

12.4: Two-way ANOVA

In all the statistical testing we've done so far, and will do in Psy 233/234, there is only one *dependent variable* (DV) — we have been/are doing univariate statistics.

And so far, in all the tests we've seen there has only been one *independent variable* (IV). For the t -tests the IV is group or population with only two values^[1] 1 and 2. In one-way ANOVA the single IV has k (number of groups) values. Also, so far, the IV has been a discrete variable (that will change when we get to regression). The graph to keep in mind for the one-way ANOVA is a profile graph as shown in Figure 12.1.

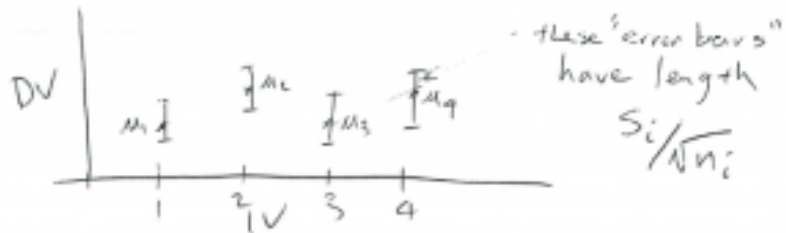
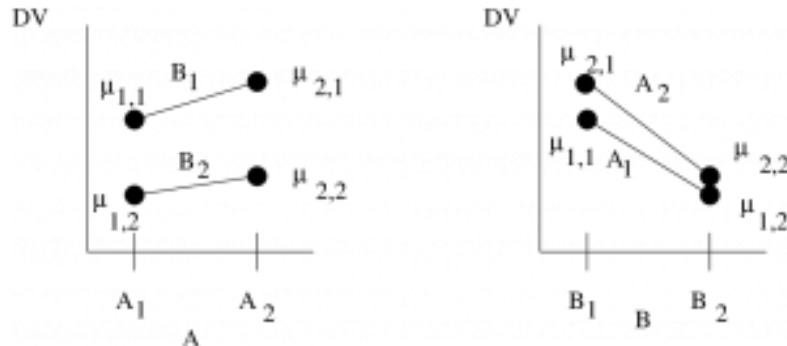


Figure 12.1: The profile plot is a good way to think of one-way ANOVA data with the IV on the x -axis and the DV on the y -axis. A one-way ANOVA tests the hypothesis: are all the means μ_i equal to each other? (An actual data profile graph can only have sample values \bar{x}_i , we show a kind of confidence interval plot here.)

With two-way ANOVA you have **two IVs**. Let's call the two IVs A and B . Each IV in two-way ANOVA is called a *factor*. A and B can each have several values (or "levels"). To introduce concepts, let's stick with the case where each of A and B have only 2 values: A_1 and A_2 for A , and B_1 and B_2 for B . This is the 2×2 ANOVA case, where the 2 tells you how many levels are in each factor. If, for example, A had 4 levels (values) and B had 3 levels then you'd have a 4×3 ANOVA. Let's stick with the 2×2 case for now.

There are several ways to think of a two-way ANOVA. Let's start with two-dimensional profile plots for a 2×2 ANOVA :



The profile plot can be done in one of two ways. The y axis represents the DV in both cases. On the left, the x axis represents the IV A and the two values of the other IV, B , are represented as lines. On the right, the x axis represents the IV B and the two values of the other IV, A , are represented as lines. Look closely at the plots. The dots represent the population values^[2] with $\mu_{i,j}$ being the value of the population labelled by $A = i$, $B = j$. The means in the two plots are exactly the same. Each combination of IVs, i, j , defines a *treatment group*. For a 2×2 ANOVA there are four treatment groups.

Two way ANOVA supposedly had one of its first applications to agriculture. So, to fix ideas, let's take our two IV's, also known as two **factors** as :

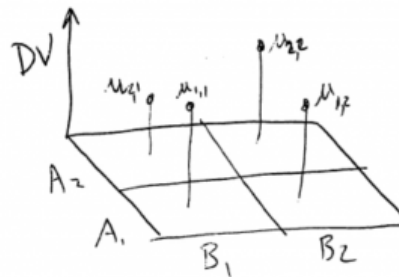
A = Plant food type
 B = Soil type

Then, with two levels for each factor, we can visualize the setup as fields where you would grow plants :

Plant food type	A_2	$\mu_{2,1}$	$\mu_{2,2}$
	A_1	$\mu_{1,1}$	$\mu_{1,2}$
		B_1	B_2
		Soil type	

Each field, or treatment group, is also known as a *cell*.

Now, let's use the x and y axes to represent the IVs (B and A in this case). Then we can use the z axis to represent the DV in a 3D plot :

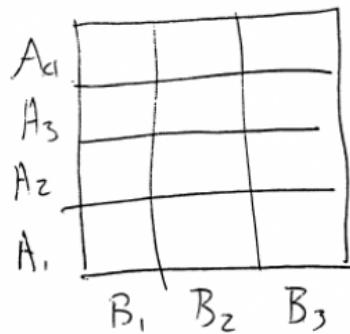


This makes sense. A two-way ANOVA has three variables, two IVs and one DV, so the data are 3D data and the plot above shows how those data appear in 3D space. If you look at the 3D plot from the front you see the profile plot with B on the x axis. If you look at the 3D plot from the (right) side, you see the profile plot with A on the x axis.

We've focused on 2×2 designs. But the two IVs can have any number of discrete values or levels. For example, a 3×2 cell diagram would look like :

A_3		
A_2		
A_1		
	B_1	B_2

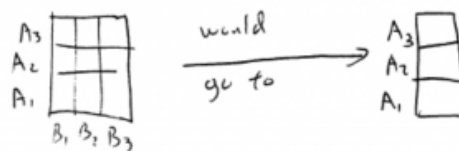
And a 4×3 design would look like :



Now that we understand what kind of data we have, it's time to move onto hypothesis testing. In two-way ANOVA there are three hypotheses to test :

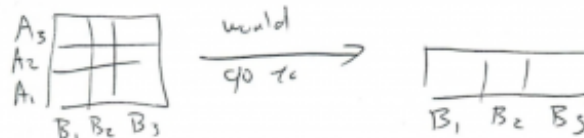
1. Is there a “main effect” of A ?
2. Is there a “main effect” of B ?
3. Is there an *interaction* of $A \times B$?

In all cases, H_0 is that there is no effect or interaction. As we will see, each hypothesis is a one-way ANOVA of the two-way data suitably collapsed into a one-way design. Let's begin with the main effect of A . The hypothesis is equivalent to collapsing the design across B :



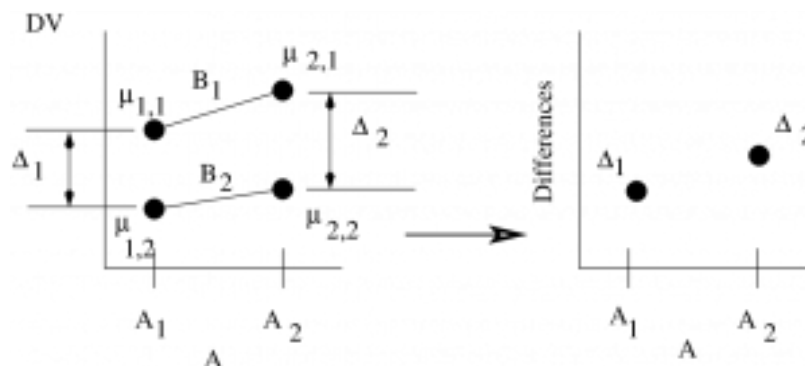
and then doing a one-way ANOVA with the one IV equal to A . The collapse is done by averaging over B which is the same as removing the cell boundaries between the B cells and only categorizing the data by the A levels.

The hypothesis for the main effect of B is similarly equivalent to collapsing across A :



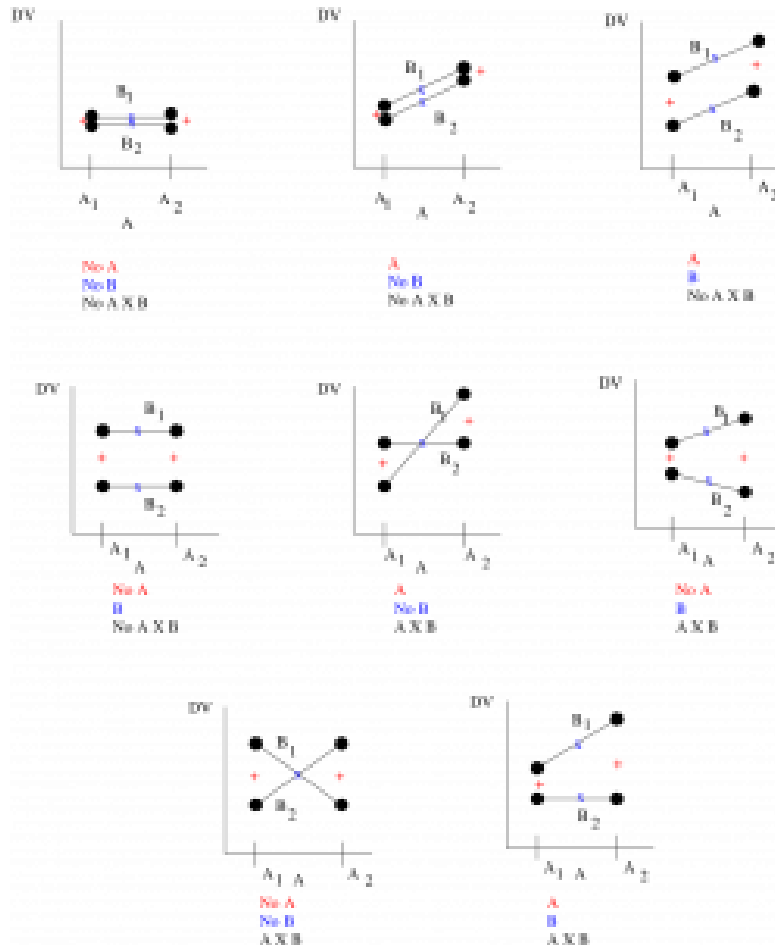
and then doing a one-way ANOVA with the one IV equal to B .

The hypothesis test for the interaction is a one-way ANOVA on the “difference of differences”. The idea in interpreting a significant^[3] interaction is that the effect of changing IV A *depends on* the effect of changing B . Let's see how the differences arise in a 2×2 ANOVA :



The interaction tests if there is a significant difference between the two differences $\Delta_1 = (\mu_{1,1} - \mu_{1,2})$ and $\Delta_2 = (\mu_{2,1} - \mu_{2,2})$. Note that I could have set up the differences on the profile plot with B on the x axis. It does not matter, the resulting one-way ANOVA turns out to be the same. For $A \times 2$ or $2 \times B$ ANOVAs you get more than two differences to compare with a one-way ANOVA. With a generic $A \times B$ ANOVA you need to take a mean of differences to compare to each other. The interpretation of a generic $A \times B$ can be tricky.

The interpretation of 2×2 interactions is, however, pretty straightforward. You need to consider all the possible outcome of the 2×2 ANOVA with its three hypotheses. Since any hypothesis can be significant or not we have $2^3 = 8$ possible outcomes^[4]. Let's look at *generic* cases of all the combinations of the outcomes of a 2×2 ANOVA using A , B and $A \times B$ to denote that H_0 has been rejected and significant effects have been found :



The first thing to remember about these diagrams is that they are for *interpretation* — for step 5 of our hypothesis testing procedure. You have to do the actual hypothesis test with three F_{test} statistics to decide which case you have. Secondly, note that the graphs are *generic*. In statistics numbers are fuzzy. That is, every mean is fuzzy by a standard deviation. So think of the dots on the graphs as fuzzy balls and that the lines do not have to go to the centers of the fuzzy balls. Now look at the + and x symbols. The + symbols show what happens when you collapse the design over B to see the main effect of A ; what is left are the two averages^[5] for A_1 and A_2 . The two + means are then compared with a one-way ANOVA (essentially a t -test since $t^2_{\nu} = F_{1,\nu}$) to see if there is a main effect of A . Similarly the x symbols show what happens when you collapse the design across A to see the main effect of B . The means for B_1 and B_2 will be halfway^[6] along the B_1 and B_2 lines. The two x means are then compared with a one-way ANOVA to see if there is a main effect of B . Finally let's look at the interactions. There are four cases in the diagrams that show interactions. In two cases the diagrams have crossed lines where the differences at either end are the negative of each other (and so are different) and in two cases the magnitudes of the differences are different. Looking at all the cases we see that there will be an interaction if the lines are not *statistically* parallel. The concept of statistically parallel is important here. Your actual data profile plot may not look like it has parallel lines but there will be no significant interaction if the lines are not statistically distinguishable from being parallel — this is the information that the hypothesis test gives you.

Before we move on, let's consider post-hoc testing for two-way ANOVAs. This usually means comparing means pairwise cell by cell. As with one-way ANOVA that would mean also finding a suitable correction for the p -value if t -tests are used. We won't cover post-hoc testing for two-way ANOVAs in any detail here except to point out that post hoc testing for a 2×2 ANOVA is redundant. A 2×2 ANOVA is essentially three t tests. If there is any interesting cell by cell difference, there will be an interaction. With a 2×2 ANOVA comparing cells is an interpretation problem, not one of statistical testing. The post-hoc test for a 2×2 ANOVA is really to figure out what generic profile plot matches your data.

Next, let's look at the ANOVA table for a two-way ANOVA. It looks like :

Source	Sum of Squares	Degrees of Freedom	Mean Square	F_{test}
A	SS_A	$\nu_A = a - 1$	MS_A	F_A
B	SS_B	$\nu_B = b - 1$	MS_B	F_B
$A \times B$	$SS_{A \times B}$	$\nu_{A \times B} = (a - 1)(b - 1)$	$MS_{A \times B}$	$F_{A \times B}$
Within (error)	SS_W	$\nu_W = ab(n - 1)$	MS_W	
Totals	SS_T	N-1		

The two-way ANOVA table is very similar to the one-way ANOVA table except that there is now one line for each of the three hypotheses (three signals) plus a line that essentially quantifies the noise. We could also add another column for the p -value of the three effects. In the degrees of freedom formula, a is the number of levels for the A factor and b is the number of levels for the B factor. The formula for ν_W in the table is for a *balanced design* that has the same number, $n_{i,j} = n$, of data points in each cell (i, j) . The total number of data points in a balanced design is $N = abn$. For a generic design, $N = \sum_{i=1}^a \sum_{j=1}^b n_{i,j}$ and $\nu_W = \sum_{i=1}^a \sum_{j=1}^b (n_{i,j} - 1)$.

The formulae in the other columns are the same for any ANOVA table: $MS = SS/\nu$ for each line, or effect, and $F_{\text{effect}} = MS_{\text{effect}} / MS_W$. Explicitly:

$$MS_A = \frac{SS_A}{\nu_A} \quad MS_B = \frac{SS_B}{\nu_B} \quad MS_{A \times B} = \frac{SS_{A \times B}}{\nu_{A \times B}} \quad MS_W = \frac{SS_W}{\nu_W} \quad (12.4.1)$$

and the F test statistics are

$$F_A = \frac{MS_A}{MS_W} \quad F_B = \frac{MS_B}{MS_W} \quad F_{A \times B} = \frac{MS_{A \times B}}{MS_W}. \quad (12.4.2)$$

For the critical statistics, which you look up in the [F Distribution Table](#), the degrees of freedom to use are $\nu_A = a - 1$, $\nu_B = b - 1$, $\nu_{A \times B} = (a - 1)(b - 1)$ and $\nu_W = ab(n - 1)$.

Now all we need are the formulae for the sums of squares. These sums of squares formulae, and the two-way ANOVA that you are responsible for in this class are for a *between subjects* design. That is, the samples for each cell are *independent*, every data point is from a different individual. We also assume homoscedasticity, $\sigma_{i,j}^2 = \sigma^2$ for all cells (i, j) . Now to the SS formulae, we'll just give them for a balanced design^[7]. To do this we need to label the data points this way: use $x_{i,j,k}$ where i and j label the cell ($1 \leq i \leq a$) and ($1 \leq j \leq b$) and k labels the data point within the cell ($1 \leq k \leq n$). First we define a "correction term", C , to keep the formulae simple:

$$C = \frac{1}{N} \left(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n x_{i,j,k} \right)^2 \quad (12.4.3)$$

With this, the formulae for the sums of squares in a balanced design two-way ANOVA are:

$$\begin{aligned}
 SS_T &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n x_{i,j,k}^2 - C \\
 SS_A &= \frac{1}{bn} \sum_{i=1}^a \left(\sum_{j=1}^b \left(\sum_{k=1}^n x_{i,j,k} \right) \right)^2 - C \\
 SS_B &= \frac{1}{an} \sum_{j=1}^b \left(\sum_{i=1}^a \left(\sum_{k=1}^n x_{i,j,k} \right) \right)^2 - C \\
 SS_{A \times B} &= \frac{1}{n} \left(\sum_{i=1}^a \sum_{j=1}^b \left(\sum_{k=1}^n x_{i,j,k} \right)^2 \right) - C - SS_A - SS_B \\
 SS_W &= SS_T - SS_A - SS_B - SS_{A \times B}
 \end{aligned}$$

Relax, you won't have to chug your way through these sums of squares formulae in an exam. That would be way too tedious even if you are comfortable with all those summation signs. But we will take a look at using them in an example where we set up cell diagrams and use marginal sums to help us along. On an exam, you will be able to simply read the values for the sums of squares from an SPSS ANOVA table output.

Example 12.4 : A researcher wishes to see whether the type of gasoline used and the type of automobile driven have any effect on gasoline consumption. Two types of gasoline, regular and high octane, will be used and two types of automobiles, two-wheel drive and four-wheel drive, will be used in each group. There will be two automobiles in each group for a total of eight automobiles used. The data, in cell form are (the DV is miles per gallon) :

		Type of Automobile (B)	
		2-Wheel	4-Wheel
Gas (A)	Regular	26.7 25.2	28.6 29.3
	High Octane	32.3 32.8	26.1 24.2

Using a two-way ANOVA at $\alpha = 0.05$ test the effects of gasoline and automobile types on gas millage.

Solution :

0.Data Reduction.

Here we will calculate the sums of squares, SS_A , SS_B , $SS_{A \times B}$ and SS_W (and SS_T) by hand using marginal sums. Again, in an exam you will be given the sums of squares. But we will see marginal sums again when we do χ^2 contingency tables in [Chapter 15](#).

		B ₁	B ₂					
A ₁	A ₂	26.7	28.6	A ₁	51.9	57.9	109.8	
		25.2	29.3		65.1	50.3	115.4	
		32.3	26.1			117.0	108.2	225.2
		32.8	24.2			marginal sums		

Beginning with the data on the left, sum each cell to give the numbers on the right. In summing each cell you are computing the terms $\sum_{k=1}^n x_{i,j,k} = \sum_{k=1}^2 x_{i,j,k}$, $1 \leq i, j \leq 2$ in the sum of squares equations on page 234. (Note that these sums are n times the means, $\bar{x}_{i,j}$, of the cells, which is what the two-way ANOVA compares: $\sum_{k=1}^n x_{i,j,k} = n\bar{x}_{i,j}$.) Next, compute the marginal sums,

the **sums of the rows**, on the far right, and the **sums of the columns**, on the bottom. Then compute the **grand sum**, the sum of everything, which is the the sum of the marginal sums on the right which equals the sums on the bottoms (which should be equal — a check). The marginal sums show up in the second inner brackets in the sums of squares formula. Notice that the **sums of the rows** collapse the design across B to give a one-way ANOVA for A (main effect of A) and the **sums of the columns** collapse the design across A to give a one-way ANOVA for B (main effect of B). With the marginal sums we compute:

$$C = \frac{1}{N} \left(\sum \sum \sum x_{i,j,k} \right)^2 = \frac{8}{1} = 6339.38$$

$$\begin{aligned} SS_T &= \sum \sum \sum x_{i,j,k}^2 - C \quad (\text{This one's not from the marginal sums.}) \\ &= (26.7^2 + 25.2^2 + 28.6^2 + 29.3^2 + 32.3^2 + 32.8^2 + 26.1^2 + 24.2^2) - 6339.38 \\ &= 6410.36 - 6339.38 \\ &= 70.98 \end{aligned}$$

$$\begin{aligned} SS_A &= \frac{1}{bn} \sum_{i=1}^a \left(\sum_{j=1}^b \left(\sum_{k=1}^n x_{i,j,k} \right) \right)^2 - C \\ &= \frac{1}{(2)(2)} [(109.8)^2 + (115.4)^2] - 6339.38 \\ &= 6343.30 - 6339.38 \\ &= 3.92 \end{aligned}$$

$$\begin{aligned} SS_B &= \frac{1}{an} \sum_{j=1}^b \left(\sum_{i=1}^a \left(\sum_{k=1}^n x_{i,j,k} \right) \right)^2 - C \\ &= \frac{1}{(2)(2)} [(117.0)^2 + (108.2)^2] - 6339.38 \\ &= 6349.06 - 6339.38 \\ &= 9.68 \end{aligned}$$

$$\begin{aligned} SS_{A \times B} &= \frac{1}{n} \left(\sum \sum \left(\sum x_{i,j,k} \right)^2 \right) - C - SS_A - SS_B \quad (\text{No marginal sums.}) \\ &= \frac{1}{2} (51.9^2 + 57.9^2 + 65.1^2 + 50.3^2) - 6339.38 - 3.92 - 9.68 \\ &= 6407.06 - 6339.38 - 3.93 - 9.68 \\ &= 54.08 \end{aligned}$$

$$\begin{aligned} SS_W &= SS_T - SS_A - SS_B - SS_{A \times B} \\ &= 70.89 - 3.92 - 9.68 - 54.08 \\ &= 3.30 \end{aligned}$$

There. Now the sums of squares are ready for computing the test statistics. At this point you can start making your ANOVA table to keep track of your calculations. Here we'll see the ANOVA table at the last step.

1. Hypotheses.

H_0 : No main effect of A . (Changing gas type doesn't change mileage.)

H_1 : Main effect of A .

H_0 : No main effect of B . (Changing auto type doesn't change mileage.)

H_1 : Main effect of B .

H_0 : No interaction $A \times B$.

H_1 : Interaction $A \times B$. (The effect of gas type on mileage depends on auto type.)

2. Critical statistics.

There are three of them, one for each hypothesis pair. Use the **F Distribution Table** with the α labelling the table equal to the test $\alpha = 0.05$ since there are no such things as one and two tailed tests for ANOVA. From the **F Distribution Table** find:

For A :

$$\begin{aligned}\nu_1 &= a - 1 = 2 - 1 = 1 \text{ (d.f.N)} \\ \nu_2 &= ab(n - 1) = (2)(2)(2 - 1) = 4 \text{ (d.f.D.)} \\ F_{\text{crit}} &= 7.71\end{aligned}$$

For B :

$$\begin{aligned}\nu_1 &= b - 1 = 2 - 1 = 1 \text{ (d.f.N)} \\ \nu_2 &= ab(n - 1) = (2)(2)(2 - 1) = 4 \text{ (d.f.D.)} \\ F_{\text{crit}} &= 7.71\end{aligned}$$

For $A \times B$:

$$\begin{aligned}\nu_1 &= (a - 1)(b - 1) = (2 - 1)(2 - 1) = 1 \text{ (d.f.N)} \\ \nu_2 &= ab(n - 1) = (2)(2)(2 - 1) = 4 \text{ (d.f.D.)} \\ F_{\text{crit}} &= 7.71\end{aligned}$$

The critical statistics are all the same for a 2×2 ANOVA ($\nu_1 = 1$ for all the hypotheses pairs — essentially three t -tests because $t^2_{\nu} = F_{1,\nu}$). For bigger designs, the critical statistics will, in general, be different for each hypothesis pair.

3. Test statistics.

Use the sums of squares to compute:

$$MS_A = \frac{SS_A}{a - 1} = \frac{3.920}{2 - 1} = 3.920 \quad (12.4.4)$$

$$MS_B = \frac{SS_B}{b - 1} = \frac{9.680}{2 - 1} = 9.680 \quad (12.4.5)$$

$$MS_{A \times B} = \frac{SS_{A \times B}}{(a - 1)(b - 1)} = \frac{54.080}{(2 - 1)(2 - 1)} = 54.080 \quad (12.4.6)$$

$$MS_W = \frac{SS_W}{ab(n - 1)} = \frac{3.300}{4} = 0.825 \quad (12.4.7)$$

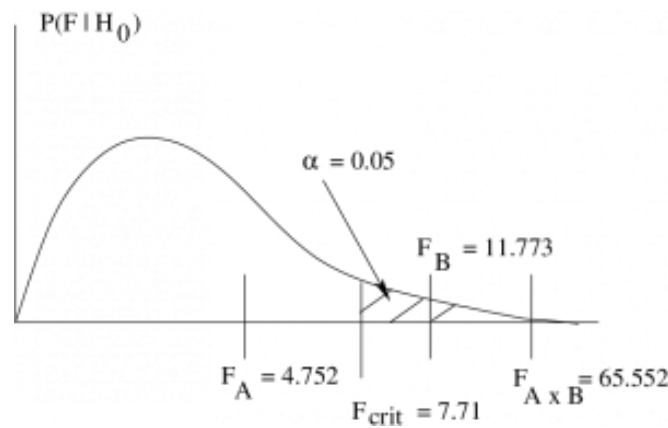
$$F_A = \frac{MS_A}{MS_W} = \frac{3.920}{0.825} = 4.752 \quad (12.4.8)$$

$$F_B = \frac{MS_B}{MS_W} = \frac{9.680}{0.825} = 11.773 \quad (12.4.9)$$

$$F_{A \times B} = \frac{MS_{A \times B}}{MS_W} = \frac{54.080}{0.825} = 65.552 \quad (12.4.10)$$

4. Decision.

In general, we need three diagrams, but in this case all the critical statistics are the same so we can draw :



So :

- For A , do not reject H_0 , there is no main effect of A .
- For B , reject H_0 , there is a main effect of B .
- For $A \times B$, reject H_0 , there is an interaction.

5. Interpretation.

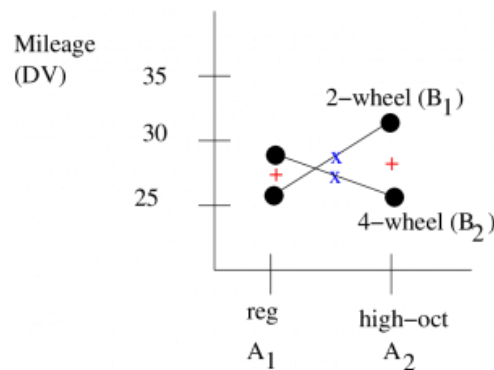
Simply put, at $\alpha = 0.05$ there is no effect of gas type (factor A) on mileage; there is an effect of auto type (factor B) on gas mileage and; there is an interaction between gas type and mileage, the change in mileage with auto type depends on the gas type used. We'll look at the profile plot to see if we really understand what this means but first we should complete the ANOVA table :

Source	Sum of Squares	Degrees of Freedom	Mean Square	F_{test}	p
A (gas)	3.92	1	3.92	F_A	$p > 0.05$
B (auto)	9.68	1	9.68	F_B	$p < 0.05$
$A \times B$	54.08	1	54.08	$F_{A \times B}$	$p < 0.05$
Within (error)	3.30	4	0.825		
Totals	70.98	7			

To draw the profile plot, we need to do one more data reduction :

		2-wheel B_1	4-wheel B_2
regular	A_1	25.95	28.95
high-octane	A_2	32.55	25.15
		cell means	

So the profile plot (without error bars — but remember the numbers are fuzzy) is :



To interpret this fully, remember the rules from previously about collapsing by looking at the midpoints between the two A values and the midpoint of the lines. Here we see that averaged over auto types, it looks like there is no difference in gas mileage between gas type. That conclusion is statistically confirmed by the fact that we found no main effect of factor A , gas — the two $+$ values are not significantly different. The centres of the lines marked by the x are, however, significantly different because we found a main effect of B , auto type. And the nature of the statistically significant interaction is obvious, the gas mileage can go up or down when you change gas types depending on what kind of car you drive. Switching from regular gas to high octane gas will improve your mileage if you drive a 4-wheel drive car but the mileage will get worse if you drive a 2-wheel drive car.

□

You will see bigger designs than a 2×2 . Collapsing across A or B in those larger designs to get to one-way ANOVAs is conceptually straightforward. The interaction is trickier, but the idea of an interaction existing when there are statistically non-parallel lines still holds. The 2×2 ANOVA is essentially three t -tests. This makes the 2×2 ANOVA powerful and easy to interpret. As always, in statistics simpler is more powerful. We will take a brief quantitative look at statistical power in [Chapter 13](#) but qualitatively, simpler is more powerful.

1. For the single sample t -test the two values of the IV were the population of interest and a hypothetical population representing H_0 having the mean k . ↩
2. Population values are used for this illustration. When you plot profile plots like this you will use sample means. ↩
3. Remember that "significant" means "reject H_0 ". ↩
4. Remember the counting rule! ↩
5. If the cell sizes, $n_{i,j}$, are all the same then the average is exactly halfway between the dots. ↩
6. For equal cell sizes. For unequal cell sizes the x will still be somewhere along the line. ↩
7. Of course, if you're using SPSS you don't need to restrict yourself to a balanced design. SPSS knows the generic SS formulae. ↩

This page titled [12.4: Two-way ANOVA](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Gordon E. Sarty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.