

7.4: ANOVA on Real Data

We've covered many fundamentals about the ANOVA, how to calculate the necessary values to obtain an F -statistic, and how to interpret the F -statistic along with its associated p -value once we have one. In general, you will be conducting ANOVAs and playing with F 's and p 's using software that will automatically spit out the numbers for you. It's important that you understand what the numbers mean, that's why we've spent time on the concepts. We also recommend that you try to compute an ANOVA by hand at least once. It builds character, and let's you know that you know what you are doing with the numbers.

But, we've probably also lost the real thread of all this. The core thread is that when we run an experiment we use our inferential statistics, like ANOVA, to help us determine whether the differences we found are likely due to chance or not. In general, we like to find out that the differences that we find are not due to chance, but instead due to our manipulation.

So, we return to the application of the ANOVA to a real data set with a real question. This is the same one that you will be learning about in the lab. We give you a brief overview here so you know what to expect.

Tetris and bad memories

Yup, you read that right. The research you will learn about tests whether playing Tetris after watching a scary movie can help prevent you from having bad memories from the movie (James et al. 2015). Sometimes in life people have intrusive memories, and they think about things they'd rather not have to think about. This research looks at one method that could reduce the frequency of intrusive memories.

Here's what they did. Subjects watched a scary movie, then at the end of the week they reported how many intrusive memories about the movie they had. The mean number of intrusive memories was the measurement (the dependent variable). This was a between-subjects experiment with four groups. Each group of subjects received a different treatment following the scary movie. The question was whether any of these treatments would reduce the number of intrusive memories. All of these treatments occurred after watching the scary movie:

1. No-task control: These participants completed a 10-minute music filler task after watching the scary movie.
2. Reactivation + Tetris: These participants were shown a series of images from the trauma film to reactivate the traumatic memories (i.e., reactivation task). Then, participants played the video game Tetris for 12 minutes.
3. Tetris Only: These participants played Tetris for 12 minutes, but did not complete the reactivation task.
4. Reactivation Only: These participants completed the reactivation task, but did not play Tetris.

For reasons we elaborate on in the lab, the researchers hypothesized that the **Reactivation+Tetris** group would have fewer intrusive memories over the week than the other groups.

Let's look at the findings. Note you will learn how to do all of these steps in the lab. For now, we just show the findings and the ANOVA table. Then we walk through how to interpret it.

```

ta.table)
plot2)
packageStartupMessages(library(dplyr))
- fread(
  /stats.libretexts.org/@api/deki/files/10605/Jamesetal2015Experiment2.csv")
condition <- as.factor(all_data$Condition)
all_data$Condition) <- c("Control",
  "Reactivation+Tetris",
  "Tetris_only",
  "Reactivation_only")

# and SEs
all_df <- all_data %>%
  dplyr::group_by(Condition) %>%
  dplyr::summarise(means= mean(Days_One_to_Seven_Number_of_Intrusions),
    SEs = sd(Days_One_to_Seven_Number_of_Intrusions)/sqrt(length(Days_One_to_Seven_Number_of_Int
    rusions)))

# plot
ggplot(all_df, aes(x=Condition, y=means))+
  (stat="identity", aes(fill=Condition))+ # add means
  geom_bar(aes(ymin=means-SEs, # add error bars
    ymax=means+SEs), width=.1) +
  geom_point(data=all_data, aes(x=Condition, y=Days_One_to_Seven_Number_of_Intrusions), alpha=.5)+
  theme(alpha=.25)+
  ggtitle("Intrusive Memories (Mean for Week)")

```

restart & run all

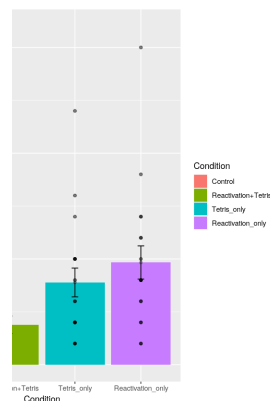


Figure 7.4.2: Mean number of intrusive memories per week as a function of experimental treatments.

OOooh, look at that. We did something fancy. You are looking at the the data from the four groups. The height of each bar shows the mean intrusive memories for the week. The dots show the individual scores for each subject in each group (useful to the spread of the data). The error bars show the standard errors of the mean. What can we see here? Right away it looks like there is some support for the research hypothesis. The green bar, for the Reactivation + Tetris group had the lowest mean number of intrusive memories. Also, the error bar is not overlapping with any of the other error bars. This implies that the mean for the Reactivation + Tetris group is different from the means for the other groups. And, this difference is probably not very likely by chance.

We can now conduct the ANOVA on the data to ask the omnibus question. If we get an F -value with an associated p -value of less than .05 (the alpha criterion set by the authors), then we can reject the hypothesis of no differences. Let's see what happens:

```

library(data.table)
library(xtable)
all_data <- fread(
  "https://stats.libretexts.org/@api/deki/files/10605/Jamesetal2015Experiment2.csv")
all_data$Condition <- as.factor(all_data$Condition)
levels(all_data$Condition) <- c("Control",
  "Reactivation+Tetris",
  "Tetris_only",
  "Reactivation_only")

aov_out<-aov(Days_One_to_Seven_Number_of_Intrusions ~ Condition, all_data)
summary_out<-summary(aov_out)
knitr::kable(xtable(summary_out))

```

run restart restart & run all

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Condition	3	114.8194	38.27315	3.794762	F)" style="vertical-align:middle;">0.0140858
Residuals	68	685.8333	10.08578	NA	F)" style="vertical-align:middle;">NA

We see the ANOVA table, it's up there. We could report the results from the ANOVA table like this:

There was a significant main effect of treatment condition, $F(3, 68) = 3.79$, $MSE = 10.08$, $p=0.014$.

We called this a significant effect because the p -value was less than 0.05. In other words, the F -value of 3.79 only happens 1.4% of the time when the null is true. Or, the differences we observed in the means only occur by random chance (sampling error) 1.4% of the time. Because chance rarely produces this kind of result, the researchers made the inference that chance DID NOT produce their differences, instead, they were inclined to conclude that the Reactivation + Tetris treatment really did cause a reduction in intrusive memories. That's pretty neat.

Comparing means after the ANOVA

Remember that the ANOVA is an omnibus test, it just tells us whether we can reject the idea that all of the means are the same. The F-test (synonym for ANOVA) that we just conducted suggested we could reject the hypothesis of no differences. As we discussed before, that must mean that there are some differences in the pattern of means.

Generally after conducting an ANOVA, researchers will conduct follow-up tests to compare differences between specific means. We will talk more about this practice throughout the textbook. There are many recommended practices for follow-up tests, and there is a lot of debate about what you should do. We are not going to wade into this debate right now. Instead we are going to point out that you need to do something to compare the means of interest after you conduct the ANOVA, because the ANOVA is just the beginning...It usually doesn't tell you what you want to know. You might wonder why bother conducting the ANOVA in the first place...Not a terrible question at all. A good question. You will see as we talk about more complicated designs, why ANOVAs are so useful. In the present example, they are just a common first step. There are required next steps, such as what we do next.

How can you compare the difference between two means, from a between-subjects design, to determine whether or not the difference you observed is likely or unlikely to be produced by chance? We covered this one already, it's the independent t -test. We'll do a couple t -tests, showing the process.

Control vs. Reactivation+Tetris

What we really want to know is if Reactivation+Tetris caused fewer intrusive memories...but compared to what? Well, if it did something, the Reactivation+Tetris group should have a smaller mean than the Control group. So, let's do that comparison:

```
library(data.table)
library(ggplot2)
suppressPackageStartupMessages(library(dplyr))
all_data <- fread(
  "https://stats.libretexts.org/@api/deki/files/10605/Jamesetal2015Experiment2.csv")
all_data$Condition <- as.factor(all_data$Condition)
levels(all_data$Condition) <- c("Control",
  "Reactivation+Tetris",
  "Tetris_only",
  "Reactivation_only")

comparison_df <- all_data %>%
  filter(Condition %in% c('Control', 'Reactivation+Tetris')==TRUE)
t.test(Days_One_to_Seven_Number_of_Intrusions ~ Condition,
  comparison_df,
  var.equal=TRUE)
```

run

restart

restart & run all

Two Sample t-test

```
data: Days_One_to_Seven_Number_of_Intrusions by Condition
t = 2.9893, df = 34, p-value = 0.005167
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.031592 5.412852
sample estimates:
    mean in group Control mean in group Reactivation+Tetris
      5.111111           1.888889
```

We found that there was a significant difference between the control group ($M=5.11$) and Reactivation + Tetris group ($M=1.89$), $t(34) = 2.99$, $p=0.005$.

Above you just saw an example of reporting another t -test. This sentences does an OK job of telling the reader everything they want to know. It has the means for each group, and the important bits from the t -test.

More important, as we suspected the difference between the control and Reactivation + Tetris group was likely not due to chance.

Control vs. Tetris_only

Now we can really start wondering what caused the difference. Was it just playing Tetris? Does just playing Tetris reduce the number of intrusive memories during the week? Let's compare that to control:

```
library(data.table)
suppressPackageStartupMessages(library(dplyr))
all_data <- fread(
  "https://stats.libretexts.org/@api/deki/files/10605/Jamesetal2015Experiment2.csv")
all_data$Condition <- as.factor(all_data$Condition)
levels(all_data$Condition) <- c("Control",
  "Reactivation+Tetris",
  "Tetris_only",
  "Reactivation_only")

comparison_df <- all_data %>%
  filter(Condition %in% c('Control', 'Tetris_only')==TRUE)
t.test(Days_One_to_Seven_Number_of_Intrusions ~ Condition,
  comparison_df,
  var.equal=TRUE)
```

[run](#)[restart](#)[restart & run all](#)

Two Sample t-test

```
data: Days_One_to_Seven_Number_of_Intrusions by Condition
t = 1.0129, df = 34, p-value = 0.3183
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.230036  3.674480
sample estimates:
 mean in group Control mean in group Tetris_only
      5.111111          3.888889
```

Here we did not find a significant difference. We found that no significant difference between the control group ($M=5.11$) and Tetris Only group ($M=3.89$), $t(34) = 2.99$, $p=0.318$.

So, it seems that not all of the differences between our means are large enough to be called statistically significant. In particular, the difference here, or larger, happens by chance 31.8% of the time.

You could go on doing more comparisons, between all of the different pairs of means. Each time conducting a t -test, and each time saying something more specific about the patterns across the means than you get to say with the omnibus test provided by the ANOVA.

Usually, it is the pattern of differences across the means that you as a researcher are primarily interested in understanding. Your theories will make predictions about how the pattern turns out (e.g., which specific means should be higher or lower and by how much). So, the practice of doing comparisons after an ANOVA is really important for establishing the patterns in the means.

This page titled 7.4: ANOVA on Real Data 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.