

7.2: One-factor ANOVA

The one-factor ANOVA is sometimes also called a between-subjects ANOVA, an independent factor ANOVA, or a one-way ANOVA (which is a bit of a misnomer as we discuss later). The critical ingredient for a one-factor, between-subjects ANOVA, is that you have one independent variable, with at least two-levels. When you have one IV with two levels, you can run a t -test. You can also run an ANOVA. Interestingly, they give you almost the exact same results. You will get a p -value from both tests that is identical (they are really doing the same thing under the hood). The t -test gives a t -value as the important sample statistic. The ANOVA gives you the F -value (for Fisher, the inventor of the test) as the important sample statistic. It turns out that t^2 equals F , when there are only two groups in the design. They are the same test. Side-note, it turns out they are all related to Pearson's r too (but we haven't written about this relationship yet in this textbook).

Remember that t is computed directly from the data. It's like a mean and standard error that we measure from the sample. In fact it's the mean difference divided by the standard error of the sample. It's just another descriptive statistic isn't it.

The same thing is true about F . F is computed directly from the data. In fact, the idea behind F is the same basic idea that goes into making t . Here is the general idea behind the formula, it is again a ratio of the effect we are measuring (in the numerator), and the variation associated with the effect (in the denominator).

$$\text{name of statistic} = \frac{\text{measure of effect}}{\text{measure of error}}$$
$$F = \frac{\text{measure of effect}}{\text{measure of error}}$$

The difference with F , is that we use variances to describe both the measure of the effect and the measure of error. So, F is a ratio of two variances.

Remember what we said about how these ratios work. When the variance associated with the effect is the same size as the variance associated with sampling error, we will get two of the same numbers, this will result in an F -value of 1. When the variance due to the effect is larger than the variance associated with sampling error, then F will be greater than 1. When the variance associated with the effect is smaller than the variance associated with sampling error, F will be less than one.

Let's rewrite in plainer English. We are talking about two concepts that we would like to measure from our data. 1) A measure of what we can explain, and 2) a measure of error, or stuff about our data we can't explain. So, the F formula looks like this:

$$F = \frac{\text{Can Explain}}{\text{Can't Explain}}$$

When we can explain as much as we can't explain, $F = 1$. This isn't that great of a situation for us to be in. It means we have a lot of uncertainty. When we can explain much more than we can't we are doing a good job, F will be greater than 1. When we can explain less than what we can't, we really can't explain very much, F will be less than 1. That's the concept behind making F .

If you saw an F in the wild, and it was .6. Then you would automatically know the researchers couldn't explain much of their data. If you saw an F of 5, then you would know the researchers could explain 5 times more than the couldn't, that's pretty good. And the point of this is to give you an intuition about the meaning of an F -value, even before you know how to compute it.

Computing the F -value

Fisher's ANOVA is very elegant in my opinion. It starts us off with a big problem we always have with data. We have a lot of numbers, and there is a lot of variation in the numbers, what to do? Wouldn't it be nice to split up the variation into kinds, or sources. If we could know what parts of the variation were being caused by our experimental manipulation, and what parts were being caused by sampling error, we would be making really good progress. We would be able to know if our experimental manipulation was causing more change in the data than sampling error, or chance alone. If we could measure those two parts of the total variation, we could make a ratio, and then we would have an F value. This is what the ANOVA does. It splits the total variation in the data into two parts. The formula is:

Total Variation = Variation due to Manipulation + Variation due to sampling error

This is a nice idea, but it is also vague. We haven't specified our measure of variation. What should we use?

Remember the sums of squares that we used to make the variance and the standard deviation? That's what we'll use. Let's take another look at the formula, using sums of squares for the measure of variation:

$$SS_{\text{total}} = SS_{\text{Effect}} + SS_{\text{Error}}$$

SS Total

The total sums of squares, or SS_{Total} is a way of thinking about all of the variation in a set of data. It's pretty straightforward to measure. No tricky business. All we do is find the difference between each score and the grand mean, then we square the differences and add them all up.

Let's imagine we had some data in three groups, A, B, and C. For example, we might have 3 scores in each group. The data could look like this:

```
suppressPackageStartupMessages(library(dplyr))
scores <- c(20,11,2,6,2,7,2,11,2)
groups <- as.character(rep(c("A","B","C"), each=3))
diff <- scores - mean(scores)
diff_squared <- diff^2
df <- data.frame(groups,scores,diff, diff_squared)
df$groups <- as.character(df$groups)
df <- df %>%
  rbind(c("Sums",colSums(df[1:9,2:4]))) %>%
  rbind(c("Means",colMeans(df[1:9,2:4])))
knitr::kable(df)
```

run restart restart & run all

groups	scores	diff	diff_squared
A	20	13	169
A	11	4	16
A	2	-5	25
B	6	-1	1
B	2	-5	25
B	7	0	0
C	2	-5	25
C	11	4	16
C	2	-5	25
Sums	63	0	302
Means	7	0	33.5555555555556

The data is organized in long format, so that each row is a single score. There are three scores for the A, B, and C groups. The mean of all of the scores is called the Grand Mean. It's calculated in the table, the Grand Mean = 7.

We also calculated all of the difference scores from the Grand Mean. The difference scores are in the column titled `diff`. Next, we squared the difference scores, and those are in the next column called `diff_squared`.

Remember, the difference scores are a way of measuring variation. They represent how far each number is from the Grand Mean. If the Grand Mean represents our best guess at summarizing the data, the difference scores represent the error between the guess and each actual data point. The only problem with the difference scores is that they sum to zero (because the mean is the balancing point in the data). So, it is convenient to square the difference scores, this turns all of them into positive numbers. The size of the squared difference scores still represents error between the mean and each score. And, the squaring operation exacerbates the differences as the error grows larger (squaring a big number makes a really big number, squaring a small number still makes a smallish number).

OK fine! We have the squared deviations from the grand mean, we know that they represent the error between the grand mean and each score. What next? SUM THEM UP!

When you add up all of the individual squared deviations (difference scores) you get the sums of squares. That's why it's called the sums of squares (SS).

Now, we have the first part of our answer:

$$SS_{\text{total}} = SS_{\text{Effect}} + SS_{\text{Error}}$$

$$SS_{\text{total}} = 302$$

and

$$302 = SS_{\text{Effect}} + SS_{\text{Error}}$$

What next? If you think back to what you learned about algebra, and solving for X, you might notice that we don't really need to find the answers to both missing parts of the equation. We only need one, and we can solve for the other. For example, if we found SS_{Effect} , then we could solve for SS_{Error} .

SS Effect

SS_{Total} gave us a number representing all of the change in our data, how all the scores are different from the grand mean.

What we want to do next is estimate how much of the total change in the data might be due to the experimental manipulation. For example, if we ran an experiment that causes change in the measurement, then the means for each group will be different from other. As a result, the manipulation forces change onto the numbers, and this will naturally mean that some part of the total variation in the numbers is caused by the manipulation.

The way to isolate the variation due to the manipulation (also called effect) is to look at the means in each group, and calculate the difference scores between each group mean and the grand mean, and then sum the squared deviations to find SS_{Effect} .

Consider this table, showing the calculations for SS_{Effect} .

```
suppressPackageStartupMessages(library(dplyr))
scores <- c(20,11,2,6,2,7,2,11,2)
means <- c(11,11,11,5,5,5,5,5,5)
groups <- as.character(rep(c("A","B","C"), each=3))
diff <- means - mean(scores)
diff_squared <- diff^2
df <- data.frame(groups,scores,means,diff, diff_squared)
df$groups <- as.character(df$groups)
df <- df %>%
  rbind(c("Sums",colSums(df[1:9,2:5]))) %>%
  rbind(c("Means",colMeans(df[1:9,2:5])))
knitr::kable(df)
```

run restart restart & run all

groups	scores	means	diff	diff_squared
A	20	11	4	16
A	11	11	4	16
A	2	11	4	16
B	6	5	-2	4
B	2	5	-2	4
B	7	5	-2	4
C	2	5	-2	4
C	11	5	-2	4
C	2	5	-2	4

groups	scores	means	diff	diff_squared
Sums	63	63	0	72
Means	7	7	0	8

Notice we created a new column called `means`. For example, the mean for group A was 11. You can see there are three 11s, one for each observation in row A. The means for group B and C happen to both be 5. So, the rest of the numbers in the means column are 5s.

What we are doing here is thinking of each score in the data from the viewpoint of the group means. The group means are our best attempt to summarize the data in those groups. From the point of view of the mean, all of the numbers are treated as the same. The mean doesn't know how far off it is from each score, it just knows that all of the scores are centered on the mean.

Let's pretend you are the mean for group A. That means you are an 11. Someone asks you "hey, what's the score for the first data point in group A?". Because you are the mean, you say, I know that, it's 11. "What about the second score?"...it's 11... they're all 11, so far as I can tell..."Am I missing something..."", asked the mean.

Now that we have converted each score to it's mean value we can find the differences between each mean score and the grand mean, then square them, then sum them up. We did that, and found that the $SS_{\text{Effect}} = 72$.

SS_{Effect} represents the amount of variation that is caused by differences between the means. I also refer to this as the amount of variation that the researcher can explain (by the means, which represent differences between groups or conditions that were manipulated by the researcher).

Notice also that $SS_{\text{Effect}} = 72$, and that 72 is smaller than $SS_{\text{total}} = 302$. That is very important. SS_{Effect} by definition can never be larger than SS_{total} .

SS Error

Great, we made it to SS Error. We already found SS Total, and SS Effect, so now we can solve for SS Error just like this:

$$SS_{\text{total}} = SS_{\text{Effect}} + SS_{\text{Error}}$$

switching around:

$$SS_{\text{Error}} = SS_{\text{total}} - SS_{\text{Effect}}$$

$$SS_{\text{Error}} = 302 - 72 = 230$$

We could stop here and show you the rest of the ANOVA, we're almost there. But, the next step might not make sense unless we show you how to calculate SS_{Error} directly from the data, rather than just solving for it. We should do this just to double-check our work anyway.

```
suppressPackageStartupMessages(library(dplyr))
scores <- c(20,11,2,6,2,7,2,11,2)
means <-c(11,11,11,5,5,5,5,5,5)
groups <- as.character(rep(c("A","B","C"), each=3))
diff <-means-scores
diff_squared <-diff^2
df<-data.frame(groups,scores,means,diff, diff_squared)
df$groups<-as.character(df$groups)
df <- df %>%
  rbind(c("Sums",colSums(df[1:9,2:5]))) %>%
  rbind(c("Means",colMeans(df[1:9,2:5])))
knitr::kable(df)
```

run restart restart & run all

groups	scores	means	diff	diff_squared
--------	--------	-------	------	--------------

groups	scores	means	diff	diff_squared
A	20	11	-9	81
A	11	11	0	0
A	2	11	9	81
B	6	5	-1	1
B	2	5	3	9
B	7	5	-2	4
C	2	5	3	9
C	11	5	-6	36
C	2	5	3	9
Sums	63	63	0	230
Means	7	7	0	25.5555555555556

Alright, we did almost the same thing as we did to find SS_{Effect} . Can you spot the difference? This time for each score we first found the group mean, then we found the error in the group mean estimate for each score. In other words, the values in the *diff* column are the differences between each score and it's group mean. The values in the *diff_squared* column are the squared deviations. When we sum up the squared deviations, we get another Sums of Squares, this time it's the SS_{Error} . This is an appropriate name, because these deviations are the ones that the group means can't explain!

Degrees of freedom

Degrees of freedom come into play again with ANOVA. This time, their purpose is a little bit more clear. *Dfs* can be fairly simple when we are doing a relatively simple ANOVA like this one, but they can become complicated when designs get more complicated.

Let's talk about the degrees of freedom for the SS_{Effect} and SS_{Error} .

The formula for the degrees of freedom for SS_{Effect} is

$df_{\text{Effect}} = \text{Groups} - 1$, where Groups is the number of groups in the design.

In our example, there are 3 groups, so the df is $3 - 1 = 2$. You can think of the df for the effect this way. When we estimate the grand mean (the overall mean), we are taking away a degree of freedom for the group means. Two of the group means can be anything they want (they have complete freedom), but in order for all three to be consistent with the Grand Mean, the last group mean has to be fixed.

The formula for the degrees of freedom for SS_{Error} is

$df_{\text{Error}} = \text{scores} - \text{groups}$, or the number of scores minus the number of groups. We have 9 scores and 3 groups, so our *df* for the error term is $9 - 3 = 6$. Remember, when we computed the difference score between each score and its group mean, we had to compute three means (one for each group) to do that. So, that reduces the degrees of freedom by 3. 6 of the difference scores could be anything they want, but the last 3 have to be fixed to match the means from the groups.

Mean Squared Error

OK, so we have the degrees of freedom. What's next? There are two steps left. First we divide the *SSes* by their respective degrees of freedom to create something new called Mean Squared Error. Let's talk about why we do this.

First of all, remember we are trying to accomplish this goal:

$$F = \frac{\text{measure of effect}}{\text{measure of error}}$$

We want to build a ratio that divides a measure of an effect by a measure of error. Perhaps you noticed that we already have a measure of an effect and error! How about the SS_{Effect} and SS_{Error} . They both represent the variation due to the effect, and the leftover variation that is unexplained. Why don't we just do this?

$$\frac{SS_{\text{Effect}}}{SS_{\text{Error}}}$$

Well, of course you could do that. What would happen is you can get some really big and small numbers for your inferential statistic. And, the kind of number you would get wouldn't be readily interpretable like a t value or a z score.

The solution is to normalize the SS terms. Don't worry, normalize is just a fancy word for taking the average, or finding the mean. Remember, the SS terms are all sums. And, each sum represents a different number of underlying properties.

For example, the SS_{Effect} represents the sum of variation for three means in our study. We might ask the question, well, what is the average amount of variation for each mean...You might think to divide SS_{Effect} by 3, because there are three means, but because we are estimating this property, we divide by the degrees of freedom instead ($\# \text{ groups} - 1 = 3 - 1 = 2$). Now we have created something new, it's called the MSE_{Effect} .

$$MSE_{\text{Effect}} = \frac{SS_{\text{Effect}}}{df_{\text{Effect}}}$$

$$MSE_{\text{Effect}} = \frac{72}{2} = 36$$

This might look alien and seem a bit complicated. But, it's just another mean. It's the mean of the sums of squares for the effect. If this reminds you of the formula for the variance, good memory. The MSE_{Effect} is a measure variance for the change in the data due to changes in the means (which are tied to the experimental conditions).

The SS_{Error} represents the sum of variation for nine scores in our study. That's a lot more scores, so the SS_{Error} is often way bigger than than SS_{Effect} . If we left our SS es this way and divided them, we would almost always get numbers less than one, because the SS_{Error} is so big. What we need to do is bring it down to the average size. So, we might want to divide our SS_{Error} by 9, after all there were nine scores. However, because we are estimating this property, we divide by the degrees of freedom instead ($\text{scores} - \text{groups} = 9 - 3 = 6$). Now we have created something new, it's called the MSE_{Error} .

$$MSE_{\text{Error}} = \frac{SS_{\text{Error}}}{df_{\text{Error}}}$$

$$MSE_{\text{Error}} = \frac{230}{6} = 38.33$$

Calculate F

Now that we have done all of the hard work, calculating F is easy:

$$F = \frac{\text{measure of effect}}{\text{measure of error}}$$

$$F = \frac{MSE_{\text{Effect}}}{MSE_{\text{Error}}}$$

$$F = \frac{36}{38.33} = .939$$

Done!

The ANOVA TABLE

You might suspect we aren't totally done here. We've walked through the steps of computing F . Remember, F is a sample statistic, we computed F directly from the data. There were a whole bunch of pieces we needed, the dfs , the SS es, the MSE s, and then finally the F .

All of these little pieces are conveniently organized by ANOVA tables. ANOVA tables look like this:

```
library(xtable)
suppressPackageStartupMessages(library(dplyr))
scores <- c(20,11,2,6,2,7,2,11,2)
means <- c(11,11,11,5,5,5,5,5,5)
groups <- as.character(rep(c("A","B","C"), each=3))
diff <- means - scores
```

```
diff_squared <- diff^2
df<-data.frame(groups,scores,means,diff, diff_squared)
df$groups<-as.character(df$groups)
df <- df %>%
  rbind(c("Sums",colSums(df[1:9,2:5]))) %>%
  rbind(c("Means",colMeans(df[1:9,2:5])))

aov_out<-aov(scores~ groups, df[1:9,])
summary_out<-summary(aov_out)
knitr::kable(xtable(summary_out))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
groups	2	72	36.00000	0.9391304	F)" style="vertical-align:middle;" class="lt-stats-7927">0.4417359
Residuals	6	230	38.33333	NA	F)" style="vertical-align:middle;" class="lt-stats-7927">NA

You are looking at the print-out of an ANOVA summary table from R. Notice, it had columns for *Df*, *SS* (Sum Sq), *MSE* (Mean Sq), *F*, and a *p*-value. There are two rows. The **groups** row is for the Effect (what our means can explain). The **Residuals** row is for the Error (what our means can't explain). Different programs give slightly different labels, but they are all attempting to present the same information in the ANOVA table. There isn't anything special about the ANOVA table, it's just a way of organizing all the pieces. Notice, the MSE for the effect (36) is placed above the MSE for the error (38.333), and this seems natural because we divide 36/38.33 in or to get the *F*-value!

This page titled 7.2: One-factor ANOVA is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Matthew J. C. Crump via source content that was edited to the style and standards of the LibreTexts platform.