction is found to be non-significant, then the model reduces to:

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + \epsilon_{ijk} \tag{5.1.1a.2}$$

Here we can see that the response variable is simply a function of adding the effects of the two factors.

Example $5.1.1a.1$: Glucose in Blood Serum

As an example, (adapted from Kuehl, 2000), let's look at a study designed to evaluate two chemical methods used for assaying the amount of glucose in blood serum. A large volume of blood serum served as a starting point for the experiment. The blood serum was divided into three portions, each of which was 'doped' or augmented by adding an additional amount of glucose. Three doping levels were used. Samples of the doped serum were then assayed for glucose concentration by one of two chemical methods. This type of ‘doping’ experiment is commonly used to compare the sensitivity of assay methods.

The amount of glucose detected in each sample was recorded and is presented in the table below.

	Chemical Assay Method					
	Method 1			Method 2		
Doping Level	1	2	3	1	2	3
	46.5	138.4	180.9	39.8	132.4	176.8
	47.3	144.4	180.5	40.3	132.4	173.6
	46.9	142.7	183	41.2	130.3	174.9

Solution

The model was run as a two-factor factorial and produced the following results:

Type 3 Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
method	1	263.733889	263.733889	Var(Residual) + Q(method, method*doping)	MS(Residual)	12	98.35	<.0001
doping	2	57026	28513	Var(Residual) + Q(doping, method*doping)	MS(Residual)	12	10632.5	<.0001

Type 3 Analysis of Variance								
method*doping	2	13.821111	6.910556	Var(Residual) + Q(method*doping)	MS(Residual)	12	2.58	0.1172
Residual	12	32.180000	2.681667	Var(Residual)				

Here we can see that the interaction of *method*doping* was not significant (*p-value* > 0.05) at a 5% level. We drop the interaction effect from the model and run the additive model. The resulting ANOVA table is:

The Mixed Procedure								
Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
method	1	263.733889	263.733889	Var(Residual)+Q(method, method)	MS(Residual)	14	80.26	<.0001
doping	2	57026	28513	Var(Residual) + Q(doping,doping)	MS(Residual)	14	8677.63	<.0001
1Residual	14	46.001111	3.285794	Var(Residual)				

The Error SS is now 46.001, which is the sum of the interaction SS and the error SS of the model with the interaction. The df values were also added the same way. This example shows that any term not included in the model gets added into the error term, which may erroneously inflate the error especially if the impact of excluded term on the response is not negligible.

The Error SS is now 46.001, which is the sum of the interaction SS and the error SS of the model with the interaction. The df values were also added the same way. This example shows that any term not included in the model gets added into the error term, which may erroneously inflate the error especially if the impact of excluded term on the response is not negligible.

method Least Squares Means								
method	Estimate	Standard Error	DF	t Value	Pr >|t|	Alpha	Lower	Upper
1	123.40	0.6042	14	204.23	<.0001	0.05	122.10	124.70
2	115.74	0.6042	14	191.56	<.0001	0.05	114.45	117.04

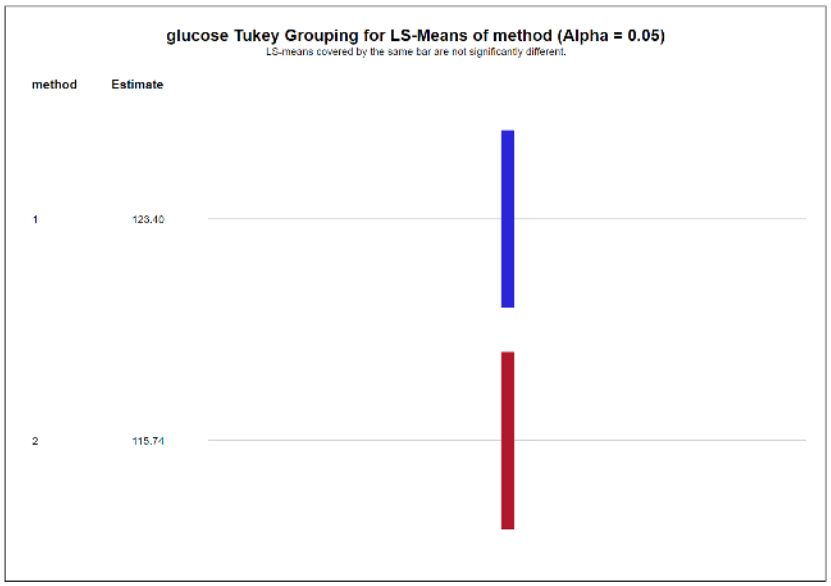


Figure 5.1.1a. 1 : Glucose Tukey grouping for LS-Means of method.

doping Least Squares Means

Doping	Estimate	Standard Error	DF	t Value	Pr >|t|	Alpha	Lower	Upper
1	43.67	0.7400	14	59.01	<.0001	0.05	42.08	45.25
2	136.77	0.7400	14	184.81	<.0001	0.05	135.18	138.35
3	178.28	0.7400	14	240.92	<.0001	0.05	176.70	179.87

Here, we can see that the response variable, the amount of glucose detected in a sample, is the overall mean **PLUS** the effect of the method used **PLUS** the effect of the glucose amount added to the original sample. (Hence, the additive nature of this model!)

This page titled 5.1.1a: The Additive Model (No Interaction) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.1.2: Two-Factor Factorial - Greenhouse Example (Minitab)

For Minitab, we also need to convert the data to a stacked format (Lesson 4 2 way Stacked Dataset). Once we do this, we will need to use a different set of commands to generate the ANOVA. We use...

`Stat` > `ANOVA` > `General Linear Model` > `Fit General Linear Model`

and get the following dialog box:

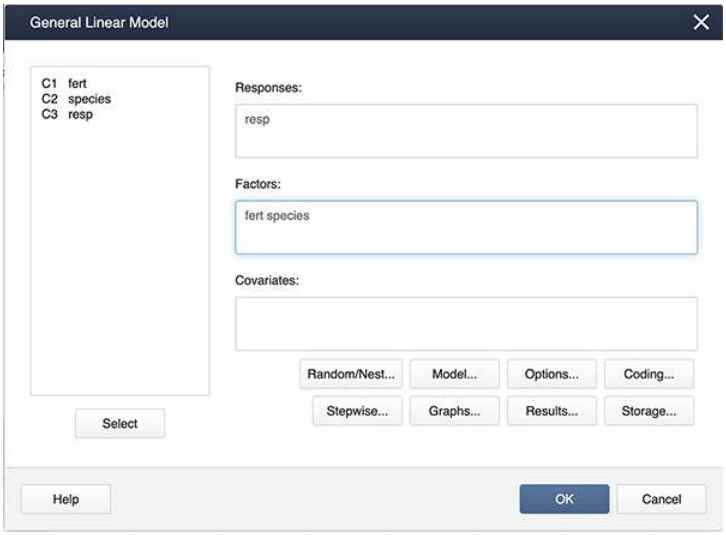


Figure 5.1.2.1: General Linear Model pop-up window.

Click on **Model**…, hold down the shift key and highlight both factors. Then click on the **Add** box to add the interaction to the model.

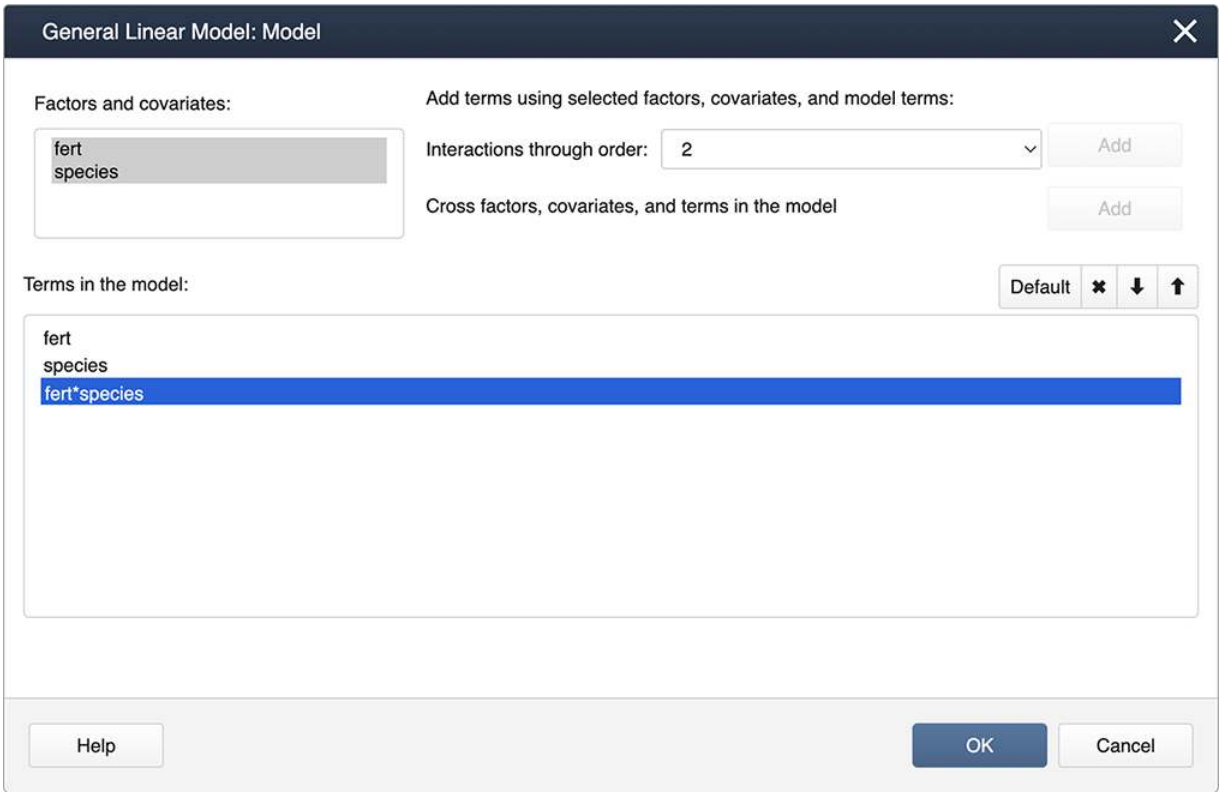


Figure 5.1.2.2: General Linear Model: Model pop-up window.

These commands will produce the ANOVA results below which are similar to the output generated by SAS (shown in the previous section).

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-value	P-value
fert	3	745.44	248.479	73.10	0.000
species	1	236.74	236.741	69.65	0.000
fert*species	3	50.58	16.861	4.96	0.005
Error	40	135.97	3.399		
Total	47	1168.73			

Following the ANOVA run, you can generate the mean comparisons by

`Stat` > `ANOVA` > `General Linear Model` > `Comparisons`

Then specify the *fert*species* interaction term for the comparisons by checking the box.

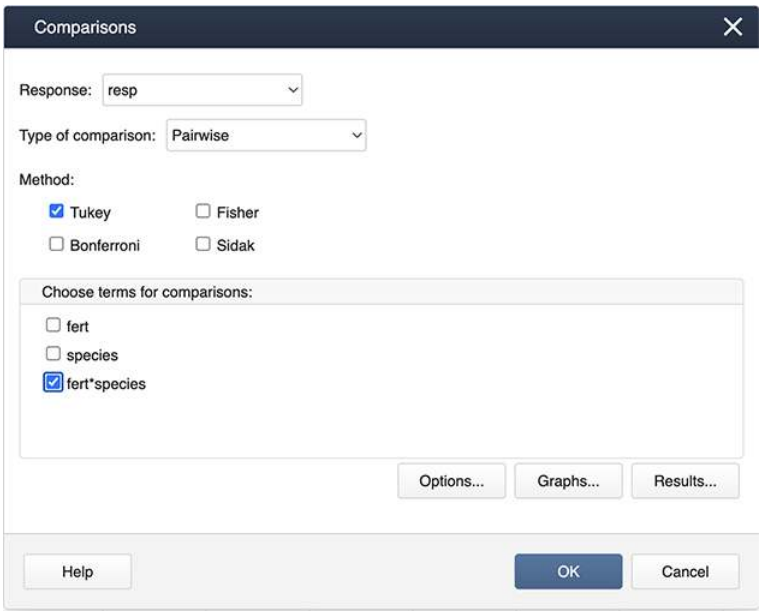


Figure 5.1.2.3: Comparisons pop-up window.

Then choose **Graphs** to get the following dialog box, where "Interval plot for difference of means" should be checked.

Figure 5.1.2.4: Comparisons: Graphs pop-up window.

The outputs are shown below.

Grouping Information Using the Tukey Method and 95% Confidence

fert	species	N	Mean	Grouping
f3	SppB	6	37.0667	A
f1	SppB	6	31.6167	B
f2	SppB	6	30.0500	B
f3	SppA	6	29.2000	B C
f1	SppA	6	28.6000	B C
f2	SppA	6	25.8667	C D
control	SppB	6	23.7000	D E
control	SppA	6	21.0000	E

Means that do not share a letter are significantly different.

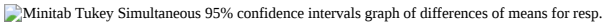


Figure 5.1.2.5: Tukey simultaneous 95% confidence intervals.

This page titled 5.1.2: Two-Factor Factorial - Greenhouse Example (Minitab) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.1.3: Two-Factor Factorial - Greenhouse Example (R)

- Load the greenhouse data.
- Produce a boxplot to plot the differences in heights for each species organized by fertilizer.
- Produce a “means plot” (interval plot) to view the differences in heights for each species organized by fertilizer.
- Obtain the ANOVA table with interaction.
- Obtain Tukey’s multiple comparisons CIs, grouping, and plot.

1. Load the greenhouse data by using the following commands:

```
setwd("~/path-to-folder/")
greenhouse_2way_data <-read.table("greenhouse_2way_data.txt",header=T)
attach(greenhouse_2way_data)
```

2. Produce the Boxplot by using the following commands:

```
library("ggpubr")
boxplot(height ~ species*fertilizer, data = greenhouse_2way_data,
xlab = "Species", ylab = "Plant Height",
main="Distribution of Plant Height by Species",
frame = TRUE)
```

Figure 5.1.3.1: Boxplot of plant height distribution by species.

3. Produce the means plot (interval plot) by using the following commands:

```
library("gplots")
plotmeans(height ~ interaction(species,fertilizer), data = greenhouse_2way_data,conne
xlab = "Fertilizer*species", ylab = "Plant Height",
main="Means Plot with 95% CI")
```

Figure 5.1.3.2: Means plot for plant height vs fertilizer*species.

4. Obtain the ANOVA table with interaction by using the following commands:

```
anova<-aov(height~fertilizer+species+fertilizer*species,greenhouse_2way_data)
summary(anova)
#                    Df Sum Sq Mean Sq F value   Pr(>F)
# fertilizer          3  745.4  248.48   73.10 2.77e-16 ***
# species             1  236.7  236.74   69.64 2.71e-10 ***
# fertilizer:species  3   50.6   16.86    4.96  0.00508 **
# Residuals          40  136.0    3.40
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

5. Obtain Tukey multiple comparisons of means with 95% family-wise confidence level by using the following commands:

```
library(multcomp)
library(multcompView)
tukey_multiple_comparisons<-TukeyHSD(anova,conf.level=0.95,ordered=TRUE)
```

```
tukey_multiple_comparisons
Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered
Fit: aov(formula = height ~ fertilizer + species + fertilizer * species, data = green
$fertilizer
diff       lwr       upr     p adj
f2-control  5.608333 3.5908095  7.625857 0.0000000
f1-control  7.758333 5.7408095  9.775857 0.0000000
f3-control 10.783333 8.7658095 12.800857 0.0000000
f1-f2       2.150000 0.1324762  4.167524 0.0328745
f3-f2       5.175000 3.1574762  7.192524 0.0000002
f3-f1       3.025000 1.0074762  5.042524 0.0013828
$species
diff      lwr      upr p adj
SppB-SppA 4.441667 3.365986 5.517348     0
$`fertilizer:species`
diff        lwr       upr     p adj
control:SppB-control:SppA  2.700000 -0.7025601  6.102560 0.2100548
f2:SppA-control:SppA       4.866667  1.4641065  8.269227 0.0010962
f1:SppA-control:SppA       7.600000  4.1974399 11.002560 0.0000003
f3:SppA-control:SppA       8.200000  4.7974399 11.602560 0.0000001
f2:SppB-control:SppA       9.050000  5.6474399 12.452560 0.0000000
f1:SppB-control:SppA      10.616667  7.2141065 14.019227 0.0000000
f3:SppB-control:SppA      16.066667 12.6641065 19.469227 0.0000000
f2:SppA-control:SppB       2.166667 -1.2358935  5.569227 0.4721837
f1:SppA-control:SppB       4.900000  1.4974399  8.302560 0.0009970
f3:SppA-control:SppB       5.500000  2.0974399  8.902560 0.0001745
f2:SppB-control:SppB       6.350000  2.9474399  9.752560 0.0000138
f1:SppB-control:SppB       7.916667  4.5141065 11.319227 0.0000001
f3:SppB-control:SppB      13.366667  9.9641065 16.769227 0.0000000
f1:SppA-f2:SppA            2.733333 -0.6692268  6.135893 0.1979193
f3:SppA-f2:SppA            3.333333 -0.0692268  6.735893 0.0584747
f2:SppB-f2:SppA            4.183333  0.7807732  7.585893 0.0072041
f1:SppB-f2:SppA            5.750000  2.3474399  9.152560 0.0000832
f3:SppB-f2:SppA           11.200000  7.7974399 14.602560 0.0000000
f3:SppA-f1:SppA            0.600000 -2.8025601  4.002560 0.9991227
f2:SppB-f1:SppA            1.450000 -1.9525601  4.852560 0.8685338
f1:SppB-f1:SppA            3.016667 -0.3858935  6.419227 0.1150225
f3:SppB-f1:SppA            8.466667  5.0641065 11.869227 0.0000000
f2:SppB-f3:SppA            0.850000 -2.5525601  4.252560 0.9922487
f1:SppB-f3:SppA            2.416667 -0.9858935  5.819227 0.3344595
f3:SppB-f3:SppA            7.866667  4.4641065 11.269227 0.0000001
f1:SppB-f2:SppB            1.566667 -1.8358935  4.969227 0.8173904
f3:SppB-f2:SppB            7.016667  3.6141065 10.419227 0.0000019
f3:SppB-f1:SppB            5.450000  2.0474399  8.852560 0.0002022
```

We can see the mean differences for fertilizer combinations, for the two species and for all fertilizer*species combinations. By using the confidence intervals or the p-values we can conclude which of these combinations are significant or not.

6. Obtain Tukey grouping by using the following commands:

```
tukey_grouping<-multcompLetters4(anova,tukey_multiple_comparisons)
print(tukey_grouping)
$fertilizer
f3      f1      f2 control
"a"     "b"     "c"     "d"
$species
SppB SppA
"a"  "b"
$`fertilizer:species`
f3:SppB f1:SppB f2:SppB f3:SppA f1:SppA f2:SppA control:SppB control:SppA
"a"     "b"     "b"    "bc"     "bc"     "cd"      "de"          "e"
```

7. Obtain a plot of differences in mean response for fertilizer*species combinations by using the following commands:

```
par(mar=c(4.1,13,4.1,2.1))
plot(tukey_multiple_comparisons,las=2)
detach(greenhouse_2way_data)
```

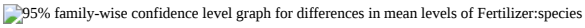


Figure 5.1.3.3: Graph of differences in mean levels of fertilizer:species, showing 95% family-wise confidence levels.

This page titled 5.1.3: Two-Factor Factorial - Greenhouse Example (R) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.1.3a: The Additive Model

- Load the glucose in blood serum data.
- Obtain the ANOVA table with interaction.
- Obtain the ANOVA table without interaction.
- Obtain estimators and CIs for means for each treatment level.
- Obtain Tukey's multiple comparisons CIs and grouping.

1. Load the glucose in blood serum data by using the following commands:

```
setwd("~/path-to-folder/")
glucose_data <- read.table("glucose_data.txt",header=T)
attach(glucose_data)
```

2. Obtain the ANOVA table with interaction by using the following commands:

```
anova<-aov(glucose ~ factor(method) + factor(doping) + factor(method)*factor(doping),
summary(anova)
Df Sum Sq Mean Sq   F value   Pr(>F)
factor(method)                1    264     264    98.347 3.92e-07 ***
factor(doping)                2  57026   28513 10632.526  < 2e-16 ***
factor(method):factor(doping)  2     14       7     2.577    0.117
Residuals                     12     32       3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Here we can see that the interaction term is not significant, and we can drop it from the model. Also, notice that I have defined method and doping as factors since they have numeric values.

3. Obtain the ANOVA table without interaction by using the following commands:

```
anova1<-aov(glucose ~ factor(method) + factor(doping),data=glucose_data)
summary(anova1)
Df Sum Sq Mean Sq F value   Pr(>F)
factor(method)  1    264     264   80.27 3.58e-07 ***
factor(doping)  2  57026   28513 8677.63  < 2e-16 ***
Residuals      14     46       3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The Error SS is now 46.001, which is the sum of the interaction SS and the error SS of the model with the interaction. The df values were also added the same way.

4. Obtain estimators and CIs for means for each treatment level by using the following commands:

```
library(lsmeans)
lsmeans(anova1,"method")
method lsmean    SE df lower.CL upper.CL
1    123 0.604 14      122      125
2    116 0.604 14      114      117
Results are averaged over the levels of: doping
```

```
Confidence level used: 0.95
lsmeans(anova1,"doping")
doping lsmean   SE df lower.CL upper.CL
1   43.7 0.74 14     42.1     45.3
2  136.8 0.74 14    135.2    138.4
3  178.3 0.74 14    176.7    179.9
Results are averaged over the levels of: method
Confidence level used: 0.95
```

5. Obtain Tukey's multiple comparisons CIs and grouping by using the following commands:

```
tukey_multiple_comparisons<-TukeyHSD(anova1,conf.level=0.95,ordered=TRUE)
tukey_multiple_comparisons
Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered
Fit: aov(formula = glucose ~ factor(method) + factor(doping), data = glucose_data)
$`factor(method)`
diff      lwr      upr p adj
1-2 7.655556 5.822828 9.488283 4e-07
$`factor(doping)`
diff       lwr       upr p adj
2-1  93.10000  90.36089  95.83911     0
3-1 134.61667 131.87755 137.35578     0
3-2  41.51667  38.77755  44.25578     0
tukey_grouping<-multcompLetters4(anova1,tukey_multiple_comparisons)
print(tukey_grouping)
$`factor(method)`
1   2
"a" "b"
$`factor(doping)`
3   2   1
"a" "b" "c"
detach(glucose_data)
```

This page titled 5.1.3a: The Additive Model is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.2: Nested Treatment Design

When setting up a multi-factor study, sometimes it is not possible to cross the factor levels. In other words, because of the logistics of the situation, we may not be able to have each level of treatment be combined with each level of another treatment.

Here is an example:

A research team interested in the lifestyle of high school students conducted a study to compare the activity levels of high school students across the 3 geographic regions in the United States, Northeast (NE), Midwest (MW), and the West (W). The study also included the comparison of activity levels among cities within each region. Two school districts were chosen from two major cities from each of these 3 regions and the response variable, the average number of exercise hours per week for high school students for each school district was recorded.

A diagram to illustrate the treatment design can be set up as follows. Here, the subscript i identifies the regions, and the subscript j indicates the cities:

Factor A (Region) i		**Factor B (City)** j		**Average**
		1	**2**	
NE		30	18	
		35	20	
	Average	$\bar{Y}_{11.} = 32.5$	$\bar{Y}_{12.} = 19$	$\bar{Y}_{1..} = 25.75$
MW		10	20	
		9	22	
	Average	$\bar{Y}_{21.} = 9.5$	$\bar{Y}_{22.} = 21$	$\bar{Y}_{2..} = 15.25$
W		18	4	
		19	6	
	Average	$\bar{Y}_{31.} = 18.5$	$\bar{Y}_{32.} = 5$	$\bar{Y}_{3..} = 9.5$
			Average	$\bar{Y}_{...} = 16.83$

The table above shows the data obtained: the grand mean, the marginal means which are the treatment level means, and finally, the cell means. The cell means are the averages of the two school district mean activity levels for each combination of *Region* and *City*.

This example drives home the point that the levels of the second factor (*City*) cannot practically be crossed with the levels of the first factor (*Region*) as cities are specific or unique to regions. Note that the cities are identified as 1 or 2 within each region. But it is important to note that city 1 in the Northeast is not the same as city 1 in the Midwest. The concept of nesting does come in useful to describe this type of situation and the use of parentheses is appropriate to clearly indicate the nesting of factors. To indicate that the *City* is nested within the factor *Region*, the notation: *City(Region)* will be used. Here, *City* is the nested factor and *Region* is the nesting factor.

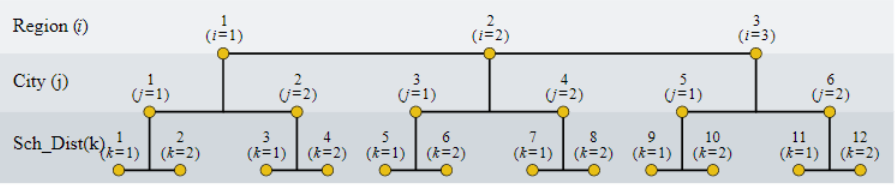


Figure 5.2.1: Diagram of the levels of treatment design.

We can partition the deviations as before into the following components:

$$\underbrace{Y_{ijk}-\bar{Y}_{...}}_{\text{Total deviation}}=\underbrace{\bar{Y}_{i..}-\bar{Y}_{...}}_{\text{A main effect}}+\underbrace{\bar{Y}_{ij.}-\bar{Y}_{i..}}_{\text{Specific B effect when A at the } i^{th} \text{ level}}+\underbrace{Y_{ijk}-\bar{Y}_{ij.}}_{\text{Residual}} \tag{5.2.1}$$

Source	d.f.
Region	$(a-1)=2$
City (Region)	$a(b-1)=3$
Error	$ab(n-1)=6$
Total	$N-1=11$

The statistical model follows as:

$$Y_{ijk}=\mu+\alpha_i+\beta_{j(i)}+\epsilon_{ijk} \tag{5.2.2}$$

\[

where: μ is a constant

α_i are constants subject to the restriction $\sum \alpha_i=0$

$\beta_{j(i)}$ are constants subject to the restriction $\sum \beta_{j(i)}=0$ for all i

ϵ_{ijk} are independent $N\left(0,\sigma^2\right)$

$i=1,\ldots,a;\ j=1,\ldots,b;\ k=1,\ldots,n$

We will want to test the following Null Hypotheses:

For Factor A

$$H_0:\ \mu_{\text{Northeast}}=\mu_{\text{Midwest}}=\mu_{\text{West}} \text{ vs. } H_A:\ \text{Not all equal} \tag{5.2.3}$$

For Factor B

When stating the Null Hypothesis for Factor B, the nested effect, alternative notation has to be used.

Up to this point, we have been stating Null Hypotheses in terms of the means (e.g. $H_0:\ \mu_1=\mu_2=\ldots=\mu_k$), but we can alternatively state a Null Hypothesis in terms of the parameters for that treatment in the model. For example, for the nesting factor A, we could also state the Null Hypothesis as

$$H_0:\ \alpha_{\text{Northeast}}=\alpha_{\text{Midwest}}=\alpha_{\text{West}}=0 \text{ or } H_0:\ \text{all } \alpha_i=0 \tag{5.2.4}$$

For the nested factor B, the Null Hypothesis should differentiate between the nesting and the nested factors, because we are evaluating the nested factor within the levels of the nesting factor.

So for the nested factor (*City*, nested within *Region*), we have the Null Hypothesis.

$$H_0:\ \text{all } \beta_{j(i)}=0 \text{ vs. } H_A:\ \text{not all } \beta_{j(i)}=0 \text{ for } j=1,2, \tag{5.2.5}$$

The F-tests can then proceed as usual using the ANOVA results. The first two columns of the ANOVA table should be as follows on the next page.

Note

1. There is no interaction between a nested factor and its nesting factor.
2. The nested factors always have to be accompanied by their nested factor. This means that the effect B does not exist and B(A) represents the effect of B within the factor A
3. df of B(A) = df of B + df of A*B (This is simply a mathematically correct identity and may not be of much practical use, as effects B(A) and A*B cannot coexist)
4. The residual effect of any ANOVA model is a nested effect - the replicate effect nested within the factor level combinations. Recall that the replicates are considered homogeneous and so any variability among them serves to estimate the model error.

This page titled 5.2: Nested Treatment Design is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.2.1: Nested Model in SAS

Here is the SAS code to run the ANOVA model for the hours of exercise for high school students example discussed in lesson 5.2:

```
data Nested_Example_data;
infile datalines delimiter=',';
input Region $ City $ ExHours;
datalines;
    NE,NY,30
    NE,NY,35
    NE,Pittsburgh,18
    NE,Pittsburgh,20
    MW,Chicago,10
    MW,Chicago,9
    MW,Detroit,20
    MW,Detroit,22
    W,LA,18
    W,LA,19
    W,Seattle,4
    W,Seattle,6
;

/*to run the nested ANOVA model*/
proc mixed data=Nested_Example_data method=type3;
    class Region City;
    model ExHours = Region City(Region);
    store nested1;
run;

/*to obtain the resulting multiple comparison results*/
ods graphics on;
proc plm restore=nested1;
    lsmeans Region / adjust=tukey plot=meanplot cl lines;
    lsmeans City(Region) / adjust=tukey plot=meanplot cl lines;
run;
```

When we run this SAS program, here is the output that we are interested in:

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Region	2	424.666667	212.333333	Var(Residual)+Q(Region, City(Region))	MS(Residual)	6	65.33	<.0001

Type 3 Analysis of Variance								
City(Region)	3	496.750000	165.583333	Var(Residual)+Q(City(Region))	MS(Residual)	6	50.95	0.0001
Residual	6	19.500000	3.250000	Var(Residual)				

Type 3 Test of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr>F
Region	2	6	65.33	<.0001
City(Region)	3	6	50.95	0.0001

The *p*-values above indicate that both *Region* and *City(Region)* are statistically significant. The plots and charts below obtained from the Tukey option specify the means which are significantly different.

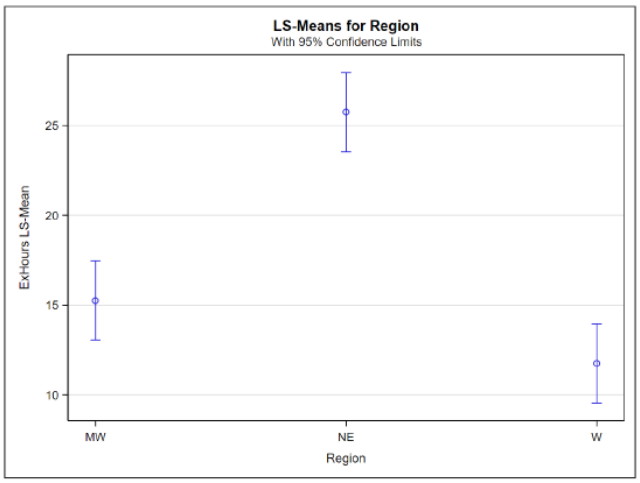


Figure 5.2.1.1: Mean hours of exercise by Region with 95% CIs

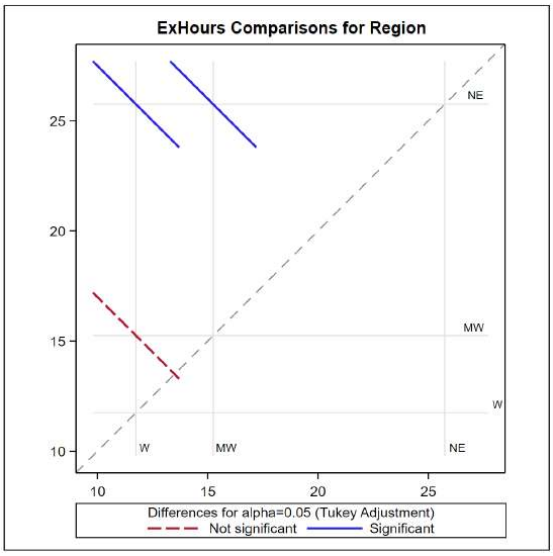


Figure 5.2.1.2: Diffogram for Mean Comparisons by Region

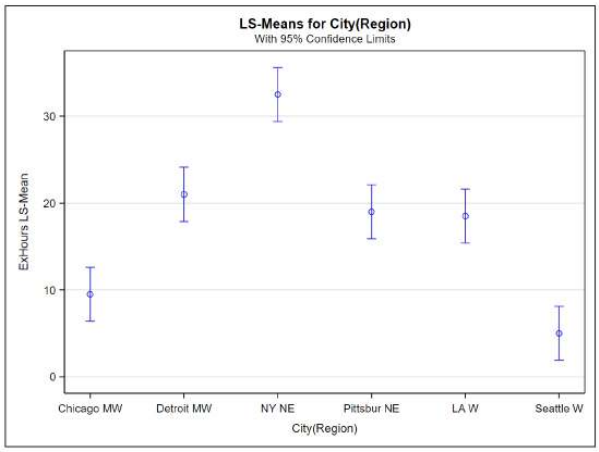


Figure 5.2.1.3: Mean hours of exercise by City(Region) with 95% CIs

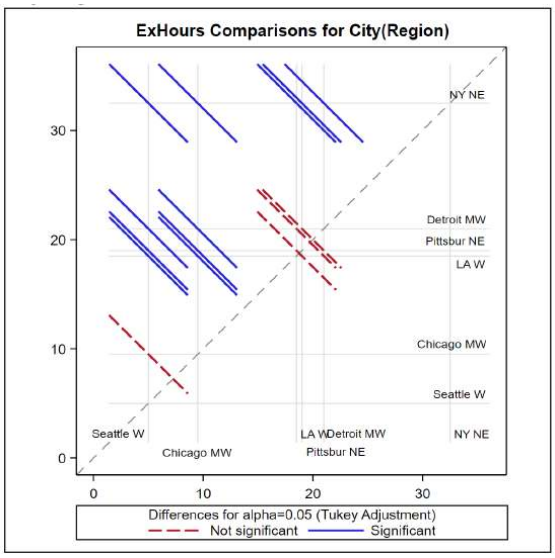


Figure 5.2.1.4: Diffogram for Mean Comparisons by City(Region)

The exercise hours on average are statistically higher in the northeastern region compared to the midwest and the west while the average exercise hours of these two regions are not significantly different.

Also, the comparison of the means between cities indicates that the high schoolers in New York city exercise significantly more than the other cities in the study. The exercise levels are similar among Detroit, Pittsburgh, and LA, while exercise levels of high schoolers in Chicago and Seattle are similar but significantly lower than all other cities in the study.

These grouping observations are further confirmed by the lines plots below.

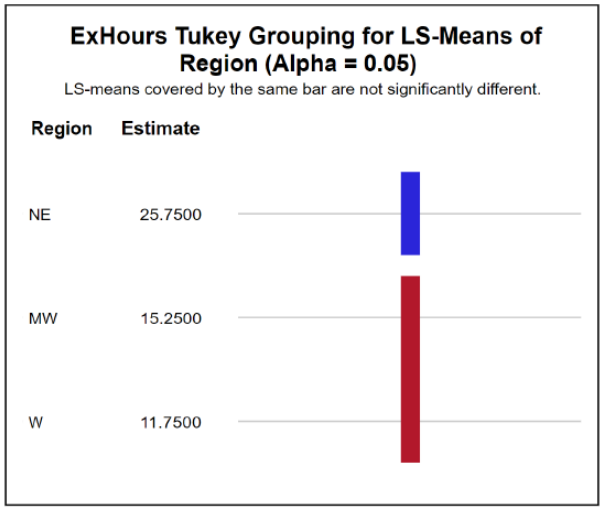


Figure 5.2.1.5: Line plot for multiple comparisons of means for Regions.

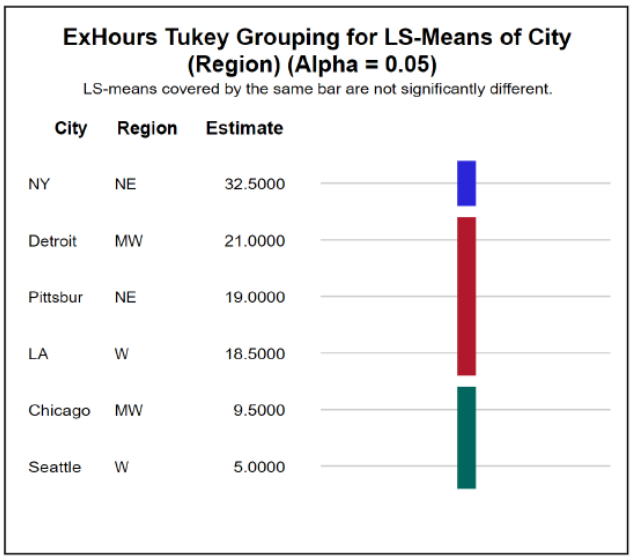


Figure 5.2.1.6: Line plot for multiple comparisons of means for Cities.

This page titled 5.2.1: Nested Model in SAS is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.2.2: Nested Model in Minitab

In Minitab, for the following (Nested Example Data):

`Stat` > `ANOVA` > `General Linear Model` > `Fit General Linear Model`

Enter the factors 'Region' and 'City' in the Factors box, then click on **Random/Nest**...Here is where we specify the nested effect of City in Region.

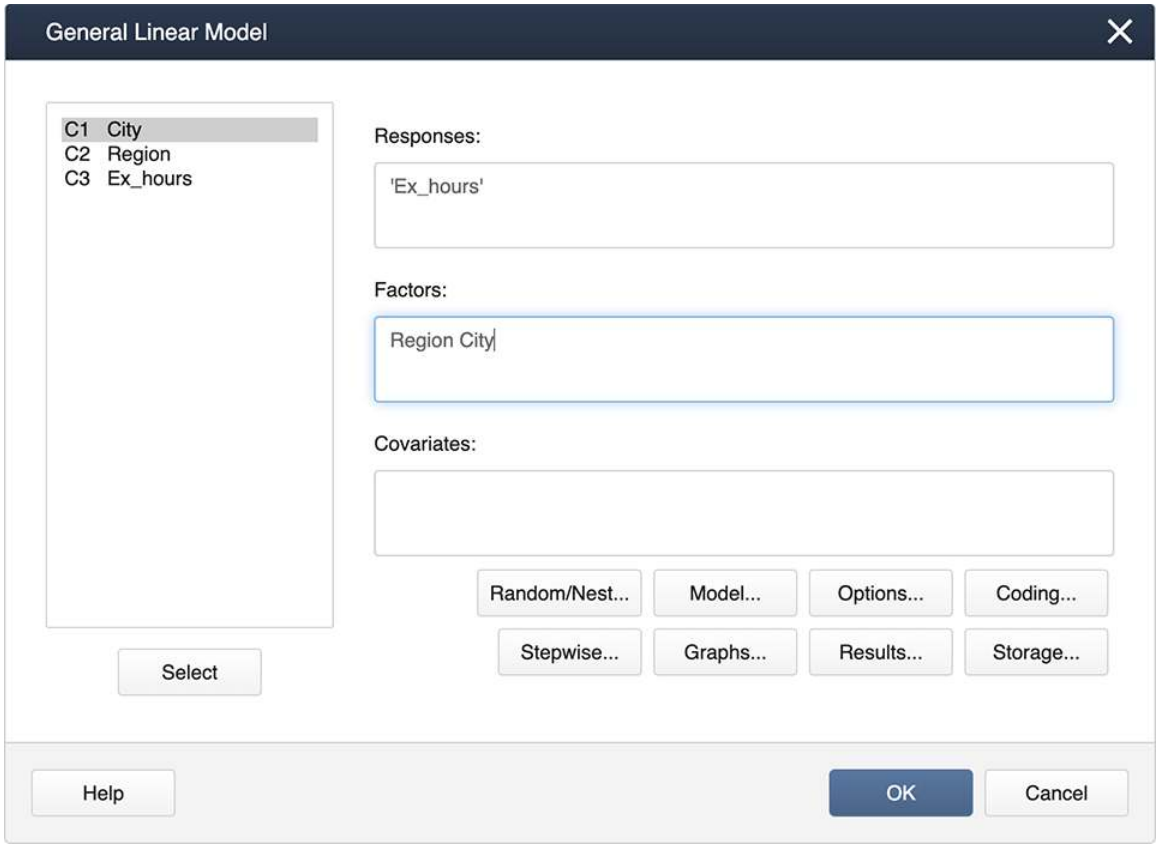


Figure 5.2.2.1: General Linear Model pop-up window.

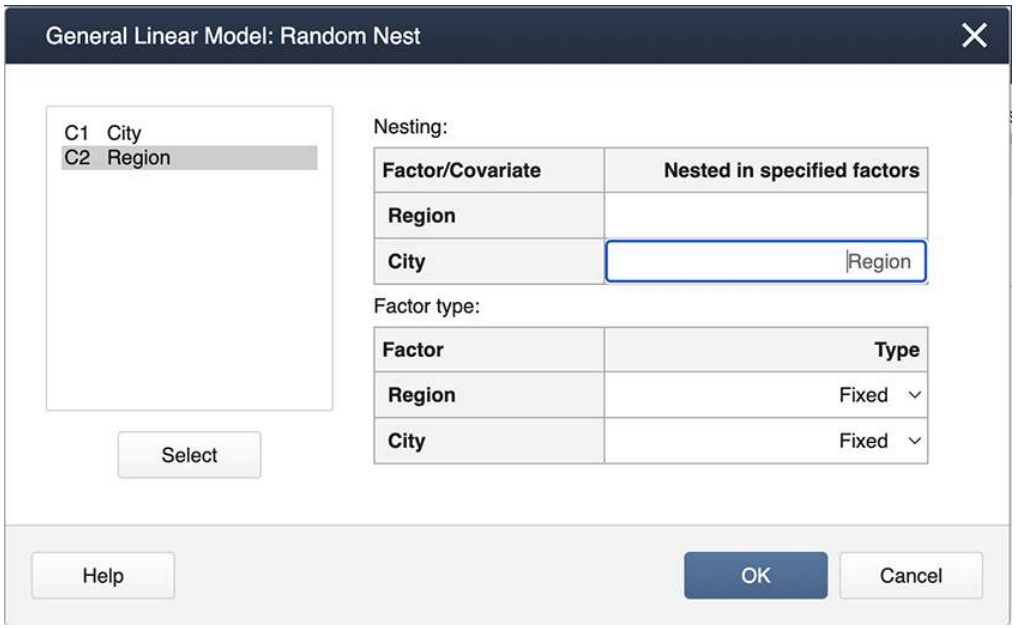


Figure 5.2.2.2: Random Nest pop-up window.

The output is shown below.

Factor Information

General Linear Model: response versus School, Instructor

Factor	Type	Levels	Values
Region	Fixed	2	1,2
City(Region)	Fixed	6	Atlanta(1), Chicago(1), SanFran(1), Atlanta(2), Chicago(2), SanFran(2)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F	P
Region	1	108.00	108.000	15.43	0.008
City(Region)	4	616.00	154.000	22.00	0.001
Error	6	42.00	7.000		
Total	11	766.00			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.64575	94.52%	89.95%	78.07%

Following the ANOVA run, you can generate the mean comparisons by

`Stat` > `ANOVA` > `General Linear Model` > `Comparisons`

Then specify "Region" and "City(Region)" for the comparisons by checking the boxes.

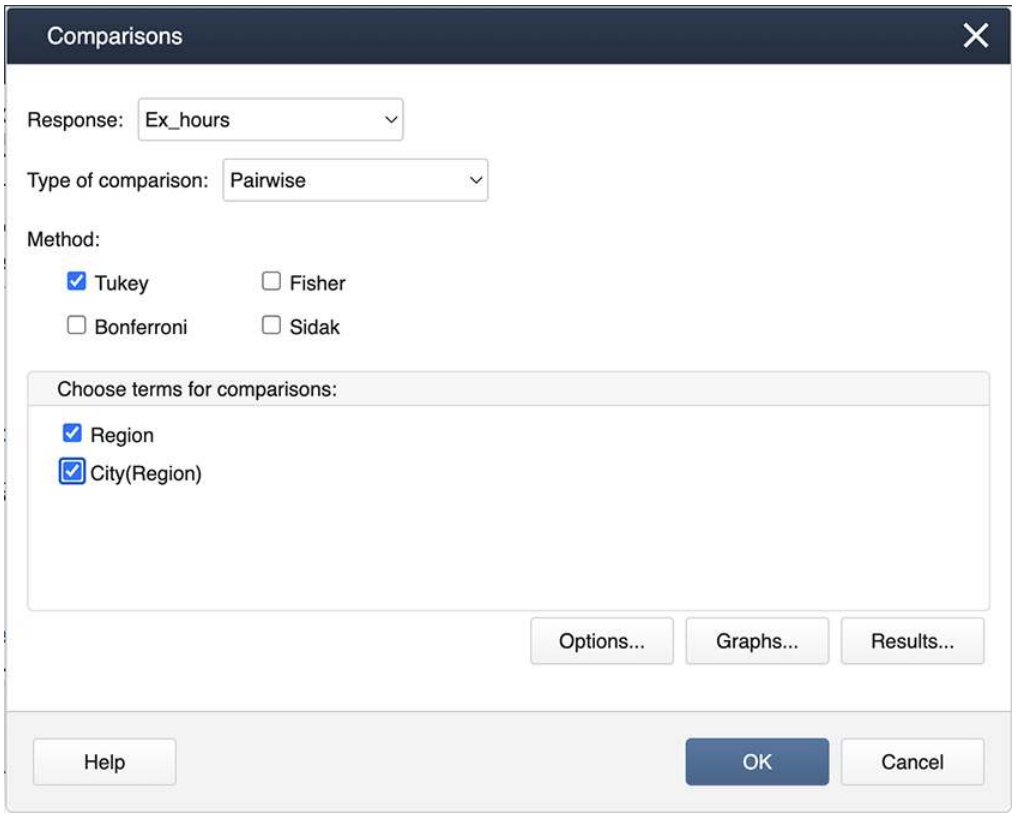


Figure 5.2.2.3: Comparisons pop-up window.

Then choose **Graphs** to get the following dialog box, where "Interval plot for difference of means" should be checked.

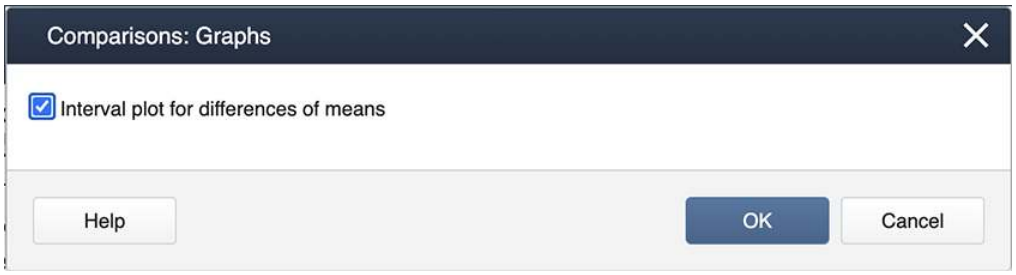


Figure 5.2.2.4: Comparisons: Graphs pop-up window.

The outputs are as follows.

Comparison for Ex_hours

Tukey Pairwise Comparisons: Region

Grouping Information Using Tukey Method and 95% Confidence

Region	N	Mean	Grouping
1	6	18	A
2	6	12	B

Means that do not share a letter are significantly different.

Figure 5.2.2.5: Tukey simultaneous 95% CIs differences of means graph for Ex_hours, by Region.

Tukey Pairwise Comparisons: (City)Region

Grouping Information Using Tukey Method and 95% Confidence

City(Region)	N	Mean	Grouping			
Atlanta (1)	2	27.0	A			
Chicago(2)	2	20.0	A	B		
SanFran(1)	2	18.5	A	B	C	
Atlanta(2)	2	12.5		B	C	D
Chicago(1)	2	8.5			C	D
SanFran(2)	2	3.5				D

Means that do not share a letter are significantly different.

Figure 5.2.2.6: Tukey simultaneous 95% CIs differences of means graph for Ex_hours, by City(Region).

This page titled 5.2.2: Nested Model in Minitab is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.2.3: Nested Model in R

- Load the Exercise Hours data.
- Obtain the ANOVA table for the nested treatment design.
- Obtain estimators and CIs for means for each region and city.
- Obtain means plot for region and city within the region.
- Obtain Tukey's multiple comparisons CIs.

1. Load the Exercise Hours data by using the following commands:

```
setwd("~/path-to-folder/")
ex_hours_data <- read.table("ex_hours_data.txt",header=T)
attach(ex_hours_data)
```

2. Obtain the ANOVA table for the nested treatment design by using the following commands:

```
nested<-aov(Ex_hours ~ Region+Region/City,data=ex_hours_data)
summary(nested)
# Df Sum Sq Mean Sq F value Pr(>F)
# Region       2  424.7  212.33   65.33 8.46e-05 ***
# Region:City  3  496.8  165.58   50.95 0.000116 ***
# Residuals    6   19.5    3.25
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

3. Obtain estimators and CIs for means for each region and city by using the following commands:

```
library(lsmeans)
lsmeans(nested,"Region")
# Region lsmean    SE df lower.CL upper.CL
# MW       15.2 0.901  6    13.04     17.5
# NE       25.8 0.901  6    23.54     28.0
# W        11.8 0.901  6     9.54     14.0
#Results are averaged over the levels of: City
#Confidence level used: 0.95
lsmeans(nested,"City")
#City       Region lsmean   SE df lower.CL upper.CL
# Chicago    MW        9.5 1.27  6     6.38    12.62
# Detroit    MW       21.0 1.27  6    17.88    24.12
# NY         NE       32.5 1.27  6    29.38    35.62
# Pittsburgh NE       19.0 1.27  6    15.88    22.12
# LA         W        18.5 1.27  6    15.38    21.62
# Seattle    W         5.0 1.27  6     1.88     8.12
#Confidence level used: 0.95
```

4. Obtain means plot for region and city within region by using the following commands:

```
library(plotrix)
region_means<-as.data.frame(lsmeans(nested,"Region"))
```

```
plotCI(x = region_means$lsmean,y = NULL ,li = region_means$lower.CL, ui = region_mean
axis(1, at=1:3, labels=region_means$Region)
```

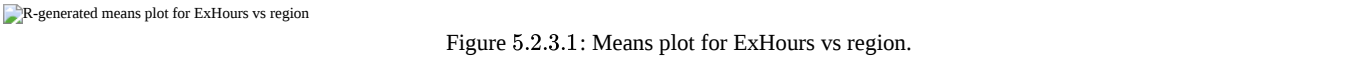


Figure 5.2.3.1: Means plot for ExHours vs region.

```
city_means<-as.data.frame(lsmeans(nested,"City"))
City_Region<-paste(city_means$City,city_means$Region)
plotCI(x = city_means$lsmean,y = NULL ,li = city_means$lower.CL, ui = city_means$uppe
axis(1, at=1:6, labels=City_Region)
```

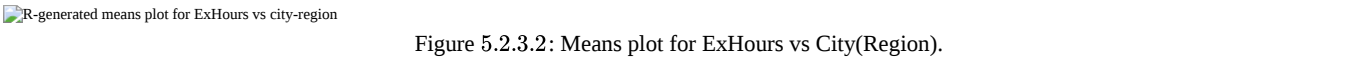


Figure 5.2.3.2: Means plot for ExHours vs City(Region).

5. Obtain Tukey's multiple comparisons CIs by using the following commands:

```
library(multcomp)
library(multcompView)
tukey_multiple_comparisons_region<-TukeyHSD(nested,"Region",conf.level=0.95,ordered=T
tukey_multiple_comparisons_region
  Tukey multiple comparisons of means
    95% family-wise confidence level
    factor levels have been ordered
Fit: aov(formula = Ex_hours ~ Region + Region/City, data = ex_hours_data)
# $Region
#      diff        lwr       upr     p adj
#MW-W   3.5 -0.4112978  7.411298 0.0747598
#NE-W  14.0 10.0887022 17.911298 0.0000836
plot(tukey_multiple_comparisons_region)
```

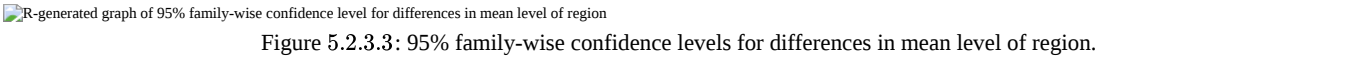


Figure 5.2.3.3: 95% family-wise confidence levels for differences in mean level of region.

```
tukey_multiple_comparisons_city<-TukeyHSD(nested,"Region:City",conf.level=0.95,ordere
cities<-as.data.frame(na.omit(tukey_multiple_comparisons_city$"Region:City"))
cities

#                          diff         lwr       upr        p adj
# MW:Chicago-W:Seattle      4.5 -4.96579743 13.965797 0.5867601138
# W:LA-W:Seattle           13.5  4.03420257 22.965797 0.0087623039
# NE:Pittsburgh-W:Seattle  14.0  4.53420257 23.465797 0.0072411812
# MW:Detroit-W:Seattle     16.0  6.53420257 25.465797 0.0035459602
# NE:NY-W:Seattle          27.5 18.03420257 36.965797 0.0001761692
# W:LA-MW:Chicago           9.0 -0.46579743 18.465797 0.0626471065
# NE:Pittsburgh-MW:Chicago  9.5  0.03420257 18.965797 0.0491884424
# MW:Detroit-MW:Chicago    11.5  2.03420257 20.965797 0.0198221594
# NE:NY-MW:Chicago         23.0 13.53420257 32.465797 0.0004610102
# NE:Pittsburgh-W:LA        0.5 -8.96579743  9.965797 1.0000000000
# MW:Detroit-W:LA           2.5 -6.96579743 11.965797 0.9752059356
# NE:NY-W:LA               14.0  4.53420257 23.465797 0.0072411812
```

```
# MW:Detroit-NE:Pittsburgh  2.0 -7.46579743 11.465797 0.9960158169
# NE:NY-NE:Pittsburgh      13.5  4.03420257 22.965797 0.0087623039
# NE:NY-MW:Detroit         11.5  2.03420257 20.965797 0.0198221594

library(plotrix)
city_diff<-as.character(c("
MW:Chicago-W:Seattle","W:LA-W:Seattle", "NE:Pittsburgh-W:Seattle","MW:Detroit-W:Seatt
par(mar=c(8, 4, 2, 2) + 0.1)
plotCI(x = cities$diff,y = NULL ,li = cities$lwr, ui = cities$upr, xaxt = "
n",ylab="Differences of Means",xlab="")
abline(h=0)
axis(1, at=1:15, labels=city_diff,las = 2, cex.axis = 0.8)
```

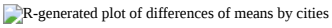


Figure 5.2.3.4: Differences of means by cities plot.

This page titled 5.2.3: Nested Model in R is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.3: Crossed-Nested Designs

Multi-factor studies can involve factor combinations in which factors are crossed and/or nested. These treatment designs are based on the extensions of the concepts discussed so far.

Consider an example (from Canavos and Koutrouvelis, 2009) where machines in an assembly process are evaluated for assembly times. There were three factors of interest: Machine ID (1, 2, or 3), Configuration (1 or 2), and Power level (1, 2, or 3).

3-factor table		**Machine (A)**					
		1		**2**		**3**	
Configuration (B)		**1**	**2**	**1**	**2**	**1**	**2**
	1	10.2	4.2	12.0	4.1	13.1	4.1
		13.1	5.2	13.5	6.1	12.9	6.1
Power (C)	**2**	16.2	8.0	12.6	4.0	12.9	2.2
		16.9	9.1	14.6	6.1	13.7	3.8
	3	13.8	2.5	12.9	3.7	11.8	2.7
		14.9	4.4	15.0	5.0	13.5	4.1

It turns out that each machine can be operated at each power level, and so these factors can be crossed. Also, each configuration can be operated at each power level and so these factors also are crossed. But the configurations (1 or 2) are unique to each machine. As a result, the configuration is nested within the machine.

The statistical model contains both crossed and nested effects and is:

$$Y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{j(i)k} + \epsilon_{ijk} \tag{5.3.1}$$

with the ANOVA table as follows:

Source	**df**
Factor A	$a-1$
Factor B(A)	$a(b-1)$
Factor C	$c-1$
AC	$(a-1)(c-1)$
CB(A)	$a(b-1)(c-1)$
Error	$abc(n-1)$
Total	$N-1=(nabc)-1$

Notice that the two main effects, *Machine* and *Power*, are included in the model along with their interaction effect. The nested relationship of *Configuration* within *Machine* is represented by the *Configuration(Machine)* term and the crossed relationship between *Configuration* and *Power* is represented by their interaction effect.

Notice that the main effect *Configuration* and the crossed effect *Configuration* × *Machine* are not included in the model. This is consistent with the facts that a nested effect cannot be represented as the main effect and also that a nested effect cannot interact with its nesting effect.

This page titled 5.3: Crossed-Nested Designs is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.4: Try It!

Exercise 5.4.1: CO_2 Emissions

To study the variability in CO2 emission rate by global regions 4 countries: US, Britain, India, and Australia were chosen. From each country, 3 major cities were chosen and the emission rates for each month for the year 2019 were collected.

1. What type of model is this?

   1. nested
   2. cross-nested
   3. factorial

2. How many factors?

   1. 4
   2. 3
   3. 2

3. The replicates are...

   1. 12 months of 2019
   2. countries US, Britain, India, and Australia
   3. major cities in US, Britain, India, and Australia

4. The residual effect in the ANOVA model is...

   1. country*city*month
   2. month(city(country))
   3. month(country*city)

5. How many degrees of freedom?

   1. 66
   2. 132
   3. 88

Show Answers and Explanations

Answers

Q1 - 1. nested

Q2 - 3. 2 factors

Q3 - 1. 12 months of 2019

Q4 - 2. month(city(country)

Q5 - 2. 132

Explanations

Residual effect (or error term) is the month(city(country), which is the nested effect of "month", the replicate, within the combinations of the two factors "country" and "city". One way to double-check this answer is to verify if the df values are the same.

$$\begin{aligned}\text{error term df} &= \text{total df} - (\text{sum df values of the model terms}) \\ &= (144-1) - (\text{country df} + \text{city(country) df}) \\ &= 143 - (3 + 2*4) \\ &= 143 - 11 = 132\end{aligned}$$

$\text{df for city(country)} = 11(12) = 132$

See the Section 5.2 ANOVA table for df formula.

Exercise 5.4.2

A military installation is interested in evaluating the speed of reloading a large gun. Two methods of reloading are considered, and 3 groups of cadets were evaluated (slight, average, and heavy individuals). Three teams were set up within each group and they wanted to identify the fastest team within each group to go on to a demonstration for the military officials. Each team performed the reloading with each method two times (two replications).

1. Identify (i.e. name) the treatment design.

1. nested
2. cross-nested
3. factorial

2. They started to construct the ANOVA table which is given below. Given that there are a total of 36 observations in the dataset, there seems to be a missing source of variation in the analysis. What is this source of variation?

Source	df
Method	1
Group	2
Method*Group	2
Team (Group)	6

1. Team*group*method
2. Team (Group)*method
3. Team*Group

3. How many degrees of freedom are associated with the error term?

1. 6
2. 24
3. 18

Show Answers

Q1 - 2. cross-nested

Q2 - 2. Team (Group)*method

Q3 - 3. 18

Exercise 5.4.3: GPA Comparisons

The GPA comparison of four popular majors—biology, business, engineering, and psychology—between males and females is of interest. For 6 semesters, the average GPA of each of these majors for male and female students was computed.

1. What type of model is this?

1. nested
2. cross-nested
3. crossed

2. How many factors?

1. 4
2. 3
3. 2

3. The replicates are...

1. semesters
2. majors
3. gender

4. The residual effect in the ANOVA model is...

1. major*gender*semester
2. semester(gender*major)
3. semester(major(gender))

5. How many degrees of freedom?

1. 48
2. 40
3. 2

Show Answers and Explanations

Answers

Q1 - 3. crossed

Q2 - 3. 2 factors

Q3 - 1. semesters

Q4 - 2. semester(gender*major)

Q5 - 2. 40

Explanations

Residual effect (or error term) is semester (gender*major). The error term is the nested effect of "semester", the replicate nested within gender*major, which is the "combined effect" of the factors. One way to double-check is to verify if df values are the same.

$$\begin{aligned}\text{error term df} &= \text{total df} - (\text{sum df values of the model terms}) \\ &= (48-1) - (\text{major df} + \text{gender df} + \text{major*gender df}) \\ &= 47 - (3+1+3*1) \\ &= 40\end{aligned}$$

$\text{df for semester(major*gender)} = 5 * 8 = 40$

See the Section 5.2 ANOVA table for df formula.

This page titled 5.4: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.5: Chapter 5 Summary

In this lesson, we discussed important elements of the "Treatment Design," one of the two components of an "Experimental Design." We are now familiar with the main effects and interaction effects of a factorial design.

In a full factorial design, the experiment is carried out at every factor level combination. Most factorial studies do not go beyond a two-way interaction, and if a two-way interaction is significant, the mean response values should be compared among different combinations of the two factors rather than among the single factor levels. In other words, the focus should be on response vs. interaction effect rather than response vs. main effects. An Interaction plot is a useful graphical tool to understand the extent of interactions among factors (or treatments) with parallel lines indicating no interaction.

In a nested design, the experiment need not be conducted at every combination of levels in all factors. Given two factors in a nested design, there is a distinction between the nested and the nesting factor. The levels of the nested factor may be unique to each level of the nesting factor. Therefore, the comparison of the nested factor levels should be made within each level of the nesting factor—a fact that should be kept in mind when stating null and alternative hypotheses for the nested factor(s), and also when writing programming code.

This page titled 5.5: Chapter 5 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

5.6: Treatment Design Summary (Optional Enrichment Material)

In an effort to summarize how to think about sums of squares and degrees of freedom and how this translates into a model that can be implemented in SAS, Dr. Rosenberger walks you through this process in the videos below. Pay attention to the subscripts and these are the keys to understanding this material.

Part One

Part Two

This page titled 5.6: Treatment Design Summary (Optional Enrichment Material) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

CHAPTER OVERVIEW

6: Random Effects and Introduction to Mixed Models

Overview

So far, in our discussion of treatment designs, we have made the (unstated) assumption that the treatment levels were chosen intentionally by the researcher as dictated by his/her specific interests. The scope of inference in this situation is limited to the specific (or fixed) levels used in the study. However, this is not always the case. Sometimes, treatment levels may be a (random) sample of possible levels, and the scope of inference is to a larger population of all possible levels.

If it is clear that the researcher is interested in comparing specific, chosen levels of treatment, that treatment is called a ***fixed effect***. On the other hand, if the levels of the treatment are a sample of a larger population of possible levels, then the treatment is called a ***random effect***.

Learning Objectives

Upon completion of this lesson, you should be able to:

1. Extend the treatment design to include random effects.
2. Understand the basic concepts of random-effects models.
3. Calculate and interpret the intraclass correlation coefficient.
4. Combining fixed and random effects in the mixed model.
5. Work with mixed models that include both fixed and random effects.

This page titled 6: Random Effects and Introduction to Mixed Models is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Random Effects

When a treatment (or factor) is a random effect, the model specifications together with relevant null and alternative hypotheses will have to be changed. Recall the cell means model defined in Chapter 4 for the fixed effect case, which has the model equation:

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{6.1.1}$$

where μ_i are parameters for the treatment means.

For the ***single factor random effects model*** we have:

$$Y_{ij} = \mu_i + \epsilon_{ij} \tag{6.1.2}$$

where μ_i and ϵ_{ij} are independent random variables such that $\mu_i \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma_\mu^2\right)$ and $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right)$. Here, $i = 1, 2, \ldots, T$ and $j = 1, 2, \ldots, n_i$, where $n_i \equiv n$ if balanced.

Notice that the random effects ANOVA model is similar in appearance to the fixed effects ANOVA model. However, the treatment mean μ_i's are constant in the fixed-effect ANOVA model, whereas in the random-effects ANOVA model the treatment mean μ_i's are random variables.

Note that the expected mean response, in the random effects model stated above, is the same at every treatment level and equals μ.

$$E\left(Y_{ij}\right) = E\left(\mu_i + \epsilon_{ij}\right) = E\left(\mu_i\right) + E\left(\epsilon_{ij}\right) = \mu \tag{6.1.3}$$

The variance of the response variable (say σ_Y^2) in this case can be partitioned as:

$$\sigma_Y^2 = V\left(Y_{ij}\right) = V\left(\mu_i + \epsilon_{ij}\right) = V\left(\mu_i\right) + V\left(\epsilon_{ij}\right) = \sigma_\mu^2 + \sigma_\epsilon^2 \tag{6.1.4}$$

as μ_i and ϵ_{ij} are independent random variables.

Similar to fixed effects ANOVA model, we can express the random effects ANOVA model using the factor effect representation, using $\tau_i = \mu_i - \mu$. Therefore the factor effects representation of the random effects ANOVA model would be:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \tag{6.1.5}$$

where μ is a constant overall mean, and τ_i and ϵ_{ij} are independent random variables such that $\tau_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\mu^2\right)$ and $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right)$. Here, $i = 1, 2, \ldots, T$ and $j = 1, 2, \ldots, n_i$, where $n_i \equiv n$ if balanced. Here, τ_i is the effect of the randomly selected i^{th} level.

The terms σ_μ^2 and σ_ϵ^2 are referred to as **variance components**. In general, as will be seen later in more complex models, there will be a variance component associated with each effect involving at least one random factor.

Variance components play an important role in analyzing random effects data. They can be used to verify the significant contribution of each random effect to the variability of the response. For the single factor random-effects model stated above, the appropriate null and alternative hypothesis for this purpose is:

$$H_0: \sigma_\mu^2 = 0 \text{ vs. } H_A: \sigma_\mu^2 > 0 \tag{6.1.6}$$

Similar to the fixed effects model, an ANOVA analysis can then be carried out to determine if H_0 can be rejected.

The MS and the df computations of the ANOVA table are the same for both the fixed and random-effects models. However, the computations of the F-statistics needed for hypothesis testing require some modification.

Specifically, the F statistics denominator will no longer always be the mean squared error (MSE or MSERROR) and will vary according to the effect of interest (listed in the Source column of the ANOVA table). For a random-effects model, the quantities known as ***Expected Means Squares (EMS)***, ***shown in the ANOVA table below***, can be used to identify the appropriate F-statistic denominator for a given source in the ANOVA table. These EMS quantities will also be useful in estimating the variance components associated with a given random effect. Note that the EMS quantities are in fact the population counterparts of the mean sums of squares (MS) that we are already familiar with. In SAS the `proc mixed , method=type3` option will generate the EMS column in the ANOVA table output.

Source	df	SS	MS	F	P	EMS (Expected Means Squares)
Trt						$\sigma_\epsilon^2 + n\sigma_\mu^2$
Error						σ_ϵ^2
Total						

Note

Variance components are NOT synonymous with mean sums of squares. Variance components are usually estimated by using the Method of Moments where algebraic equations, created by setting the mean sums of squares (MS) equal to the EMS for the relevant effects, are solved for the unknown variance components. For example, the variance component for the treatment in the single-factor random effects discussed above can be solved as:

$$s^2_{\text{among trts}} = \frac{MS_{trt} - MS_{error}}{n} \tag{6.1.7}$$

This is by using the two equations:

$MS_{error} = \sigma_\epsilon^2$

$MS_{trt} = \sigma_\epsilon^2 + n\sigma_\mu^2$

More about variance components...

Often the variance component of a specific effect in the model is expressed as a percent of the total variation of the variation in the response variable.

Another common application of variance components is when researchers are interested in the relative size of the treatment effect compared to the within-treatment level variation. This leads to a quantity called the ***intraclass correlation coefficient*** (ICC), defined as: \[ICC = \frac{\sigma_{\text{among trts}}^{2}}{\sigma_{\text{\text{among trts}}^{2} + \sigma_{\text{within trts}}^{2}}\]

For single random factor studies, $ICC = \frac{\sigma^2_{mu}}{\sigma^2_{mu}+\sigma^2_\epsilon}$. ICC can also be thought of as the correlation between the observations within the group (i.e. corr $\left(Y_{ij}, Y_{ij'}\right)$, where $j \neq j'$. Small values of ICC indicate a large spread of values at each level of the treatment, whereas large values of ICC indicate relatively little spread at each level of the treatment:

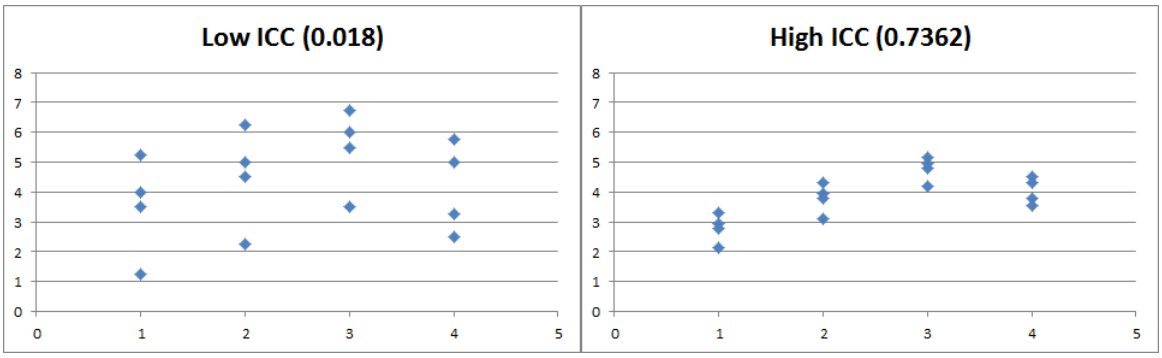


Figure 6.1.1: Dot plots for data sets with low and high ICC values.

This page titled 6.1: Random Effects is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.2: Battery Life Example

Consider a study of Battery Life, measured in hours, where 4 brands of batteries are evaluated using 4 replications in a completely randomized design (Battery Data):

Brand A	Brand B	Brand C	Brand D
110	118	108	117
113	116	107	112
108	112	112	115
115	117	108	119

A reasonable question to ask in this study would be, *should the brand of the battery be considered a fixed effect or a random effect*?

If the researchers were interested in comparing the performance of the specific brands they chose for the study, then we have a fixed effect.

But if the researchers were actually interested in studying the overall variation in battery life, so that the results would be applicable to all brands of batteries, then they may have chosen (presumably with a random sampling process) a sample of 4 of the many brands available and tested 4 batteries of each of these brands. In this latter case, the battery brand would add a dimension of variability to battery life and can be considered a random effect.

Now, let us use SAS `proc mixed;` to compare the results of battery brand as a fixed vs. random effect:

A. Fixed Effect model:

Type 3 Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Brand	3	141.687500	47.229167	Var(Residual) + Q(Brand)	MS(Residual)	12	6.21	0.0086
Residual	12	91.250000	7.604167	Var(Residual)	.	.	.	.

B. Random Effect model:

Type 3 Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Brand	3	141.687500	47.229167	Var(Residual) + 4 Var(Brand)	MS(Residual)	12	6.21	0.0086
Residual	12	91.250000	7.604167	Var(Residual)	.	.	.	.

Covariance Parameter Estimates	
Cov Parm	Estimate

Covariance Parameter Estimates	
Cov Parm	Estimate
Brand	9.9063
Residual	7.6042

We can verify the estimated variance component (arrow above) for the random treatment effect as:

$$s^2_{\text{among trts}} = \frac{MS_{trt} - MS_{error}}{n} = \frac{47.229 - 7.604}{4} = 9.9063 \tag{6.2.1}$$

With this, we can calculate the ICC as

$$ICC = \frac{9.9063}{9.9063 + 7.604} = 0.5657 \tag{6.2.2}$$

The key points in comparing these two ANOVAs are 1) *the scope of inference* and 2) *the hypothesis being tested.* For a fixed effect, the scope of inference is restricted to only 4 brands chosen for comparison and the Null hypothesis is a statement of equality of means. In contrast, as a random effect, the scope of inference is the larger population of battery brands and the Null hypothesis is a statement that the variance due to battery brand is 0.

Using R

? R: Single Random Effect

- Load the battery life data.
- Obtain the ANOVA for a single random effect.

Show Detailed Steps

1. Load the battery life data by using the following commands:

```
setwd("~/path-to-folder/")
battery_data <- read.table("battery_data.txt",header=T)
attach(battery_data)
```

2. Obtain the ANOVA for a single random effect by using the following commands:

```
library(lmerTest)
library(lme4)
battery_anova<-lmer(lifetime ~ (1 | trt),battery_data)
summary(battery_anova)
Linear mixed model fit by REML. t-tests use Satterthwaites method ['lmerModLmerT
Formula: lifetime ~ (1 | trt)
   Data: battery_data
REML criterion at convergence: 81.3
#Scaled residuals:
#     Min       1Q   Median       3Q      Max
#-1.35317 -0.69070  0.07355  0.69665  1.34279
#Random effects:
# Groups   Name        Variance Std.Dev.
# trt      (Intercept) 9.906    3.147
# Residual             7.604    2.758
#Number of obs: 16, groups:  trt, 4
```

```
#Fixed effects:
#            Estimate Std. Error      df t value Pr(>|t|)
#(Intercept)  112.938      1.718   3.000   65.73 7.76e-06 ***
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#confint(battery_anova)
#                  2.5 %     97.5 %
#.sig01        0.6530752   7.166913
#.sigma        1.9371621   4.374014
#(Intercept) 109.1585596 116.716437
```

Note that the command `lmer()` gives the ANOVA table only for the fixed effects. Therefore, in this example, since there are no fixed effects, we won't get the ANOVA table. In the "Random effects" section of the output, under the column variance we get the estimates for σ_α^2 and σ^2, which are equal to 9.906 and 7.604 respectively. In the "Fixed effects" section under the column estimate, we get the estimate of mu, or the overall mean, which is equal to 112.938. With the command `confint()` we will get confidence intervals for the standard deviations and the overall mean. If you take the square of the lower and upper bounds, you will get a confidence interval for the model variances.

Alternatively, we can use the command `aov()` which gives a partial ANOVA table.

```
battery_anova1<-aov(lifetime~Error(trt),battery_data)
summary(battery_anova1)
#Error: trt
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals  3  141.7   47.23
#Error: Within
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals 12  91.25   7.604
detach(battery_data)
```

Note that both of these commands in R don't give the F-values and p-values for the tests. Therefore, these must be done manually.

This page titled 6.2: Battery Life Example is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.3: Random Effects in Factorial and Nested Designs

Random effects can appear in both factorial and nested designs. By inspecting the EMS quantities, we can determine the appropriate F-statistic denominator for a given source. Let us look at two-factor studies.

Factorial Design

Recall the *Greenhouse* example in section 5.1.1. In this example, there were two crossed factors (*fert* and *species*). We treated both factors as fixed and the SAS `proc mixed` ANOVA table was as follows:

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
fert	3	745.437500	248.479167	Var(Residual) + Q(fert,fert*species)	MS(Residual)	40	73.10	<.0001
species	1	236.740833	236.740833	Var(Residual) + Q(species,fert*species)	MS(Residual)	40	69.65	<.0001
fert*species	3	50.584167	16.861389	Var(Residual) + Q(fert*species)	MS(Residual)	40	4.96	0.0051
Residual	40	135.970000	3.399250	Var(Residual)	.	.	.	.

If we inspect the EMS quantities in the output, we see that the correct denominator for all F-tests when both factors are fixed in the 2-factor crossed study is Error Mean Squares.

Now let us consider a case in which both factors A and B are random effects in the factorial design (i.e. factors A and B are crossed, and both are random effects). The expected mean squares for each of the source of variations in the ANOVA model would be as follows:

Source	**EMS**
A	$\sigma^2 + nb\sigma^2_{\alpha} + n\sigma^2_{\alpha\beta}$
B	$\sigma^2 + na\sigma^2_{\beta} + n\sigma^2_{\alpha\beta}$
A × B	$\sigma^2 + n\sigma^2_{\alpha\beta}$
Error	σ^2
Total	

The F-tests following from the EMS above would be:

Source	EMS	F
A	$\sigma^2 + nb\sigma^2_\alpha + n\sigma^2_{\alpha\beta}$	MSA / MSAB
B	$\sigma^2 + na\sigma^2_\beta + n\sigma^2_{\alpha\beta}$	MSB / MSAB
A × B	$\sigma^2 + n\sigma^2_{\alpha\beta}$	MSAB / MSE
Error	σ^2	
Total		

Here we can see the ramifications of having random effects. In fixed-effects models, the denominator for the F-statistics in significance testing was the mean square error (MSE). In random-effects models, however, we may have to choose different denominators depending on the term we are testing.

The F-statistic for testing the significance of a given effect, in general, is the ratio of the two MS values with MS of the effect as the numerator, and the denominator MS is chosen such that the F-statistic equals 1 if H_0 is true and greater than 1 if H_a is true.

Following this logic, we can see that when testing for the interaction effect of 2 random factors, the correct denominator is the error mean squares. Therefore the test statistic for testing $A \times B$ is $\frac{MSAB}{MSE}$. However, when we are testing for the main effect of factor A, the correct denominator would be $MSAB$.

Recall that the EMS quantities are the population counterparts for the MS values which actually are sample statistics. Examination of EMS expressions can therefore be used to choose the correct denominator for an F-statistic utilized for testing significance and will be discussed in detail in Section 6.7.

Nested Design

In the case of a nested design, where factor B is nested within the levels of factor A and both are random effects, the expected mean squares for each of the source of variations in the ANOVA model would be as follows:

Source	EMS
A	$\sigma^2 + bn\sigma^2_\alpha + n\sigma^2_\beta$
B(A)	$\sigma^2 + n\sigma^2_\beta$
Error	σ^2
Total	

The F-tests follow from the EMS above:

Source	EMS	F
A	$\sigma^2 + bn\sigma^2_\alpha + n\sigma^2_\beta$	MSA / MSB(A)
B(A)	$\sigma^2 + n\sigma^2_\beta$	MSB(A) / MSE
Error	σ^2	
Total		

Using R

Greenhouse Data - Two Random Effects with Interaction

- Load the greenhouse data.
- Obtain the ANOVA for two random effects with interaction.

Show Detailed Steps

1. Load the greenhouse data by using the following commands:

```
setwd("~/path-to-folder/")
greenhouse_2way_data <-read.table("greenhouse_2way_data.txt",header=T)
attach(greenhouse_2way_data)
```

2. Obtain the ANOVA for two random effects with interaction by using the following commands:

```
library(lmerTest)
library(lme4)
greenhouse_anova<-lmer(height ~ (1 | fertilizer) + (1 | species) + (1 | fertiliz
summary(greenhouse_anova)

Linear mixed model fit by REML. t-tests use Satterthwaites method ['lmerModLmerT
Formula: height ~ (1 | fertilizer) + (1 | species) + (1 | fertilizer:species)
    Data: greenhouse_2way_data

REML criterion at convergence: 216.7
#Scaled residuals:
#     Min        1Q    Median        3Q       Max
#-2.46787  -0.38510  0.03012  0.38780  2.63056

#Random effects:
# Groups              Name       Variance  Std.Dev.
# fertilizer:species (Intercept)    2.244  1.498
# fertilizer         (Intercept)   19.301  4.393
# species            (Intercept)    9.162  3.027
# Residual                          3.399  1.844
# Number of obs: 48, groups: fertilizer:species, 8; fertilizer, 4; species, 2

#Fixed effects:
#            Estimate Std.  Error     df t value Pr(>|t|)
#(Intercept)     28.387     3.124  2.859 9.088 0.0034 **
#---
#Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
confint(greenhouse_anova)
#                 2.5 %     97.5 %
#.sig01      0.4327681   5.482701
#.sig02      0.0000000  10.319191
#.sig03      0.0000000  11.585745
#.sigma      1.5031328   2.335330
#(Intercept) 21.1262902 35.648887
```

Note that the command `lmer()` gives the ANOVA table only for the fixed effects. Therefore, in this example, since there are no fixed effects, we won't get the ANOVA table. In the "Random effects" section of the output, under the column variance we get the estimates for $\sigma^2_{\alpha\beta}$, σ^2_{α}, σ^2_{β}, and σ^2 which are equal to 2.244, 19.301, 9.162, and 3.399 respectively. In the "Fixed effects" section under the column estimate we get the estimate of μ, or the overall mean, which is equal to 28.387.

With the command `confint()` we will get confidence intervals for the standard deviations and the overall mean. If you take the square of the lower and upper bounds, you will get a confidence interval for the model variances.

Alternatively, we can use the command `aov()` which gives a partial ANOVA table.

```
greenhouse_anova1<-aov(height~Error(fertilizer+species+fertilizer:species),green
summary(greenhouse_anova1)
#Error: fertilizer
#           Df  Sum Sq  Mean Sq  F value  Pr(>F)
#Residuals  3   745.4    248.5

#Error: species
#           Df  Sum Sq  Mean Sq  F value  Pr(>F)
#Residuals  1   236.7    236.7

#Error: fertilizer:species
#           Df  Sum Sq  Mean Sq  F value  Pr(>F)
#Residuals  3   50.58    16.86

#Error: Within
#            Df  Sum Sq  Mean Sq  F value  Pr(>F)
#Residuals  40     136    3.399
detach(greenhouse_2way_data)
```

Note that both commands in R don't give the F-values and the p-values for the tests. Therefore, these must be done manually.

This page titled 6.3: Random Effects in Factorial and Nested Designs is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.4: Special Case - Fully Nested Random Effects Design

Here, we will consider a special case of random effects models where each factor is nested within the levels of the next "order" of a hierarchy. This Fully Nested Random Effects model is similar to Russian Matryoshka dolls, where the smaller dolls are nested within the next larger one.

Consider 3 random factors A, B, and C that are hierarchically nested. That is, C is nested in (B, A) combinations and B is nested within levels of A. Suppose there are n observations made at the lowest level.

The statistical model for this case is:

$$Y_{ijkl} = \mu + \alpha_i + \beta_{i(j)} + \gamma_{k(ij)} + \epsilon_{ijkl} \tag{6.4.1}$$

where $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, $k = 1, 2, \ldots, c$ and $l = 1, 2, \ldots, n$.

We will also have $\epsilon_{ijkl} \overset{iid}{\sim} \mathcal{N}(0, \sigma_2)$, $\gamma_{k(ij)} \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2_\gamma\right)$, $\beta_{i(j)} \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2_\beta\right)$, and $\alpha_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2_\alpha\right)$.

The DFs and expected mean squares for this design would be as follows:

Source	DF	EMS	F
A	$\boldsymbol{(a-1)}$	$\sigma^2_\epsilon + n\sigma^2_\gamma + nc\sigma^2_\beta + ncb\sigma^2_\alpha$	MSA / MSB(A)
B(A)	$\boldsymbol{a(b-1)}$	$\sigma^2_\epsilon + n\sigma^2_\gamma + nc\sigma^2_\beta$	MSB(A) / MSC(AB)
C(A,B)	$\boldsymbol{ab(c-1)}$	$\sigma^2_\epsilon + n\sigma^2_\gamma$	MSC(AB) / MSE
Error	$\boldsymbol{abc(n-1)}$	σ^2_ϵ	
Total	$\boldsymbol{abcn-1}$		

In this case, each F-test we construct for the sources will be based on different denominators.

This page titled 6.4: Special Case - Fully Nested Random Effects Design is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.5: Quality Control Example

Example - Fully Nested Random Effects Model

The temperature of a process in a manufacturing industry is critical to quality control. The researchers want to characterize the sources of this variability. They choose 4 plants and 4 operators within each plant, look at 4 shifts for each operator, and then measure temperature for each of the three batches used in production.

Collected data was read into SAS and `proc mixed` procedure was used to obtain the ANOVA model.

Show SAS Code

```
data fullnest;
input Temp Plant Operator Shift Batch;
datalines;
477    1    1    1    1
472    1    1    1    2
481    1    1    1    3
478    1    1    2    1
475    1    1    2    2
474    1    1    2    3
472    1    1    3    1
475    1    1    3    2
468    1    1    3    3
482    1    1    4    1
477    1    1    4    2
474    1    1    4    3
471    1    2    1    1
474    1    2    1    2
470    1    2    1    3
479    1    2    2    1
482    1    2    2    2
477    1    2    2    3
470    1    2    3    1
477    1    2    3    2
483    1    2    3    3
480    1    2    4    1
473    1    2    4    2
478    1    2    4    3
475    1    3    1    1
472    1    3    1    2
470    1    3    1    3
460    1    3    2    1
469    1    3    2    2
472    1    3    2    3
477    1    3    3    1
483    1    3    3    2
475    1    3    3    3
476    1    3    4    1
```

```
480    1    3    4    2
471    1    3    4    3
465    1    4    1    1
464    1    4    1    2
471    1    4    1    3
477    1    4    2    1
475    1    4    2    2
471    1    4    2    3
481    1    4    3    1
477    1    4    3    2
475    1    4    3    3
470    1    4    4    1
475    1    4    4    2
474    1    4    4    3
484    2    1    1    1
477    2    1    1    2
481    2    1    1    3
477    2    1    2    1
482    2    1    2    2
481    2    1    2    3
479    2    1    3    1
477    2    1    3    2
482    2    1    3    3
477    2    1    4    1
470    2    1    4    2
479    2    1    4    3
472    2    2    1    1
475    2    2    1    2
475    2    2    1    3
472    2    2    2    1
475    2    2    2    2
470    2    2    2    3
472    2    2    3    1
477    2    2    3    2
475    2    2    3    3
482    2    2    4    1
477    2    2    4    2
483    2    2    4    3
485    2    3    1    1
481    2    3    1    2
477    2    3    1    3
482    2    3    2    1
483    2    3    2    2
485    2    3    2    3
477    2    3    3    1
476    2    3    3    2
481    2    3    3    3
```

```
479    2    3    4    1
476    2    3    4    2
485    2    3    4    3
477    2    4    1    1
475    2    4    1    2
476    2    4    1    3
476    2    4    2    1
471    2    4    2    2
472    2    4    2    3
475    2    4    3    1
475    2    4    3    2
472    2    4    3    3
481    2    4    4    1
470    2    4    4    2
472    2    4    4    3
475    3    1    1    1
470    3    1    1    2
469    3    1    1    3
477    3    1    2    1
471    3    1    2    2
474    3    1    2    3
469    3    1    3    1
473    3    1    3    2
468    3    1    3    3
477    3    1    4    1
475    3    1    4    2
473    3    1    4    3
470    3    2    1    1
466    3    2    1    2
468    3    2    1    3
471    3    2    2    1
473    3    2    2    2
476    3    2    2    3
478    3    2    3    1
480    3    2    3    2
474    3    2    3    3
477    3    2    4    1
471    3    2    4    2
469    3    2    4    3
466    3    3    1    1
465    3    3    1    2
471    3    3    1    3
473    3    3    2    1
475    3    3    2    2
478    3    3    2    3
471    3    3    3    1
469    3    3    3    2
```

```
471    3    3    3    3
475    3    3    4    1
477    3    3    4    2
472    3    3    4    3
469    3    4    1    1
471    3    4    1    2
468    3    4    1    3
473    3    4    2    1
475    3    4    2    2
473    3    4    2    3
477    3    4    3    1
470    3    4    3    2
469    3    4    3    3
463    3    4    4    1
471    3    4    4    2
469    3    4    4    3
484    4    1    1    1
477    4    1    1    2
480    4    1    1    3
476    4    1    2    1
475    4    1    2    2
474    4    1    2    3
475    4    1    3    1
470    4    1    3    2
469    4    1    3    3
481    4    1    4    1
476    4    1    4    2
472    4    1    4    3
469    4    2    1    1
475    4    2    1    2
479    4    2    1    3
482    4    2    2    1
483    4    2    2    2
479    4    2    2    3
477    4    2    3    1
479    4    2    3    2
475    4    2    3    3
472    4    2    4    1
476    4    2    4    2
479    4    2    4    3
470    4    3    1    1
481    4    3    1    2
481    4    3    1    3
475    4    3    2    1
470    4    3    2    2
475    4    3    2    3
469    4    3    3    1
```

```
477    4    3    3    2
482    4    3    3    3
485    4    3    4    1
479    4    3    4    2
474    4    3    4    3
469    4    4    1    1
473    4    4    1    2
475    4    4    1    3
477    4    4    2    1
473    4    4    2    2
471    4    4    2    3
470    4    4    3    1
468    4    4    3    2
474    4    4    3    3
483    4    4    4    1
477    4    4    4    2
476    4    4    4    3
;
proc mixed data=fullnest covtest method=type3;
class Plant Operator Shift Batch;
model temp=;
random plant operator(plant) shift(plant operator) ;
run;
```

In the SAS code, notice that there are no terms on the right-hand side of the model statement. This is because SAS uses the model statement to specify **fixed effects** only. The random statement is used to specify the random effects. The `proc mixed` procedure will perform the fully nested random effects model as specified above, and produces the following output:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Plant	3	731.515625	243.838542	Var(Residual) + 3 Var(Shift(Plant*Operator)) + 12 Var(Operator(Plant)) + 48 Var(Plant)	MS(Operator(Plant))	12	5.85	0.0106
Operator(Plant)	12	499.812500	41.651042	Var(Residual) + 3 Var(Shift(Plant*Operator)) + 12 Var(Operator(Plant))	MS(Shift(Plant*Operator))	48	1.30	0.2483

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Shift(Plant*Operato)	48	1534.916667	31.977431	Var(Residual) + 3 Var(Shift(Plant*Operato))	MS(Residual)	128	2.58	<.0001
Residual	128	1588.000000	12.406250	Var(Residual)	.	.	.	.

Covariance Parameter Estimates				
Cov Parm	Estimate	Standard Error	Z Value	Pr Z
Plant	4.2122	4.1629	1.01	0.3116
Operator(Plant)	0.8061	1.5178	0.53	0.5953
Shift(Plant*Operato)	6.5237	2.2364	2.92	0.0035
Residual	12.4063	1.5508	8.00	<.0001

The largest (and significant) variance components are: (1) the shift within a plant × operator combination and (2) the batch-to-batch variation within the shift (the residual).

Note that the *Covariance Parameter Estimates* here are in fact the variance components. SAS does not express the variance components as percentages in this procedure, but by summing the variance components for all sources to serve as the denominator, each source can be expressed as a percentage.

Because this type of model is so commonly employed, SAS also offers two other procedures to obtain the variance components results: `proc varcomp` (which stands for variance components) and `proc nested` .

The equivalent code for these procedures is as follows:

The `proc varcomp` :

```
proc varcomp data=fullnest;
class Plant Operator Shift Batch;
model temp= plant operator(plant) shift(plant operator);
run;
```

Note that the model statement for `proc varcomp` differs from the mixed procedure, in that `proc varcomp` assumes that the factors listed in the model statement are random effects.

Partial Output:

MIVQUE(0) Estimates	
Variance Component	Temp
Var(Plant)	4.21224
Var(Operator(Plant))	0.80613
Var(Shift(Plant*Operato))	6.52373

MIVQUE(0) Estimates	
Variance Component	Temp
Var(Error)	12.40625

Note that, even in this procedure we will have to use the sum for a total and calculate the percentages ourselves.

The `proc nested`

On the other hand, the `proc nested` procedure will provide the full output including the percentages:

```
proc nested data=fullnest;
class plant operator shift;
var temp;
run;
```

Partial Output:

Nested Random Effects Analysis of Variance for Variable Temp								
Variance Source	DF	Sum of Squares	F Value	Pr > F	Error Term	Mean Square	Variance Component	Percent of Total
Total	191	4354.244792				22.797093	23.948351	100.0000
Plant	3	731.515625	5.85	0.0106	Operator	243.838542	4.212240	17.5889
Operator	12	499.812500	1.30	0.2483	Shift	41.651042	0.806134	3.3661
Shift	48	1534.916667	2.58	<.0001	Error	31.977431	6.523727	27.2408
Error	128	1588.000000				12.406250	12.406250	51.8042

Calculation of the Variance Components

From the SAS output, we get the EMS coefficients. We can use those to compute the estimated variance components.

Source	MS	EMS	Variance Components	% Variation
Plant	243.84	$\sigma_\epsilon^2 + 3\sigma_\gamma^2 + 12\sigma_\beta^2 + 48\sigma_\alpha^2$	4.21	17.58
Operator(Plant)	41.65	$\sigma_\epsilon^2 + 3\sigma_\gamma^2 + 12\sigma_\beta^2$	0.806	3.37
Shift(Plant × Operator)	31.98	$\sigma_\epsilon^2 + 3\sigma_\gamma^2$	6.52	27.24
Residual	12.41	σ_ϵ^2	12.41	51.80
		Total	23.95	

One can show that MS is an unbiased estimator for EMS (using the properties of Method of Moments estimates). With that, we can algebraically solve for each variance component. Start at the bottom of the table and work up the hierarchy.

First of all, the estimated variance component for the Residuals is given:

$$\mathbf{12.41} = \hat{\sigma}^2_{\text{error}} = \hat{\sigma}^2_\epsilon$$

Then we can use this information and subtract it from the Shift(Plant × Operator) MS to get:

$$31.98 = \hat{\sigma}^2_\epsilon + 3\hat{\sigma}^2_{\gamma \text{ or Shift(Plant} \times \text{Operator)}}$$

$$\hat{\sigma}^2_\gamma = \frac{31.98 - 12.41}{3} = \mathbf{6.52}$$

Similarly, we use what we know for Error and Shift(Plant × Operator) and subtract it from the Operator(Plant) MS to get:

$$\begin{aligned} 41.65 &= \hat{\sigma}^2_\epsilon + 3\hat{\sigma}^2_\gamma + 12\hat{\sigma}^2_{\beta \text{ or Operator(Plant)}} \\ &= 31.98 + 12\hat{\sigma}^2_\beta \\ \sigma^2_\beta &= \frac{41.65 - 31.98}{12} \\ &= \mathbf{0.806} \end{aligned}$$

Our total = 12.41 + 6.52 + 0.806 + 4.21 = 23.95

Then, dividing each variance component by the total (in this case 23.95) gives the % values shown in the output from SAS `proc nested` .

This page titled 6.5: Quality Control Example is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.5.1: Using Minitab

Minitab has a separate program just for this type of analysis for our example (Quality Data), under:

`Stat` > `ANOVA` > `Fully Nested ANOVA`

and you specify the model in the boxes provided:

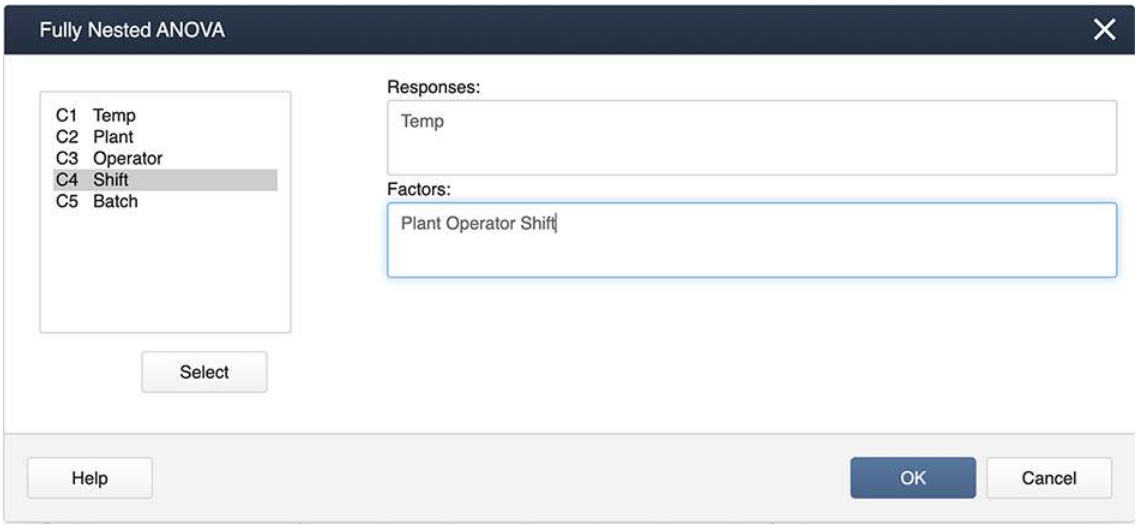


Figure 6.5.1.1: Fully Nested ANOVA pop-up window.

The output you get is very comprehensive and includes the variance components expressed as percentages.

Nested ANOVA: Temp versus Plant, Operator, Shift

Analysis of Variance for Temp

Source	DF	SS	MS	F	P
Plant	3	731.5156	243.8385	5.854	0.011
Operator	12	499.8125	41.6510	1.303	0.248
Shift	48	1534.9167	31.9774	2.578	0.000
Error	128	1588.0000	12.4062		
Total	191	4354.2448			

Variance Components

Source	Var Comp.	# of Total	StDev
Plant	4.212	17.59	2.052
Operator	0.806	3.37	0.898
Shift	6.524	27.24	2.554
Error	12.406	51.80	3.522
Total	23.948		4.894

Expected Mean Squares

1	Plant	1.00(4) + 3.00(3) + 12.00(2) + 48.00(1)
2	Operator	1.00(4) + 3.00(3) + 12.00(2)

3	Shift	1.00(4) + 3.00(3)
4	Error	1.00(4)

This page titled 6.5.1: Using Minitab is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.5.2: Using R

R Fully Nested Random Effects Model

- Load the data.
- Obtain the ANOVA for the fully nested random effects.

1. Load the data by using the following commands:

```
setwd("~/path-to-folder/")
fullnest_data <- read.table("fullnest_data.txt",header=T)
attach(fullnest_data)
```

2. Obtain the ANOVA for the fully nested random effects by using the following commands:

```
library(lmerTest)
library(lme4)
random_fullnest<-lmer(Temp ~ (1 | Plant) + (1 | Plant:Operator) +
(1 | Plant:(Operator:Shift)) ,fullnest_data)
summary(random_fullnest)
Linear mixed model fit by REML. t-tests use Satterthwaites method ['lmerModLmerTest']
Formula: Temp ~ (1 | Plant) + (1 | Plant:Operator) + (1 | Plant:(Operator:Shift))
   Data: fullnest_data
REML criterion at convergence: 1097.2

#Scaled residuals:
#     Min       1Q   Median       3Q      Max
#-2.78620 -0.61163  0.00414  0.56721  1.99397

#Random effects:
# Groups                 Name        Variance Std.Dev.
# Plant:(Operator:Shift) (Intercept)  6.5237  2.5542
# Plant:Operator         (Intercept)  0.8061  0.8979
# Plant                  (Intercept)  4.2123  2.0524
# Residual                           12.4063  3.5223
# Number of obs: 192, groups:  Plant:(Operator:Shift), 64; Plant:Operator, 16; Plant,

#Fixed effects:
#            Estimate Std. Error      df t value Pr(>|t|)
#(Intercept)  474.880      1.127   3.000   421.4 2.95e-08 ***
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
confint(random_fullnest)

 #                2.5 %     97.5 %
#.sig01        1.7251242   3.487550
#.sig02        0.0000000   2.475048
#.sig03        0.1192372   4.695585
```

```
#.sigma        3.1311707   4.002066
#(Intercept) 472.4015615 477.358858
```

Note that the command `lmer()` gives the ANOVA table only for the fixed effects. Therefore, in this example, since there are no fixed effects, we won't get the ANOVA table. In the "Random effects" section of the output, under the column variance, we get the estimates for σ^2_γ, σ^2_β, σ^2_α, and σ^2 which are equal to 6.5237, 0.8061, 4.2123, and 12.4063 respectively. In the "Fixed effects" section under the column estimate, we get the estimate of μ for the overall mean, which is equal to 474.880.

With the command `confint()` we will get confidence intervals for the standard deviations and the overall mean. If you take the square of the lower and upper bounds, you will get a confidence interval for the model variances.

Alternatively, we can use the command aov() which gives a partial ANOVA table.

```
random_fullnest1<-aov(Temp ~ Error(factor(Plant) + factor(Plant)/factor(Operator) + f
summary(random_fullnest1)

#Error: factor(Plant)
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals  3  731.5   243.8

#Error: factor(Plant):factor(Operator)
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals 12  499.8   41.65

#Error: factor(Plant):factor(Operator):factor(Shift)
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals 48   1535   31.98

#Error: Within
#            Df Sum Sq Mean Sq F value Pr(>F)
# Residuals 128   1588   12.41
detach(fullnest_data)
```

Note that both commands in R don't give the F-values and the p-values for the tests. Therefore, these must be done manually.

This page titled 6.5.2: Using R is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.6: Introduction to Mixed Models

Treatment designs can comprise both fixed and random effects. When we have this situation the treatment design is referred to as a mixed model. Mixed models are by far the most commonly encountered treatment designs. The three situations we now have are often referred to as Model I (fixed effects only), Model II (random effects only), and Model III (mixed) ANOVAs. In designating the effects of a mixed model as mixed or random, the following rule will be useful.

> **Rule!** Any interaction or nested effect containing at least one random factor is random.

Below are the ANOVA layouts of two basic mixed models with two factors.

Factorial

In the simplest case of a balanced mixed model, we may have two factors, A and B, in a factorial design in which factor A is a fixed effect and factor B is a random effect.

The statistical model is similar to what we have seen before:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{6.6.1}$$

where $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, and $k = 1, 2, \ldots, n$.

Here, $\sum_i \alpha_i = 0$, $\beta_j \sim \mathcal{N}\left(0, \sigma^2_\beta\right)$, $(\alpha\beta)_{i,j} \sim \mathcal{N}\left(0, \frac{a-1}{a}\sigma^2_{\alpha\beta}\right)$, $\sum_i (\alpha\beta)_{i,j} = 0$ and $\epsilon_{ijk} \sim \mathcal{N}\left(0, \sigma^2\right)$. Also, β_j, $(\alpha\beta)_{ij}$, and ϵ_{ij} are pairwise independent.

In this case, we have the following ANOVA.

Source	DF	EMS
A	$(a-1)$	$\sigma^2 + nb\frac{\sum \alpha_i^2}{a-1} + n\sigma^2_{\alpha\beta}$
B	$(b-1)$	$\sigma^2 + na\sigma^2_\beta$
A × B	$(a-1)(b-1)$	$\sigma^2 + n\sigma^2_{\alpha\beta}$
Error	$ab(n-1)$	σ^2
Total	$abn-1$	

The F-tests are set up based on the EMS column above and we can see that we have to use different denominators in testing significance for the various sources in the ANOVA table:

Source	EMS	F
A	$\sigma^2 + nb\frac{\sum \alpha_i^2}{a-1} + n\sigma^2_{\alpha\beta}$	MSA / MSAB
B	$\sigma^2 + na\sigma^2_\beta$	MSB / MSE
A × B	$\sigma^2 + n\sigma^2_{\alpha\beta}$	MSAB / MSE
Error	σ^2	
Total		

As a reminder, the null hypothesis for a fixed effect is that the α_i's are equal, whereas the null hypothesis for the random effect is that the σ^2_β's are equal to zero.

Note

The denominator for the F-test for the main effect of factor A is now the MS for the A × B interaction. For Factor B and the A × B interaction, the denominator is the MSE.

Nested

In the case of a balanced nested treatment design, where A is a fixed effect and B(A) is a random effect, the statistical model would be:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk} \tag{6.6.2}$$

where $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, and $k = 1, 2, \ldots, n$.

Here, $\sum_i \alpha_i = 0$, $\beta_{j(i)} \sim \mathcal{N}\left(0, \sigma^2_\beta\right)$, and $\epsilon_{ijk} = \mathcal{N}\left(0, \sigma^2\right)$.

We have the following ANOVA for this model:

Source	DF	EMS
A	$(a-1)$	$\sigma^2_\epsilon + n\sigma^2_{\beta(\alpha)} + bn\frac{\sum \alpha_i^2}{a-1}$
B(A)	$a(b-1)$	$\sigma^2_\epsilon + n\sigma^2_{\beta(\alpha)}$
Error	$ab(n-1)$	σ^2_ϵ
Total	$abn-1$	

Here is the same table with the F-statistics added. Note that the denominators for the F-test are different.

Source	EMS	F
A	$\sigma^2_\epsilon + n\sigma^2_{\beta(\alpha)} + bn\frac{\sum \alpha_i^2}{a-1}$	MSA / MSB(A)
B(A)	$\sigma^2_\epsilon + n\sigma^2_{\beta(\alpha)}$	MSB(A) / MSE
Error	σ^2_ϵ	
Total		

F-Calculation Facts

As can be seen from the examples above and also from sections 6.3-6.6, when significance testing in random or mixed models, the denominator of the F-statistic is no more the MSE value and has to be aptly chosen. Recall that the F-statistic for testing the significance of a given effect is the ratio with the numerator equal to the MS value of the effect, and the denominator is also an MS value of an effect included in the ANOVA model. Furthermore, the F-statistic has a non-central distribution when H_a is true and a central F-distribution when H_0 is true.

The non-centrality parameter of the non-central F distribution when H_a is true depends on the type of effect (fixed vs random), and equals \sum_{i=1}^{T} \alpha_{i}^{2}\) for a fixed effect and σ^2_{trt} for a random effect. Here $\alpha_i = \mu_i - \mu$, where μ_i $(i = 1, 2, \ldots, T)$ is the i^{th} level of the fixed effect and μ is the overall mean while σ^2_{trt} is the variance component associated with the random effect. Also, MS under true H_a equals to MS under true H_0 plus non-centrality parameter, so that

$$F\text{-statistic} = \frac{\text{MS when } H_0 \text{ is true} + \text{non-centrality parameter}}{\text{MS when } H_0 \text{ is true}} \tag{6.6.3}$$

The above identity can be used to identify the correct denominator (also called the error term) with the aid of EMS expressions displayed in the ANOVA table.

> **Rule!** The F-statistic denominator is the MS value of the source which has an EMS containing all EMS terms in the effect except the non-centrality parameter.

This page titled 6.6: Introduction to Mixed Models is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.7: Mixed Model Example

Consider the experimental setting in which the investigators are interested in comparing the classroom self-ratings of teachers. They created a tool that can be used to self-rate the classrooms. The investigators are interested in comparing the Eastern vs. Western US regions, and the type of school (Public vs. Private). Investigators chose 2 teachers randomly from each combination and each teacher submits scores from 2 classes that they teach.

You can download the data at **Schools Data**.

If we carefully disseminate the information in the setup, we see that the US region makes sense as a fixed effect, and so does the type of school. However, the investigators are probably not interested in testing for significant differences among individual teachers they recruited for the study; more realistically, they would be interested in how much variation there is among teachers (a random effect).

For this example, we can use a *mixed model* in which we model *teacher* as a random effect nested within the factorial fixed treatment combinations of *Region* and *School type*.

This page titled 6.7: Mixed Model Example is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.7.1: Using Minitab

In Minitab, specifying the mixed model is a little different.

In **`Stat`** > **`ANOVA`** > **`General Linear Model`** > **`Fit General Linear Model`**

we complete the dialog box:

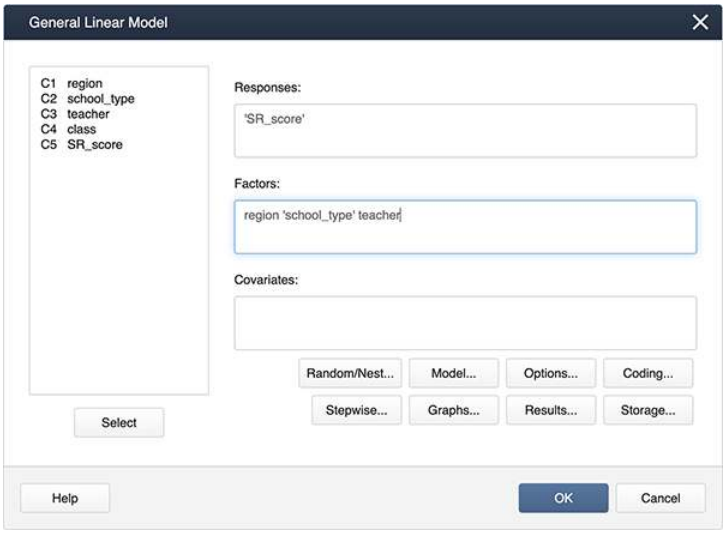


Figure 6.7.1.1: General Linear Model pop-up window.

We can create interaction terms under **Model**… by selecting "region" and "school_type" and clicking **Add**.

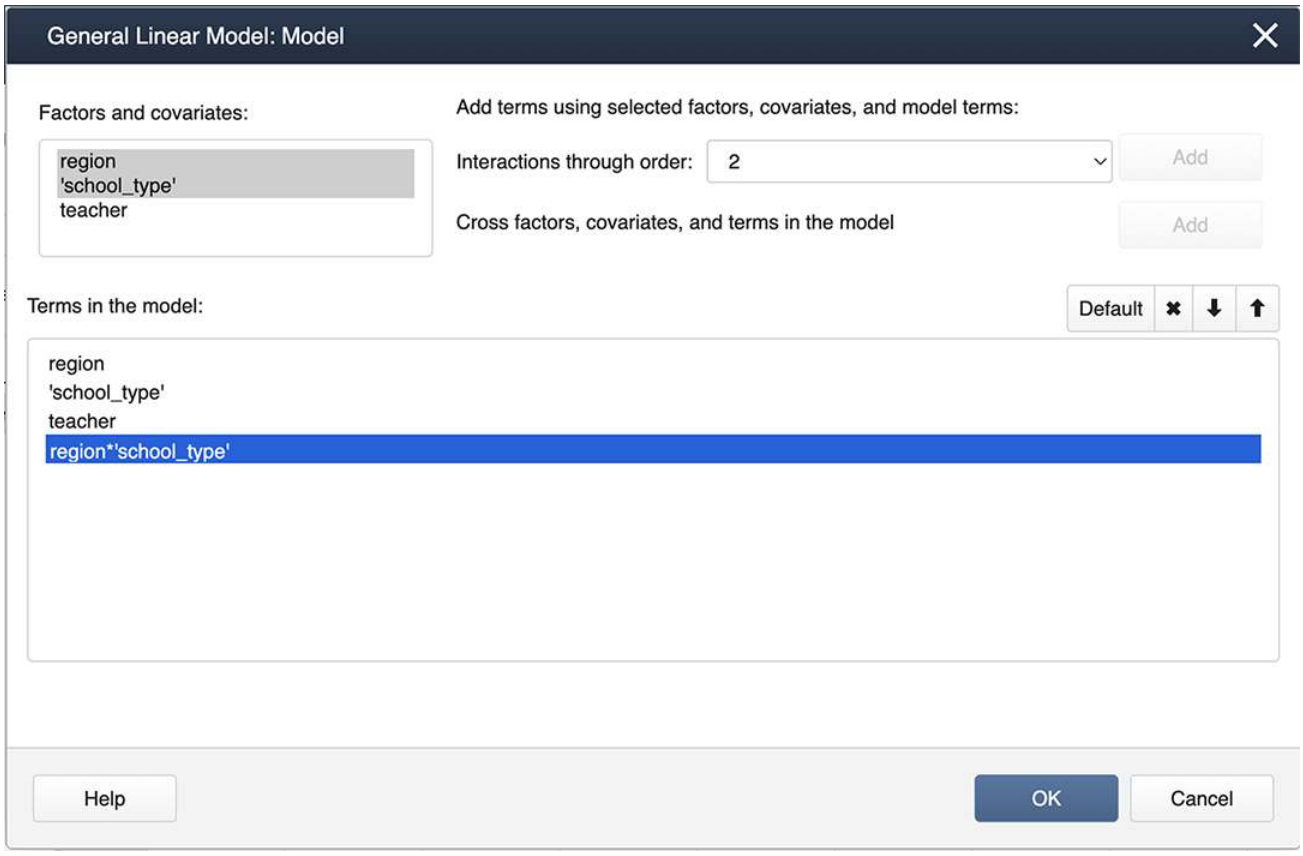


Figure 6.7.1.2: General Linear Model: Model pop-up window.

Finally, we create nested terms and effects are random under **Random/Nest…**:

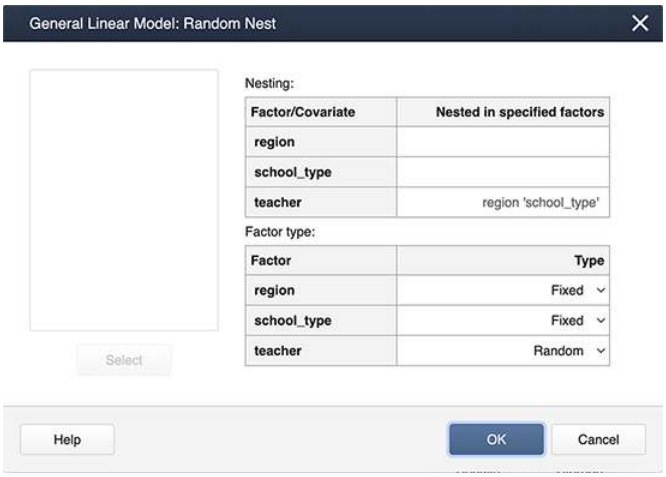


Figure 6.7.1.3: General Linear Model: Random Nest pop-up window.

Minitab Output for the mixed model:

Factor Information

Factor	Type	Levels	Values
region	Fixed	2	EastUS, WestUS
school_type	Fixed	2	Private, Public
teacher(region school_type)	Random	8	1(EastUS,Private), 2(EastUS,Private,) 1(EastUS,Public), 2(EastUS, Public), 1(WestUS, Private), 2(WestUS, Private), 1(WestUS,Public), 2(WestUS,Public)

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F-Value	P-Value
region	1	564.06	564.06	564.06	24.07	0.008
school_type	1	76.56	76.56	76.56	3.27	0.145
region*school_type	1	264.06	264.06	264.06	11.27	0.028
teacher(region schoo_type)	4	93.75	93.75	23.44	5.00	0.026
Error	8	37.50	37.50	4.69		
Total	15	1035.94				

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.16506	96.38%	93.21%	85.52%

Minitab's results are in agreement with SAS `Proc Mixed` .

This page titled 6.7.1: Using Minitab is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.7.2: Using SAS

In SAS we would set up the ANOVA as:

```
proc mixed data=school covtest method=type3;
    class Region SchoolType Teacher Class;
    model sr_score = Region SchoolType Region*SchoolType;
    random Teacher(Region*SchoolType);
    store out_school;
run;
```

In SAS `proc mixed` , we see that the fixed effects appear in the model statement, and the nested random effect appears in the random statement.

We get the following partial output:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Region	1	564.062500	564.062500	Var(Residual) + 2 Var(Teach(Region*School)) + Q(Region,Region*SchoolType)	MS(Teach(Region*School))	4	24.07	0.0080
SchoolType	1	76.562500	76.562500	Var(Residual) + 2 Var(Teach(Region*School)) + Q(SchoolType,Region*SchoolType)	MS(Teach(Region*School))	4	3.27	0.1450
Region*SchoolType	1	264.062500	264.062500	Var(Residual) + 2 Var(Teach(Region*School)) + Q(Region*SchoolType)	MS(Teach(Region*School))	4	11.27	0.0284
Teach(Region*School)	4	93.750000	23.437500	Var(Residual) + 2 Var(Teach(Region*School))	MS(Residual)	8	5.00	0.0257
Residual	8	37.500000	4.687500	Var(Residual)	.	.	.	.

The results for hypothesis tests for the fixed effects appear as:

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Region	1	4	24.07	0.0080
SchoolType	1	4	3.27	0.1450
Region*SchoolType	1	4	11.27	0.0284

Given that the *Region*SchoolType* interaction is significant, the `PLM` procedure along with the `lsmeans` statement can be used to generate the Tukey mean comparisons and produce the groupings chart and the plots to identify what means differ significantly.

```
ods graphics on;
proc plm restore=out_school;
lsmeans Region*SchoolType / adjust=tukey plot=meanplot cl lines;
run;
```

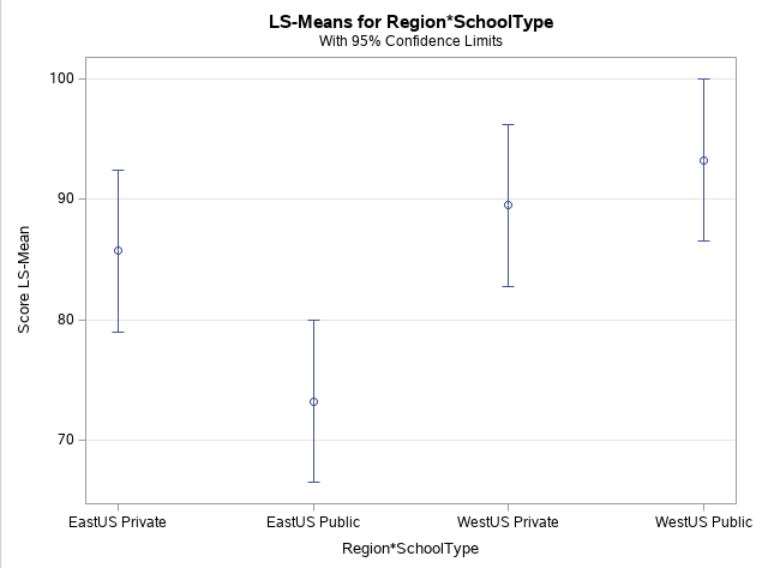


Figure 6.7.2.1: Plot of score LS-means for Region*SchoolType, with 95% confidence limits.

Differences of Region*SchoolType Least Squares Means Adjustment for Multiple Comparisons: Tukey														
Region	SchoolType	_Region	_SchoolType	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
EastUS	Private	EastUS	Public	12.5000	3.4233	4	3.65	0.0217	0.0703	0.05	2.9955	22.0045	-1.4356	26.4356
EastUS	Private	WestUS	Private	-3.7500	3.4233	4	-1.10	0.3349	0.7109	0.05	-13.2545	5.7545	-17.6856	10.1856

Differences of Region*SchoolType Least Squares Means Adjustment for Multiple Comparisons: Tukey													
Region	SchoolType	_Region	_SchoolType	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
EastUS	Private	WestUS	Public	-7.5000	3.4233	4	-2.19	0.0936	0.2677	0.05	-17.0045	2.0045	-21.4356	6.4356
EastUS	Public	WestUS	Private	-16.2500	3.4233	4	-4.75	0.0090	0.0301	0.05	-25.7545	-6.7455	-30.1856	-2.3144
EastUS	Public	WestUS	Public	-20.0000	3.4233	4	-5.84	0.0043	0.0146	0.05	-29.5045	-10.4955	-33.9356	-6.0644
WestUS	Private	WestUS	Public	-3.7500	3.4233	4	-1.10	0.3349	0.7109	0.05	-13.2545	5.7545	-17.6856	10.1856

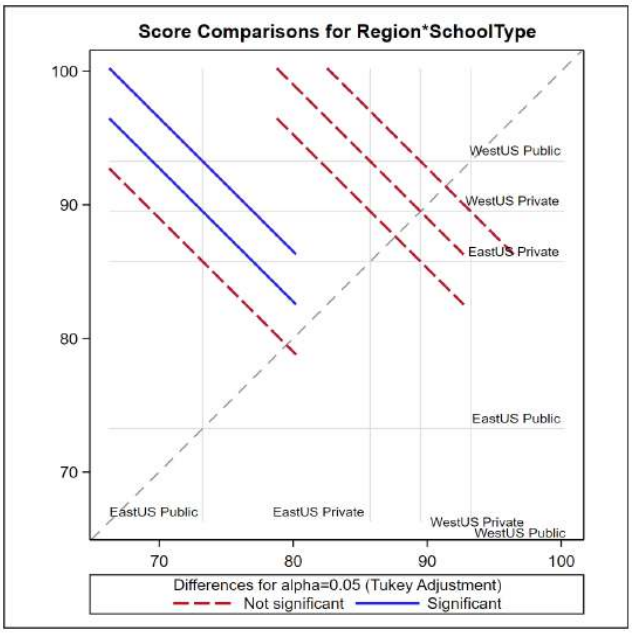


Figure 6.7.2.2: Diffogram of score comparisons for Region*SchoolType.

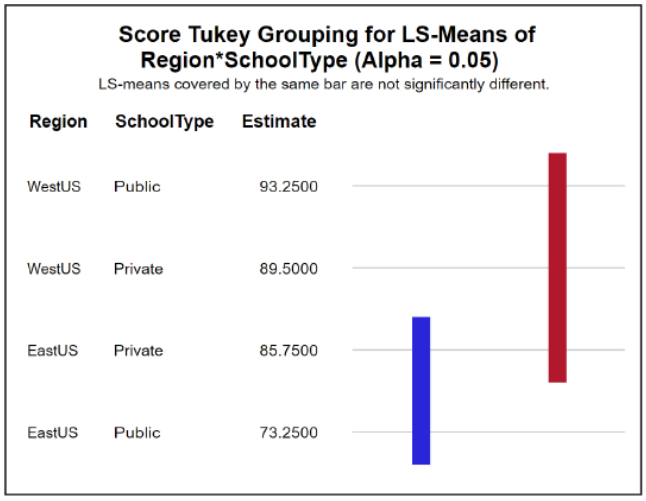


Figure 6.7.2.3: Score Tukey grouping for LS-means of Region*SchoolType.

From the results, it is clear that the mean self-rating scores are highest for the public school in the west region. The difference mean scores for public schools in the west region is significantly different from the mean scores for public schools in the east region as well as the mean scores for private schools in the east region.

The `covtest` option produces the results needed to test the significance of the random effect, *Teach(Region*SchoolType)* in terms of the following null and alternative hypothesis:

$$H_0 : \sigma^2_{teacher} = 0 \text{ vs. } H_a : \sigma^2_{teacher} > 0$$

However, as the following display shows, `covtest` option uses the Wald Z test, which is based on the z-score of the sample statistic and hence is appropriate only for large samples—specifically, when the number of random effect levels is sufficiently large. Otherwise, this test may not be reliable.

Covariance Parameter Estimates

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
Teach(Region*School)	9.3750	8.3689	1.12	0.2626
Residual	4.6875	2.3438	2.00	0.0228

Therefore, in this case, as the number of teachers employed is few, Wald's test may not be valid. It is more appropriate to use the ANOVA F-test for *Teacher(Region*SchoolType)*. Note that the results from the ANOVA table suggest that the effects of the teacher within the region and school type are significant (Pr > F = 0.0257), whereas the results based on Wald's test suggest otherwise (since the p-value is 0.2626).

This page titled 6.7.2: Using SAS is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.7.3: Using R

R - Mixed Effects Models

- Load the schools data.
- Obtain the ANOVA for the mixed effects model.
- Obtain estimators and CIs for means for each combination of region and school type.
- Obtain a means plot for each combination of region and school type.
- Obtain Tukey's multiple comparisons CIs.

1. Load the schools data by using the following commands:

```
setwd("~/path-to-folder/")
schools_data <- read.table("schools_data.txt",header=T)
attach(schools_data)
```

2. Obtain the ANOVA for the mixed effects model by using the following commands:

```
library(lmerTest)
library(lme4)
mixed_schools<-lmer(SR_score ~ region + school_type + region:school_type + (1 | teach
summary(mixed_schools) # Partial output
#Random effects:
# Groups                       Name        Variance Std.Dev.
# (region:school_type):teacher (Intercept) 9.375    3.062
# Residual                                 4.687    2.165
# Number of obs: 16, groups:  (region:school_type):teacher, 8
anova(mixed_schools)
#Type III Analysis of Variance Table with Satterthwaites method
#                   Sum Sq Mean Sq NumDF DenDF F value   Pr(>F)
#region            112.812 112.812     1     4 24.0667 0.008011 **
#school_type        15.312  15.312     1     4  3.2667 0.144986
#region:school_type  52.812  52.812     1     4 11.2667 0.028395 *
#---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that the command lmer() gives the ANOVA table only for the fixed effects. Therefore, in this example, since there are fixed effects, we get the ANOVA table with their F values and p-values.

In the "Random effects" section of the output, under the column variance, we get the estimates for σ_γ^2 and σ^2 which are equal to 9.375 and 4.687 respectively.

Alternatively, we can use the command `aov()` which gives a partial ANOVA table.

```
mixed_schools1<-aov(SR_score ~ region + school_type + region*school_type + Error((reg
summary(mixed_schools1)
#Error: region
#       Df Sum Sq Mean Sq
#region  1  564.1   564.1
#Error: school_type
#            Df Sum Sq Mean Sq
```

```
#school_type  1  76.56   76.56
#Error: region:school_type
#                   Df Sum Sq Mean Sq
#region:school_type  1  264.1   264.1
#Error: region:school_type:teacher
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals  4  93.75   23.44
#Error: Within
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals  8   37.5   4.688
```

3. Obtain estimators, CIs , and multiple comparisons CIs for means for each combination of region and school type by using the following commands:

```
library(emmeans)
pairwise_conf_intervals<-emmeans(mixed_schools,list(pairwise~region:school_type),adju
CI<-confint(pairwise_conf_intervals)
$`emmeans of region, school_type`
# region school_type emmean   SE df lower.CL upper.CL
# EastUS Private       85.8 2.42  4     79.0     92.5
# WestUS Private       89.5 2.42  4     82.8     96.2
# EastUS Public        73.2 2.42  4     66.5     80.0
# WestUS Public        93.2 2.42  4     86.5    100.0
#Degrees-of-freedom method: kenward-roger
#Confidence level used: 0.95
$`pairwise differences of region, school_type`
# 1                              estimate   SE df lower.CL upper.CL
# EastUS Private - WestUS Private    -3.75 3.42  4   -17.69    10.19
# EastUS Private - EastUS Public     12.50 3.42  4    -1.44    26.44
# EastUS Private - WestUS Public     -7.50 3.42  4   -21.44     6.44
# WestUS Private - EastUS Public     16.25 3.42  4     2.31    30.19
# WestUS Private - WestUS Public     -3.75 3.42  4   -17.69    10.19
# EastUS Public - WestUS Public     -20.00 3.42  4   -33.94    -6.06
#Degrees-of-freedom method: kenward-roger
#Confidence level used: 0.95
#Conf-level adjustment: tukey method for comparing a family of 4 estimates
```

4. Obtain means plot for each combination of region and school type by using the following commands:

```
library(plotrix)
region_means<-as.data.frame(CI$`emmeans of region, school_type`)
region<-region_means$region
school_type<-region_means$school_type
region_school_type<-paste(region,school_type)
plotCI(x=region_means$emmean,y=NULL,li=region_means$lower.CL,ui=region_means$upper.CL
axis(1,at=1:4,labels=region_school_type)
```

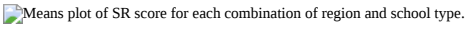

Means plot of SR score for each combination of region and school type.

Figure 6.7.3.1: SR scores mean plot for Region*SchoolType.

5. Obtain Tukey's multiple comparisons plot by using the following commands:

```
diff_comp<-as.data.frame(CI$`pairwise differences of region, school_type`)
diff_reg_sch<-diff_comp[,1]
plotCI(x=diff_comp$estimate,y=NULL,li=diff_comp$lower.CL,ui=diff_comp$upper.CL,xaxt="
abline(h=0)
axis(1,at=1:6,labels=diff_reg_sch,las=1,cex.axis=0.6)
detach(schools_data)
```

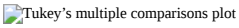

Tukey's multiple comparisons plot

Figure 6.7.3.2: Tukey comparisons differences of means plot.

This page titled 6.7.3: Using R is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.8: Complexity Happens

From what we have discussed so far, we see that even for the simplest multi-factor studies (i.e. those involving only two factors), there are many possibilities of treatment designs generated by either factor being fixed or random, and factors being crossed or nested.

For any of these possibilities, we can carry out the hypothesis tests using the EMS expressions to identify the correct denominator for the relevant F-statistics.

Crossed				
Source	d.f.	A fixed, B fixed	A fixed, B random	A random, B random
A	$a-1$	$\sigma^2+nb\frac{\sum\alpha_i^2}{a-1}$	$\sigma^2+nb\frac{\sum\alpha_i^2}{a-1}+n\sigma^2_{\alpha\beta}$	$\sigma^2+nb\sigma^2_\alpha+n\sigma^2_{\alpha\beta}$
B	$b-1$	$\sigma^2+na\frac{\sum\beta_i^2}{b-1}$	$\sigma^2+na\sigma^2_\beta$	$\sigma^2+na\sigma^2_\beta+n\sigma^2_{\alpha\beta}$
A×B	$(a-1)(b-1)$	$\sigma^2+n\frac{\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$	$\sigma^2+n\sigma^2_{\alpha\beta}$	$\sigma^2+n\sigma^2_{\alpha\beta}$
		σ^2	σ^2	σ^2

Nested				
Source	d.f.	A fixed, B fixed	A fixed, B random	A random, B random
A	$a-1$	$\sigma^2+bn\frac{\sum\alpha_i^2}{a-1}$	$\sigma^2+bn\frac{\sum\alpha_i^2}{a-1}+n\sigma^2_{\beta(\alpha)}$	$\sigma^2+bn\sigma^2_\alpha+n\sigma^2_{\beta(\alpha)}$
B(A)	$a(b-1)$	$\sigma^2+n\frac{\sum\sum\beta_{j(i)}^2}{a(b-1)}$	$\sigma^2+n\sigma^2_{\beta(\alpha)}$	$\sigma^2+n\sigma^2_{\beta(\alpha)}$
Error	$ab(n-1)$	σ^2	σ^2	σ^2

This page titled 6.8: Complexity Happens is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.9: Try It!

Exercise 6.9.1

Three teaching methods were to be compared to teach computer science in high schools. Nine different schools were chosen randomly and each teaching method was assigned to 3 randomly chosen schools so that each school implemented only one teaching method. The response that was used to compare the 3 teaching methods was the average score for each high school.

Show data Lesson6_1ex1

```
data Lesson6_ex1;
input mtd school score semester $;
datalines;
1 1 68.11 Fall
1 1 68.11 Fall
1 1 68.21 Fall
1 1 78.11 Spring
1 1 78.11 Spring
1 1 78.19 Spring
1 2 59.21 Fall
1 2 59.13 Fall
1 2 59.11 Fall
1 2 70.18 Spring
1 2 70.62 Spring
1 2 69.11 Spring
1 3 64.11 Fall
1 3 63.11 Fall
1 3 63.24 Fall
1 3 63.21 Spring
1 3 64.11 Spring
1 3 63.11 Spring
2 1 84.11 Fall
2 1 85.21 Fall
2 1 85.15 Fall
2 1 85.11 Spring
2 1 83.11 Spring
2 1 89.21 Spring
2 2 93.11 Fall
2 2 95.21 Fall
2 2 96.11 Fall
2 2 95.11 Spring
2 2 97.27 Spring
2 2 94.11 Spring
2 3 90.11 Fall
2 3 88.19 Fall
2 3 89.21 Fall
2 3 90.11 Spring
2 3 90.11 Spring
```

```
    2 3 92.21 Spring
    3 1 74.2 Fall
    3 1 78.14 Fall
    3 1 74.12 Fall
    3 1 87.1 Spring
    3 1 88.2 Spring
    3 1 85.1 Spring
    3 2 74.1 Fall
    3 2 73.14 Fall
    3 2 76.21 Fall
    3 2 72.14 Spring
    3 2 76.21 Spring
    3 2 75.1 Spring
    3 3 80.12 Fall
    3 3 79.27 Fall
    3 3 81.15 Fall
    3 3 85.23 Spring
    3 3 86.14 Spring
    3 3 87.19 Spring
    ;
```

1. Using the information about the teaching method, school, and score only, the school administrators conducted a statistical analysis to determine if the teaching method had a significant impact on student scores. Perform a statistical analysis to confirm their conclusion.
2. If possible, perform any other additional statistical analyses.

Show Solution in SAS

1. To confirm their conclusion, a model with only the two factors, teaching method and school was used, with school nested within the teaching method.

Input:

```
data Lesson6_ex1;
    input mtd school score semester $;
    datalines;
    1 1 68.11 Fall
    1 1 68.11 Fall
    1 1 68.21 Fall
    1 1 78.11 Spring
    1 1 78.11 Spring
    1 1 78.19 Spring
    1 2 59.21 Fall
    1 2 59.13 Fall
    1 2 59.11 Fall
    1 2 70.18 Spring
    1 2 70.62 Spring
    1 2 69.11 Spring
    1 3 64.11 Fall
```

```
    1 3 63.11 Fall
    1 3 63.24 Fall
    1 3 63.21 Spring
    1 3 64.11 Spring
    1 3 63.11 Spring
    2 1 84.11 Fall
    2 1 85.21 Fall
    2 1 85.15 Fall
    2 1 85.11 Spring
    2 1 83.11 Spring
    2 1 89.21 Spring
    2 2 93.11 Fall
    2 2 95.21 Fall
    2 2 96.11 Fall
    2 2 95.11 Spring
    2 2 97.27 Spring
    2 2 94.11 Spring
    2 3 90.11 Fall
    2 3 88.19 Fall
    2 3 89.21 Fall
    2 3 90.11 Spring
    2 3 90.11 Spring
    2 3 92.21 Spring
    3 1 74.2 Fall
    3 1 78.14 Fall
    3 1 74.12 Fall
    3 1 87.1 Spring
    3 1 88.2 Spring
    3 1 85.1 Spring
    3 2 74.1 Fall
    3 2 73.14 Fall
    3 2 76.21 Fall
    3 2 72.14 Spring
    3 2 76.21 Spring
    3 2 75.1 Spring
    3 3 80.12 Fall
    3 3 79.27 Fall
    3 3 81.15 Fall
    3 3 85.23 Spring
    3 3 86.14 Spring
    3 3 87.19 Spring
    ;
proc mixed data=lesson6_ex1 method=type3;
class mtd school;
model score = mtd;
random school(mtd);
store results1;
```

```
run;

proc plm restore=results1;
lsmeans mtd / adjust=tukey plot=meanplot cl lines;
run;
```

Partial outputs:

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
mtd	2	4811.400959	2405.700480	Var(Residual) + 6 Var(school(mtd)) + Q(mtd)	MS(school(mtd))	6	16.50	0.0036
school(mtd)	6	875.059744	145.843291	Var(Residual) + 6 Var(school(mtd))	MS(Residual)	45	10.13	<.0001
Residual	45	647.972350	14.399386	Var(Residual)	.	.	.	.

The p-value of .0036 indicates that the scores vary significantly among the 3 teaching methods and confirms the school administrators' conclusion. As the teaching method was significant, the Tukey procedure was conducted to determine the significantly different pairs among the 3 teaching methods. The results of the Tukey procedure shown below indicate that the mean scores of teaching methods 2 and 3 are not statistically significant and that the teaching method 1 mean score is statistically lower than the mean scores of the other two.

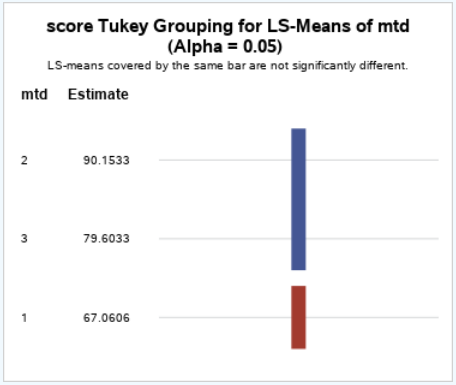


Figure 6.9.a1: LS-means of mtd score Tukey grouping.

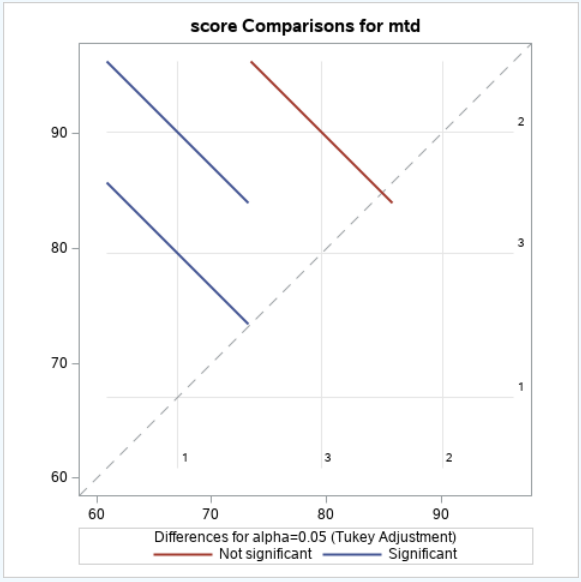


Figure 6.9.a2: Diffogram of score comparisons for mtd with Tukey adjustment.

2. Using the additional code shown below, an ANOVA was conducted including semester also as a possible fixed effect.

```
proc mixed data=lesson6_ex1 method=type3;
class mtd school semester ;
model score = mtd semester mtd*semester;
random school(mtd) semester*school(mtd);
store results2;
run;

proc plm restore= results2;
lsmeans mtd semester / adjust=tukey plot=meanplot cl lines;
run;
```

The p-values indicate that both these main effects are statistically significant, but not their interaction. The Tukey procedure indicates that the significances of paired comparisons for the teaching method remain the same. Between the two semesters, the scores are statistically higher in the spring compared to the fall.

Note

The output writes semester*school(mtd) as school*semester(mtd), probably due to arranging effects in alphabetical order.

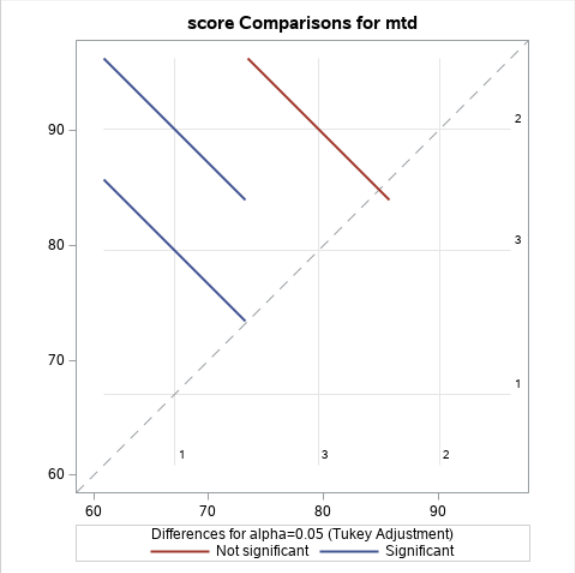


Figure 6.9.a2: Diffogram of score comparisons for mtd with Tukey adjustment.

semester Least Squares Means

semester	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Fall	76.6370	1.8265	6	41.96	<.0001	0.05	72.1677	81.1063
Spring	81.2411	1.8265	6	44.48	<.0001	0.05	76.7718	85.7104

Show Solution in Minitab

1. Choose **Stat -> ANOVA -> General Linear Model**

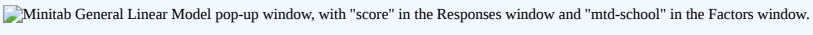


Figure 6.9.b1: Minitab General Linear Model pop-up window.

Then, click Random/Nest:

Figure 6.9.*b*2: General Linear Model: Random/Nest pop-up window.

Output:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
mtd	2	4811.4	2405.70	16.50	0.004
school(mtd)	6	875.1	145.84	10.13	0.000
Error	45	648.0	14.40		
Total	53	6334.4			

Conclusion

The p-value of .004 indicates that mtd is statistically significant, which implies that the mean score from all 3 teaching methods is not the same, thus confirming the school administrators' claim. Note that in the Minitab General Linear Model, the Tukey procedure or any other paired comparisons are not available.

2. Choose **Stat -> ANOVA -> General Linear Model**

Figure 6.9.*b*3: Minitab General Linear Model pop-up window.

Then click Random/Nest.

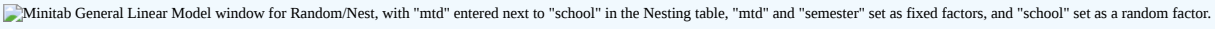


Figure 6.9.*b*4: General Linear Model: Random/Nest pop-up window.

Hit OK and then click Model

Figure 6.9.*b*5: General Linear Model: Model pop-up window.

Select the effects mtd, semester, and school(mtd), and then click Add.

Figure 6.9.*b*6: General Linear Model: Model pop-up window, with selected effects.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
mtd	2	4811.40	2405.70	16.50	0.004
semester	1	286.17	286.17	8.34	0.028
school(mtd)	6	875.06	145.84	4.25	0.051
mtd*semester	2	85.70	42.85	1.25	0.352
school(mtd)*semester	6	205.85	34.31	17.58	0.000
Error	36	70.25	1.95		
Total	53	6334.43			

Conclusion

The p-values indicate that both main effects, mtd and semester, are statistically significant, but not their interaction. Note that in the Minitab General Linear Model procedure, paired comparisons are not available.

Exercise 6.9.2

Type 3 Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	F Value	Pr > F
	2	4811.400959	2405.700480	Var(Residual) + 6 Var(A*B) + Q(A)	11.38	0.0224
	2	29.274959	14.637480	Var(Residual) + 6 Var(A*B) + 18 Var(B)	0.07	0.9342
	4	845.784785	211.446196	Var(Residual)+ 6 Var(A*B)	14.68	<.0001
Residual	45	647.972350	14.399386	Var(Residual)		

Use the ANOVA table above to answer the following.

1. Name the fixed and random effects.
2. Complete the Source column of the ANOVA table above.
3. How many observations are included in this study?
4. How many replicates are there?
5. Write the model equation.
6. Write the hypotheses that can be tested with the expression for the appropriate F-statistic.

Show Solution

1. Name the fixed and random effects.

Fixed: A with 3 levels. In the EMS column, Q(A) reveals that A is fixed and the df indicates that it has 3 levels. Note that any factor that has a quadratic form associated with it is fixed and Q(A) is the quadratic form associated with A. This actually equals $\sum_{i=1}^{3} \alpha_i^2$, where $i = 1, 2, 3$ are the treatment effects; it is non-zero if the treatment means are significantly different.

Random: B is random as indicated by the presence of Var(B), The effect of factor B is studied by sampling 3 cases (see df value for B).

- A*B is random as any effect involving a random factor is random.
- The residual is also random as indicated by the presence of the Var(residual) in the EMS column.

2. Complete the Source column of the ANOVA table above.

Use the EMS column and start from the bottom row. The bottom-most has only var(*residual) and therefore the effect on the corresponding Source is residual. The next row up has var(A*B) in the additional term indicating that the corresponding source is A*B, etc.

Type 3 Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	F Value	Pr > F

Type 3 Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	F Value	Pr > F
A	2	4811.400959	2405.700480	Var(Residual) + 6 Var(A*B) + Q(A)	11.38	0.0224
B	2	29.274959	14.637480	Var(Residual) + 6 Var(A*B) + 18 Var(B)	0.07	0.9342
A*B	4	845.784785	211.446196	Var(Residual)+ 6 Var(A*B)	14.68	<.0001
Residual	45	647.972350	14.399386	Var(Residual)	.	.

3. How many observations are included in this study?

$N-1=2+2+4+45=53$, so $N=54$.

4. How many full replicates are there?

Let r=number of replicates. Then N = number of levels of A times number of levels of B times r = $3\times3\times r$. Therefore, $9\times r=54$, which gives $r=6$.

5. Write the model equation.

$$y_{ijk}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\epsilon_{ijk} \quad \text{where } i,j=1,2,3 \text{ and } k=1,2,\ldots,6$$

6. Write the hypotheses that can be tested with the F-statistic information.

	Effect A	Effect B	Effect A*B
Hypotheses	$H_0:\alpha_i=0$ for all i vs. $H_a:\alpha_i\neq0$ for at least one $i=1,2,3$ Note that $\sum_{i=1}^{3}\alpha_i^2$ is the non-centrality parameter of the F-statistics if H_a is true.	$H_0:\sigma_\beta^2=0$ vs. $H_a:\sigma_\beta^2>0$	$H_0:\sigma_{\alpha\beta}^2=0$ vs. $H_a:\sigma_{\alpha\beta}^2>0$
F Statistic	$\frac{2405.700480}{211.446196}=11.377$ with 2 and 4 degrees of freedom	$\frac{14.63480}{211.446196}=0.0692$ with 2 and 4 degrees of freedom	$\frac{211.446916}{14.399386}=14.685$ with 4 and 45 degrees of freedom

This page titled 6.9: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

6.10: Chapter 6 Summary

Random effects of an ANOVA model, represent measurements arising from a larger population and are assumed to be $\mathcal{N}\left(\mu, \sigma_\tau^2\right)$. In other words, the levels or groups of the random effect that are observed can be considered as a sample from an original population. Random effects can also be **subject effects**. Consequently, in public health, a random effect is referred to as the **subject-specific effect**.

As all the levels of a random effect have the same mean, its significance is measured in terms of the variance with $H_0 : \sigma_\tau^2 = 0$ vs. $H_a : \sigma_\tau^2 > 0$. Note also that any interaction effect involving at least one random effect is also a random effect. Due to the added variability incurred by each random effect, the variance of the response now will have several components which are called **variance components**. In the most basic case, with only one single factor and no fixed effects, this compound variance of the response will be $\sigma_Y^2 = \sigma_\tau^2 + \sigma_\epsilon^2$, where σ_τ^2 is the variance component associated with the random factor. The **intra-class correlation (ICC)**, defined in terms of the variance components, is a useful indicator of the high or low variability within groups (or subjects).

Mixed models, as introduced in section 6.7, include both fixed and random effects. Throughout the lesson, we learned how EMS quantities can be used to determine the correct F-test to test the hypotheses associated with the effects. EMS quantities can be thought of as the population counterparts of the Mean sums of squares (MS), which are computable for each source in the ANOVA table.

This page titled 6.10: Chapter 6 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

CHAPTER OVERVIEW

7: Randomization Design Part I

Objectives

Upon completion of this chapter, you should be able to:

1. Understand the importance of randomization design, the second component of experimental design, and how it impacts on our interpretation of results.
2. Identify any blocking factors and the randomization design used in a study.
3. Use statistical software to obtain the randomization design that assigns the treatment levels to the experimental units schematically.
4. Gain experience in utilizing statistical software to analyze data obtained from a given experimental design.

Previously in the course, we have referenced how experimental design drives the statistical model to be fitted. Recall that in Chapter 5, we discussed the two components of the experimental design that accounts for two aspects of a study.

- The treatment design component, which was addressed in Chapters 5 and 6, describes the treatment levels of interest, treatment type (fixed vs. random), and also the relationship of treatments with each other (crossed vs. nested).
- The randomization design component takes into account the treatment design aspects and also the physical layout of the study setting, including other influencing factors such as confounding (or blocking) variables.

In our discussions of treatment designs, we looked at experimental data in which there were multiple observations made at the treatment applications. We referred to these loosely as *replicates*. In this lesson, we will work formally with these multiple observations and how they are to be collected. This brings us to the right-hand side of the schematic diagram portraying the randomization design component:

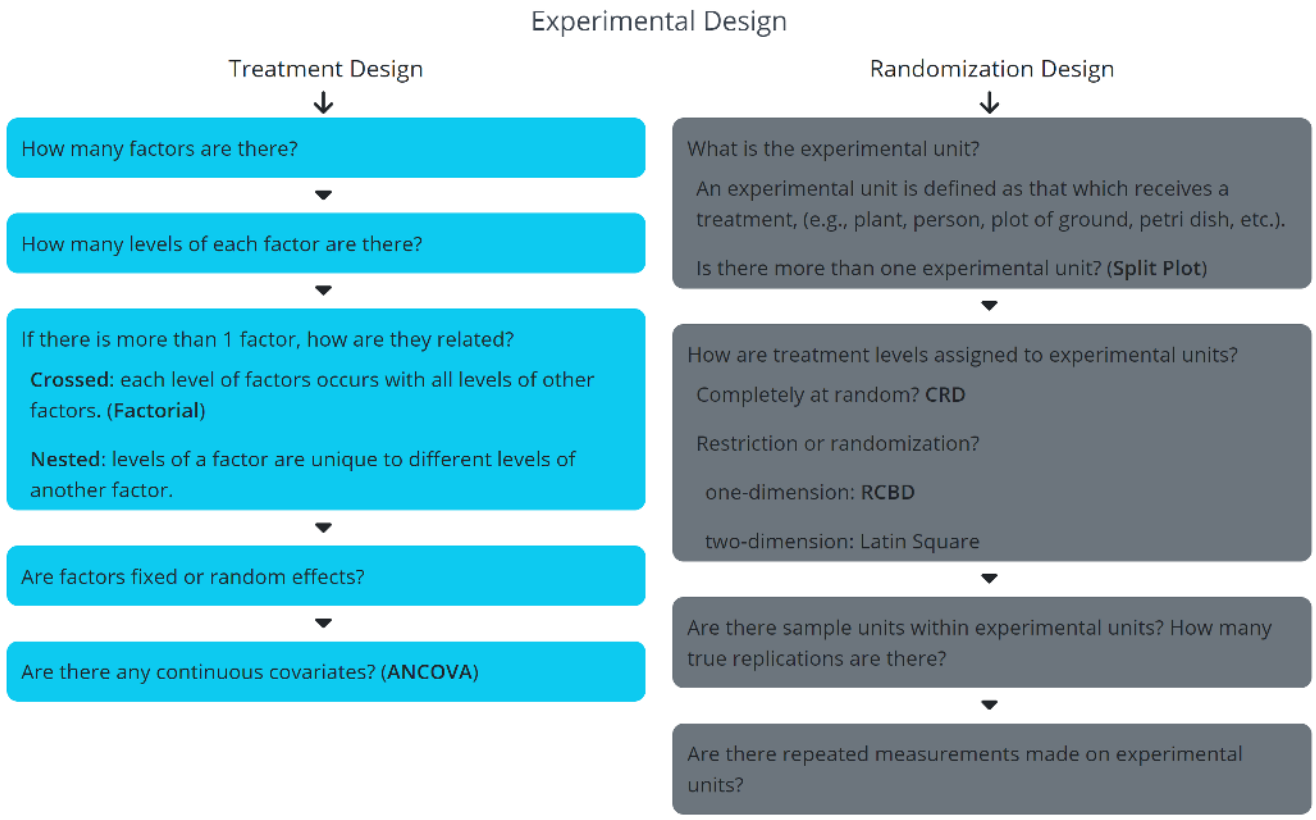


Figure 7.1: Steps of treatment design and randomization design in experimental design.

As can be seen in the diagram above, the treatment design addresses specific characteristics of the experimental factors under study. The randomization design addresses how the treatments are assigned to experimental units. Overall, the experimental design sets the stage in collecting data systematically and also dictates the statistical model to be used and the ANOVA-related calculations.

This page titled 7: Randomization Design Part I is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Experimental Unit and Replication

An **experimental unit** is an item (or physical entity) that receives the treatment. Identifying the experimental unit can be a trivial task in most experiments, but there can be exceptions.

> For example...
>
> Consider a situation where the effect of polluted stream water on fish lesions is to be studied. Two aquaria, each with 50 fish, are used for the study. The water treatment (polluted vs. control) is randomly assigned to each of the aquaria. After 30 days, the number of lesions from randomly caught 10 fish from each aquarium was counted. The treatment design is a single-factor design with 2 levels of water treatment, and a one-way ANOVA can be run on the data. But what is the experimental unit?

Going back to our definition, the experimental unit is the entity that receives the treatment. In this case, we have applied a water treatment to each aquarium. The fish are not the experimental units. In order for individual fish to be experimental units, somehow the investigators would have to take one fish at a time and apply the treatment independently to each fish. This would be impractical from a logistics standpoint and was not done. Instead, the water treatment levels were applied to the entire aquarium, and so the experimental unit is an aquarium with 50 fish.

Now we can determine what constitutes a replication of the experiment. Each time the full set of treatment levels (2 levels in our example) is applied, we have a complete replication. In the experiment described here, there is only one replication, a situation often described as an **un-replicated study**.

The individual fish that were caught and counted for lesions are **sampling units**. Sampling units are the entities from which the observations are recorded. Traditionally, to obtain a correct ANOVA, mean values of the sampling units have to be computed for each experimental unit before the calculation of the treatment SS. Failure to recognize sampling units can result in a serious problem: **pseudo-replication**. Pseudo-replication results from treating each sampling unit as if it were an experimental unit and inflating the error degrees of freedom. By artificially increasing the error df, we reduce the MSE and produce a larger (incorrect) F-statistic.

This page titled 7.1: Experimental Unit and Replication is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

7.2: Completely Randomized Design

After identifying the experimental unit and the number of replications that will be used, the next step is to assign the treatments (i.e. factor levels or factor level combinations) to experimental units.

In a completely randomized design, treatments are assigned to experimental units at random. This is typically done by listing the treatments and assigning a random number to each.

In the greenhouse experiment discussed in Chapter 1, there was a single factor (fertilizer) with 4 levels (i.e. 4 treatments), six replications, and a total of 24 experimental units (each unit a potted plant). Suppose the image below is the Greenhouse Floor plan and bench that was used for the experiment (as viewed from above).

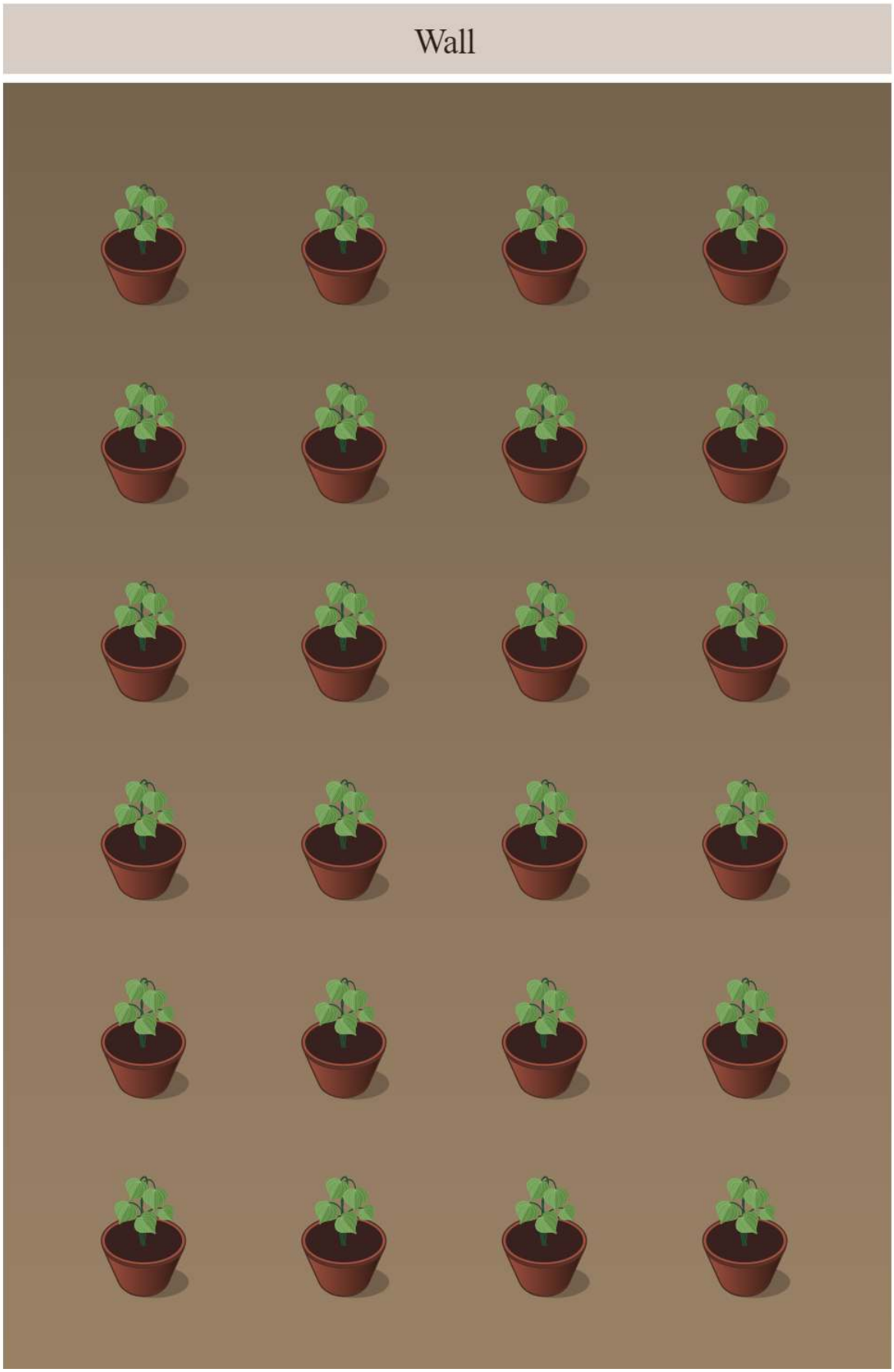

Wall

Figure 7.2.1: Greenhouse floor plan, showing arrangement of the 24 plants.

We need to be able to randomly assign each of the treatment levels to 6 potted plants. To do this, assign physical position numbers on the bench for placing the pots.

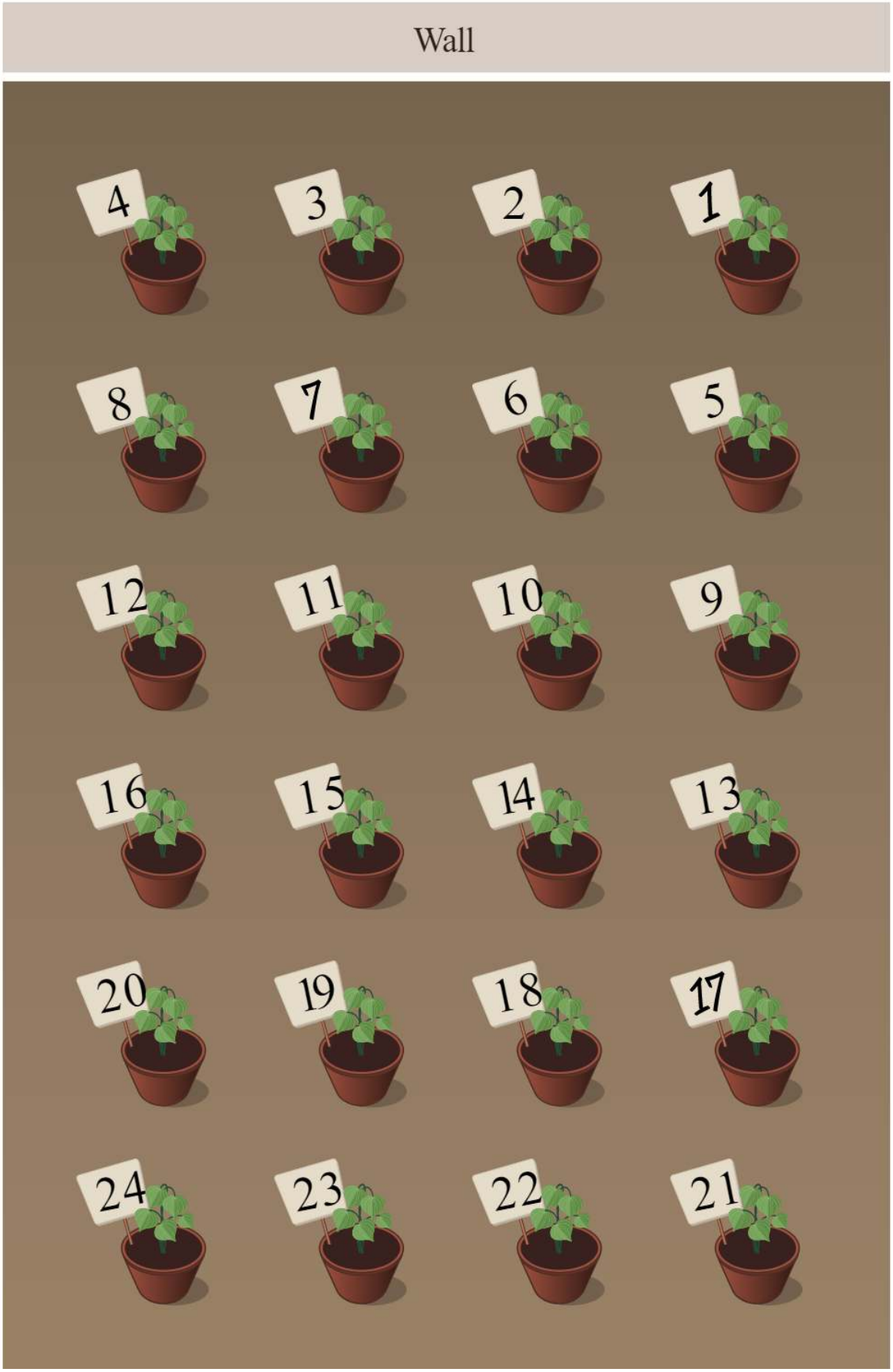

Wall
4
3
2
1
8
7
6
5
12
11
10
9
16
15
14
13
20
19
18
17
24
23
22
21

Figure 7.2.2: Greenhouse floor plan, with the plant locations numbered in a grid pattern.

Using Technology

Minitab Example

Steps in Minitab

In Minitab, this assignment can be done by manually creating two columns: one with each treatment level repeated 6 times (order not important) and the other with a position number 1 to N, where N is the total number of experimental units to be used (i.e. $N = 24$ in this example). The third column will store the treatment assignment.

	C1-T	C2	C3	C4
	Fert	position	Fert_trt	
1	F1	1		
2	F1	2		
3	F1	3		
4	F1	4		
5	F1	5		
6	F1	6		
7	F2	7		
8	F2	8		
9	F2	9		
10	F2	10		
11	F2	11		

Figure 7.2.a1: Entering treatments, position, and treatment assignment information in Minitab.

Next, select **Calc > Sample from Columns**, fill in the dialog box as seen below, and click **OK**.

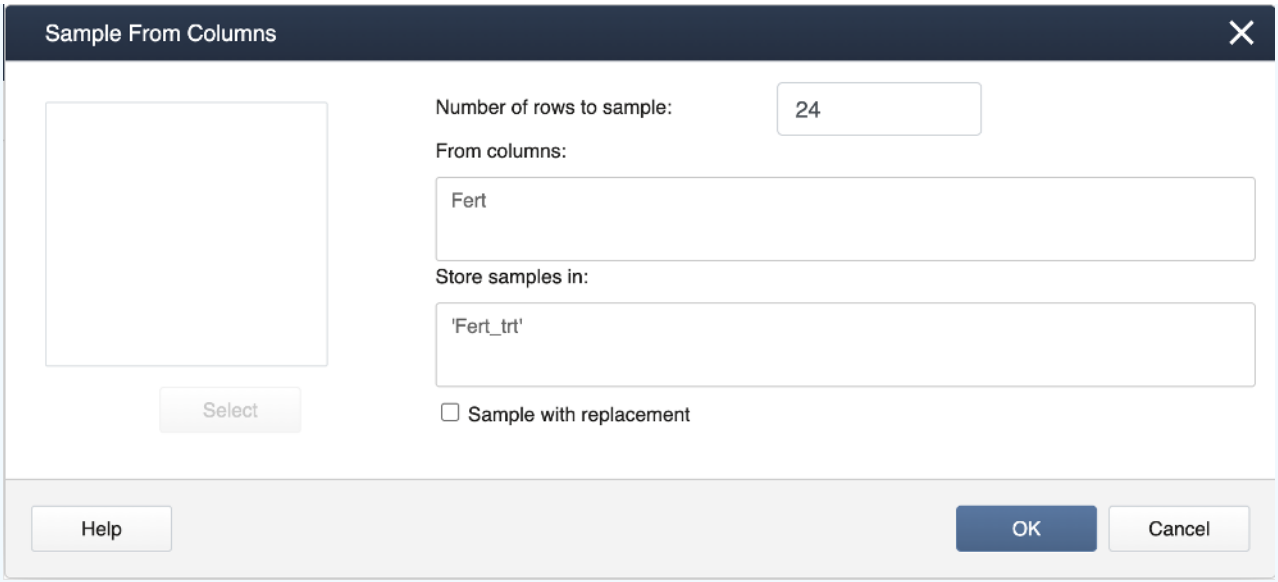


Figure 7.2.a2: Minitab Sample from Columns pop-up window.

Note!

Be sure to have the "Sample with Replacement" box unchecked so that all treatment levels will be assigned to the same number of pots, giving rise to a proper completely randomized design for a specified number of replicates.

This will result in a completely random assignment.

	C1-T	C2	C3-T
	Fert	position	Fert_trt
1	F1	1	F2
2	F1	2	Control
3	F1	3	F2
4	F1	4	F3
5	F1	5	F2
6	F1	6	F2
7	F2	7	Control
8	F2	8	F3
9	F2	9	F3
10	F2	10	F1
11	F2	11	F1

Figure 7.2.*a*3: Minitab spreadsheet showing the random treatment assignment for each plant position.

This assignment can then be used to apply the treatment levels appropriately to pots on the greenhouse bench.

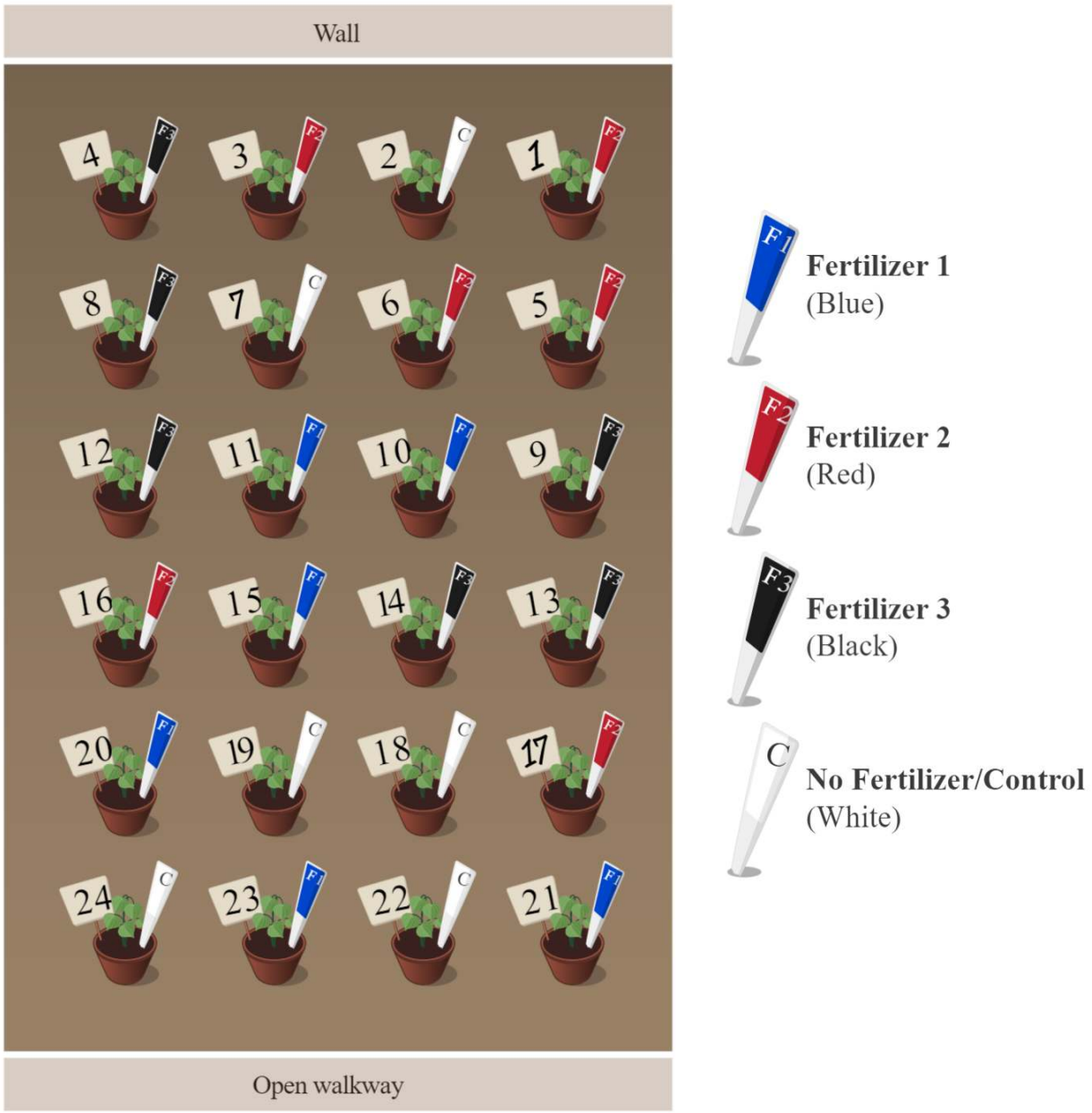


Figure 7.2.*a*4: Plants in the greenhouse with their appropriate randomly assigned fertilizer treatment levels.

? SAS Example

Steps in SAS

To make the assignments in SAS we can utilize the SAS `surveyselect` procedure as below:

```
proc surveyselect data=greenhouse out=trtassignment outrandom
method=srs
samprate=1;
run;
```

The output would be as below. In practice, it is recommended to specify a seed to ensure the results are reproducible.

Obs	Fertilizer
1	F3
2	F2
3	Con
4	F2
5	F3
6	Con
7	F2
8	F2
9	F3
10	F1
11	F1
12	F3
13	F2
14	F1
15	F3
16	F3
17	F1
18	Con
19	Con
20	F2
21	Con
22	F1
23	Con
24	F1

? R Example

Steps in R

Completely Randomized Design

To randomly assign treatment levels to each of our plants we can use the following commands:

```
sample(treatment)
[1] "F3"      "F2"      "F1"      "F2"      "F3"      "F1"      "Control" "F2"
[10] "F3"      "F2"      "Control" "F3"      "F1"      "F1"      "F2"      "Cont
[19] "F1"      "Control" "F3"      "Control" "Control" "F1"
```

This means that the first experimental unit will get Fertilizer 3, the second experimental unit will get Fertilizer 2, etc.

Randomized Complete Block Design

Obtain the block design. Load the greenhouse data and obtain the ANOVA table.

To obtain the block design we can use the following commands:

```
library(blocksdesign)
block_design<-blocks(4,6,6)$Design
obs<-c(1:24)
block<-block_design[,1]
plant<-rep(c(1:4),6)
treatment<-block_design[,3]
data.frame(cbind(obs,block,plant,treatment))
#    obs block plant treatment
# 1    1     1     1         4
# 2    2     1     2         1
# 3    3     1     3         3
# 4    4     1     4         2
# 5    5     2     1         1
# 6    6     2     2         4
# 7    7     2     3         3
# 8    8     2     4         2
# 9    9     3     1         3
# 10  10     3     2         1
# 11  11     3     3         4
# 12  12     3     4         2
# 13  13     4     1         1
# 14  14     4     2         4
# 15  15     4     3         2
# 16  16     4     4         3
# 17  17     5     1         3
# 18  18     5     2         2
# 19  19     5     3         1
# 20  20     5     4         4
# 21  21     6     1         2
# 22  22     6     2         1
# 23  23     6     3         4
# 24  24     6     4         3
```

To load the greenhouse data and obtain the ANOVA table (`lmer()` and `aov(`)) we use the following commands:

```
setwd("~/path-to-folder/")
greenhouse_RCBD_data <- read.table("greenhouse_RCBD_data.txt",header=T)
attach(greenhouse_RCBD_data)
library(lmerTest)
library(lme4)
greenhouse_RCBD_anova<-lmer(Height ~ Fertilizer + (1 | factor(Block)),greenhouse
anova(greenhouse_RCBD_anova)
#Type III Analysis of Variance Table with Satterthwaites method
#            Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
```

```
#Fertilizer 251.44  83.813     3    15  162.96 1.144e-11 ***
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
greenhouse_RCBD_anova1<-aov(Height~Fertilizer+Error(factor(Block)),greenhouse_RC
summary(greenhouse_RCBD_anova1)
#Error: factor(Block)
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals  5  53.32   10.66
#Error: Within
#           Df Sum Sq Mean Sq F value   Pr(>F)
#Fertilizer  3 251.44   83.81     163 1.14e-11 ***
#Residuals  15   7.72    0.51
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

For comparison the ANOVA table for the completely randomized design is given below:

```
greenhouse_CRD_anova<-aov(Height~Fertilizer,greenhouse_RCBD_data)
summary(greenhouse_CRD_anova)
#            Df Sum Sq Mean Sq F value   Pr(>F)
#Fertilizer   3 251.44   83.81   27.46 2.71e-07 ***
#Residuals   20  61.03    3.05
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
detach(greenhouse_RCBD_data)
```

This page titled 7.2: Completely Randomized Design is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

7.3: Restriction on Randomization - RCBD

A completely randomized design (CRD) for the greenhouse experiment is reasonable, provided the positions on the bench are equivalent. In reality, this is rarely the case. In this setting, for example, some micro-environmental variation can be expected due to the glass wall on one end, and the open walkway at the other end of the bench.

A powerful alternative to the CRD is to restrict the randomization process to form blocks. Blocks, in a physical setting such as in this example, are usually set up at right angles to suspected gradients in variation.

In a block design, general blocks are formed such that the experimental units are expected to be homogeneous within a block and heterogeneous between blocks. The number of experimental units within a block is called its block size.

In a randomized complete block design (RCBD), each block is of the same size and is equal to the number of treatments (i.e. factor levels or factor level combinations). Furthermore, each treatment will be randomly assigned to exactly one experimental unit within every block. So we think of the data in the greenhouse example in terms of RCBD, we will have 6 blocks each with block size equal to 4, the number of fertilizer levels.

To establish an RCBD for this data, the assignments of fertilizer levels to the experimental units (the potted plants) have to be done within each block separately.

Using SAS

To obtain the block design in SAS, we can use the following code:

```
proc plan ordered ;
factors Block=6 Plant=4;
treatments Fertilizer=4 random;
output out=rcb block
cvals=('Block 1' 'Block 2' 'Block 3' 'Block 4' 'Block 5' 'Block 6');
run;
proc format;
value FertFmt
1 = "F1"
2 = "F2"
3 = "F3"
4 = "Con";
run;
proc print data=rcb;
format Fertilizer FertFmt.;
run;
```

The output we obtain would be as follows:

Obs	Block	Plant	Fertilizer
1	Block 1	1	F3
2	Block 1	2	F2
3	Block 1	3	Con
4	Block 1	4	F1
5	Block 2	1	F1
6	Block 2	2	F3

Obs	Block	Plant	Fertilizer
7	Block 2	3	F2
8	Block 2	4	Con
9	Block 3	1	F2
10	Block 3	2	Con
11	Block 3	3	F3
12	Block 3	4	F1
13	Block 4	1	F2
14	Block 4	2	F3
15	Block 4	3	F1
16	Block 4	4	Con
17	Block 5	1	F3
18	Block 5	2	F1
19	Block 5	3	Con
20	Block 5	4	F2
21	Block 6	1	Con
22	Block 6	2	F2
23	Block 6	3	F3
24	Block 6	4	F1

Using Minitab

To obtain the design in Minitab, we do the following.

For Block 1, manually create two columns: one with each treatment level and the other with a position number 1 to n, where n is the block size (i.e. $n = 4$ in this example). The third column will store the assignment of fertilizer levels to the experimental units.

	C1-T	C2-T	C3	C4
		Fert	Position	Assign
1	(Block 1)	F1	1	
2		F2	2	
3		F3	3	
4		Control	4	

Figure 7.3.1: Columns for entering Block 1 data in Minitab.

Next, select **`Calc`** > **`Sample from Columns`** > fill in the dialog box as seen below, and click **OK**.

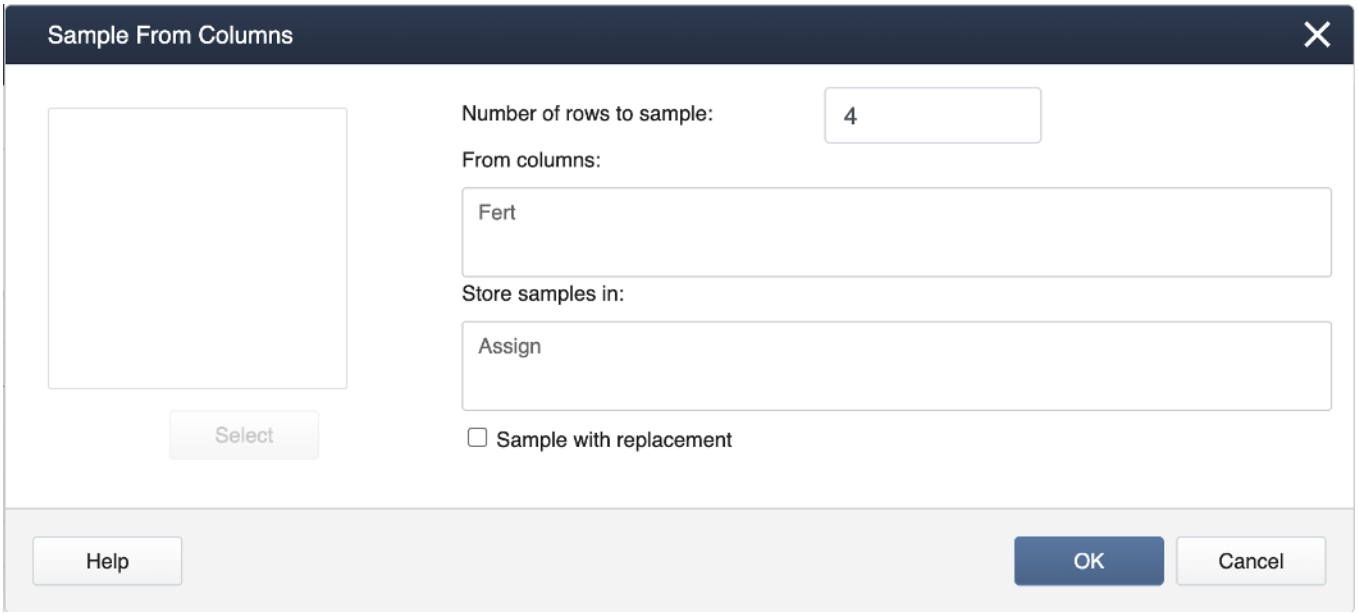


Figure 7.3.2: Minitab Sample from Columns pop-up window.

Here, the number of rows to be specified is our block size (and number of treatment levels), which yields a random assignment from Block 1.

	C1-T	C2-T	C3	C4-T
		Fert	Position	Assign
1	(Block 1)	F1	1	F1
2		F2	2	Control
3		F3	3	F2
4		Control	4	F3

Figure 7.3.3: Random treatment level assignments for positions in Block 1.

The same process should be repeated for the remaining blocks. The key element is that each treatment level or treatment combination appears in each block (forming complete blocks), and is assigned at random within each block.

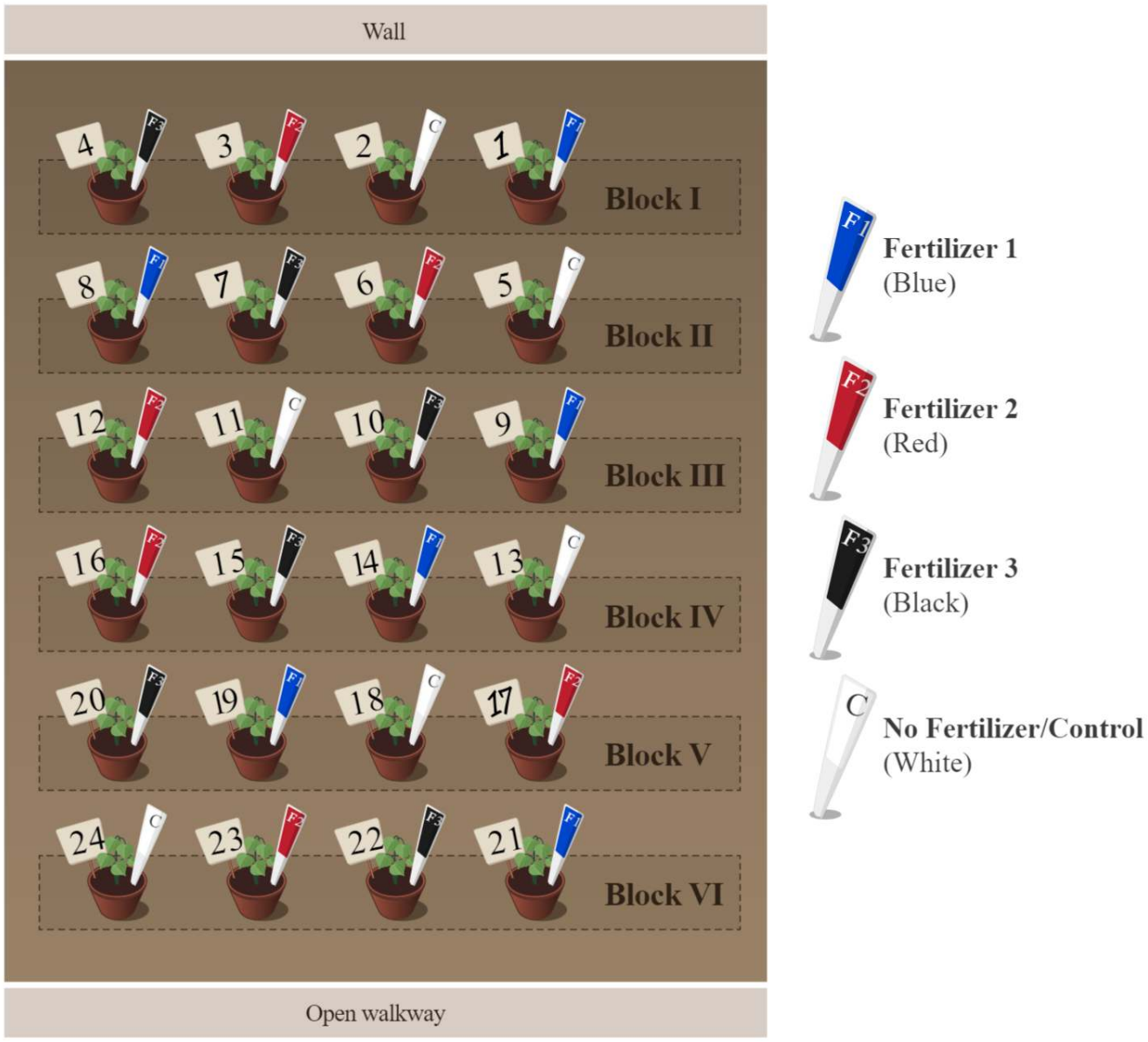


Figure 7.3.4: Greenhouse floor plan divided into blocks, with random treatment level assignment to plant positions within each block.

Blocks are usually treated as random effects, as they would represent the population of all possible blocks. In other words, the mean comparison among blocks is not of interest. But the variation between blocks has to be incorporated into the model and will be partitioned out of the Error Mean squares of the CRD, resulting in a smaller MSE for testing hypotheses about treatments.

The statistical model corresponding to the RCBD is similar to the two-factor studies with one observation per cell (i.e. we assume the two factors do not interact).

Here is Dr. Shumway stepping through this experimental design in the greenhouse.

Video 7.3.1: Demonstrating RCBD in the greenhouse.

Once we collect the data for this experiment, we can use SAS to analyze the data and obtain the results.

We will consider the greenhouse experiment with one factor of interest (Fertilizer). We also have the identifications for the blocks. In this example, we consider *Fertilizer* as a *fixed* effect (as we are only interested in comparing the 4 fertilizers we chose for the study) and *Block* as a *random* effect.

Therefore the statistical model would be

$$Y_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij} \tag{7.3.1}$$

where $i = 1, 2, \ldots, 6$ and $j = 1, 2, 3, 4$. ρ_i and ϵ_{ij} are independent variables such that $\rho_i \sim \mathcal{N}\left(0, \sigma_\rho^2\right)$ and $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right)$.

Let us read the data into SAS and obtain the `proc summary` output.

```
data RCBD_oneway;
input block Fert $ Height;
datalines;
1     Control     19.5
2     Control     20.5
3     Control     21
4     Control     21
5     Control     21.5
6     Control     22.5
1     F1     25
2     F1     27.5
3     F1     28
4     F1     28.6
5     F1     30.5
6     F1     32
1     F2     22.5
2     F2     25.2
3     F2     26
4     F2     26.5
```

```
5     F2     27
6     F2     28
1     F3     27.5
2     F3     28
3     F3     29.2
4     F3     29.5
5     F3     30
6     F3     31
;
proc summary data=RCBD_oneway;
class block fert;
var height;
output out=output1 mean=mean stderr=se;
run;
proc print data=output1;
```

The `proc summary` output would be as follows. We see that the first line in the table with `_TYPE_=0` identification is the estimated overall mean (i.e. $\bar{y}_{..}$). The estimated treatment means (i.e. $\bar{y}_{.j}$) are displayed with `_TYPE_=1` identification and the estimated block means are displayed with `_TYPE_=2` identification. Since we only have one observation per treatment within each block, we cannot estimate the standard error using the data.

Obs	block	Fert	_TYPE_	_FREQ_	mean	se
1	.		0	24	26.1667	0.75238
2	.	Control	1	6	21.0000	0.40825
3	.	F1	1	6	28.6000	0.99499
4	.	F2	1	6	25.8667	0.77531
5	.	F3	1	6	29.2000	0.52599
6	1		2	4	23.6250	1.71239
7	2		2	4	25.3000	1.71221
8	3		2	4	26.0500	1.80808
9	4		2	4	26.4000	1.90657
10	5		2	4	27.2500	2.06660
11	6		2	4	28.3750	2.13478
12	1	Control	3	1	19.5000	.
13	1	F1	3	1	25.0000	.
14	1	F2	3	1	22.5000	.
15	1	F3	3	1	27.5000	.
16	2	Control	3	1	20.5000	.
17	2	F1	3	1	27.5000	.
18	2	F2	3	1	25.2000	.
19	2	F3	3	1	28.0000	.

Obs	block	Fert	_TYPE_	_FREQ_	mean	se
20	3	Control	3	1	21.0000	.
21	3	F1	3	1	28.0000	.
22	3	F2	3	1	26.0000	.
23	3	F3	3	1	29.2000	.
24	4	Control	3	1	21.0000	.
25	4	F1	3	1	28.6000	.
26	4	F2	3	1	26.5000	.
27	4	F3	3	1	29.5000	.
28	5	Control	3	1	21.5000	.
29	5	F1	3	1	30.5000	.
30	5	F2	3	1	27.0000	.
31	5	F3	3	1	30.0000	.
32	6	Control	3	1	22.5000	.
33	6	F1	3	1	32.0000	.
34	6	F2	3	1	28.0000	.
35	6	F3	3	1	31.0000	.

To run the model in SAS we can use the following code:

```
/* RCBD */
proc mixed data=RCBD_oneway method=type3;
class block fert;
model height=fert;
random block;
run;
```

We obtain the ANOVA table below for the RCBD.

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Fert	3	251.440000	83.813333	Var(Residual) + Q(Fert)	MS(Residual)	15	162.96	<.0001
block	5	53.318333	10.663667	Var(Residual) + 4 Var(block)	MS(Residual)	15	20.73	<.0001
Residual	15	7.715000	0.514333	Var(Residual)	.	.	.	.

For comparison, let us obtain the ANOVA table for the CRD for the same data. We use the following SAS commands:

```
/* CRD for comparison */
proc mixed data=RCBD_oneway method=type3;
class fert;
model height=fert;
run;
```

The CRD ANOVA table for our data would be as follows:

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Fert	3	251.440000	83.813333	Var(Residual) + Q(Fert)	MS(Residual)	20	27.46	<.0001
Residual	20	61.033333	3.051667	Var(Residual)	.	.	.	.

Comparing the two ANOVA tables, we see that the MSE in RCBD has decreased considerably in comparison to the CRD. This reduction in MSE can be viewed as the partition in SSE for the CRD (61.033) into SSBlock + SSE (53.32 + 7.715, respectively). The potential reduction in SSE by blocking is offset to some degree by losing degrees of freedom for the blocks. But more often than not, is worth it in terms of the improvement in the calculated F-statistic. In our example, we observe that the F-statistic for the treatment has increased considerably for RCBD in comparison to CRD. It is reasonable to assume that the result from the RCBD is more valid than that from the CRD as the MSE value obtained after accounting for the block to block variability is a more accurate representation of the random error variance.

This page titled 7.3: Restriction on Randomization - RCBD is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

7.4: Blocking in 2 Dimensions - Latin Square

The fundamental idea of blocking can be extended to more dimensions. However, the full use of multiple blocking variables in a complete block design usually requires many experimental units. Latin Square design can be useful when we want to achieve blocking simultaneously in two directions with a limited number of experimental units.

The limitation is that the Latin Square experimental layout will only be possible if:

$$\text{number of Row blocks} = \text{number of Column blocks} = \text{number of treatment levels} \tag{7.4.1}$$

The experimental design process begins with a *Standard Latin Square*. These have the treatment levels ordered across the first row and first column. For example, a single factor with three levels (A, B, C) to be blocked in two directions could begin with this standard 3×3 square:

A	B	C
B	C	A
C	A	B

To randomize, first randomly permute the order of the rows and produce a new square.

B	C	A
C	A	B
A	B	C

Then randomly permute the order of the columns to yield the final square for the experimental layout.

C	A	B
A	B	C
B	C	A

This process assures that any row or column will have all treatment levels. To obtain the design in SAS we can use:

```
proc plan;
factors Row=4 ordered Col=4 ordered / noprint;
treatments Treatment=4 cyclic;
output out=LatinSquare
    Row cvals=('RowBlock 1' 'RowBlock 2' 'RowBlock 3' 'RowBlock 4') random
    Col cvals=('ColBlock 1' 'ColBlock 2' 'ColBlock 3' 'ColBlock 4') random
    Treatment nvals=(1 2 3 4) random;
run;
```

The ANOVA for the Latin Square is a direct extension of the RCBD with random blocking effects. The SAS `random` statement has to be modified accordingly to incorporate both blocking factors and with the assumption of no interaction between them (because of only one observation for each cell). For example, we could use the following SAS code to estimate the model:

```
proc mixed data=LatinSquare method=type3;
 class Row Col Treatment;
 model Response = Treatment;
```

```
 random Row Col;
run;
```

Using R

To obtain a Latin Square Design for four treatments we can use the following commands:

```
library(magic)
latin_square_design<-rlatin(4)
# latin_square_design
#      [,1] [,2] [,3] [,4]
# [1,]    3    1    2    4
# [2,]    4    2    1    3
# [3,]    2    4    3    1
# [4,]    1    3    4    2
```

This page titled 7.4: Blocking in 2 Dimensions - Latin Square is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

7.5: Try It!

Exercise 7.5.1

A poultry experiment was run to investigate the effect of diet and antibiotics on egg production. They evaluated 2 diets of interest and 2 specific antibiotics that are on the market. The feed and antibiotic were combined and used to fill the feeding trays in barns. They chose 3 poultry farms at random and randomly assigned the combinations of diet and antibiotic to 4 barns within each farm. Total egg production by the chickens was recorded after 4 weeks.

a. What is the experimental design (hint: think about the randomization process)?
b. Identify which factors are fixed and which are random.

Show Solution

a) RCBD

b) Fixed factors: Diet and Antibiotic; Random factor: Farms

Exercise 7.5.2

A commercial farmer is studying the corn yield of two fertilizer types at 2 different temperature levels. He strips his cornfield into 20 strips. Each fertilizer type and temperature level combination is then assigned to 5 of the randomly chosen strips.

a. What is the Treatment design?
b. What is the Randomization design?

Show Solution

a) 2×2 factorial with fertilizer types and temperature levels, each having 2 levels

b) CRD with 5 replicates

Exercise 7.5.3

An investigator wants to run an experiment in a Latin square design evaluating 5 levels of a treatment (labeled A, B, C, D, and E) and included the layout in a research proposal that you are reviewing. Identify any problems you see and suggest how to revise the design.

A	B	C	D	E
B	C	D	**B**	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

Show Solution

Column 4, row 2, **B** should be **E** to satisfy the property that *each treatment occurs only once in each row and once in each column*. In addition, the rows and columns need to be independently randomized to produce the actual layout of the Latin square for the experimental plan.

This page titled 7.5: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

7.6: Chapter 7 Summary

This chapter introduced us to Randomization Design, which provides the scheme of how treatment levels can be assigned to experimental units. The specific designs discussed are CRD, RCBD, and Latin Square Design. An RCBD is employed to account for a blocking factor, or a nuisance variable, which is not of interest but may have an impact on the response. Likewise, a Latin square design is helpful in the presence of two such blocking variables. In an RCBD, with no replicates, the interaction between the treatment and the blocking variable is assumed to be negligible and the Mean Square(MS) value of this interaction serves as the estimate of the error variance which turns out to be the denominator of the F-statistic for testing treatment significance. The next chapter will introduce us to another widely used design called split-plot design.

This page titled 7.6: Chapter 7 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

CHAPTER OVERVIEW

8: Randomization Design Part II

Objectives

Upon completion of this chapter, you should be able to:

1. Recognize multiple experimental units in an experimental design.
2. Understand the structure of split-plot ANOVA.
3. Utilize split-plots administered in RCBD experiments.
4. Utilize split-plots administered in CRD experiments.
5. Extend the split-plot concept to analyze split-split-plot designs.

Sometimes multi-factor experiments use multiple (different) experimental units for the different factors in the experiment. To visualize this, think of applying multiple treatments in a sequence. The levels of the first factor are applied to experimental units using specific randomization and then the levels of a second factor are applied to sub-units within the application of the first factor. In other words, the experimental unit used for the application of the first factor has been split, forming the experimental units for the application of the second-factor levels.

Split-plot designs accommodate the above scheme in assigning two factors appropriately to their experimental units. They are extremely common and typically result from logistical restrictions, practicality, or efficiency. Though sometimes split-plots and their experimental unit set up are difficult to recognize, understanding the correct structure is necessary for the implementation of ANOVA.

Split-plots occur most commonly in two experimental designs applied for the first factor: the CRD and RCBD. The ANOVA differs between these two, and this chapter focuses on both types. Split-plots can be extended to accommodate multiple splits by sub-unit subdivision. For example, a split-split-plot experimental design can be achieved with three stages of randomization for three treatments when there are three types of experimental units with two sub-divisions.

This page titled 8: Randomization Design Part II is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Split-Plot Design in RCBD

Recall the Randomized Complete Block Design (RCBD) we discussed in Chapter 7. In RCBD, general blocks are formed such that the experimental units are expected to be homogenous within a block and heterogeneous between blocks.

For example. suppose we are studying the effect of irrigation amount (I_1 and I_2) and fertilizer type (A and B) on crop yield. We have 4 treatments in this experiment. Suppose we want to have at least 2 replicates and have two large lands that can be used for the experiment. In RCBD, we can split each land into 4 fields and can apply the 4 treatments randomly to each field. Here *lands* are blocks and *fields* are the experimental units.

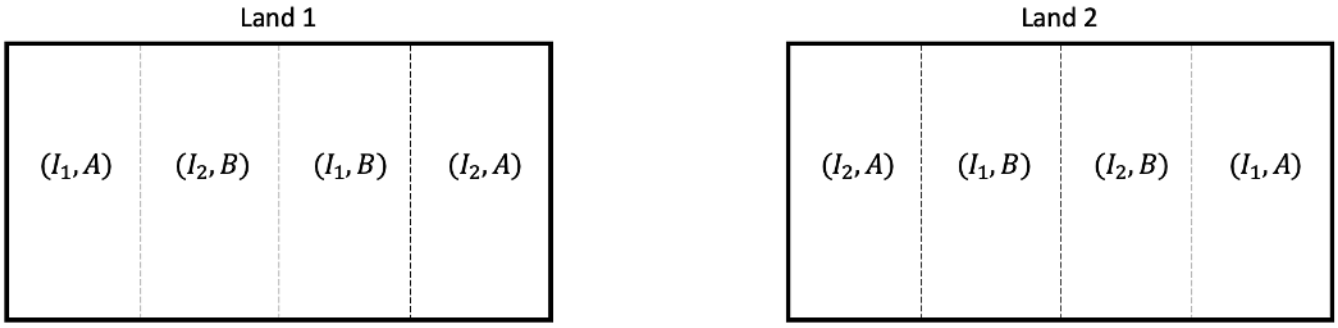


Figure 8.1.1: Lands divided into 4 fields each, each field assigned one of the 4 random treatments.

In this example, we have assumed that managing levels of irrigation and fertilizer require the same effort. Now suppose varying the level of irrigation is difficult on a small scale and it makes more sense to apply irrigation levels to larger areas of land.

In such situations, we can divide each land into two large fields (whole plots) and apply irrigation amounts to each field randomly. And then divide each of these large fields into smaller fields (subplots) and apply fertilizer randomly within the whole plots.

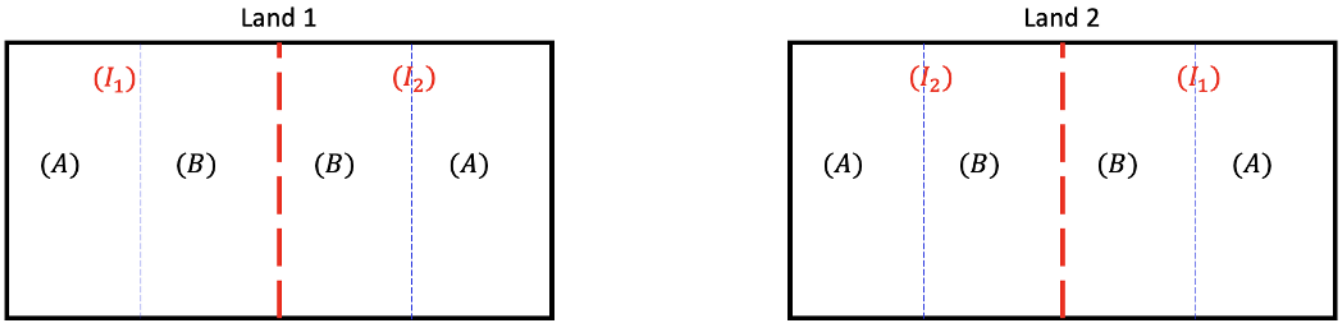


Figure 8.1.2: Lands divided into 2 plots for irrigation, with each plot divided into 2 fields for fertilizer treatment.

In this strategy, each land contains two whole plots and irrigation amount is assigned to each whole plot randomly using RCBD (i.e. lands are treated as blocks and irrigation amount is assigned randomly within each block to the whole plots). Each whole plot contains two subplots and fertilizer type is assigned to each subplot using RCBD (i.e. whole plots are treated as blocks and fertilizer type is assigned randomly within each whole plot to the subplots).

When some factors are more difficult to vary than others at the levels of experimental units, it is more efficient to assign more difficult-to-change factors to larger units (**whole plots**) and then apply the easier-to-change factor to smaller units (**subplots**). This is known as the ***split-plot*** design.

As an example (adapted from Hicks, 1964), consider an experiment where an electrical component is subjected to 4 different temperatures for 3 different amounts of time. If the investigators desire 3 replications for each of the 12 temperature and time combinations (i.e. 12 treatments), a basic CRD or an RCBD (with a suitable blocking factor that would generate the replicates) will require as many as 36 attempts of testing.

Instead, the experimentation can be modified as follows to reduce effort and time. Regarding ovens as blocks, 3 ovens can be set to each of the 4 different temperature settings and then investigators can take out randomly selected components at the 3 different times of interest.

In this setting, temperatures are assigned randomly within each oven (i.e. an oven is treated as a block) and within each temperature, the baking times are assigned randomly to components. We have two RCBD sub-experiments: whole plot levels (temperatures) are assigned as RCBD within the oven and subplots levels (baking time) are assigned as RCBD within whole plot levels.

The data (Bake Time Data) were:

		Oven Temperature (°F)			
Rep	Baking Time (min)	580	600	620	640
I	5	217	158	229	223
	10	233	138	186	227
	15	175	152	155	156
II	5	188	126	160	201
	10	201	130	170	181
	15	195	147	161	172
III	5	162	122	167	182
	10	170	185	181	201
	15	213	180	182	199

It is important to notice that in a split-plot design, randomization is a two-stage process. Levels of one factor (say, factor A) are randomized over the whole plots within each block, and the levels of the other factor (say, factor B) are randomized over the subplots within each whole plot. This restriction in randomization results in two different error terms: one appropriate for comparisons at the whole plot level and one appropriate for comparisons at the subplot level.

The appropriate error for whole plot level in split-plot RCBD is whole plot factor $\times$ block interaction. In other words, the analysis at the whole plot level is essentially of a one-way ANOVA with blocking (i.e. one observation per block-treatment combination). From the perspective of the whole plot, the subplots are simply subsamples and it is reasonable to average them when testing the whole plot effects (i.e. factor A effects).

The subplot factor (i.e. factor B) is always compared within the whole plot factor.

Source	DF
Blocks	$r-1$
Factor A	$a-1$
Whole plot Error	$(r-1)(a-1)$
Factor B	$b-1$
$A \times B$	$(a-1)(b-1)$
Subplot Error	$a(r-1)(b-1)$
Total	$rab-1$

The statistical model associated with the split-plot design with whole plots arranged as RCBD is

$$Y_{ijk} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{8.1.1}$$

where γ_k for $k = 1, \ldots, r$ are block effects, α_i for $i = 1, \ldots, a$ are factor A effects, and β_j for $j = 1, \ldots, b$ are factor B effects.

SAS Example

Steps in SAS

In SAS, we could specify the model with the following statements:

```
proc mixed data=BakeTimeData method=type3;
class oven temp time;
model resp=temp time temp*time;
random oven oven*temp;
run;
```

This will generate the ANOVA table as shown below.

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
temp	3	12494	4164.768519	Var(Residual) + 3 Var(oven*temp) + Q(temp,temp*time)	MS(oven*temp)	6	14.09	0.0040
time	2	566.222222	283.111111	Var(Residual) + Q(time,temp*time)	MS(Residual)	16	0.46	0.6418
temp*time	6	2600.444444	433.407407	Var(Residual) + Q(temp*time)	MS(Residual)	16	0.70	0.6551
oven	2	1962.722222	981.361111	Var(Residual) + 3 Var(oven*temp) + 12 Var(oven)	MS(oven*temp)	6	3.32	0.1070
oven*temp	6	1773.944444	295.657407	Var(Residual) + 3 Var(oven*temp)	MS(Residual)	16	0.48	0.8162
Residual	16	9933.333333	620.833333	Var(Residual)	.	.	.	.

The ANOVA table can be rearranged to the following to make it easier to understand the whole plot and subplot analyses.

Source	DF	Expected Mean Square

Source	DF	Expected Mean Square
(Whole Plots)		
oven	2	Var(Residual) + 3 Var(block*temp) + 12 Var(oven)
temp	3	Var(Residual) + 3 Var(oven*temp) + Q(temp, temp*time)
oven*temp	6	Var(Residual) + 3 Var(oven*temp)
(Subplots)		
time	2	Var(Residual) + Q(time, temp*time)
temp*time	6	Var(Residual) + Q(temp*time)
Residual	16	Var(Residual)

Notice that the correct error term for the F-test of the treatment applied to whole plots is the block $\times$ whole plot factor (assuming blocks are a random effect).

Note!

One might wonder about the terms block $\times$ subplot factor and block $\times$ whole plot factor $\times$ subplot factor . With these terms in the model, we will not be able to retrieve the residual (the error DF will be zero). If repeat observations are made within the split-plots, then a separate error term can be estimated. However, it is important to keep in mind that tests of replication effects are not of interest, but are being isolated in the ANOVA to reduce the error variance. As a result, the model that is usually run in this design drops out the block $\times$ subplot factor and block $\times$ whole plot factor $\times$ subplot factor terms, and combine these interactions with the true error variance to obtain a working error term.

? R Example

Steps in R

Load the bake time data and obtain the ANOVA table by using the following commands:

```
setwd("~/path-to-folder/")
baketime_data <- read.table("baketime_data.txt",header=T)
attach(baketime_data)
baketime_anova<-aov(resp ~ factor(temp) + factor(time) + factor(temp):factor(tim
summary(baketime_anova)
#Error: factor(oven)
#          Df Sum Sq Mean Sq F value Pr(>F)
#Residuals  2   1963   981.4
#Error: factor(oven):factor(temp)
#             Df Sum Sq Mean Sq F value Pr(>F)
#factor(temp)  3  12494    4165   14.09  0.004 **
#Residuals     6   1774     296
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#Error: Within
#                          #Df Sum Sq Mean Sq F value Pr(>F)
```

```
#factor(time)                2    566   283.1   0.456  0.642
#factor(temp):factor(time)   6   2600   433.4   0.698  0.655
#Residuals                  16   9933   620.8
detach(baketime_data)
```

This page titled 8.1: Split-Plot Design in RCBD is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

8.2: Split-Plot Design in CRD

Recall the irrigation amount and fertilizer type example we discussed in the previous section. We had two large lands and managing the irrigation amount was harder on a smaller scale; we assigned the irrigation amount within each land to whole plots using an RCBD.

Now suppose in this case, instead of two large lands, we had 4 large fields. Irrigation amount is still a factor that is difficult to control. In that case, we can assign the irrigation amount randomly using a CRD for the 4 whole plots. Then each whole plot can be divided into smaller fields (subplots) and we can assign fertilizer type randomly within each whole plot.

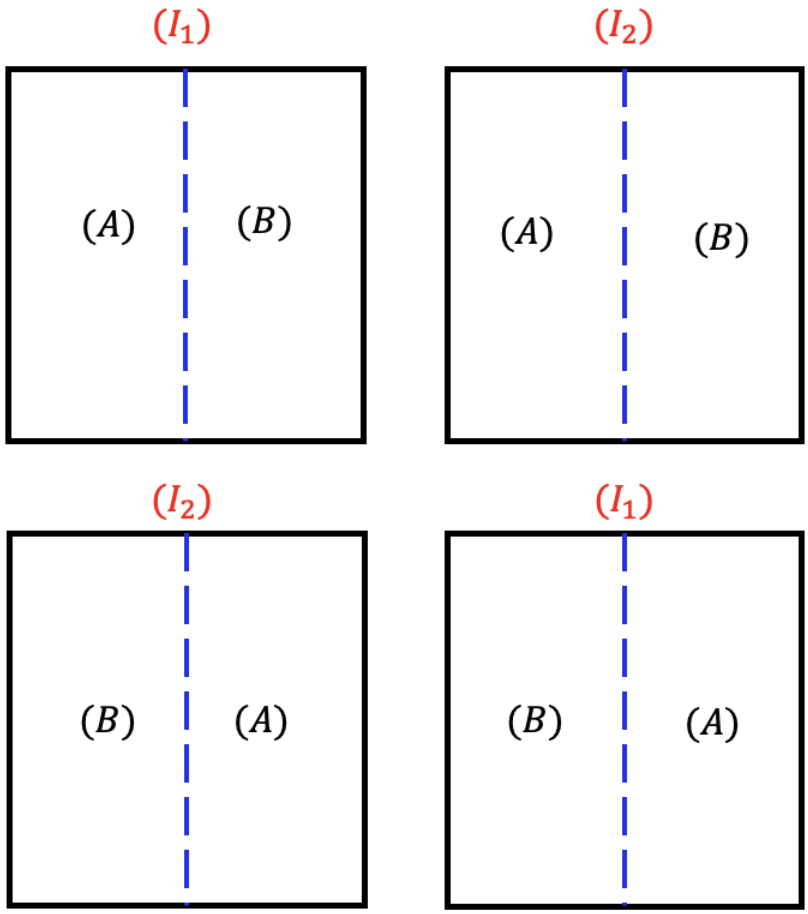


Figure 8.2.1: Large fields assigned irrigation amounts using CBD, with fertilizer type randomly assigned to subplots within each field.

Within the whole plot, the subplots are always arranged in an RCBD. The difference between split-plot in RCBD and split-plot in CRD is how the whole plot factor is randomized.

Example:

Consider a study in which the experimenters are interested in two factors: irrigation (Factor A at 2 levels) and seed type (Factor B at 2 levels), and they are crossed to form a factorial treatment design. The seed treatment can be easily applied at a small scale, but the irrigation treatment is problematic. Irrigating one plot may influence neighboring plots, and furthermore, the irrigation equipment is most efficiently used in a large area. As a result, the investigators want to apply the irrigation to a large *whole plot* and then split the whole plot into 2 smaller *subplots* in which they can apply the seed treatment levels.

In the first step, the levels of the irrigation treatment are applied to four experimental (fields) to end up with 2 replications:

Field 1	Field 2	Field 3	Field 4
A2	A1	A1	A2

Following that, the fields are split into two subplots and a level of Factor B is randomly applied to subplots within each application of the Irrigation treatment:

Field 1	Field 2	Field 3	Field 4
A2 B2	A1 B1	A1 B2	A2 B1
A2 B1	A1 B2	A1 B1	A2 B2

In this design, the whole plot treatments (i.e factor A, irrigation) are arranged in a CRD and the subplot treatments (i.e. factor B, seed type) are arranged within whole plots in an RCBD.

If we carefully think about this, we see that the replicates (i.e. fields) are nested within the whole factor levels. For example, fields 2 and 3 are nested within level A_1, and fields 1 and 4 are nested within level A_2. So the variability due to replicates is nested within the whole factor.

The statistical model for the design is:

$$Y_{ijk} = \mu + \alpha_i + \gamma_{k(i)} + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{8.2.1}$$

where $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, and $k = 1, 2, \ldots, r$ where a is the number of levels in factor A, b is the number of levels in factor B and r is the number of replicates.

As discussed in section 8.1, from the perspective of whole plots (i.e. Factor A, irrigation), the subplots are simply subsamples and it is reasonable to average them when testing the whole plot effects. If the values of the subplots within each whole plot are average, the resulting design is CRD, and the error term in a simple CRD is the $\text{replication(whole factor)}$. Therefore, for split-plot in CRD, the whole plot errors are computationally equivalent to $\text{replication(whole factor)}$, but in order to use it, we must explicitly extract it from the error term and put it in the model.

The ANOVA table, in this case, would look like this:

Source	DF	Expected Mean Square	Error Term
(Whole Plots)			
A	1	Var(Residual) + 2Var(Replicate(A)) + Q(A, A*B)	MS(Replicate(A))
Replicate(A)	2	Var(Residual) + 2Var(Replicate(A))	
(Subplots)			
B	1	Var(Residual) + Q(B, A*B)	MS(Residual)
A*B	1	Var(Residual) + Q(A*B)	MS(Residual)
Residual	2	Var(Residual)	

Using Technology

SAS Example

In SAS, the code would be:

```
proc mixed data=example_8_2 method=type3;
class factorA factorB field;
model resp=factorA factorB factorA*factorB;
random field(factorA);
run;
```

Minitab Example

In Minitab the "field(FactorA)" term would need to be constructed in the **Random/Nest**… options box under the **STAT > ANOVA > General Linear Model > Fit the General Linear Model**.

R Example

In R use the following code:

```
anova<-aov(resp ~ factorA + factorB + factorA:factorB + Error(factorA/replicate),data
summary(anova)
```

This page titled 8.2: Split-Plot Design in CRD is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

8.3: Split-Split-Plot Design

The idea of split-plots can easily be extended to multiple splits. In a 3-factor factorial, for example, it is possible to assign Factor A to whole plots, then Factor B to subplots within the applications of Factor A, and then split the experimental units used for Factor B into sub-subplots to receive the levels of Factor C.

For a fixed effect factorial treatment design in an RCBD (with blocks, levels of Factor A, levels of Factor B, and levels of Factor C), the split-split-plot would produce the following table:

Source	d.f.
(Whole plots)	
Block	$r-1$
Factor A	$a-1$
Whole plot error	$(r-1)(a-1)$
(Subplots)	
Factor B	$b-1$
A × B	$(a-1)(b-1)$
Subplot error	$a(r-1)(b-1)$
(Sub-subplots)	
Factor C	$c-1$
A × C	$(a-1)(c-1)$
B × C	$(b-1)(c-1)$
A × B × C	$(a-1)(b-1)(c-1)$
Sub-subplot error	\(ab(r - 1)(c - 1)\)
Total	$(rabc)-1$

The model is specified as we did earlier for the split-plot in RCBD, retaining only the interactions involving replication where they form denominators for F-tests for factor effects. For the model above, we would need to include the block, block × A, and block × A × B terms in the random statement in SAS. In SAS, Block × A × B would automatically include the Block × B effect SS and *df*. All other interactions involving replications and factor C would be included in the residual error term. The block × A term is often referred to as "Error a" ("Whole plot error" in the table), the Block × A × B term as "Error b" ("Subplot error" in the table), and the residual error as "Error c" ("Sub-subplot error" in the table) because of their roles as the denominator in the F-tests.

This page titled 8.3: Split-Split-Plot Design is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

8.4: Try It!

Exercise 8.4.1

Researchers are investigating the effect of storage temperature on bacterial growth for two types of seafood. They set up the experiment to evaluate 3 storage temperatures. There were 9 storage units that were available, and so they randomly selected 3 storage units to be used for each storage temperature, and both seafood types were stored in each unit. After 2 weeks, bacterial counts were made. After taking a logarithmic transformation of the counts, they produced the following ANOVA:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
temp	2	107.656588	53.828294	Var(Residual) + 2 Var(unit(temp)) + Q(temp, temp*seafood)
seafood	1	3.713721	3.713721	Var(Residual) + Q(seafood, temp*seafood)
temp*seafood	2	2.647594	1.323797	Var(Residual) + Q(temp*seafood)
unit(temp)	6	44.050650	7.341775	Var(Residual) + 2 Var(unit(temp))
Residual	6	5.590873	0.931812	Var(Residual)

a) For each factor, indicate whether it is a fixed or random effect.

b) Identify the treatments and describe (in words) the treatment design.

c) Describe the randomization used.

d) Compute the F-statistic for the temperature effect in the ANOVA, and determine significance for the effect.

Show Solution

a) temp=fixed, seafood=fixed, storage unit=random

b) Temperature and Seafood, factorial design. Each seafood type is combined with each temperature level in the experiment.

c) Split-plot in a CRD. Temperature levels were assigned (randomly) to storage units. Then the storage unit set at a given temperature is split to accommodate each of the two seafood types.

d) $F_{Temperature} = 53.83/7.342 = 7.3318 F_{critical} = 5.14$, so reject H_0.

Exercise 8.4.2

Answer the questions based on the following output:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square
group	3	6429.388333	2143.129444	Var(Residual) + 3 Var(blk*group) + Q(group,group*tech_int)
tech_int	2	881.408750	440.704375	Var(Residual) + Q(tech_int,group*tech_int)
group*tech_int	6	207.507917	34.584653	Var(Residual) + Q(group*tech_int)
blk	3	408.985000	136.328333	Var(Residual) + 3 Var(blk*group) + 12 Var(blk)
blk*group	9	466.543333	51.838148	Var(Residual) + 3 Var(blk*group)
Residual	24	595.696667	24.820694	Var(Residual)

a) For each factor, indicate whether it is a fixed or random effect

b) Identify the treatments and describe (in words) the treatment design.

c) Describe (in words) the randomization used.

d) Compute the F-statistic for each effect in the ANOVA, and determine significance (i.e., compare $F_{calculated}$ to $F_{critical}$ for each effect).

Show Solution

a) group = fixed, tech_int = fixed, blk = random

b) group and tech_int, crossed for a factorial treatment design

c) Split-plot in a RCBD, with *group* as the whole plot treatment and *tech_int* as the subplot treatment with *blk* as the blocking factor.

d) group: $F = \dfrac{2143.129444}{51.838148} = 41.3427$, $F_{critical} = 3.86$, reject H_0

tech_int: $F = \dfrac{440.704375}{24.820694} = 17.7555$, $F_{critical} = 3.40$, reject H_0

group $\times$ tech_int: $F = \dfrac{34.584653}{24.820694} = 1.3934$, $F_{critical} = 2.51$, do not reject H_0

blk: $F = \dfrac{136.3283}{51.8381} = 2.6299$, $F_{critical} = 3.86$, do not reject H_0

? Exercise 8.4.3

1. An experimenter wants to compare the yield of three varieties of oats at four different levels of manure. Suppose 6 farmers agree to participate in the experiment and each farmer will designate 3 fields from their farms for the experiment.

a. What is the treatment design?
b. What is the randomization design?

Show Solution

a) **Treatment design**: 3×4 factorial with oat variety and manure levels as factors having 3 and 4 levels respectively

b) **Randomization design**: Three oats varieties will be randomly assigned to the 3 fields from each farm using RCBD with farms as blocks. Four manure levels are then randomized within each field using an RCBD. So the randomization design is a split-plot in RCBD.

2. In an agricultural setting, an experimenter is applying one of two irrigation methods randomly to 6 plots where all plots are similar in moisture, soil type, slope, fertility, etc. Each plot is then subdivided into 5 portions and 5 levels of nitrogen fertilizer are applied randomly to these portions.

a. What is the treatment design?
b. What is the randomization design?

Show Solution

a) **Treatment design**: 2×5 factorial with irrigation method and fertilizer levels as factors having 2 and 5 levels respectively

b) **Randomization design**: Split-plot in CRD with the whole factor as irrigation method and subplot factor as fertilizer level

3. A survey was conducted among 100 high schoolers who were potential athletes to learn about their preferences on financial benefits. The sample consisted of an equal number of male and female students and 3 incentive types were offered: a 20% tuition reduction for all 4 years; a 50% tuition reduction in the first year, but renewable based on freshman GPA; and full room and board for all 4 years.

a. What is the treatment design?
b. What is the randomization design?

Show Solution

a) **Treatment design**: A single factor study with 3 levels; the factor of interest is incentive type

b) **Randomization design**: RCBD with gender as the blocking factor

This page titled 8.4: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

8.5: Chapter 8 Summary

In this chapter, we discussed split-plot designs with the special feature of having two types of experimental units: whole plots into which the whole plot treatments are assigned and the subplots into which the subplot treatments are assigned.

The whole plot assignment can be either according to a CRD or an RCBD, and depending on this design type, the overall design is called a split-plot in either CRD or RCBD. Note that in either case, the denominator of the F-statistic for testing the whole plot factor is not MSE, but equals the MS of $\mathrm{replicate}(\mathrm{A})$ and MS of $\mathrm{block} \times \mathrm{A}$ respectively.

This page titled 8.5: Chapter 8 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics.

CHAPTER OVERVIEW

9: ANCOVA Part I

Objectives

Upon completion of this chapter, you should be able to:

1. Be familiar with the basics of the General Linear Model (GLM) necessary for ANCOVA implementation.
2. Develop the ANCOVA procedure by extending the ANOVA methodology to include a continuous predictor.
3. Carry out the testing sequences for ANCOVA with equal and unequal slopes.

The analysis of covariance (ANCOVA) procedure is used when the statistical model has both quantitative and qualitative predictors and is based on the concepts of the General Linear Model (GLM). In ANCOVA, we will combine the concepts applicable to categorical factors learned so far in this course with the principles and foundations of regression, applicable to continuous predictors learned in STAT 501.

In this chapter, we will address the classic case of ANCOVA where the ANOVA model is extended to include the linear effect of a continuous variable, known as the covariate. In the next chapter, we will generalize the ANCOVA model to include the quadratic and cubic effects of the covariate as well.

You might find it interesting that when SAS first came out they had PROC ANOVA and PROC REGRESSION and that was it. Then people asked, "What about the case when you have categorical factors and you want to do an ANOVA but now you have this other variable, a continuous variable, that you can use as a covariate to account for extraneous variability in the response?" So, SAS came out with PROC GLM, which is the general linear model. With PROC GLM you could take the continuous regression variable and pop it into the ANOVA model and it runs. Or, conversely, if you are running a regression and you have a categorical predictor like gender, you could include it into the regression model and it runs. The general linear model handles both the regression and the categorical variables in the same model. There is no PROC ANCOVA in SAS, but there is PROC MIXED. PROC GLM had problems when it came to random effects and was effectively replaced by PROC MIXED. The same sort of process can be seen in Minitab and accounts for the multiple tabs under Stat > ANOVA and Stat > Regression. In SAS PROC MIXED or in Minitab's General Linear Model, you have the capacity to include covariates and correctly work with random effects. But enough about history; let's get to this lesson.

Introduction to Analysis of Covariance (ANCOVA)

A "classic" ANOVA tests for differences in mean responses to categorical factor (treatment) levels. When there is heterogeneity in experimental units, sometimes restrictions on the randomization (blocking) can improve the accuracy of significance testing results. In some situations, however, the opportunity to construct blocks may not exist, but there may be a continuous variable that may be causing the heterogeneity in the experimental units. Such sources of extraneous variability are referred to as "covariates", and historically have been also termed "nuisance" or "concomitant" variables.

Note that an ANCOVA model is formed by including a continuous covariate in an ANOVA model. As the continuous covariate enters the model as a regression variable, an ANCOVA requires a few additional steps that should be combined with the ANOVA procedure.

This page titled 9: ANCOVA Part I is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Role of the Covariate

To illustrate the role the covariate has in the ANCOVA, let's look at a hypothetical situation wherein investigators are comparing the salaries of male vs. female college graduates. A random sample of 5 individuals for each gender is compiled, and a simple one-way ANOVA is performed:

Males	Females
78	80
43	50
103	30
48	20
80	60

$H_0: \mu_{\text{males}} = \mu_{\text{females}}$

? SAS Example

Using SAS

SAS coding for the One-way ANOVA:

```
data ancova_example;
input gender $ salary;
datalines;
m 78
m 43
m 103
m 48
m 80
f 80
f 50
f 30
f 20
f 60
;
proc mixed data=ancova_example method=type3;
class gender;
model salary=gender;
run;
```

Here is the output we get:

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
gender	1	8	2.11	F">0.1840

Minitab Example

Using Minitab

To perform a one-way ANOVA test in Minitab, you can first open the data (ANCOVA Example Minitab Data) and then select **Stat** > **ANOVA** > **One Way…**

In the pop-up window that appears, select salary as the Response and gender as the Factor.

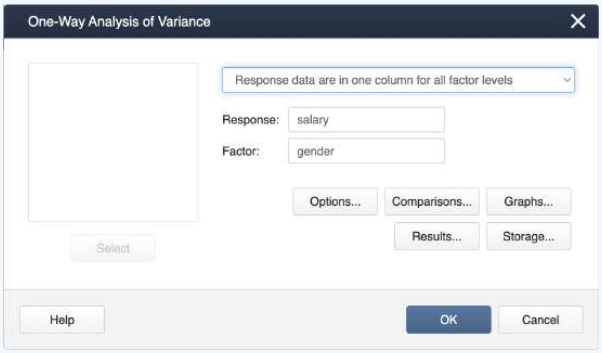


Figure 9.1.1: Minitab One-Way Analysis of Variance window

Click **OK**, and the output is as follows.

Analysis of Variance

Source	DF	SS	SS	F-Value	P-Value
gender	1	1254	1254	2.11	0.184
Error	8	4745	593		
Total	9	6000			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
24.3547	20.91%	11.02%	0.00%

R Example

Using R

Tasks:

- Load the ANCOVA example data.
- Obtain the ANOVA table.
- Plot the data.

1. Load the ANCOVA example data and obtain the ANOVA table by using the following commands:

```
setwd("~/path-to-folder/")
ancova_example_data <- read.table("ancova_example.txt",header=T)
attach(ancova_example_data)
ancova<-aov(salary ~ gender,ancova_example_data)
summary(ancova)
#            Df  Sum Sq  Mean Sq  F value  Pr(>F)
#gender       1    1254   1254.4    2.115   0.184
#Residuals    8    4745    593.1
```

2. Plot for the data, salary by gender, by using the following commands:

```
library(ggplot2)
myplot<-ggplot(ancova_example_data, aes(x = gender, y = salary)) +  geom_point()
myplot + theme_bw() + theme(panel.border = element_blank(), panel.grid.major = e
panel.grid.minor = element_blank(), axis.line = element_line(colour = "black"))
```

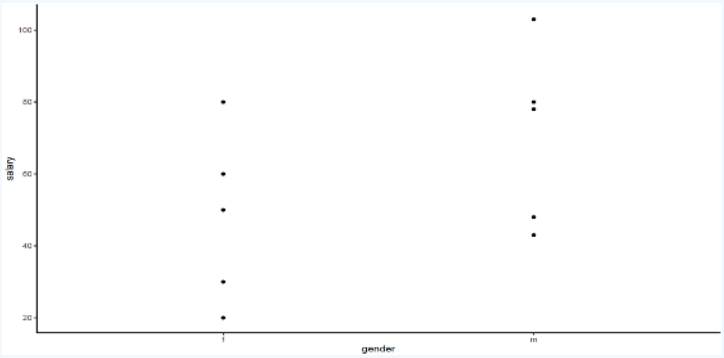


Figure 9.1.2: Gender and salary plot

3. Plot for the data, salary vs years, by using the following commands:

```
plot(years,salary, xlab="Years after graduation", ylab="Salary(Thousands)",pch=2
abline(lm(salary~years,data=ancova_example_data))
detach(ancova_example_data)
```

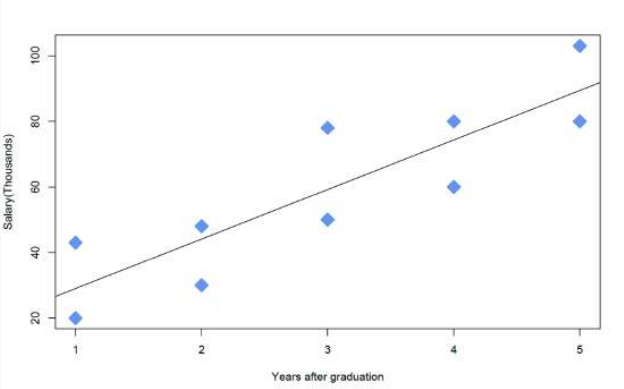


Figure 9.1.3: Plot of salary vs years

Because the p-value > α (=0.05), they can't reject the H_0.

A plot of the data shows the situation:

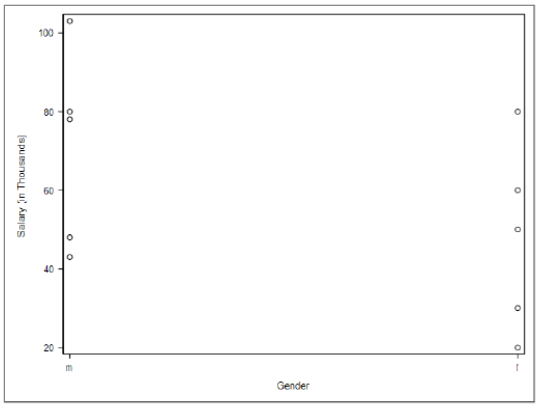


Figure 9.1.4: Plot of salary vs gender

However, it is reasonable to assume that the length of time since graduation from college is also likely to influence one's income. So more appropriately, the duration since graduation, a continuous variable, should be also included in the analysis, and the required data is shown below.

Females		**Males**	
Salary	years	Salary	years
80	5	78	3
50	3	43	1
30	2	103	5
20	1	48	2
60	4	80	4

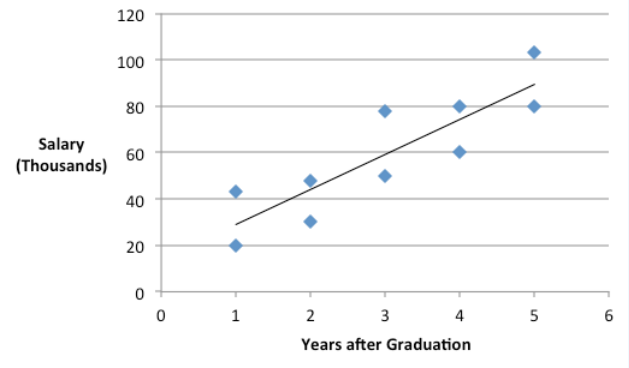


Figure 9.1.5: Plot of salary vs years since graduation

The plot above indicates an upward linear trend between salary and the number of years since graduation, which could be a marker for experience and/or postgraduate education. The fundamental idea of including a covariate is to take this trend into account and to "control" it effectively. In other words, including the covariate in the ANOVA will make the comparison between Males and Females after accounting for the covariate.

This page titled 9.1: Role of the Covariate is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

9.2: ANCOVA in the GLM Setting - The Covariate as a Regression Variable

In this section, we will develop the statistical ANCOVA, which by definition is a general linear model that includes both ANOVA (categorical) predictors and regression (continuous) predictors. The simple linear regression model is:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{9.2.1}$$

where β_0 and β_1 are the intercept and the slope of the line, respectively. The significance of a regression is equivalent to testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ using the F statistic: $\frac{MS(Regr)}{MSE}$ where $MS(Regr)$ is the mean sum of squares for regression and MSE is the mean squared error. In this case of a simple linear regression, this test is equivalent to a t-test.

Now, in adding the regression variable to our one-way ANOVA model, we can envision a notational problem. In the balanced one-way ANOVA, we have the grand mean (μ), but now we also have the intercept β_0.

To get around this, we can use

$$X^* = X_{ij} - \bar{X} \tag{9.2.2}$$

and get the following as an expression of our covariance model:

$$Y_{ij} = \mu + \tau_i + \gamma X^* + \epsilon_{ij} \tag{9.2.3}$$

Note that the above model fits into the general linear model (GLM) and the Type III (model fit) sums of squares for the treatment levels in this model are being corrected (or adjusted) for the regression relationship. This has the effect of evaluating the treatment levels "on the same playing field", that is, comparing the means of the treatment levels at the mean value of the covariate. This process effectively removes the variation due to the covariate that may otherwise be attributed to treatment level differences.

This page titled 9.2: ANCOVA in the GLM Setting - The Covariate as a Regression Variable is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

9.3: Steps in ANCOVA

First, we need to confirm that for at least one of the treatment groups there is a significant regression relationship with the covariate. Otherwise, including the covariate in the model won't improve the estimation of treatment means.

Then, we need to make sure that the regression relationship of the response with the covariate has the same slope for each treatment group. Graphically, this means that the regression line at each factor level has the same slope and therefore the lines are all parallel. Depending on the outcome of the test for equal slopes, we have two alternative ways to finish up the ANCOVA:

1. Fit a common slope model and adjust the treatment SS for the presence of the covariate
2. Evaluate the differences in means at least three levels of the covariate

These steps are illustrated in the following two sections and are diagrammed below:

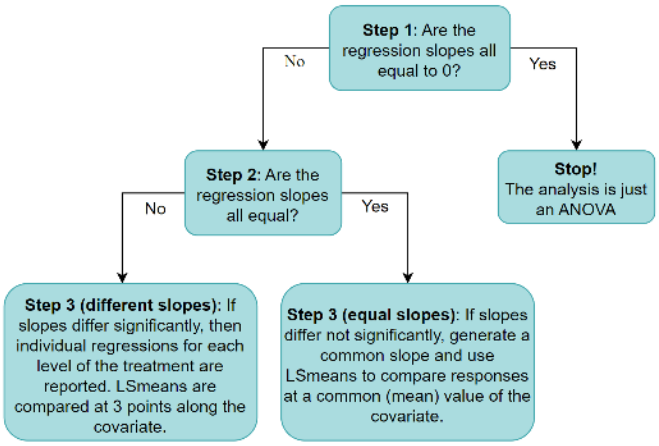


Figure 9.3.1: Flowchart for the ANCOVA

Note

The figure above is presented as a guideline and does require some subjective judgment. Small sample sizes, for example, may result in none of the individual regressions in step 1 being statistically significant. Yet the inclusion of the covariate in the model may still be advantageous, as pooling the data will increase the number of observations when fitting the joint model. Exploratory data analysis and regression diagnostics also will be useful.

This page titled 9.3: Steps in ANCOVA is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

9.4: Using Technology - Equal Slopes Model

Using Technology

SAS Example

Using our Salary example using the data in the table below, we can run through the steps for the ANCOVA.

Females		**Males**	
Salary	Years	Salary	Years
80	5	78	3
50	3	43	1
30	2	103	5
20	1	48	2
60	4	80	4

Steps in SAS

Step 1: Are all regression slopes = 0?

A simple linear regression can be run for each treatment group, Males and Females.

Running these procedures using statistical software we get the following:

Males

Use the following SAS code:

```
data equal_slopes;
input gender $ salary years;
datalines;
m 78 3
m 43 1
m 103 5
m 48 2
m 80 4
f 80 5
f 50 3
f 30 2
f 20 1
f 60 4
;
proc reg data=equal_slopes;
where gender='m';
model salary=years;
title 'Males';
run; quit;
```

And here is the output that you get:

The REG Procedure Mode1:: MODEL1 Dependent Variable: salary	
Number of Observations Read	5
Number of Observations Used	5

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2310.40000	2310.40000	44.78	F" class=" ">0.0068
Error	3	154.80000	51.60000		F" class=" ">
Corrected Total	4	2465.20000			

Females

Use the following SAS code:

```
data equal_slopes;
input gender $ salary years;
datalines;
m 78 3
m 43 1
m 103 5
m 48 2
m 80 4
f 80 5
f 50 3
f 30 2
f 20 1
f 60 4
;
proc reg data=equal_slopes;
where gender='f';
model salary=years;
title 'Females';
run; quit;
```

And here is the output for this run:

The REG Procedure Mode1:: MODEL1 Dependent Variable: salary	
Number of Observations Read	5
Number of Observations Used	5

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2250.00000	2250.00000	225.00	F" class=" ">0.0006
Error	3	30.00000	10.00000		F" class=" ">
Corrected Total	4	2280.00000			F" class=" ">

In both cases, the simple linear regressions are significant, so the slopes are not = 0.

Step 2: Are the slopes equal?

We can test for this using our statistical software.

In SAS we now use `proc mixed` and include the covariate in the model.

We will also include a "treatment × covariate" interaction term and the significance of this term answers our question. If the slopes differ significantly among treatment levels, the interaction p-value will be < 0.05.

If the slopes differ significantly among treatment levels, the interaction p-value will be < 0.05.

```
data equal_slopes;
input gender $ salary years;
datalines;
m 78 3
m 43 1
m 103 5
m 48 2
m 80 4
f 80 5
f 50 3
f 30 2
f 20 1
f 60 4
;
proc mixed data=equal_slopes;
class gender;
model salary = gender years gender*years;
run;
```

Note

In SAS, we specify the treatment in the class statement, indicating that these are categorical levels. By NOT including the covariate in the class statement, it will be treated as a continuous variable for regression in the model statement.

The Mixed Procedure Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
years	1	6	148.06	F" class=" "><.0001
gender	1	6	7.01	F" class=" ">0.0381

Proc…	years*gender	1	6	0.01	F" class=" ">0.9384

So here we see that the slopes are equal and in a plot of the regressions, we see that the lines are parallel.

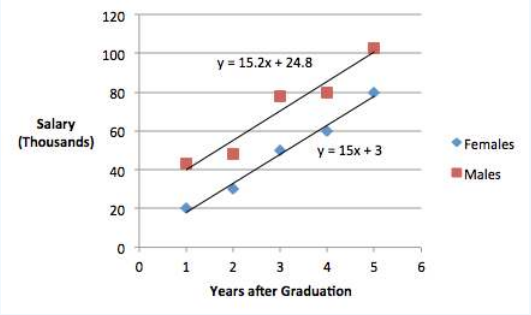


Figure 9.4.a1: Parallel lines of best fit

To obtain the plot in SAS, we can use the following SAS code:

SAS code:

```
ods graphics on;
proc sgplot data=equal_slopes;
styleattrs datalinepatterns=(solid);
reg y=salary x=years / group=gender;
run;
```

Step 3: Fit an Equal Slopes Model

We can now proceed to fit an Equal Slopes model by removing the interaction term. Again, we will use our statistical software SAS.

```
data equal_slopes;
input gender $ salary years;
datalines;
m 78 3
m 43 1
m 103 5
m 48 2
m 80 4
f 80 5
f 50 3
f 30 2
f 20 1
f 60 4
;
proc mixed data=equal_slopes;
class gender;
model salary = gender years;
lsmeans gender / pdiff adjust=tukey;
/* Tukey unnecessary with only two treatment levels */
title 'Equal Slopes Model';
run;
```

We obtain the following results:

The Mixed Procedure Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
years	1	7	172.55	F" class=" "><.0001
gender	1	7	47.46	F" class=" ">0.0002

Least Squares Means						
Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
gender	f	48.0000	2.2991	7	20.88	|t|" class=" "> <.0001
gender	m	70.4000	2.2991	7	30.62	|t|" class=" "> <.0001

In SAS, the model statement automatically creates an intercept, and so the ANCOVA model is technically over-parameterized. To get the slopes and intercepts for the covariate directly, we have to re-parameterize the model. This entails suppressing the intercept (`noint`), and then specifying that we want the solutions, (`solution`), to the model. Here is what the SAS code looks like for this:

```
data equal_slopes;
input gender $ salary years;
datalines;
m 78 3
m 43 1
m 103 5
m 48 2
m 80 4
f 80 5
f 50 3
f 30 2
f 20 1
f 60 4
;
proc mixed data=equal_slopes;
class gender;
model salary = gender years / noint solution;
ods select SolutionF;
title 'Equal Slopes Model';
run;
```

Here is the output:

Solution for Fixed Effects						
Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|

Solution for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Fix… gender	f	2.7000	4.1447	7	0.65	|t|" class=" ">0.5356
Fix… gender	m	25.1000	4.1447	7	6.06	|t|" class=" ">0.0005
Fix… years		15.1000	1.1495	7	13.14	|t|" class=" ">.0001

In the first section of the output above is reported a separate intercept for each gender, the 'Estimate' value for each gender, and a common slope for both genders, labeled 'Years'.

Thus, the estimated regression equation for Females is $\hat{y} = 2.7 + 15.1(\text{Years})$, and for Males it is $\hat{y} = 25.1 + 15.1(\text{Years})$.

To this point in this analysis, we can see that 'gender' is now significant. By removing the impact of the covariate, we went from

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
of … gender	1	8	2.11	F" class=" ">0.1840

(without covariate consideration)

to

gender	1	7	47.46	0.0002

(adjusting for the covariate)

Minitab Example

Using our Salary example and the data in the table below, we can run through the steps for the ANCOVA. On this page, we will go through the steps using Minitab.

Females		**Males**	
Salary	Years	Salary	Years
80	5	78	3
50	3	43	1
30	2	103	5
20	1	48	2
60	4	80	4

Steps in Minitab

Step 1: Are all regression slopes = 0?

A simple linear regression can be run for each treatment group, Males and Females. To perform regression analysis on each gender group in Minitab, we will have to subdivide the salary data manually and separately, saving the male data into the Male Salary Dataset and the female data into the Female Salary dataset.

Running these procedures using statistical software we get the following:

Males

Open the Male dataset in the Minitab project file (Male Salary Dataset).

Then, from the menu bar, select **Stat** > **Regression** > **Regression** > **Fit Regression Model**

In the pop-up window, select salary into Response and years into Predictors as shown below.

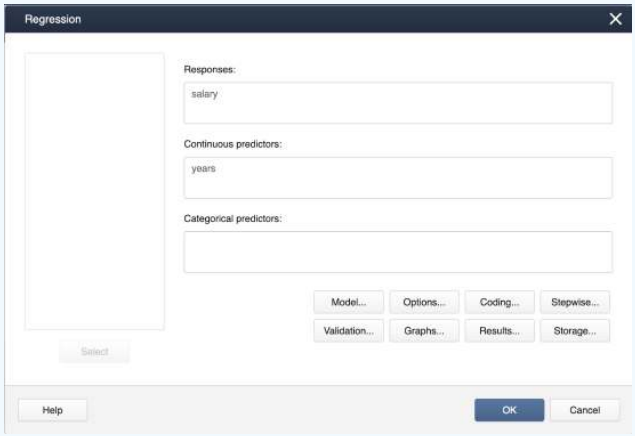


Figure 9.4.*b*1: Minitab Regressions pop-up window

Click **OK**, and Minitab will output the following.

Regression Analysis: Salary versus years

Regression Equation: salary = 24.8 + 15.2 years

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	24.80	7.53	3.29	0.046	
years	15.20	2.27	6.69	0.007	1.00

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
7.18331	R-Sq = 93.7%	91.6%	85.94%

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Regression	1	2310.4	2310.40	44.78	0.007
years	1	2310.4	2310.40	44.78	0.007
Residual Error	3	154.8	51.6		
Total	4	2465.2			

Females

Open Minitab dataset Female Salary Dataset. Follow the same procedure as was done for the Male dataset and Minitab will output the following:

Regression Analysis: Salary versus years

Regression Equation: salary = 3.00 + 15.00 years

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.00	3.32	0.90	0.432	
years	15.00	1.00	15.00	0.001	1.00

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
3.16228	98.68%	98.25%	95.92%

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Regression	1	2250.0	2250.0	225.00	0.001
years	1	2250.0	2250.0	225.00	0.001
Residual Error	3	30.0	10.0		
Total	4	2280.0			

In both cases, the simple linear regressions are significant, so the slopes are not = 0.

Step 2: Are the slopes equal?

We can test for this using our statistical software. In Minitab, we must now use GLM (general linear model) and be sure to include the covariate in the model. We will also include a "treatment x covariate" interaction term and the significance of this term is what answers our question. If the slopes differ significantly among treatment levels, the interaction *p*-value will be < 0.05.

First, open the dataset in the Minitab project file Salary Dataset. Then, from the menu select **`Stat`** > **`ANOVA`** > **`General Linear Model`** > **`Fit General Linear Model`**

In the dialog box, select salary into Responses, gender into Factors, and years into Covariates.

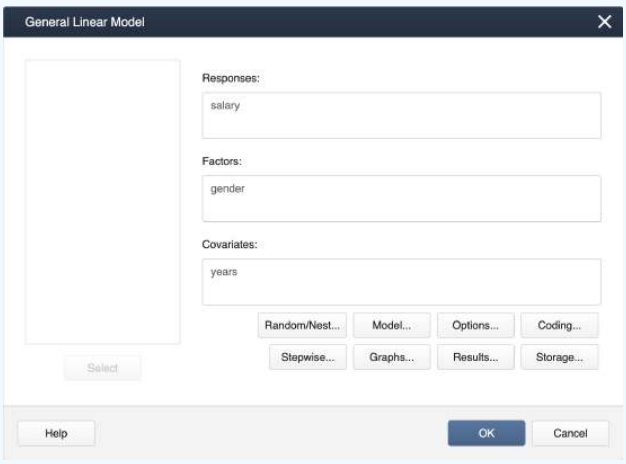


Figure \(\
PageIndex
{b2}\): Minitab
GLM
pop-up selections

To add the interaction term, first click **Model**…. Then, use the shift key to highlight gender and years, and click **Add**. Click **OK**, then **OK** again, and Minitab will display the following output.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
year	1	4560.20	4560.20	148.06	0.000
gender	1	216.02	216.02	7.01	0.038
years*gender	1	0.20	0.20	0.01	0.938
Error	6	184.80	30.80		
Total	9	5999.60			

It is clear the interaction term is not significant. This suggests the slopes are equal. In a plot of the regressions, we can also see that the lines are parallel.

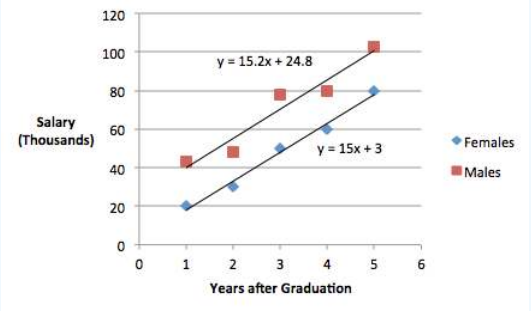


Figure 9.4.*b*3: Parallel lines of best fit

Step 3: Fit an Equal Slopes Model

We can now proceed to fit an Equal Slopes model by removing the interaction term. This can be easily accomplished by starting again with **`STAT` > `ANOVA` > `General Linear Model` > `Fit General Linear Model`**

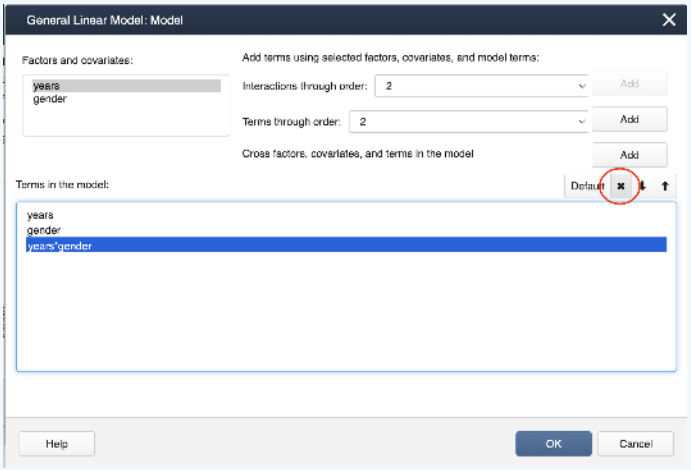


Figure 9.4.*b*4: Removing the `years*gender` term from the model

Click **OK**, then **OK** again, and Minitab will display the following output.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
year	1	4560.20	4560.20	172.55	0.000
gender	1	1254.4	1254.40	47.46	0.000
Error	7	185.0	26.43		
Total	9	5999.6			

To generate the mean comparisons select **`STAT`** > **`ANOVA`** > **`General Linear Model`** > **`Comparisons...`** and fill in the dialog box as seen below.

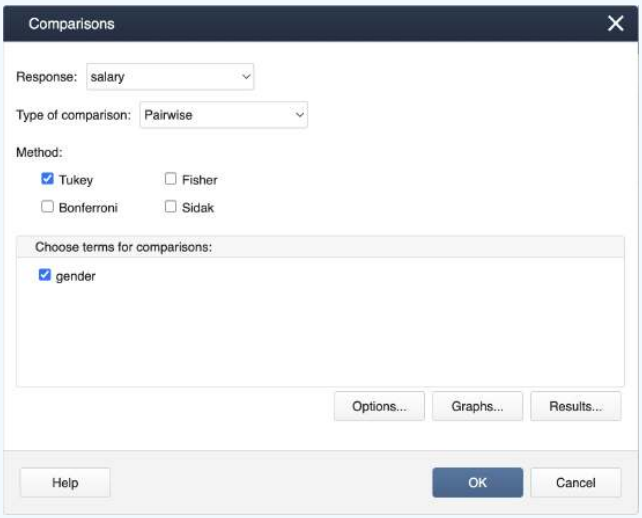


Figure 9.4.*b5*: Comparisons window selections

Click **OK** and Minitab will produce the following output.

Comparison of salary

Tukey Pairwise Comparisons: gender
Grouping information Using the Tukey Method and 95% Confidence

gender	N	Mean	Grouping
Male	5	70.4	A
gender	5	48.0	B

Means that do not share a letter are significantly different.

? R Example

Steps for the ANCOVA for the Salary example in R:

- Run a simple linear model for each treatment group.
- Testing whether the slopes are equal.
- Plot the regression lines.
- Fit an equal slopes model.

Steps in R

1. Run a simple linear model for each treatment group (males and females) by using the following commands:

Males

```
males_regression <- lm(salary~years,data=subset(equal_slopes_data,gender=="m"))
anova(males_regression)
#Analysis of Variance Table
#Response: salary
#          Df  Sum Sq  Mean Sq  F value    Pr(>F)
#years      1  2310.4   2310.4   44.775  0.006809 **
#Residuals  3   154.8     51.6
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#summary(males_regression)$coefficients
#              Estimate  Std. Error    t value     Pr(>|t|)
#(Intercept)       24.8    7.533923   3.291778  0.046016514
#years             15.2    2.271563   6.691427  0.006808538
```

Females

```
females_regression <- lm(salary~years,data=subset(equal_slopes_data,gender=="f")
anova(females_regression)
#Analysis of Variance Table
#Response: salary
#          Df  Sum Sq  Mean Sq  F value     Pr(>F)
#years      1    2250     2250      225  0.0006431 ***
#Residuals  3      30       10
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#  summary(females_regression)$coefficients
#              Estimate  Std. Error   t value       Pr(>|t|)
#(Intercept)          3    3.316625  0.904534   0.4323889978
#years               15    1.000000 15.000000   0.0006431193
```

2. Test whether the slopes are equal by using the following commands:

```
ancova_model<-lm(salary ~ gender + years + gender:years,equal_slopes_data)
anova(ancova_model)
Analysis of Variance Table
Response: salary
             Df  Sum Sq  Mean Sq  F value     Pr(>F)
gender        1  1254.4   1254.4  40.7273  0.0006961 ***
years         1  4560.2   4560.2 148.0584  1.874e-05 ***
gender:years  1     0.2      0.2   0.0065  0.9383948
Residuals     6   184.8     30.8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

With a p-value of 0.9383948 in the interaction term (`gender*years`), we can conclude that the slopes are equal.

3. Plot the regression line for males and females by using the following commands:

```
plot(years,salary, xlab="Years after graduation", ylab="Salary(Thousands)",pch=2
abline(males_regression)
abline(females_regression)
text(locator(1),"y=15.2x+24.8",col="red")
text(locator(1),"y=15x+3",col="blue")
```

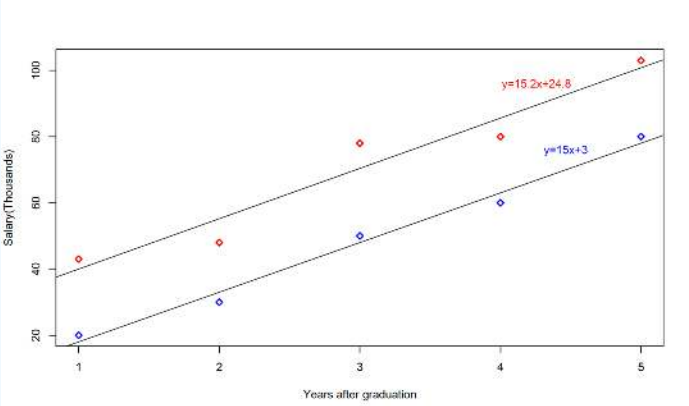


Figure 9.4.*c*1: Regression lines for male and female data

4. Fit an equal slopes model by using the following commands:

```
equal_slopes_model<-lm(salary ~ gender + years,equal_slopes_data)
anova(equal_slopes_model)
#Analysis of Variance Table
#Response: salary
#          Df  Sum Sq  Mean Sq  F value     Pr(>F)
#gender     1  1254.4   1254.4   47.464  0.0002335 ***
#years      1  4560.2   4560.2  172.548  3.458e-06 ***
#Residuals  7   185.0     26.4
#---
#Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can see that gender is significant now. To estimate the two regression lines, we need the following output:

```
summary(equal_slopes_model)$coefficients
#Coefficients:
#            Estimate  Std. Error  t value  Pr(>|t|)
#(Intercept)    2.700       4.145    0.651  0.535560
#genderm       22.400       3.251    6.889  0.000234
#years         15.100       1.150   13.136  3.46e-06
detach(equal_slopes_data)
```

The estimate for the years (15.1) is the slope of the models. The intercept for females is 2.7 and the intercept for males is 2.7+22.4=25.1

Thus, the estimated regression equation for females is $y = 15.1x + 2.7$ and for males it's $y = 15.1x + 25.1$.

This page titled 9.4: Using Technology - Equal Slopes Model is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

9.5: Using Technology - Unequal Slopes Model

SAS Example

If the data collected in the example study were instead as follows:

Females		**Males**	
Salary	Years	Salary	Years
80	5	42	1
50	3	112	4
30	2	92	3
20	1	62	2
60	4	142	5

We would see in **Step 2** of the ANCOVA that we do have a significant treatment × covariate interaction.

Steps for ANCOVA

Using this SAS program with the new data shown below.

```
data unequal_slopes;
input gender $ salary years;
datalines;
m 42 1
m 112 4
m 92 3
m 62 2
m 142 5
f 80 5
f 50 3
f 30 2
f 20 1
f 60 4
;
proc mixed data=unequal_slopes;
class gender;
model salary=gender years gender*years;
title 'Covariance Test for Equal Slopes';
/*Note that we found a significant years*gender interaction*/
/*so we add the lsmeans for comparisons*/
/*With 2 treatments levels we omitted the Tukey adjustment*/
lsmeans gender/pdiff at years=1;
lsmeans gender/pdiff at years=3;
lsmeans gender/pdiff at years=5;
run;
```

We get the following output:

Type 3 Test of Fixed Effects				
Effect	Num DF	De DF	F Value	Pr > F
years	1	6	800.00	F"><.0001
gender	1	6	6.55	F">0.0430
years*gender	1	6	50.00	F">0.0004

Generating Covariate Regression Slopes and Intercepts

```
data unequal_slopes;
input gender $ salary years;
datalines;
m 42 1
m 112 4
m 92 3
m 62 2
m 142 5
f 80 5
f 50 3
f 30 2
f 20 1
f 60 4
;
proc mixed data=unequal_slopes;
class gender;
model salary=gender years gender*years / noint solution;
ods select SolutionF;
title 'Reparmeterized Model';
run;
```

Output:

Solution for Fixed Effects						
Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
gender	f	3.0000	3.3166	6	0.90	|t|">0.4006
gender	m	15.0000	3.3166	6	4.52	|t|">0.0040
years		25.0000	1.0000	6	25.00	|t|"><.0001
years*gender	f	-10.0000	1.4142	6	-7.07	|t|">0.0004
years*gender	m	0	.	.	.	|t|">.

Here the intercepts are the Estimates for effects labeled "gender" and the slopes are the Estimates for the effect labeled "years*gender". Thus, the regression equations for this unequal slopes model are:

$$\text{Females} \quad \hat{y} = 3.0 + 15(\text{Years}) \tag{9.5.1}$$

$$\text{Males} \quad \hat{y} = 15 + 25(\text{Years}) \tag{9.5.2}$$

The slopes of the regression lines differ significantly and are not parallel:

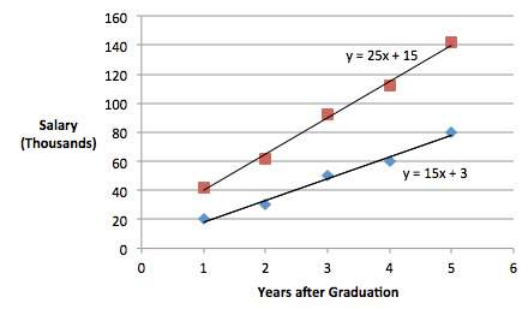


Figure 9.5.a1: Non-parallel regression lines of salary vs years since graduation

And here is the output:

Differences of Least Squares Means								
Effect	gender	_gender	years	Estimate	Standard Error	DF	t Value	Pr > |t|
gender	f	m	1.00	-22.000	3.4641	6	-6.35	|t|">0.0007
gender	f	m	3.00	-42.000	2.0000	6	-21.00	|t|">< .0001
gender	f	m	5.00	-62.000	3.4641	6	-17.90	|t|">< .0001

In this case, we see a significant difference at each level of the covariate specified in the `lsmeans` statement. The magnitude of the difference between males and females differs (giving rise to the interaction significance). In more realistic situations, a significant treatment × covariate interaction often results in significant treatment level differences at certain points along the covariate axis.

? Minitab Example

Steps in Minitab

When we re-run the program with the new dataset Salary-new Data, we find a significant interaction between gender and years.

To do this, open the Minitab dataset Salary-new Data.

Go to **`Stat`** > **`ANOVA`** > **`General Linear model`** > **`Fit General Linear Model`** and follow the same sequence of steps as in the previous section. In Step 2, Minitab will display the following output.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
years	1	8000.0	8000.0	800.00	0.000
gender	1	65.5	65.45	6.55	0.043
years*gender	1	500.0	500.0	50.00	0.000
Error	6	60.0	10.00		
Total	9	12970.0			

It is clear the interaction term is significant and should not be removed. This suggests the slopes are not equal. Thus, the magnitude of the difference between males and females differs (giving rise to the interaction significance).

R Example

Steps:

- Fit an unequal slopes model.
- Plot the regression lines.

Steps in R

1. Fit an unequal slopes model by using the following commands:

```
setwd("~/path-to-folder/")
unequal_slopes_data <- read.table("unequal_slopes.txt",header=T)
attach(unequal_slopes_data)
unequal_slopes_model<-lm(salary ~ gender + years + gender:years,unequal_slopes_d
anova(unequal_slopes_model)
#Analysis of Variance Table
#Response: salary
#             Df Sum Sq Mean Sq F value     Pr(>F)
#gender        1   4410    4410     441  7.596e-07 ***
#years         1   8000    8000     800  1.293e-07 ***
#gender:years  1    500     500      50  0.0004009 ***
#Residuals     6     60      10
#---
#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a p-value of 0.0004009 in the interaction term (`gender*years`), we can conclude that the slopes are unequal. To estimate the two regression lines, we need the following output:

```
#summary(unequal_slopes_model)$coefficients
#             Estimate  Std. Error    t value      Pr(>|t|)
#(Intercept)         3    3.316625   0.904534  4.005719e-01
#genderm            12    4.690416   2.558409  4.300074e-02
#years              15    1.000000  15.000000  5.530240e-06
#genderm:years      10    1.414214  7.071068   4.008775e-04
```

Here the intercept for females is the estimate for `intercept` and the intercept for males is the summation of the estimates `intercept + genderm` (note the letter *m* after *gender*). The slope for females is the estimate for `years` and the slope for males is the summation of the estimates `years + genderm: years` (note the letter *m* after *gender*). Thus, the regression equations for the unequal slopes model are: $y = 3 + 15x$ for females and $y = 15 + 25x$ for males.

2. Plot the regression lines by using the following commands:

```
males_regression <- lm(salary~years,data=subset(unequal_slopes_data,gender=="m")
females_regression <- lm(salary~years,data=subset(unequal_slopes_data,gender=="f
plot(years,salary, xlab="Years after graduation", ylab="Salary(Thousands)",pch=2
abline(males_regression)
abline(females_regression)
text(locator(1),"y=25x+15",col="red")
```

```
text(locator(1),"y=15x+3",col="blue")
detach(unequal_slopes_data)
```

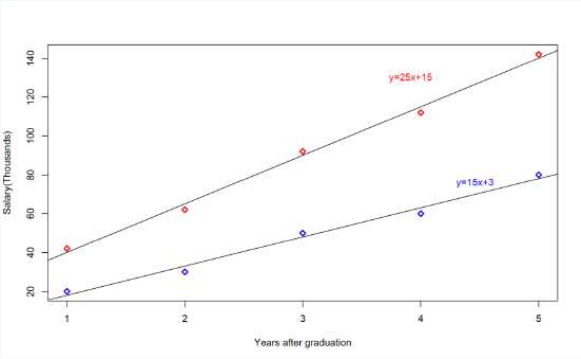


Figure 9.5.c1: Regression line plot in R

This page titled 9.5: Using Technology - Unequal Slopes Model is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

9.6: Chapter 9 Summary

This chapter introduced us to ANCOVA methodology, which accommodates both continuous and categorical predictors. The model discussed in this chapter has one categorical factor and only the linear effect of one single covariate, the continuous predictor. We noted that the fitted linear relationship between the response and the covariate results in a straight line for each factor level and the ANCOVA procedure then depends on the condition of equal slopes. One advantage of ANCOVA is the ability to examine the differences among the factor levels after adjusting for the impact of the covariate on the response.

The salary data comparing males and females after accounting for their years after college illustrated how software such as SAS and Minitab can be utilized in analyzing data using the ANCOVA procedure. In the next chapter, the ANCOVA topic will be extended to include up to a cubic polynomial as the regression model of the response vs. covariate.

This page titled 9.6: Chapter 9 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

10: ANCOVA Part II

Objectives

Upon completion of this chapter, you should be able to:

- Use ANCOVA to analyze experiments that require polynomial modeling for quantitative (numerical) predictors.
- Test hypotheses for treatment effects on polynomial coefficients.

In this chapter, we will extend our work with ANCOVA to model quantitative predictors with higher-order polynomials by utilizing orthogonal polynomial coding. Fitting a polynomial to express the impact of the quantitative predictor on the response is also called trend analysis and helps to evaluate the separate contributions of linear and nonlinear components of the polynomial. The examples discussed will illustrate how software can be used to fit higher-order polynomials within an ANCOVA model.

This page titled 10: ANCOVA Part II is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

10.1: ANCOVA with Quantitative Factor Levels

An Extended Overview of ANCOVA

Designed experiments often contain treatment levels that have been set with increasing numerical values. For example, a chemical process may be hypothesized to vary by two factors: the Reagent type (A or B), and temperature. So the researchers conducted an experiment that investigates a response at 40, 50, 60, 70, and 80 degrees (Fahrenheit) for each of the Reagent types.

You can find the data at **QuantFactorData.csv**.

If temperature is considered as a categorical factor, we can proceed as usual with a 2 × 5 factorial ANOVA to evaluate the Null Hypotheses:

$$H_0 : \mu_A = \mu_B \tag{10.1.1}$$

$$H_0 : \mu_{40} = \mu_{50} = \mu_{60} = \mu_{70} = \mu_{80} \tag{10.1.2}$$

and

$$H_0 : \text{no interaction} \tag{10.1.3}$$

Although the above hypotheses achieve the goal of comparing response means for the process carried out at different temperatures, no conclusion can be made about the trend of the response as the temperature is increased.

In general, the trend effects of a continuous predictor are modeled using a polynomial where its non-constant terms represent the different trends such as linear, quadratic, and cubic effects. These non-constant terms in the polynomial are called trend terms. The statistical significance of these trend terms can also be tested in an ANCOVA setting by adding columns representing the trend terms and their interaction effects with the categorical factor into the design matrix (X) of the General Linear Model (see Chapter 4 for the definition of a design matrix).

Note that the design matrix representing only the categorical factor contains the column of ones representing the reference factor level and other dummy variable columns representing the remaining factor levels.

Inclusion of the trend term columns will facilitate significance testing for the overall trend effects and the columns representing the interactions can be utilized to compare differences of each trend effect among the categorical factor levels.

Getting back to the chemical process example, if the quantitative property of measured temperature is used, we can carry out an ANCOVA by fitting a polynomial regression model to express the impact of temperature on the response. If a quadratic polynomial is desired, the appropriate ANCOVA design matrix can be obtained by adding two columns representing $temp$ and $temp^2$ along with the column of ones representing the reagent type A, the reference reagent category, and one dummy variable column representing the reagent type B.

The $temp$ and $temp^2$ terms allow us to investigate the linear and quadratic trends respectively. Furthermore, the inclusion of columns representing the interactions between the reagent type and the two trend terms will facilitate the testing of differences between these two trends between the two reagent types. Note also that additional columns can be added appropriately to fit a polynomial of an even higher order.

Rule

To fit a polynomial of degree n, the response should be measured at least (n+1) distinct levels of the covariate. Preliminary graphics such as scatterplots are useful in deciding the degree of the polynomial to be fitted.

Suggestion

To reduce structural multicollinearity, centering the covariate by subtracting the mean is recommended. For more details see STAT 501 - Chapter 12: Multicollinearity

The necessary software code and/or commands along with outputs and conclusions are given below.

In SAS, this process would look like this:

```
/*centering the covariate creating x^2 */
data centered_quant_factor;
set quant_factor;
x = temp-60;
x2 = x**2;
run;
proc mixed data=centered_quant_factor method=type3;
class reagent;
model product=reagent x x2 reagent*x reagent*x2;
title 'Centered';
run;
```

Notice that we specify *reagent* as a class variable, but x and x^2 enter the model as continuous variables. The regression coefficient of x and x^2 can be used to test the significance of the linear and quadratic trends for reagent type A, the reference category and the interaction term coefficients can be used if these trends differ by categorical factor level. For example, testing the null hypothesis $H_0: \beta_{Reagent*x} = 0$ where $\beta_{Reagent*x}$ is the regression coefficient of the $Reagent * x$ term is equivalent to testing that the linear effects are the same for reagent type A and B.

SAS output:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
reagent	1	3.066357	3.066357	Var(Residual) + Q(reagent)	MS(Residual)	24	2.97	F" class=" ">0.0977
x	1	97.600495	97.600495	Var(Residual) + Q(x,x*reagent)	MS(Residual)	24	94.52	F" class=" "><.0001
x2	1	88.832986	88.832986	Var(Residual) + Q(x2,x2*reagent)	MS(Residual)	24	86.03	F" class=" "><.0001
x*reagent	1	0.341215	0.341215	Var(Residual) + Q(x*reagent)	MS(Residual)	24	0.33	F" class=" ">0.5707
x2*reagent	1	0.067586	0.067586	Var(Residual) + Q(x2*reagent)	MS(Residual)	24	0.07	F" class=" ">0.8003
Residual	24	24.782417	1.032601	Var(Residual)	.	.	.	F" class=" ">.

1. The reagent effect was not significant with $p = 0.0977$
2. Only the linear and quadratic effects were significant in describing the trend in the response, and linear and quadratic effects were the same for each of the reagent types (no interactions)

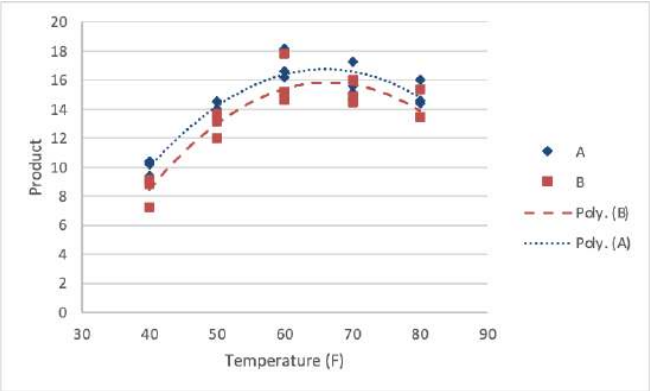


Figure 10.1.1: Graphing product vs temperature

Using R

Steps:

- Load the Quant Factor Data.
- Obtain the ANOVA table after centering the covariate and creating x^2.
- Plot the data.

Steps in R

1. Load the Quant Factor data, obtain the ANOVA table (after centering the covariate), and create x^2 by using the following commands:

```
setwd("~/path-to-folder/")
QuantFactor_data <- read.table("QuantFactorData.txt",header=T)
attach(QuantFactor_data)
temp_center<-temp-60
temp_square_center<-temp_center^2
new_data<-cbind(QuantFactor_data,temp_center,temp_square_center)
ancova_model<-lm(product ~ reagent + temp_center + temp_square_center + reagent:
anova(ancova_model)
#Analysis of Variance Table
#Response: product
#                          Df  Sum Sq  Mean Sq  F value     Pr(>F)
#reagent                    1   9.239    9.239   8.9476   0.006336 **
#temp_center                1  97.600   97.600  94.5191  8.499e-10 ***
#temp_square_center         1  88.833   88.833  86.0284  2.093e-09 ***
#reagent:temp_center        1   0.341    0.341   0.3304  0.570749
#reagent:temp_square_center 1   0.068    0.068   0.0655  0.800257
#Residuals                 24  24.782    1.033
#---
#Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Only the linear and quadratic effects were significant in describing the trend in the response, and linear and quadratic effects were the same for each of the reagent types (no interactions).

2. Plot the polynomial regression curve for reagent A and reagent B by using the following commands:

```
reagentA_regression <- lm(product ~ temp_center + temp_square_center,data=subset
reagentB_regression <- lm(product ~ temp_center + temp_square_center,data=subset
plot(temp,product,ylim=c(0,20),xlab="Temperature", ylab="Product",pch=23, col=if
lines(fitted(reagentA_regression) ~ temp, data=subset(new_data,reagent=="A"), co
lines(fitted(reagentB_regression) ~ temp, data=subset(new_data,reagent=="B"), co
text(locator(1),"reagent A",col="blue")
text(locator(1),"reagent B",col="red")
detach(QuantFactor_data)
```

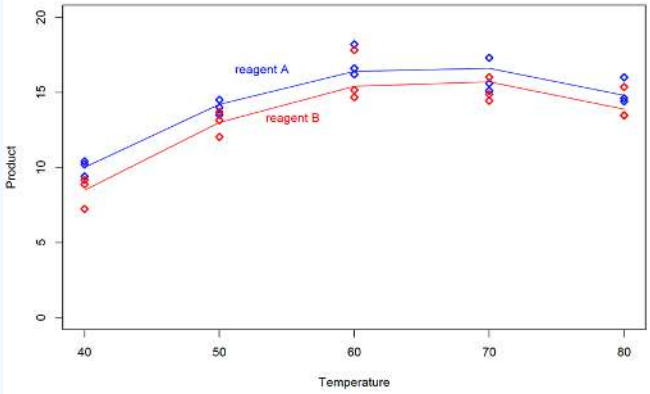


Figure 10.1.2: Graphing product vs temperature using R

This page titled 10.1: ANCOVA with Quantitative Factor Levels is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

10.2: Quantitative Predictors - Orthogonal Polynomials

Polynomial trends in the response with respect to a quantitative predictor can be evaluated by using orthogonal polynomial contrasts, a special set of linear contrasts. This is an alternative to the Regression analysis illustrated in the previous section, which may be affected by multicollinearity. Note that centering to remedy multicollinearity is effective only for quadratic polynomials. Therefore, this simple technique of trend analysis performed via orthogonal polynomial coding will prove to be beneficial for higher-order polynomials. Orthogonal polynomials have the property that the cross-products defined by the numerical coefficients of their terms add to zero.

The orthogonal polynomial coding can be applied only when the levels of quantitative predictor are equally spaced. The method is to partition the quantitative factor in the ANOVA table into independent single degrees of freedom comparisons. The comparisons are called orthogonal polynomial contrasts or comparisons.

Orthogonal polynomials are equations such that each is associated with a power of the independent variable (e.g. x, linear; x^2, quadratic; x^3, cubic, etc.). In other words, orthogonal polynomials are coded forms of simple polynomials. The number of possible comparisons is equal to $k-1$, where k is the number of quantitative factor levels. For example, if $k=3$, only two comparisons are possible allowing for testing of linear and quadratic effects.

Using orthogonal polynomials to fit the desired model to the data would allow us to eliminate collinearity and to seek the same information as simply polynomials.

A typical polynomial model of order k would be:

$$y=\beta_0+\beta_1 x+\beta_2 x^2+\cdots+\beta_k x^k+\epsilon \tag{10.2.1}$$

The simple polynomials used are $x, x^2, \ldots, x^k$. We can obtain orthogonal polynomials as linear combinations of these simple polynomials. If the levels of the predictor variable, x, are equally spaced, then one can easily use coefficient tables to determine the orthogonal polynomial coefficients that can be used to set up an orthogonal polynomial model.

If we are to fit the k^{th} order polynomial to using orthogonal contrasts coefficients, the general equation can be written as

$$y_{ij}=\alpha_0+\alpha_1 g_{1i}(x)+\alpha_2 g_{2i}(x)+\cdots+\alpha_k g_{ki}(x)+\epsilon_{ij} \tag{10.2.2}$$

where $g_{pi}(x)$ is a polynomial in x of degree $p, (p=1,2,\ldots,k)$ for the i^{th} level treatment factor and the parameter α_p depends on the coefficients β_p. Using the properties of the function $g_{pi}(x)$, one can show that the first five orthogonal polynomial are of the following form:

$$\text{Mean:} \quad g_0(x)=1 \tag{10.2.3}$$

$$\text{Linear:} \quad g_1(x)=\lambda_1\left(\frac{x-\bar{x}}{d}\right) \tag{10.2.4}$$

$$\text{Quadratic:} \quad g_2(x)=\lambda_2\left(\left(\frac{x-\bar{x}}{d}\right)^2-\left(\frac{t^2-1}{12}\right)\right) \tag{10.2.5}$$

$$\text{Cubic:} \quad g_3(x)=\lambda_3\left(\left(\frac{x-\bar{x}}{d}\right)^3-\left(\frac{x-\bar{x}}{d}\right)\left(\frac{3t^2-7}{20}\right)\right) \tag{10.2.6}$$

$$\text{Quartic:} \quad g_4(x)=\lambda_4\left(\left(\frac{x-\bar{x}}{d}\right)^4-\left(\frac{x-\bar{x}}{d}\right)^2\left(\frac{3t^2-13}{14}\right)+\frac{3\left(t^2-1\right)\left(t^2-9\right)}{560}\right) \tag{10.2.7}$$

where t = number of levels of the factor, x = value of the factor level, $\bar{x}$ = mean of the factor levels, and d = distance between factor levels.

In the next section, we will illustrate how the orthogonal polynomial contrast coefficients are generated, and the Factor SS is partitioned. This method will be required to fit polynomial regression models with terms greater than the quadratic, because even after centering there will still be multicollinearity between x and x^3 as well as between x^2 and x^4.

The following example is taken from *Design of Experiments: Statistical Principles of Research Design and Analysis* by Robert Kuehl.

Example 10.2.1: Grain Yield

The treatment design consisted of five plant densities (10, 20, 30, 40, and 50). Each of the five treatments was assigned randomly to three field plots in a completely randomized experimental design. The resulting grain yields are shown in the table below (Grain Data):

	Plant Density (x)				
	10	20	30	40	50
	12.2	16.0	18.6	17.6	18.0
	11.4	15.5	20.2	19.3	16.4
	12.4	16.5	18.2	17.1	16.6
Means $(\bar{y}_i)$	12.0	16.0	19.0	18.0	17.0

Solution

We can see that the factor levels of plant density are equally spaced. Therefore, we can use the orthogonal contrast coefficients to fit a polynomial to the response, grain yields. With $k = 5$, we can only fit up to a quartic term. The orthogonal polynomial contrast coefficients for the example are shown in Table 10.1.

Table 10.1 - Computations for orthogonal polynomial contrasts and sums of squares

Density (x)	$\bar{y}_i$	Orthogonal Polynomial Coefficients (g_{pi})				
		Mean	Linear	Quadratic	Cubic	Quartic
10	12	1	-2	2	-1	1
20	16	1	-1	-1	2	-4
30	19	1	0	-2	0	6
40	18	1	1	-1	-2	-4
50	17	1	2	2	1	1
λ_p		-	1	1	5/6	35/12
Sum = $\sum g_{pi}\bar{y}_i$		82	12	-14	1	7
Divisor = $\sum g_{pi}^2$		5	10	14	10	70
$SSP_p = r(\sum g_{pi}\bar{y}_i)^2 / \sum g_{pi}^2$		-	43.2	42.0	0.3	2.1
$\hat{\alpha}_p = \sum g_{pi}\bar{y}_i / \sum g_{pi}^2$		16.4	1.2	-1.0	0.1	0.1

As mentioned before, one can easily find the orthogonal polynomial coefficients for a different order of polynomials using pre-documented tables for equally spaced intervals. However, let us try to understand how the coefficients are obtained.

First note that the five values of x are $10, 20, 30, 40, 50$ Therefore, $\bar{x} = 30$ and the spacing $d = 10$. This means that the five values of $\frac{x-\bar{x}}{d}$ are $-2, -1, 0, 1,$ and 2.

Linear coefficients: The polynomial g_1 for linear coefficients turn out to be:

Linear Coefficient Polynomials g_1					
x	10	20	30	40	50
$(x-30)$	-20	-10	0	10	20
$\frac{(x-30)}{10}$	-2	-1	0	1	2
Linear orthogonal polynomial	$(-2)\lambda_1$	$(-1)\lambda_1$	$(0)\lambda_1$	$(1)\lambda_1$	$(2)\lambda_1$

To obtain the final set of coefficients we choose λ_1 so that the coefficients are integers. Therefore, we set $\lambda_1 = 1$ and obtain the coefficient values in Table 10.1.

Quadratic coefficients: The polynomial g_2 for linear coefficients:

Linear Coefficient Polynomials g_2					
Linear orthogonal polynomial	$\left((-2)^2 - \left(\frac{5^2-1}{12}\right)\right)\lambda_2$	$\left((-1)^2 - \left(\frac{5^2-1}{12}\right)\right)\lambda_2$	$\left((0)^2 - \left(\frac{5^2-1}{12}\right)\right)\lambda_2$	$\left((1)^2 - \left(\frac{5^2-1}{12}\right)\right)\lambda_2$	$\left((2)^2 - \left(\frac{5^2-1}{12}\right)\right)\lambda_2$
Simplified form	$(2)\lambda_2$	$(-1)\lambda_2$	$(-2)\lambda_2$	$(-1)\lambda_2$	$(2)\lambda_2$

To obtain the final set of coefficients we choose λ_2 so that the coefficients are integers. Therefore, we set $\lambda_2 = 1$ and obtain the coefficient values in Table 10.1.

Cubic coefficients: The polynomial g_3 for linear coefficients:

Linear Coefficient Polynomials g_3					
Linear orthogonal polynomial	$\left((-2)^3 - (-2)\left(\frac{3(5^2)-7}{20}\right)\right)\lambda_3$	$\left((-1)^3 - (-1)\left(\frac{3(5^2)-7}{20}\right)\right)\lambda_3$	$\left((0)^3 - (0)\left(\frac{3(5^2)-7}{20}\right)\right)\lambda_3$	$\left((1)^3 - (1)\left(\frac{3(5^2)-7}{20}\right)\right)\lambda_3$	$\left((2)^3 - (2)\left(\frac{3(5^2)-7}{20}\right)\right)\lambda_3$
Simplified form	$\left(-\frac{6}{5}\right)\lambda_3$	$\left(\frac{12}{5}\right)\lambda_3$	$(0)\lambda_3$	$\left(-\frac{12}{5}\right)\lambda_3$	$\left(\frac{6}{5}\right)\lambda_3$

Quartic coefficients: The polynomial g4 can be used to obtain the quartic coefficients in the same way as above.

Notice that each set of coefficients for contrast among the treatments since the sum of coefficients is equal to zero. For example, the quartic coefficients $(1, -4, 6, -4, 1)$ sums to zero. Using orthogonal polynomial contrasts, we can partition the treatment sums of squares into a set of additive sums of squares corresponding to orthogonal polynomial contrasts. Computations are similar to what we learned in lesson 2.5. We can use those partitions to test sequentially the significance of linear, quadratic, cubic, and quartic terms in the model to find the polynomial order appropriate for the data.

Table 10.1 shows how to obtain the sums of squares for each component and how to compute the estimates of the α_p coefficients for the orthogonal polynomial equation. Using the results in table 10.1, we have estimated orthogonal polynomial equation as:

$$\hat{y}_i = 16.4 + 1.2g_{1i} - 1.0g_{2i} + 0.1g_{3i} + 0.1g_{4i}$$

Table 10.2 summarizes how the treatment sums of squares are partitioned and their test results.

Table 10.2 - Analysis of variance for the orthogonal polynomial model relationship between plant density and grain yield.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	Pr > F
Density	4	87.60	21.90	29.28	F">.000
Error	10	7.48	0.75		F">

Contrast	DF	Contrast SS	Mean Square	F	Pr > F
Linear	1	43.20	43.20	57.75	F">.000
Quadratic	1	42.00	42.00	56.15	F">.000
Cubic	1	.30	.30	.40	F">.541
Quartic	1	2.10	2.10	2.81	F">.125

To test whether any of the polynomials are significant (i.e. $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$), we can use the global F-test where the test statistic is equal to 29.28. We see that the p-value is almost zero and therefore we can conclude that at the 5% level at least one of the polynomials is significant. Using the orthogonal polynomial contrasts we can determine which of the polynomials are useful. From table 3.5, we see that for this example only the linear and quadratic terms are useful. Therefore we can write the estimated orthogonal polynomial equation as:

$$16.4 + 1.2g_{1i} - 1.0g_{2i}$$

The polynomial relationship expressed as a function of y and x in actual units of the observed variables is more informative than when expressed in units of the orthogonal polynomial.

We can obtain the polynomial relationship using the actual units of observed variables by back-transforming using the relationships presented earlier. The necessary quantities to back-transform are $\lambda_1 = 1$, $d = 10$, $\bar{x} = 30$, and $t = 5$. Substituting these values, we obtain

$$\begin{aligned}\hat{y} &= 16.4 + 1.2g_{1i} - 1.0g_{2i} \\ &= 16.4 + 1.2(1)\left(\frac{x-30}{10}\right) - 1.0(1)\left(\left(\frac{x-30}{10}\right)^2 - \frac{5^2-1}{12}\right)\end{aligned}$$

which simplifies to

$$\hat{y} = 5.8 + 0.72x - 0.01x^2$$

Generating Orthogonal Polynomials

? Using SAS

Steps in SAS

Below is the code for generating polynomials from the IML procedure in SAS:

```
/* read the grain data set */
/* Generating Ortho_Polynomials from IML */
proc iml;
x={10 20 30 40 50};
xpoly=orpol(x,4); /* the '4' is the df for the quantitative factor */
density=x`; new=density || xpoly;
create out1 from new[colname={"density" "xp0" "xp1" "xp2" "xp3" "xp4"}];
append from new; close out1;
quit;
proc print data=out1;
run;
/* Here data is sorted and then merged with the original dataset */
proc sort data=grain;
by density;
run;
```

```
data ortho_poly; merge out1 grain;
by density;
run;
proc print data=ortho_poly;
run;
/* The following code will then generate the results shown in the
Online Lesson Notes for the Kuehl example data */
proc mixed data=ortho_poly method=type3;
class;
model yield=xp1 xp2 xp3 xp4;
title 'Using Orthog polynomials from IML';
run;
/* We can use proc glm to obtain the same results without using
IML codings, to directly obtained the same results.
Proc glm will use the orthogonal contrast coefficients directly */
proc glm data=grain;
class density;
model yield = density;
contrast 'linear' density -2 -1 0 1 2;
contrast 'quadratic' density 2 -1 -2 -1 2;
contrast 'cubic' density -1 2 0 -2 1;
contrast 'quartic' density 1 -4 6 -4 1;
run;
```

The output is:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
xp1	1	43.200000	43.200000	Var(Residual) + Q(xp1)	MS(Residual)	10	57.75	F"> < .0001
xp2	1	42.000000	42.000000	Var(Residual) + Q(xp2)	MS(Residual)	10	56.15	F"> < .0001
xp3	1	0.300000	0.300000	Var(Residual) + Q(xp2)	MS(Residual)	10	0.40	F"> 0.5407
xp4	1	2.100000	2.100000	Var(Residual) + Q(xp4)	MS(Residual)	10	2.81	F"> 0.1248
Residual	10	7.480000	7.480000	Var(Residual)				F" class=" ">

Vari…
Vari…
Vari…
Vari…

Fitting a Quadratic Model with Proc Mixed

Often we can see that only a quadratic curvature is of interest in a set of data. In this case, we can plan to simply run an order 2 (quadratic) polynomial and can easily use proc mixed (the general linear model). This method just requires centering the quantitative variable levels by subtracting the mean of the levels (30) and then creating the quadratic polynomial terms.

```
data grain;
set grain;
x=density-30;
x2=x**2;
run;
proc mixed data=grain method=type3;
class;
model yield = x x2;
run;
```

The output is:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
x	1	43.200000	43.200000	Var(Residual) + Q(x)	MS(residual)	12	52.47	F"> <.0001
x2	1	42.000000	42.000000	Var(Residual) + Q(x2)	MS(residual)	12	51.01	F"> <.0001
Residual	12	9.880000	0.823333	Var(Residual)				F" class=" ">

We can also generate the solutions (coefficients) for the model with:

```
proc mixed data=grain method=type3;
class;
model yield = x x2 / solution;
run;
```

which gives the following output for the regression coefficients:

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	18.4000	0.3651	12	50.40	|t"> <.0001
x	0.1200	0.01657	12	7.24	|t"> <.0001
x2	-0.01000	0.001400	12	-7.14	|t"> <.0001

Here we need to keep in mind that the regression was based on centered values for the predictor, so we have to back-transform to get the coefficients in terms of the original variables. This back-transform process (from Kutner et.al) is:

Regression Function in Terms of X

After a polynomial regression model has been developed, we often wish to express the final model in terms of the original variables rather than keeping it in terms of the centered variables. This can be done readily. For example, the fitted second-order model for one predictor variable that is is expressed in terms of centered values $x = X - \bar{X}$:

$$\hat{Y} = b_0 + b_1(x) + b_{11}x^2$$

because in terms of the original X variable:

$$\hat{Y} = b'_0 + b'_1 X + b'_{11} X^2$$

where:

$$\begin{aligned} b'_0 &= b_0 - b_1 \bar{X} + b_{11} \bar{X}^2 \\ b'_1 &= b_1 - 2 b_{11} \bar{X} \\ b'_{11} &= b_{11} \end{aligned}$$

In the example above, this back-transformation uses the estimates from the Solutions for Fixed Effects table above.

```
data backtransform;
bprime0=18.4-(.12*30)+(-.01*(30**2));
bprime1=.12-(2*-.01*30);
bprime2=-.01;
title 'bprime0=b0-(b1*meanX)+(b2*(meanX)2)';
title2 'bprime1=b1=2*b2*meanX';
title3 'bprime2=b2';
run;
proc print data=backtransform;
var bprime0 bprime1 bprime2;
run;
```

The output is then:

Obs	bprime0	bprime1	bprime2
1	5.8	0.72	-0.01

Note

The ANOVA results and the final quadratic regression equation here are identical to the results from the orthogonal polynomial coding approach.

Using R

- Load the Grain Data.
- Obtain the ANOVA table.
- Fit a quadratic model after centering the covariate and creating x^2. Transform back to the original variables.

Steps in R

1. Load the Grain data and obtain the ANOVA table by using the following commands:

```
setwd("~/path-to-folder/")
grain_data <- read.table("grain_data.txt",header=T)
attach(grain_data)
poly_model<-lm(yield ~ poly(density,4),data=grain_data)
summary(poly_model)
#Coefficients:
#                  Estimate  Std. Error  t value  Pr(>|t|)
#(Intercept)         16.4000      0.2233   73.441  5.35e-15 ***
```

```
#poly(density, 4)1    6.5727      0.8649    7.600  1.84e-05 ***
#poly(density, 4)2   -6.4807      0.8649   -7.493  2.08e-05 ***
#poly(density, 4)3    0.5477      0.8649    0.633     0.541
#poly(density, 4)4    1.4491      0.8649    1.676     0.125
anova(poly_model)
#Analysis of Variance Table
#Response: yield
#                 Df  Sum Sq  Mean Sq  F value    Pr(>F)
#poly(density, 4)   4   87.60   21.900   29.278  1.69e-05 ***
#Residuals         10    7.48    0.748
#---
#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By using the command `anova()` we can test whether any of the polynomials are significant (i.e. $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$. We can use the global F-test where the test statistic is equal to 29.28. We see that the p-value is almost zero, and therefore we can conclude that at the 5% level at least one of the polynomials is significant.

By using the command `summary()` we can test which contrasts are significant. For this example only the linear and quadratic terms are significant since there p-values are almost zero.

2. Fit a quadratic model after centering the covariate and creating x^2 by using the following commands:

Transform back to the original variables

```
density_center<-density-30
density_square_center<-density_center^2
new_data<-cbind(grain_data,density_center,density_square_center)
ancova_model<-lm(yield ~ density_center + density_square_center,new_data)
summary(ancova_model)
#Coefficients:
#                     Estimate  Std. Error  t value  Pr(>|t|)
#(Intercept)          18.40000     0.36511   50.396  2.44e-15 ***
#density_center        0.12000     0.01657    7.244  1.02e-05 ***
#density_square_center -0.01000     0.00140   -7.142  1.18e-05 ***
#---
#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(ancova_model)
#Analysis of Variance Table
#Response: yield
#                     Df  Sum Sq  Mean Sq  F value     Pr(>F)
#density_center        1   43.20   43.200   52.470  1.024e-05 ***
#density_square_center  1   42.00   42.000   51.012  1.177e-05 ***
#Residuals             12    9.88    0.823
#---
#Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Transform back to the original variables

The estimated coefficients for the polynomial model are 18.4, 0.12 and -0.01. Here we need to keep in mind that the regression was based on centered values for the predictor, so we have to back-transform to get the coefficients in terms of the original variables. We can do that by using the following commands:

```
b_0_prime<-18.4-0.12*30-0.01*30^2 #5.8
b_1_prime<-0.12-0.01*(-2*30) # 0.72
b_2_prime<--0.01 # -0.01
detach(grain_data)
```

For the original variables the estimated coefficients are 5.8, 0.72 and -0.01.

This page titled 10.2: Quantitative Predictors - Orthogonal Polynomials is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

10.3: Chapter 10 Summary

We've seen some of the versatility of ANCOVA in Chapter 9 and Chapter 10. In application, it's often used in ANOVA settings to adjust or "control for" a covariate that may be masking real treatment differences. In regression settings, researchers may be focused on a family of regression relationships, and are interested in testing for significant differences among regression coefficients across different groups.

These are like two sides of the same coin: in terms of model development, ANOVA and Regression approaches converge in the general linear model as ANCOVA. Mastery of ANCOVA methodology is arguably one of the most important tools to have in an applied statistician's toolbox.

This page titled 10.3: Chapter 10 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

11: Introduction to Repeated Measures

Objectives

Upon completion of this lesson, you should be able to:

- Recognize repeated measures designs in time.
- Understand the different covariance structures that can be imposed on model error.
- Use software such as SAS, Minitab, and R for fitting repeated measures ANOVA.

The focus of many studies can be expanded by introducing time also as a potential covariate. In the greenhouse example, the growth of plants can be measured weekly over a period of time, allowing time also to be included as a predictor in the statistical model. Another example is to compare the effect of two anti-cancer drugs on disease status at different intervals of time. In both these examples, the response has to be measured multiple times from the same experimental unit, hence the term "repeated measures." The repeated measurements made on the same experimental unit cannot be assumed independent which means that the model errors may not be uncorrelated anymore and the statistical model should be modified accordingly.

Two fundamental types of repeated measures are common. Repeated measures in time are the type in which experimental units receive treatment, and they are simply followed with repeated measures on the response variable over several times. In contrast, experiments can involve administering all treatment levels (in a sequence) to each experimental unit. This type of repeated measures study is called a crossover design, the topic of our next lesson.

Repeated measures are frequently encountered in clinical trials including longitudinal studies, growth models, and situations in which experimental units are difficult to acquire.

This page titled 11: Introduction to Repeated Measures is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Historical Methods

Repeated measures in time were historically handled in either a multivariate analysis setting or as a univariate split-plot in time. The focus in this course is limited only to the latter.

A split-plot in time approach looks at each subject (experimental unit) as the main plot (receiving treatment) and then is split into sub-plots (time periods). Historically, the default assumption in split-plot in time data analysis has been that the correlations among responses at different time points are the same for all treatment levels and time points (compound symmetry). However, depending on the study and nature of data, other correlation structures can be more appropriate (e.g. autoregressive lag 1).

Most of the current software facilitates the inclusion of different correlation structures which has helped in the evolution of methodology for repeated measures to accommodate the presence of different correlated structures in residuals.

This page titled 11.1: Historical Methods is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Correlated Residuals

Note

The first part of the section uses a hypothetical data set to illustrate the origin of the covariance structure, by capturing the residuals for each time point and looking at the simple correlations for pairs of time points. *Therefore, the software code used for this purpose is NOT what we would ordinarily use in conducting a repeated measures analysis* as generating the residuals of a fitted model and their variances and covariances is automatically done by software. The variances and the covariances of the residuals will be outputted as the diagonals and the off-diagonals of the variance-covariance (R block) matrix in SAS or R. Minitab currently does not accommodate various covariance structures, opting instead to treat repeated measures as "split-plot in time" (which assumes compound symmetry).

If we look at the ANOVA mixed model in general terms, we have:

$$\text{Model: response} = \text{fixed effects} + \text{random effects} + \text{errors} \tag{11.2.1}$$

In the case of repeated measures with measures taken at p number of time points, the covariance structure of the errors can be expressed as a matrix. The diagonals of this matrix are the error variances at each time point. The off-diagonals are the covariances between successive time points. In general, the variance-covariance matrix can be expressed as follows:

$$\sum_i = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \tag{11.2.2}$$

The structure shown above does not assume any specific properties of the variances and covariances and is called an **unstructured covariance structure**. Note that there are p variances and $\frac{p(p-1)}{2}$ covariances that adds to $\frac{p(p+1)}{2}$ unknown quantities which define this matrix. So, even for a small number of time points, a substantial number of parameters will have to be estimated. Therefore, in practice, specific structures are imposed to reduce the number of distinct parameters that need to be estimated, which will be discussed in Section 11.3.

To understand the correlation structure of errors, let us use SAS to generate the variance-covariance matrix of the errors for a repeated measures model using hypothetical data stored in **Repeated Measures Example Data**. The data consists of a single treatment with 3 levels. Subjects are assigned a treatment level at random (CRD) and then are measured at $p = 3$ time points. The SAS code which is given below fits a factorial model and generates the errors along with the correlations among responses taken at three time points.

```
data rmanova;
  input trt $ time subject resp;
  datalines;
   A 1 1 10
   A 1 2 12
   A 1 3 13
   A 2 1 16
   A 2 2 19
   A 2 3 20
   A 3 1 25
   A 3 2 27
   A 3 3 28
   B 1 4 12
   B 1 5 11
   B 1 6 10
```

```
   B 2 4 18
   B 2 5 20
   B 2 6 22
   B 3 4 25
   B 3 5 26
   B 3 6 27
   C 1 7 10
   C 1 8 12
   C 1 9 13
   C 2 7 22
   C 2 8 23
   C 2 9 22
   C 3 7 31
   C 3 8 34
   C 3 9 33
;
```

We can run a simple model and obtain the residuals.

```
/* 2-factor factorial for trt and time - saving residuals */
proc mixed data=rmanova method=type3;
  class trt time subject;
  model resp=trt time trt*time / ddfm=kr outpm=outmixed;
  title 'Two_factor_factorial';
run; title;
```

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	2	18	14.52	F">0.0002
time	2	18	292.72	F"><.0001
trt*time	4	18	4.67	F">0.0092

```
/* re-organize the residuals to (unstacked data for correlation) */
data one;
  set outmixed;
  where time=1; time1=resid;
  keep time1;
run;
data two; set outmixed; where time=2; time2=resid; keep time2; run;
data three; set outmixed; where time=3; time3=resid; keep time3; run;

data corrcheck; merge one two three;

proc print data=corrcheck;
run;
```

```
proc corr data=corrcheck nosimple; var time1 time2 time3; run;
```

The residuals then are:

The Print Procedure			
Obs	**time1**	**time2**	**time3**
1	-1.66667	-2.33333	-1.66667
2	0.33333	0.66667	0.33333
3	1.33333	1.66667	1.33333
4	1.00000	-2.00000	-1.00000
5	0.00000	0.00000	0.00000
6	-1.00000	2.00000	1.00000
7	-1.66667	-0.33333	-1.66667
8	0.33333	0.66667	1.33333
9	1.33333	-0.33333	0.33333

The correlations of responses between time points are:

The CORR Procedure

3 Variables:	time1 time2 time3

Pearson Correlation Coefficients, N = 9 Prob > |r| under H0: Rho=0			
	time1	time2	time3
time1	1.00000	0.19026 0.6239	0.55882 0.1178
time2	0.19026 0.6239	1.00000	0.83239 0.0054
time3	0.55882 0.1178	0.83239 0.0054	1.00000

Notice that in the above code, the repeated nature of the data is not being utilized. The "*repeated*" statement in `proc mixed` , which is used in practice, accounts for this. As in the code given below, in the repeated statement, the option of `subject=` specifies what experimental (or observational) units the repeated measures are made on. The `type=` can be used to specify one of many types of structures for these correlations. Here we specified the unstructured covariance structure and obtained the same correlations that were generated earlier with simple statistics.

```
proc mixed data=rmanova ;
    class trt time subject;
    model resp=trt time trt*time / ddfm=kr solution ;
    repeated /subject=subject(trt) type=UN rcorr;
    title 'Repeated Measures';
run; title;
```

Finding the best covariance structure is much of the work in modeling repeated measures and is usually done by considering a subset of candidate structures. These include UN (Unstructured), CS (Compound Symmetry), AR(1) (Autoregressive lag 1) – if time intervals are evenly spaced, or SP(POW) (Spatial Power) – if time intervals are unequally spaced.

Choosing the best covariance structure is based on Fit Statistics (also known as information criteria). PROC MIXED in SAS automatically generates four of such Fit Statistics measures and for this example, they are:

Fit Statistics	
-2 Res Log Likelihood	63.0
AIC (Smaller is Better)	75.0
AICC (Smaller is Better)	82.6
BIC (Smaller is Better)	76.2

Smaller or more negative values indicate a better fit to the data. The process amounts to trying various candidate structures and then selecting the covariance structure producing the smallest or most negative values. The information criteria listed above are usually similar in value, but for small sample sizes, the AICC criterion is recommended. The topic of covariance structures for a general setting is discussed in the next section.

Using R

- Load the Repeated Measures Example Data.
- Obtain the ANOVA table.
- Obtain the correlations of responses between time points.
- Obtain the results for the split-plot in time approach.
- Run the analysis as a repeated-measures ANOVA by using different covariance structures.

Steps in R

1. Load the Repeated Measures Example data and obtain the ANOVA by using the following commands:

```
setwd("~/path-to-folder/")
repeated_measures_example_data <- read.table("repeated_measures_example_data.txt
attach(repeated_measures_example_data)
rmanova<-aov(resp ~ trt + factor(time) + trt*factor(time), repeated_measures_exa
anova(rmanova)
#Analysis of Variance Table
#Response: resp
#                 Df   Sum Sq  Mean Sq   F value     Pr(>F)
#trt               2    64.52    32.26   14.5167  0.0001761 ***
#factor(time)      2  1300.96   650.48  292.7167   1.87e-14 ***
#trt:factor(time)  4    41.48    10.37    4.6667  0.0092424 **
#Residuals        18    40.00     2.22
#---
#Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

2. Obtain the correlations of responses between time points by using the following commands:

```
time_1<-c(rmanova$residuals[1:3],rmanova$residuals[10:12],rmanova$residuals[19:2
time_2<-c(rmanova$residuals[4:6],rmanova$residuals[13:15],rmanova$residuals[22:2
time_3<-c(rmanova$residuals[7:9],rmanova$residuals[16:18],rmanova$residuals[22:2
residuals<-cbind(time_1,time_2,time_3)
```

```
rownames(residuals)<-NULL
#residuals
#             time_1        time_2         time_3
#[1,] -1.666667e+00  -2.333333e+00  -1.666667e+00
#[2,]  3.333333e-01   6.666667e-01   3.333333e-01
#[3,]  1.333333e+00   1.666667e+00   1.333333e+00
#[4,]  1.000000e+00  -2.000000e+00  -1.000000e+00
#[5,] -3.885781e-16   9.436896e-16  -7.216450e-16
#[6,] -1.000000e+00   2.000000e+00   1.000000e+00
#[7,] -1.666667e+00  -3.333333e-01  -3.333333e-01
#[8,]  3.333333e-01   6.666667e-01   6.666667e-01
#[9,]  1.333333e+00  -3.333333e-01  -3.333333e-01
#cor(residuals)
#
#           time_1     time_2     time_3
#time_1  1.0000000  0.1902606  0.3290726
#time_2  0.1902606  1.0000000  0.9756655
#time_3  0.3290726  0.9756655  1.0000000
```

This page titled 11.2: Correlated Residuals is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

11.3: More on Covariance Structures

Variance Components (VC)

$$\begin{bmatrix} \sigma_1^2 & \cdots & \cdots & \vdots \\ \vdots & \sigma_1^2 & \cdots & \vdots \\ \vdots & \cdots & \sigma_1^{2\vdots} & \\ \cdots & \cdots & \cdots & \sigma_1^2 \end{bmatrix} \tag{11.3.1}$$

The variance component structure (VC) is the simplest, where the correlations of errors within a subject are presumed to be 0. This structure is the default setting in proc mixed, but is not a reasonable choice for most repeated measures designs. It is included in the exploration process to get a sense of the effect of fitting other structures.

Compound Symmetry (CS)

$$\sigma^2 \begin{bmatrix} 1.0 & \rho & \rho & \rho \\ & 1.0 & \rho & \rho \\ & & 1.0 & \rho \\ & & & 1.0 \end{bmatrix} = \begin{bmatrix} \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\ & \sigma_b^2+\sigma_e^2 & \sigma_b^2 & \sigma_b^2 \\ & & \sigma_b^2+\sigma_e^2 & \sigma_b^2 \\ & & & \sigma_b^2+\sigma_e^2 \end{bmatrix} \tag{11.3.2}$$

The simplest covariance structure that includes within-subject correlated errors is compound symmetry (CS). Here we see correlated errors between time points within subjects, and note that these correlations are presumed to be the same for each set of times, regardless of how distant in time the repeated measures are made.

First Order Autoregressive AR(1)

$$\sigma^2 \begin{bmatrix} 1.0 & \rho & \rho^2 & \rho^3 \\ & 1.0 & \rho & \rho^2 \\ & & 1.0 & \rho \\ & & & 1.0 \end{bmatrix} \tag{11.3.3}$$

The autoregressive (Lag 1) structure considers correlations to be highest between adjacent times, and a systematically decreasing correlation with increasing distance between time points. For one subject, the error correlation between time 1 and time 2 would be $\rho^{t_2-t_1}$. Between time 1 and time 3 the correlation would be less, and equal to $\rho^{t_3-t_1}$. Between time 1 and 4, the correlation is lesser, as $\rho^{t_4-t_1}$, and so on. Note that this structure is only applicable for evenly spaced time intervals for the repeated measure; so that consecutive correlations are ρ raised to powers of 1, 2, 3, etc.

Spatial Power

$$\sigma^2 \begin{bmatrix} 1.0 & \rho^{\left|\frac{t_1-t_2}{t_1-t_2}\right|} & \rho^{\left|\frac{t_1-t_3}{t_1-t_2}\right|} & \rho^{\left|\frac{t_1-t_4}{t_1-t_2}\right|} \\ & 1.0 & \rho^{\left|\frac{t_2-t_3}{t_1-t_2}\right|} & \rho^{\left|\frac{t_2-t_4}{t_1-t_2}\right|} \\ & & 1.0 & \rho^{\left|\frac{t_3-t_4}{t_1-t_2}\right|} \\ & & & 1.0 \end{bmatrix} \tag{11.3.4}$$

When time intervals are not evenly spaced, a covariance structure equivalent to the AR(1) is the spatial power (SP(POW)). The concept is the same as the AR(1) but instead of raising the correlation to powers of 1, 2, 3, …, the correlation coefficient is raised to a power that is the actual difference in times (e.g. t^2-t^1 for the correlation between time 1 and time 2). It is clear that this method requires having quantitative values for the variable time in the data so that it can be specified for the calculation of the exponents in the SP(POW) structure. If an analysis is run wherein the repeated measures are equally spaced in time, the AR(1) and SP(POW) structures yield identical results.

Unstructured Covariance

$$\begin{bmatrix} \sigma_1^1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \end{bmatrix} \tag{11.3.5}$$

The Unstructured covariance structure (UN) is the most complex because it is estimating unique correlations for each pair of time points. As there are too many parameters (all distinct correlations), the estimates most times will not be computable. SAS for instance returns an error message indicating that there are too many parameters to estimate with the data.

Choosing the Best Covariance Structure

The fit statistics used for model selection can also be utilized in choosing the best covariance matrix. The model selections most commonly supported by software are -2 Res Log Likelihood, Akaike's information criterion - corrected (AICC), and Bayesian Information Criteria (BIC). These statistics are functions of the log likelihood and can be compared across different models as well as different covariance structures provided the fixed effects part is the same in each model. The smaller the criterion statistics value is, the better the model is, and if they are close, the simpler model is preferred.

BIC tends to choose simpler models compared to AICC. Choosing a model that is too simple however inflates the Type I error rate. Therefore, if controlling Type I error is of importance, AICC may be the better criterion. On the other hand, if loss of power is of more concern, BIC might be preferable (Guerin and Stroup 2000).

The MIXED procedure in SAS outputs the 3 criterion statistics when using the `type =` option in the Repeated statement.

In addition to using the above fit statistics, graphical approaches are also available, and see **Graphical Approach** for more details. Combining information from both approaches to make the final choice may also prove to be beneficial.

This page titled 11.3: More on Covariance Structures is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

11.4: Worked Example

For the example dataset in **Repeated Measures Example Data**, which we introduced in the 11.2: Correlated Residuals section, we can plot the data:

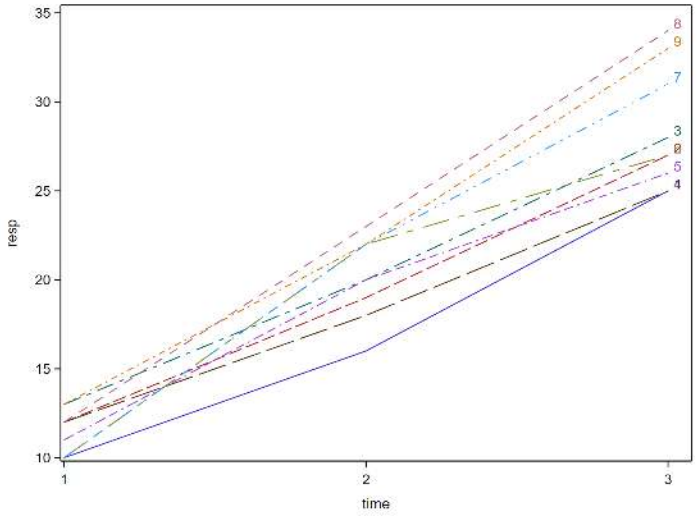


Figure 11.4.1: The values of the response plotted at each of the three time points for each of the 9 subjects.

We can obtain the results for the split-plot in time approach using the following:

```
/* Split-Plot in Time */
proc mixed data=rmanova method=type3;
  class trt time subject;
  model resp=trt time trt*time / ddfm=kr;
  random subject(trt); title 'Split-Plot in Time';
run;
```

Next, we run the analysis as a repeated-measures ANOVA, which allows us to evaluate which covariance structure fits best.

Next, we run the analysis as a repeated-measures ANOVA, which allows us to evaluate which covariance structure fits best.

```
/* Repeated Measures Approach */
/* Fitting Covariance structures: */
/* Note: the code begining with "ods output ..." for each
run of the Mixed procedure generates an output that
is tabulated at the end to enable comparison of
the candidate covariance structure */
proc mixed data=rmanova;
  class trt time subject;
  model resp=trt time trt*time / ddfm=kr;
  repeated time/subject=subject(trt) type=cs rcorr;

ods output FitStatistics=FitCS (rename=(value=CS))
 FitStatistics=FitCSp;
title 'Compound Symmetry'; run;
title ' '; run;
```

```
proc mixed data=rmanova;
  class trt time subject;
  model resp=trt time trt*time / ddfm=kr;
  repeated time/subject=subject(trt) type=ar(1) rcorr;

ods output FitStatistics=FitAR1 (rename=(value=AR1))
 FitStatistics=FitAR1p;
title 'Autoregressive Lag 1'; run;
title ' '; run;

proc mixed data=rmanova;
  class trt time subject;
  model resp=trt time trt*time / ddfm=kr;
  repeated time/subject=subject(trt) type=un rcorr;

ods output FitStatistics=FitUN (rename=(value=UN))
 FitStatistics=FitUNp;
title 'Unstructured'; run;
title ' '; run;

data fits;
  merge FitCS FitAR1 FitUN;
  by descr;
  run;
ods listing; proc print data=fits; run;
```

We get the following Summary Table:

Obs	Descr	CS	AR1	UN
1	-2 Res Log Likelihood	70.9	71.9	63.0
2	AIC (smaller is better)	74.9	75.9	75.0
3	AICC (smaller is better)	75.7	76.7	82.6
4	BIC (smaller is better)	75.3	76.3	76.2

Using the AICC as our criteria, we would choose the compound symmetry (CS) covariance structure.

The output from this would be:

Type 3 Test of Fixed Effect				
Effect	Num DF	Den DF	F Value	Pr > F
trt	2	6	7.14	F">0.0259
time	2	12	605.62	F"><.0001
trt*time	4	12	9.66	F">0.0010

Note that for this case, the *p*-values obtained here are identical to the split-plot in time approach.

Using R

Steps in R

1. Obtain the results for the split-plot in time approach by using the following commands:

```
library(lmerTest)
library(lme4)
model<-lmer(resp ~ trt + factor(time) + trt:factor(time) + (1 | factor(subject)
anova(model)
#Type III Analysis of Variance Table with Satterthwaite s method
#               Sum Sq  Mean Sq  NumDF  DenDF  F value   Pr(>F)
#trt                15        8      2      6     7.14  0.02590 *
#(time)           1301      650      2     12   605.62  8.9e-13 ***
#trt:factor(time)   41       10      4     12     9.66  0.00099 ***
#---
#Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

2. Run the analysis as a repeated-measures ANOVA by using different covariance structures. We can use the following commands:

```
library(nlme)
model_cs<-gls(resp ~ trt + factor(time) + trt*factor(time),repeated_measures_exa
model_AR<-gls(resp ~ trt + factor(time) + trt*factor(time),repeated_measures_exa
model_UN<-gls(resp ~ trt + factor(time) + trt*factor(time),repeated_measures_exa
Model_Selection <- data.frame(
c ("","-2LogLik","AIC", "BIC"),
c("CS", round(-2*summary(model_cs)$logLik,2),round(summary(model_cs)$AIC,2),roun
c("AR1", round(-2*summary(model_AR)$logLik,2),round(summary(model_AR)$AIC,2),rou
c("UN", round(-2*summary(model_UN)$logLik,2),round(summary(model_UN)$AIC,2),roun
stringsAsFactors = FALSE)
names(Model_Selection) <- c(" ", " ","","")
print(Model_Selection)
#1               CS     AR1      UN
#2 -2LogLik   80.54   82.03   69.95
#3      AIC  102.54  104.03   95.95
#4      BIC  116.79  118.29  112.79
detach(repeated_measures_example_data)
```

This page titled 11.4: Worked Example is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

11.5: Chapter 11 Summary

This lesson introduced us to the topic of repeated measures designs. The focus was on repeated measures in time where each experimental unit is assigned to exactly one treatment level the response is observed over several time periods. This means that the responses from the same experimental unit observed over time can be correlated and the model assumption of independent observations is no longer valid. Therefore, an appropriate covariance structure should be imposed to account for the correlated nature of the response, and the best is chosen based on fit statistics. Note that the AR(1) covariance structure is a possible choice only when time intervals are equally spaced. If time intervals are unequal sp(pow) has to be the alternative.

Other scenarios can result in repeated measures, not necessarily in time. The important feature is that multiple measurements are being made on the same experimental unit. A special case of this is the cross-over design wherein the treatments themselves are switched on the same experimental unit during the course of the experiment. This would be the topic of the next lesson.

This page titled 11.5: Chapter 11 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

12: Cross-over Repeated Measure Designs

Objectives

Upon completion of this lesson, you should be able to:

- Recognize a cross-over repeated measures design.
- Understand what a wash-out period is.
- Test for the significance of carry-over effects.
- Adjust treatment means to account for carry-over effects.

In this lesson, we will be discussing the basics of cross-over designs briefly. A crossover design is a repeated measures design in which each experimental unit is given each of the different treatment levels during different time periods. This means that over time each experimental unit is assigned to a specific ordered sequence of different treatment levels. This is in contrast to a repeated-measures design in time, discussed in the previous chapter, where multiple (repeat) measurements are taken through time from the same experimental unit assigned to a specific treatment level.

This page titled 12: Cross-over Repeated Measure Designs is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

12.1: Introduction to Cross-Over Designs

The simplest cross-over design is a 2-level treatment, 2-period design. If we use A and B to represent the two treatment levels, then we can build the following table to represent their administering sequences.

Sequence	Period 1	Period 2
1	A	B
2	B	A

Experimental units are randomly assigned to receive one of the two different sequences. For example, if this were a clinical trial, patients assigned to sequence 2 would be given treatment B first, then after assessment of their condition, given treatment A and their condition re-assessed.

The complicated part of the cross-over design is the potential for carry-over effects. A carry-over effect is when the response to a particular treatment level has been influenced by the previous application of a different treatment level. The presence of carry-over effects is dealt with differently by different researchers in different ways. Having a sufficiently long washout period is one way to reduce carry-over effects. A washout period is a gap in time between the application of the treatment levels such that any residual effect of a previous treatment level has been dissipated and there is no detectable carry-over effect.

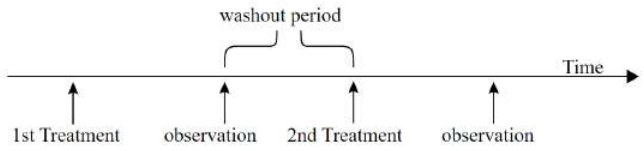


Figure 12.1.1: Timeline showing the washout period

However, there may be instances where significant carry-over effects may exist and sufficiently long washout periods may not be practically feasible. In such situations, an adjustment for carry-over effects would be appropriate during the statistical analysis.

If the treatment has only 2 levels, it is sufficient to simply include a "sequence" categorical variable in the model to assess the presence of a carry-over effect. If the sequence variable is significant, then a detectable carry-over effect exists.

With more than two treatment levels, the complexity of the analysis rises sharply. For 3 levels of treatment, 3 periods will be needed, and now we have 3! = 6 sequences to consider. What is needed in this case, in addition to a sequence variable, is a way to adjust the assessment of treatment effects for the presence of carry-over effects. This can be accomplished with a set of coded covariates in a repeated-measures ANCOVA.

This page titled 12.1: Introduction to Cross-Over Designs is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Coding for Carry-Over Covariates

The late Dr. Steve Arnold (Penn State), came up with a satisfactory solution to account for carry-over effects in the data analysis. The following example will illustrate how the procedure works. The data can be found in the textbook *Design of Experiments*, by Kuehl, as Example 16.1. Investigators want to evaluate the effect of 3 diets on Neutral Detergent Fiber (NDF) levels in steer. The three diets are administered to each steer in a sequence over 3 periods. A total of 6 sequences were used and two steers were assigned to each sequence of treatments.

The cross-over design can be summarized as:

	Period		
Sequence	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B
4	A	C	B
5	B	A	C
6	C	B	A

If we look at the first part of the dataset (Steer Data) for this example in Excel, we can see the following:

	A	B	C	D	E
1	PER	SEQ	DIET	STEER	NDF
2	1	1	A	1	50
3	1	1	A	2	55
4	1	2	B	1	44
5	1	2	B	2	51
6	1	3	C	1	35
7	1	3	C	2	41
8	1	4	A	1	54
9	1	4	A	2	58
10	1	5	B	1	50
11	1	5	B	2	55
12	1	6	C	1	41
13	1	6	C	2	46
14	2	1	B	1	61
15	2	1	B	2	63
16	2	2	C	1	42
17	2	2	C	2	45
18	2	3	A	1	55
19	2	3	A	2	56
20	2	4	C	1	48
21	2	4	C	2	51
22	2	5	A	1	57
23	2	5	A	2	59

Figure 12.2.1: First five columns of steer dataset in Excel.

We need now to add two columns to use an effect-type coding for the 3 treatment levels. We can use:

	x_1	x_2
A	1	0
B	0	1
C	-1	-1

Where x_1 and x_2 will be columns we create in the data to input for all of the rows of data. The coding values depend on which treatment level is administered during the ***previous period***. For example, if treatment A was administered in the previous period, then coding values would be $x_1 = 1, x_2 = 0$.

There will be no entries for the first period because on the first application of each treatment there are no treatments that have preceded it. Therefore a 0 is used as the coding value for both x_1 and x_2.

	A	B	C	D	E	F	G
1	PER	SEQ	DIET	STEER	NDF	x1	x2
2	1	1	A	1	50	0	0
3	1	1	A	2	55	0	0
4	1	2	B	1	44	0	0
5	1	2	B	2	51	0	0
6	1	3	C	1	35	0	0
7	1	3	C	2	41	0	0
8	1	4	A	1	54	0	0
9	1	4	A	2	58	0	0
10	1	5	B	1	50	0	0
11	1	5	B	2	55	0	0
12	1	6	C	1	41	0	0
13	1	6	C	2	46	0	0
14	2	1	B	1	61	1	0
15	2	1	B	2	63	1	0
16	2	2	C	1	42	0	1
17	2	2	C	2	45	0	1
18	2	3	A	1	55	-1	-1
19	2	3	A	2	56	-1	-1
20	2	4	C	1	48	1	0
21	2	4	C	2	51	1	0
22	2	5	A	1	57	0	1
23	2	5	A	2	59	0	1

Figure 12.2.2: Steer dataset, including x_1 and x_2, in Excel.

Looking at Period 2, sequence 1, treatment B we can refer back to the Sequence chart and see that it was preceded by treatment level A, so we assign $x_1 = 1$, and $x_2 = 0$, indicating that it was treatment A that could produce a carry-over effect here.

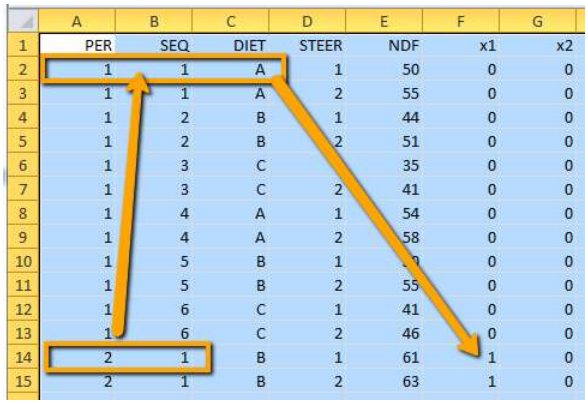


Figure 12.2.3: Identifying the carry-over effects using the spreadsheet.

The process can be repeated to define the coding variables to each entry in the dataset. The coded variables x_1 and x_2 are then entered into the general linear model as continuous covariates and LSmeans for treatments are adjusted for carry-over effects.

This page titled 12.2: Coding for Carry-Over Covariates is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

12.3: Programming for Steer Example

Using SAS

The SAS code given below will run a repeated measures ANCOVA in SAS for the Neutral Detergent Fiber levels in steer example in section 12.2.

Code

12.2.

```
data steer;
input PER   SEQ   DIET $ STEER NDF x1 x2;
datalines;
1     1     A     1     50     0     0
1     1     A     2     55     0     0
1     2     B     1     44     0     0
1     2     B     2     51     0     0
1     3     C     1     35     0     0
1     3     C     2     41     0     0
1     4     A     1     54     0     0
1     4     A     2     58     0     0
1     5     B     1     50     0     0
1     5     B     2     55     0     0
1     6     C     1     41     0     0
1     6     C     2     46     0     0
2     1     B     1     61     1     0
2     1     B     2     63     1     0
2     2     C     1     42     0     1
2     2     C     2     45     0     1
2     3     A     1     55     -1    -1
2     3     A     2     56     -1    -1
2     4     C     1     48     1     0
2     4     C     2     51     1     0
2     5     A     1     57     0     1
2     5     A     2     59     0     1
2     6     B     1     56     -1    -1
2     6     B     2     58     -1    -1
3     1     C     1     53     0     1
3     1     C     2     57     0     1
3     2     A     1     57     -1    -1
3     2     A     2     59     -1    -1
3     3     B     1     47     1     0
3     3     B     2     50     1     0
3     4     B     1     51     -1    -1
3     4     B     2     54     -1    -1
3     5     C     1     51     1     0
3     5     C     2     55     1     0
3     6     A     1     58     0     1
```

```
3     6     A     2     61     0     1
;
run;

/*Obtaining fit Statistics*/
proc mixed data=steer;
    class per seq diet steer;
    model ndf = per diet seq x1 x2/ddfm=kr;
    repeated per / subject=steer(seq) type=cs rcorr;
    ods output FitStatistics=FitCS (rename=(value=CS)) FitStatistics=FitCSp;
    title 'Compound Symmetry';
run;

proc mixed data=steer;
    class per seq diet steer;
    model ndf = per diet seq x1 x2/ddfm=kr;
    repeated per / subject=steer(seq) type=AR(1) rcorr;
    ods output FitStatistics=FitAR1 (rename=(value=AR1)) FitStatistics=FitAR1p;
    title 'Autoregressive Lag 1';
run;

proc mixed data=steer;
    class per seq diet steer;
    model ndf = per diet seq x1 x2/ddfm=kr;
    repeated per / subject=steer(seq) type=UN rcorr;
    ods output FitStatistics=FitUN (rename=(value=UN)) FitStatistics=FitUNp;
    title 'Unstructured';
run;

proc mixed data=steer;
    class per seq diet steer;
    model ndf = per diet seq x1 x2/ddfm=kr;
    repeated per / subject=steer(seq) type=CSH rcorr;
    ods output FitStatistics=FitCSH (rename=(value=CSH)) FitStatistics=FitCSHp;
    title 'HETEROGENOUS COMPOUND SYMMETRY';
run;

data fits;
    merge FitCS FitAR1 FitUN FITCSH;
    by descr;
run;
ods listing; title 'Summerized Fit Statistics'; run;
proc print data=fits; run;

/* Model Adjusting for carryover effects */
proc mixed data= steer;
    class per seq diet steer;
```

```
    model ndf = per diet seq x1 x2/ddfm=kr;
    repeated per / subject=steer(seq) type=csh;
    store out_steer;
run;

proc plm restore=out_steer;
    lsmeans diet / adjust=tukey plot=meanplot cl lines;
    ods exclude diffs diffplot;
run;

/* Reduced Model, Ignoring carryover effects */
proc mixed data= steer;
    class per seq diet steer;
    model ndf = per diet seq/ddfm=kr;
    repeated per / subject=steer(seq) type=csh;
    lsmeans diet / pdiff adjust=tukey;
run;
```

The results of the fit statistics are as follows:

Obs	Descr	CS	AR1	UN	CSH
1	-2 Res Log Likelihood	148.3	147.2	121.6	122.5
2	AIC (Smaller is Better)	152.3	151.2	133.6	130.5
3	AICC (Smaller is Better)	152.8	151.7	138.6	132.6
4	BIC (Smaller is Better)	153.2	152.1	136.5	132.5

Based on the fit statistics AIC (and also AICC and BIC), the covariance structure heterogeneous compound symmetry `(type=CSH)` was shown to be better compared to UN or CS or AR(1). Similar to the covariance structure CS, the CSH covariance structure, also has a constant correlation in the off-diagonal elements. However, the diagonal elements (the variance at each time point), can be different.

Here is the output that is generated for the full model:

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
PER	2	5.09	10.86	0.0146
DIET	2	11.6	188.52	< .001>
SEQ	5	10.9	31.96	< .0001
x1	1	11.2	17.03	0.0016
x2	1	11.2	78.85	< .0001

The Type 3 tests shown above are "model dependent," meaning that the sum of squares for each of the effects are adjusted for the other effects in the model. In this case, we have adjusted for the presence of carry-over effects. As the diet is significant, it is appropriate to generate LSmeans and the Tukey-Kramer mean comparisons for the diet factor.

DIET Least Squares Means								
DIET	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	upper
A	57.8092	1.6046	6.412	36.03	< 0.001	0.05	53.9432	61.6752
B	50.8134	1.6046	6.412	31.67	< 0.001	0.05	46.9474	54.6794
C	48.3774	1.6046	6.412	30.15	< 0.001	0.05	44.5114	52.2434

To see the adjustment on the treatment means, we can compare the LSmeans for a reduced model that does not contain the carry-over covariates.

LSmeans		
Full Model with Covariates		
Effect	DIET	Estimate
DIET	A	57.8092
DIET	B	50.8134
DIET	C	48.3774

Reduced Model (without carry-over covariates)		
Effect	DIET	Estimate
DIET	A	57.3941
DIET	B	50.9766
DIET	C	48.6292

Although the differences in the LSmeans between the two models are small in this particular example, these carry-over effect adjustments can be very important in many research situations.

Using R

- Load the Steer Data.
- Run the analysis by using different covariance structures and obtain fit statistics.

Code

- Load the Steer data, run the analysis by using different covariance structures and obtain fit statistics by using the following commands:

```
setwd("~/path-to-folder/")
steer_data <- read.table("steer_data.txt",header=T)
attach(steer_data)
library(nlme)
```

```
model_CS<-gls(NDF ~ factor(PER) + DIET + factor(SEQ) + x1 + x2,steer_data,correlat
model_AR<-gls(NDF ~ factor(PER) + DIET + factor(SEQ) + x1 + x2,steer_data,correlat
Model_Selection <- data.frame(
c ("","-2LogLik","AIC", "BIC"),
c("CS", round(-2*summary(model_CS)$logLik,2),round(summary(model_CS)$AIC,2),round(
c("AR1", round(-2*summary(model_AR)$logLik,2),round(summary(model_AR)$AIC,2),round
stringsAsFactors = FALSE)
names(Model_Selection) <- c( " ","","")
print(Model_Selection)
#1              CS    AR1
#2 -2LogLik 140.75 141.11
#3      AIC 168.75 169.11
#4      BIC 185.25 185.61
detach(steer_data)
```

This page titled 12.3: Programming for Steer Example is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

12.4: Testing the Significance of the Carry-Over Effect

To test for the overall significance of carry-over effects, we can drop the carry-over covariates (x_1 and x_2 in our example) and re-run the ANOVA. Because the reduced model is a subset of the full model that includes the covariates, we can construct a likelihood ratio test.

$$\Delta G^2 = (-2\log L_{Reduced}) - (-2\log L_{Full}) \quad \text{with } df_{Reduced} - df_{Full} \text{ degrees of freedom} \tag{12.4.1}$$

The $-2\log L$ values are provided in the SAS Fit Statistics output for each model. For our example, the SAS output for the Full model with carry-over covariates is:

Fit Statistics	
-2 Res Log Likelihood	122.5
AIC (smaller is better)	130.5
AICC (smaller is better)	132.6
BIC (smaller is better)	132.5

And for the reduced model without the carry-over covariates is:

Fit Statistics	
-2 Res Log Likelihood	136.5
AIC (smaller is better)	144.5
AICC (smaller is better)	146.4
BIC (smaller is better)	146.4

So,

$$\Delta G^2 = 136.5 - 122.5 = 14$$

and with

$$\chi^2_{.05,2} = 5.991$$

we conclude that there are significant carry-over effects.

This page titled 12.4: Testing the Significance of the Carry-Over Effect is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

12.5: Try It!

Exercise 12.5.1

Ginkgo Biloba is recognized as a herbal remedy for memory improvement. To investigate its effectiveness on memory recall, a cross-over study was planned using 3 treatments: one tablet of 120mg Ginkgo Biloba (G), one tablet of 200mg Caffeine pill (C), and sleep for 2 hours before the recall test (S). The assignment order of the 3 treatments to participants was determined by randomly assigning 12 college students to one of 6 possible sequences of the 3 treatments. The student recall capability was assessed based on a Recall score and the 3 treatments were given over 3 consecutive days. On each day, only one treatment was administered before one 1 hour of taking the recall test at 2.00 pm (the higher the recall score the better).

a) Which variable signifies the experimental unit?

Solution

Id

b) What is the washout period?

Solution

One day

c) How many periods are required?

Solution

3

d) How many replicates are there?

Solution

2

e) Perform a statistical analysis to determine if the treatments vary with regard to memory recall. The data can be found in Cross_over_Ex1.txt

Solution: Using SAS

```
DATA CROSS_OVER;
INPUT score  Seq $ PER  Id  TRT $ X1  X2;
DATALINES;
74  CGS  1  1  C  0  0
45  CGS  1  2  C  0  0
92  CSG  1  3  C  0  0
94  CSG  1  4  C  0  0
79  GCS  1  5  G  0  0
35  GCS  1  6  G  0  0
31  GSC  1  7  G  0  0
40  GSC  1  8  G  0  0
106  SCG  1  9  S  0  0
60  SCG  1  10  S  0  0
80  SGC  1  11  S  0  0
110  SGC  1  12  S  0  0
41  CGS  2  1  G  1  0
```

```
20  CGS  2  2  G  1  0
50  CSG  2  3  S  1  0
88  CSG  2  4  S  1  0
92  GCS  2  5  C  0  1
50  GCS  2  6  C  0  1
32  GSC  2  7  S  0  1
54  GSC  2  8  S  0  1
120  SCG  2  9  C  -1  -1
80  SCG  2  10  C  -1  -1
75  SGC  2  11  G  -1  -1
55  SGC  2  12  G  -1  -1
64  CGS  3  1  S  0  1
30  CGS  3  2  S  0  1
55  CSG  3  3  G  -1  -1
55  CSG  3  4  G  -1  -1
76  GCS  3  5  S  1  0
50  GCS  3  6  S  1  0
38  GSC  3  7  C  -1  -1
66  GSC  3  8  C  -1  -1
85  SCG  3  9  G  1  0
40  SCG  3  10  G  1  0
88  SGC  3  11  C  0  1
86  SGC  3  12  C  0  1
;
RUN;

proc mixed data=CROSS_OVER;
  class PER TRT SEQ  ID;
  model SCORE=PER  TRT SEQ  X1 X2 / ddfm=kr;
  repeated PER /subject=ID(SEQ) type=cs rcorr;
  ods output FitStatistics=FitCS (rename=(value=CS)) FitStatistics=FitCSp;
  title 'Compound Symmetry';
run;
title ' '; run;

proc mixed data=CROSS_OVER;
  class PER TRT SEQ  ID;
  model SCORE=PER  TRT SEQ  X1 X2 / ddfm=kr;
  repeated PER /subject=ID(SEQ) type=AR(1) rcorr;
  ods output FitStatistics=FitAR1 (rename=(value=AR1)) FitStatistics=FitAR1p;
  title 'Autoregressive Lag 1';
run;
title ' '; run;

proc mixed data=CROSS_OVER;
  class PER TRT SEQ  ID;
  model SCORE=PER  TRT SEQ  X1 X2 / ddfm=kr;
```

```
  repeated PER /subject=ID(SEQ) type=UN rcorr;
  ods output FitStatistics=FitUN (rename=(value=UN)) FitStatistics=FitUNp;
  title 'Unstructured';
run;
title ' '; run;

proc mixed data=CROSS_OVER;
  class PER TRT SEQ  ID;
  model SCORE=PER  TRT SEQ  X1 X2 / ddfm=kr;
  repeated PER /subject=ID(SEQ) type=CSH  rcorr;
  ods output FitStatistics=FitCSH (rename=(value=CSH)) FitStatistics=FitCSHp;
  title 'HETEROGENOUS COMPOUND SYMMETRY';
run;
title ' '; run;

data fits;
  merge FitCS FitAR1 FitUN FITCSH;
  by descr;
run;
ods listing; proc print data=fits; run;
```

The above code was used to obtain the fit statistics for different covariance structures and the AICC (AIC and BIC) values indicate that CS is the best covariance structure. Hence, the remaining analysis was done using CS.

Obs	Descr	CS	AR1	UN	CSH
1	-2 Res Log Likelihood	215.3	219.1	212.7	214.7
2	AIC (Smaller is Better)	219.3	223.1	224.7	222.7
3	AICC (Smaller is Better)	219.9	223.7	229.7	224.8
4	BIC (Smaller is Better)	220.3	224.1	227.6	224.6

```
/* Model Adjusting for carryover effects */
proc mixed data= CROSS_OVER;
  class per TRT SEQ ID;
  model SCORE=PER  TRT SEQ  X1 X2 / ddfm=kr;
  repeated PER /subject=ID(SEQ) type=cs rcorr;
  store out_CROSS_OVER;
run;

proc plm restore=out_CROSS_OVER;
  lsmeans TRT / adjust=tukey plot=meanplot cl lines;
  ods exclude diffs diffplot;
run;
```

```
/* Reduced Model, Ignoring carryover effects */
proc mixed data= CROSS_OVER;
  class per TRT seq  ID;
  model SCORE=PER  TRT SEQ   / ddfm=kr;
  repeated PER /subject=ID(SEQ) type=cs rcorr;
  lsmeans TRT / pdiff adjust=tukey;
run;
```

Full Model: with carryover effect

Fit Statistics	
-2 Res Log Likelihood	215.3
AIC (Smaller is Better)	219.3
AICC (Smaller is Better)	219.9
BIC (Smaller is Better)	220.3

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
PER	2	18	3.12	0.0688
TRT	2	18	18.03	<.0001
Seq	5	6.1	1.46	0.3259
X1	1	18	0.10	0.7565
X2	1	18	0.18	0.6768

Reduced Model: without carryover effect

Fit Statistics	
-2 Res Log Likelihood	224.2
AIC (Smaller is Better)	228.2
AICC (Smaller is Better)	228.7
BIC (Smaller is Better)	229.1

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
PER	2	20	3.36	0.0552
TRT	2	20	23.70	<.0001
Seq	5	6	1.52	0.3101

The test statistic below tests for the significance of the carry over effect. $\Delta^2 = (-2\log L_{Reduced}) - (-2\log L_{Full})\, with \backslash (df_{Full} - df_{Reduced}$ degrees of freedom.

$\Delta^2 = (224.2 - 215.3) = 8.9$. This exceeds the critical Chi-Square of 5.991 $\left(\chi^2_{.05,2}\right)$ indicating that model with carryover effect is more appropriate and will be used to base the final conclusions.

In the full model output, the Treatment is the only significant factor, so LSmeans and comparisons are generated only for the treatment effect. The results of the Tukey comparison procedure indicate that treatments C and S are not significantly different, but G is significantly lower, indicating that both sleep for 2 hours and caffeine are similarly effective in improving recall capability and are superior to Ginkgo biloba.

TRT Least Squares Means

TRT	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
C	76.7222	6.2382	8.572	12.30	<.0001	0.05	62.5024	90.9421
G	50.4306	6.2382	8.572	8.08	<.0001	0.05	36.2107	64.6504
S	67.5139	6.2382	8.572	10.82	<.0001	0.05	53.2940	81.7337

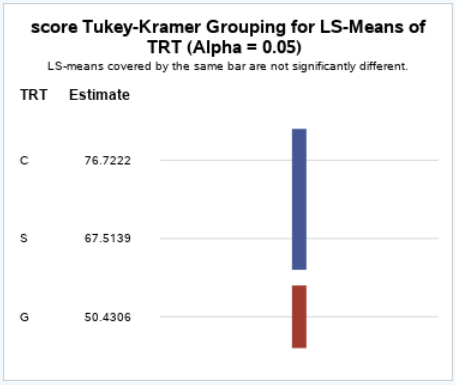


Figure 12.5.1: Scores for Tukey-Kramer grouping

This page titled 12.5: Try It! is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

12.6: Chapter 12 Summary

In this lesson, we discussed the second type of repeated measures designs, namely cross-over designs wherein the treatments themselves are switched on the same experimental unit during the course of the experiment. One concern is the presence of carryover effects caused due to previous applications of different treatment levels. Carryover effects can be reduced by imposing a wash-out period in between the application of different treatment levels on the same experimental unit or by utilizing a repeated measures ANCOVA model that includes coding covariates representing the carry-over effects.

This page titled 12.6: Chapter 12 Summary is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Penn State's Department of Statistics via source content that was edited to the style and standards of the LibreTexts platform.

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Analysis of Variance and Design of Experiments

Webpages: 111

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC 4.0: 98.2% (109 pages)
- Undeclared: 1.8% (2 pages)

By Page

Detailed Licensing

Overview

Title: Analysis of Variance and Design of Experiments

Webpages: 111

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC 4.0: 98.2% (109 pages)
- Undeclared: 1.8% (2 pages)

By Page