Trick or Treat: Using ANOVA to help your kids get the most candy

**PLEASE NOTE:** For the in-class discuss of this exercise, you’ll be uploading your R Markdown file to Canvas and another student will download it and try to knit it. Your classmate will then provide you feedback on your work. Keep in mind that another person on another computer should be able to run your code. Include sufficient comments to clarify what you are doing in case they get stuck, load the packages you used directly in the code, include a note about which packages you need at the top of the file, etc.

ANOVA is often the method of choice for experimental data where people are placed in different groups with different interventions and we measure some outcome. While t-tests are useful in experiments with only two conditions and a continuous outcome, each t-test has a chance of error associated with it (α). The overall *familywise* or *experimentwise* α can become too high if t-tests are used across three or more groups.

ANOVA is useful in experiments (or other data situations) where there are more than two conditions/groups and a continuous outcome. For example, you could use ANOVA to compare average BMI across four groups in a weight-loss study: (Group 1) an increased physical activity group, (Group 2) a decreased calorie group, (Group 3) a group that increased physical activity and decreased calories, and (Group 4) a control group that did nothing.

**Today’s experiment:** It’s Halloween and your kids (or your nieces, nephews, neighbors, etc) are in a contest to see who can collect the most individual pieces of candy (not the biggest, but the most). You saw your neighbors buying three different types of M&Ms to hand out and you want to give the kids the best chance to win by advising them which type of candy to choose.

The candy you saw your neighbors buying were:

1. Regular M&Ms
2. Peanut M&Ms
3. Caramel M&Ms

To help your kids get more than the other kids and to figure out which to pick:

1. Measure and record the total number of pieces of candy in your packages. Write your numbers on the board along with the type of M&Ms. Count *at least* 3 bags per person.
2. Develop an ANOVA model to determine if there is a significant difference in average number of pieces total by M&M type.
3. If there is a significant difference in number of pieces across types, conduct post-hoc tests and planned comparisons to determine where the significant differences lie (i.e., which of the three types of candy has significantly different number of pieces from other types)?

**Step 1: Collect and enter the data**

Open and count your candy. Refrain from eating any until the class data has all been recorded on the board…

Bag 1

Type:

Number:

Blue number:

Bag 2

Type:

Number:

Blue number:

Bag 3

Type:

Number:

Blue number:

Bag 4 (optional)

Type:

Number:

Blue number:

**Write your data on the board. Work with a partner to enter the class data into R Markdown using the R Markdown template in GitHub for this week. Check the basketball workshop for instructions if needed for making a data frame.**

**Step 2: Why not use a bunch of t-tests?**

T-tests are used to compare the means of two groups. The results of a t-test tell you whether the means are significantly different from one another. Why not use this strategy for the candy data? Compare mean number of M&Ms for Plain to Peanut, compare Peanut M&Ms to Caramel M&Ms, and Plain M&Ms to Caramel M&Ms….

*The problem is that* ***each t-test*** *comes with an α, or a small probability that we are wrong when we reject the null hypothesis.*

*If you combine these small probabilities across many t-tests, you get a* ***much larger probability*** *that you are wrong somewhere along the way.*

*This larger probability is called the experimentwise or familywise error.*

Use the following formula to determine what your familywise error would be for the candy data, if you chose to use t-tests.

1. Calculate the number of t-tests by substituting the number of groups (in this case 4) into the following equation (k represents the number of groups):
2. Substitute the number of t-tests from above into the familywise error rate equation:

Fill in the interpretation:

The familywise error rate for comparing 3 groups using a set of t-tests is \_\_\_\_\_\_\_\_\_, indicating a \_\_\_\_\_\_\_\_\_\_% chance of being wrong somewhere in this set of analyses.

**Step 3: Writing hypotheses and exploring the data**

As is customary before developing a complex model, examine your data to determine what you might find. An ANOVA model is designed to determine whether there is a difference among means of several groups. First, write your null and alternate hypotheses:

**H0: The mean number of candy pieces per package is equal across different candy types.**

**HA: Not all means are equal.**

A boxplot and descriptive statistics can aid in comparing the distribution of data in each of the groups.

* Finish the ggplot() code in the Markdown file to make a boxplot with candy type on the x-axis and number of M&Ms on the y-axis.
  + Choose one of the color palettes from the list at <https://ggplot2.tidyverse.org/reference/scale_brewer.html> and add the name of the palette you choose inside the quotes of the scale\_fill\_brewer layer of your graph
  + Choose a graph theme from <https://ggplot2.tidyverse.org/reference/ggtheme.html> and add it to your plot
  + Give the graph a title by adding text describing the plot between the quote marks of the subtitle = “” argument
  + After examining the plot, write a comment starting with # in the code chunk about any differences you saw among the means
* Fill in the CreateTableOne() code to make a table of means by group. Use the type of M&Ms as the strata variable to see means by group.

**Step 4: Checking the assumptions for ANOVA**

There are 3 assumptions for ANOVA:

1. Independence of observations
   1. Is each bag of M&Ms independent of the other bags?
2. Normal distribution of the outcome *within groups*
   1. Fill in the ggplot() code for the density plot with the palette and theme you are using and run the code to check visually if groups are normally distributed. Use a comment starting with # inside the code chunk to make a note about the whether or not the assumption was met.
3. Equal variances *within groups* 
   1. Levene’s Test is widely used to test the assumption of equal variances. The null hypothesis is that the variances are equal while the alternate is that at least two of the variances are different. The leveneTest() function can be used to conduct the Levene’s Test. A p-value less than .05 rejects the null hypothesis of equal variance and indicates this assumption is *not met*. Use a comment starting with # inside the code chunk to make a note about the whether or not the assumption was met.

If your data met all three assumptions, you can report your model as unbiased, which means that it can be generalized to the population. If you did not meet all three assumptions, your model is biased and applies only to the sample examined in the study or you will need to choose a different model to generalize to the population.

**Step 5: Conducting ANOVA**

**[in this case continue whether-or-not you met assumptions, usually you’d stop and do something else if you failed one or more assumptions]**

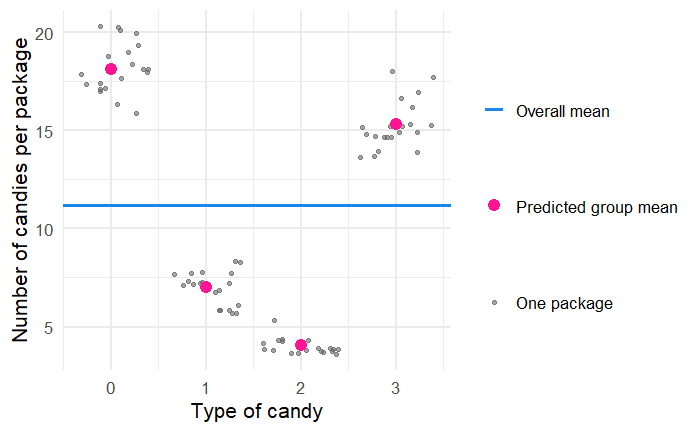
Run the code for the ANOVA in the Markdown file. As in linear regression, the F-test is used to test the null hypothesis for the full model. ANOVA is actually a *special case of linear regression where the predictors are categorical only*, so the F is computed the same way as it was for linear regression, as a ratio of explained to unexplained variance. The technical

Where:

* i is an individual
* n is the number of observations
* k is the number of groups
* y is the outcome variable

So, the numerator represents the summed difference between each group mean and the overall mean and the denominator represents the summed difference between each individual and their group mean. The ratio, then, represents the how much variation there is between the groups by how much variation there is within the groups.

The main difference between ANOVA and other linear regressions is that, instead of a slanted line through the data with a y-intercept and slope, there is a predicted mean for each group. The model does something like this:



Based on the F and its associated p-value in your ANOVA table from R, choose one of the statements below and type it *as a comment* into your document in the code chunk with the ANOVA code, include all the relevant numbers from the ANOVA.

Based on a significant F-test [F(\_\_,\_\_)=\_\_\_; p <.05)], I reject the null hypothesis. Not all the candy types have the same mean number of pieces per package.

Based on a non-significant F-test [F(\_\_,\_\_)=\_\_\_; p=n.s.)], I fail to reject the null hypothesis. The mean number of pieces are not significantly different across candy types.

So….that’s it…..that’s essentially ANOVA. ANOVA is an **omnibus** test like chi-squared, where all you end up knowing is that there is a significant difference among the means….somewhere….but you don’t know where.

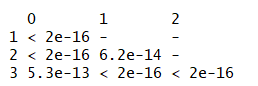
**Step 6: After ANOVA**

**[If your ANOVA were non-significant you’d stop here, in this case try the next section regardless of whether the F statistic was significant or not]**

If your ANOVA is significant, there are two ways to find out where the significant differences lie:

1. **Post-hoc tests**: Testing each pair of groups to see if they are different, like a set of t-tests.
2. **Planned comparisons**: Pre-planned comparisons of groups or sets of groups.

*Post-hoc tests* in ANOVA are used to compare each pair of groups using a t-test or something similar to a t-test but adjusting the p-value so that it is much more difficult to reach statistical significance. For example, the Bonferroni post-hoc test conducts a t-test for each pair of categories (e.g., plain vs. peanut, peanut vs. caramel, etc) and then multiplies the p-value from that t-test by the number of t-tests there are. So, in this case, there would be 3 t-tests to compare each pair of groups, so each p-value would be multiplied by 3. The output from Bonferroni is a matrix of p-values, like this:



The rows and columns will be named with the categories of your categorical variable so you can find the p-value that is, for example, from the t-test comparing plain to peanut. In the example output, you can find the Bonferroni adjusted p-value from a t-test of category 0 and category 3: 5.3e-13. This is still a super tiny p-value even though it has been multiplied by how ever many t-tests there were. So, this indicates that the mean of the outcome for category 0 is statistically significantly different from the mean of the outcome for category 3.

**Conduct a Bonferroni post-hoc test and write a comment in your Markdown file that briefly describes the results.**

*Planned comparisons* have many benefits and are generally considered a much better way to identify where the significant differences lie. To use planned comparisons, you’ll need to have some idea ahead of time of the groups you want to compare. For example, since plain M&Ms seem smaller than peanut or caramel, it might make sense to compare the mean from plain to the combined means of peanut and caramel together.

To write a planned comparison, you first need to know what order the categories of your variable are in since the comparison will use these categories to create the contrast.

The general rules:

1. The contrast should add up to 0 (we’ll use 2+(-1) +(-1) =0)
2. The groups being compared should be grouped by positives and negatives (2 is positive and the -1 are negative, so the groups compared are First category vs. Second and Third categories together)

|  |  |  |  |
| --- | --- | --- | --- |
|  | First category | Second category | Third category |
| Contrast 1 | 2 | -1 | -1 |

The way this would be entered into R would be:

contrast1 <- c(2, -1, -1)

This contrast is then linked to the categorical variable by the contrast() function:

contrasts(x = mm.data$type.mm) <- contrast1

Once this is done, re-run the ANOVA as shown in the Markdown file and then apply the contrast. The output will now give you statistical significance for the overall model AND for the contrast that compared the groups that you specified, like this:

Df Sum Sq Mean Sq F value Pr(>F)

type.mm 4 221301 55325 43.30 < 2e-16\*\*\*

type.mm: Plain vs. other 1 64411 64411 50.41 1.97e-12\*\*\*

Residuals 1404 1793757 1278

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

6 observations deleted due to missingness

Review the p-value as usual and report something like this *as a comment in the code chunk with the contrast*, but with your numbers:

A planned comparison comparing the mean number of candies for plain M&M type packages with the mean number of candies for the other two groups combined was statistically significant (F(1, 64411) = 50.41; p < .05).

When you report combined groups, it is good to be able to report descriptive statistics with the results to add some context. Yes, we know the means are different (or are not different), but what ARE the means? Which is bigger? Which is smaller? What the heck is going on?

You can do this using mutate to recode the type.mm variable combining categories and summarizing the results. Try it in your code. Be sure to change the labels so they make sense with the order of the categories in your data.

**After you examine the results, write a short report of the results of the study you just did under the “Study results” heading in the markdown file, the methods you used, and the results. Include at least:**

* Group means and standard deviations for number of candies per package for each of the three candy types (either in sentences or re-run the table code and add a table number/title to the table)
* Results of assumption checking (mention all three assumptions, how you checked, and whether the data met the assumption)
* ANOVA results (include F statistic with d.f. and p-value)
* Post-hoc results
* Planned comparison results including the means of the groups compared
* A graph showing the distribution of M&Ms in the three groups
* An evidence-based statement explaining which candy (or candies) you would tell your kids to pick in order to have the largest number of candy pieces.

**\*\*If you use code to include a table or figure in your report, use the “echo = FALSE” statement in the chunk set-up {r echo = FALSE} so that the code does not show, only the resulting table or figure.\*\***

**Knit your document to see the results! Your goal is a short report that reports the results of this study without any R code showing.**