# Complete Example (Between Subjects – Factorial)

New to this section:

* Levene’s Test: as with one-way ANOVAs, you will get a Levene’s test for homogeneity. Unlike Mauchly’s test for repeated measures, you get one test examining the variances by condition.
* Main effects: interpreting each IV on its own, ignoring the effect of the other IVs.
* Interaction: interpreting the IVs together, seeing if the conditions are significantly different OR if the pattern of data across levels is different for the other IV.

Chart of ANOVA Analysis:

|  |  |  |  |
| --- | --- | --- | --- |
|  | ANOVA | | |
|  | Main Effect | Main Effect | Interaction |
| If levels > 2  And significant | Independent t-test  Bonferroni correction | Independent t-test  Bonferroni correction | SPLIT one IV column  Independent t-test  Bonferroni correction |
| If levels = 2 | Interpret means | Interpret means |

If the interaction is significant, often people ignore any analyses with the main effects:

* This procedure reduces Type 1 error because you are running less post hoc tests.
* You are interested in the interaction anyway, so why only interpret one variable at a time?

We looked at two years worth of athletic spending data for four different sports. Are there differences across sports and years in spending?

**Datafile:** bn 2 anova.csv

**IVs:**

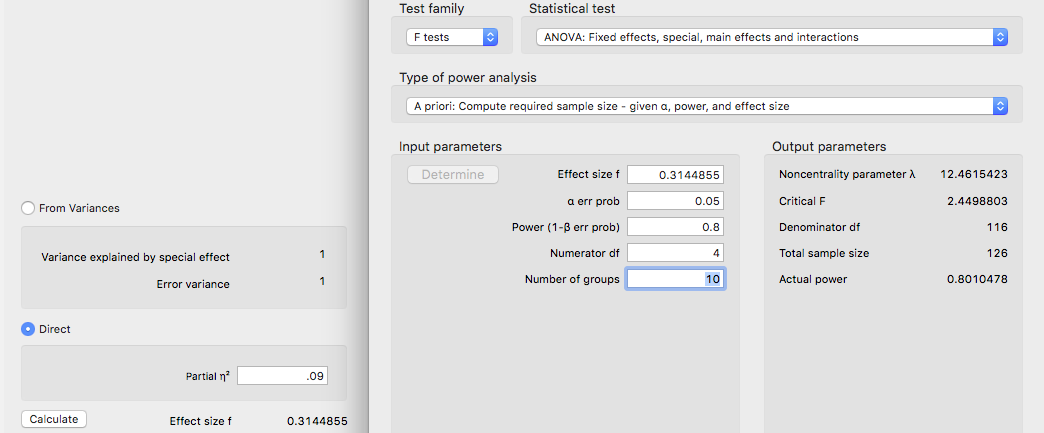
* Years:
  + Levels: 2007 versus 2008
* Type of Sport:
  + Levels: Basketball, baseball, volleyball, football, soccer

**DV:**

* Money: Spending transactions for each individual sport, represented in dollar amounts.

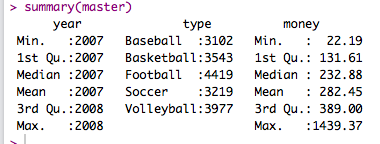
**Power:**

1. Open Gpower!
   1. Test family: F-test
   2. Statistical Test: ANOVA fixed effect, special, main effects and interactions
   3. Estimate an effect size: click determine 🡪 click direct 🡪 use eta square sizes you think might be accurate, remember small, medium, and large estimates from the notes.
   4. Alpha = .05
   5. Power (1-beta .20) = .80
   6. Numerator df:
      1. For main effects: number of levels – 1
      2. For interactions: (number of levels – 1)\*(number of levels – 1)
      3. Generally, you want to estimate based on interactions because that’s the purpose of the experiment.
   7. Number of groups:
      1. For the main effects: number of levels
      2. For the interactions: number of conditions
2. Let’s estimate the following:
   1. Medium effect size (eta = .09)
   2. Interaction component (2-1)\*(5-1)
   3. Conditions from our current study (2\*5)
3. Says we needed to run 126 people to find a significant effect with a small effect size.

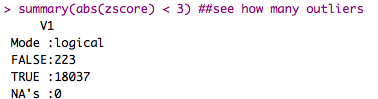


**Assumptions:**

1. Accuracy:
   1. Use the summary(*dataset name*) function to get the basic information for the data.
   2. Let’s check out minimum and maximum:
      1. This data should be factored and not go below zero.
      2. Just looking at the condition columns, we are ok because nothing is below zero or over 100.
         1. The min and max are ok, but we should factor the year column or we’ll have issues later.

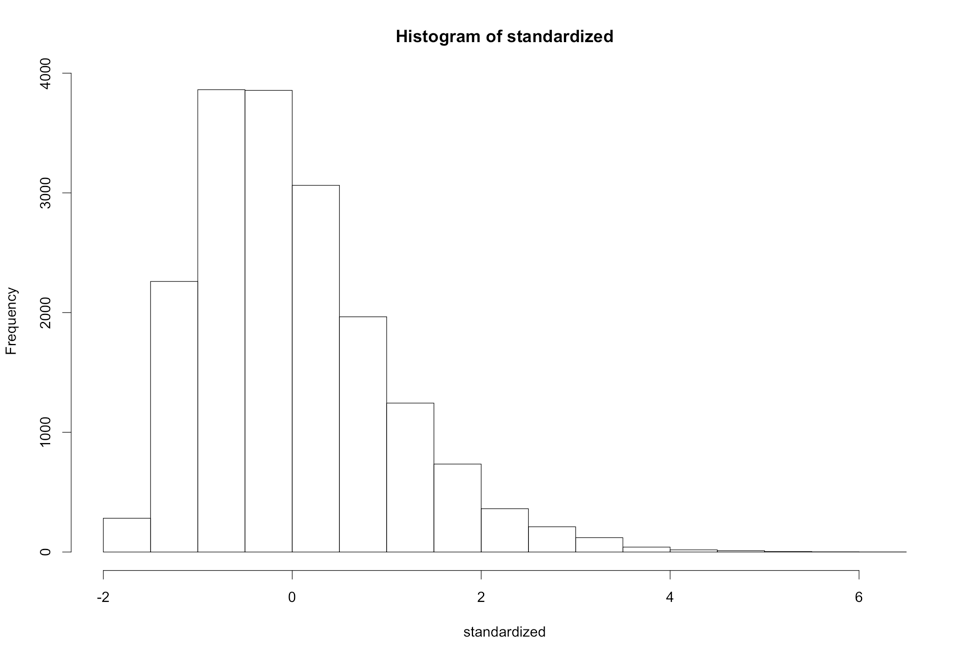


1. Missing:
   1. With the summary function, I can also see that I don’t have any missing data, because there are no NA values shown. Therefore, I can skip the missing data step.
   2. Even if there was missing data, remember that any missing data ends up being more than 5% for each participant in an ANOVA. Therefore, they should normally get excluded.
2. Outliers:
   1. One main issue with univariate (that means one DV) between-subjects ANOVA, is that we have only one column to screen for outliers since the rest are assigned by the experimenter (that’s the IV).
   2. Therefore, we cannot use Mahalanobis because it requires there be at least two columns.
   3. Instead, we will use the scale() function to create Z-scores.
      1. Code: zscore = scale(*dataset$column*)
      2. That will save you a set of z scores for the data. Z scores are the same idea as Mahalanobis – the distance from the center, but now they are the distance from the mean of the one variable you are working with.
   4. What cut off should you use? It’s always the absolute value of 3, which is the *p* < .001 cut off for z-scores.
   5. In this data set, we shouldn’t have any outliers, remember that FALSE is bad.
      1. summary(abs(zscores) < 3)
      2. We have 223 outliers, likely because the spending was much higher or much lower than the average.

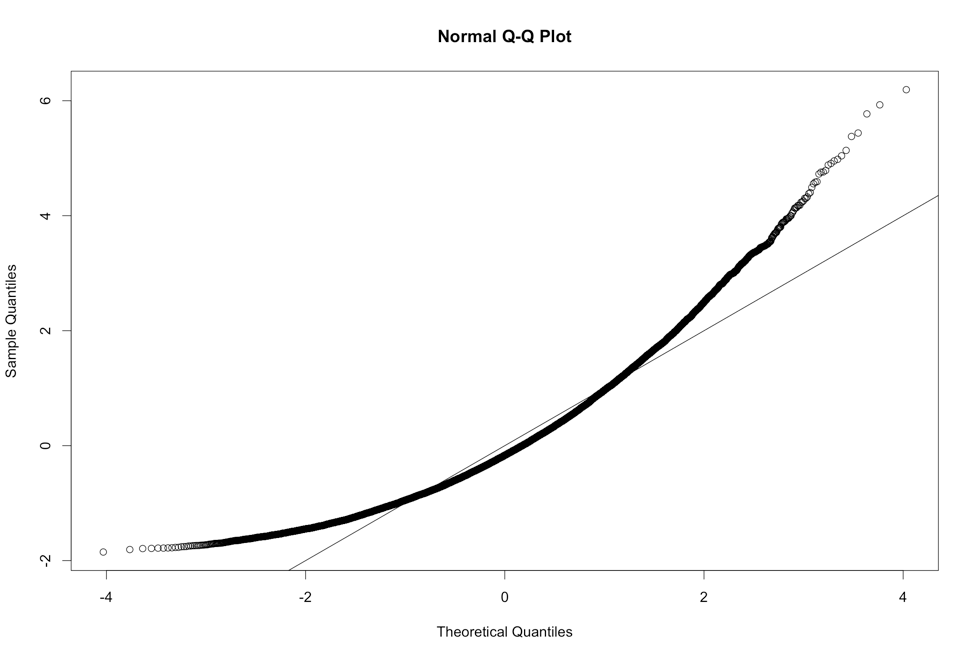


* 1. Exclude them in the same way we did for Mahalanobis.
     1. noout = subset(notypos, abs(zscore) < 3)
     2. The abs(*column*) creates the absolute value of z score, so you don’t forget to exclude the negatives.
     3. This code keeps all the people who are less than 3.

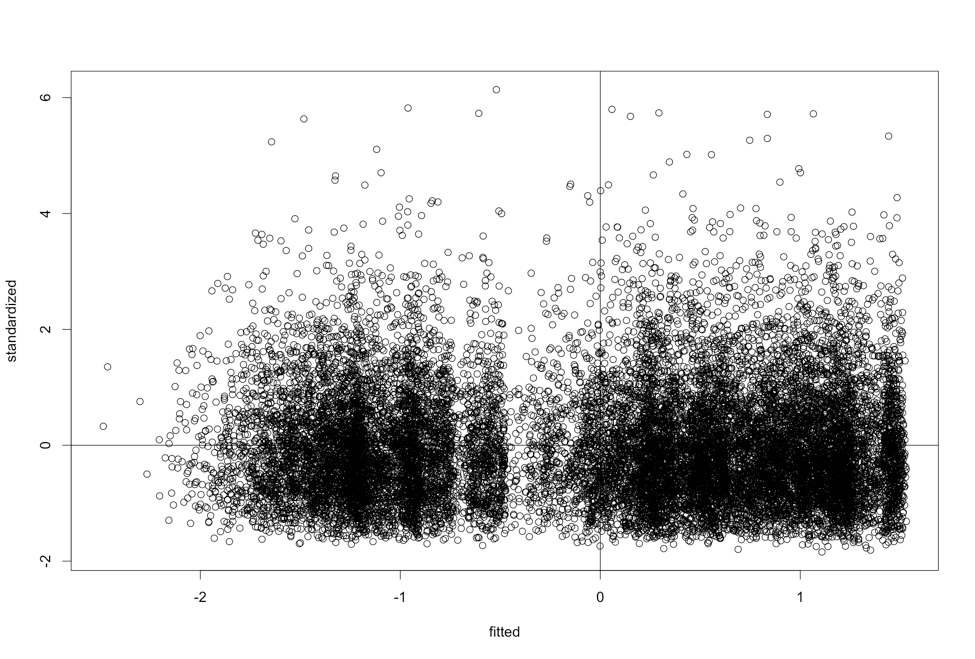
1. Additivity
   1. Not necessary, since we have only one continuous variable.
2. Set up the rest of the assumptions:
   1. Make a random variable:
      1. random = rchisq(nrow(*dataset*), 7)
   2. Run a fake regression:
      1. fake = lm(random~., data = *dataset*)
   3. Create the standardized residuals:
      1. standardized = rstudent(fake)
   4. Create the fitted values:
      1. fitted = scale(fake$fitted.values)
3. Normality:
   1. hist(standardized)
   2. Most of the data is between -2 and 2 and is centered over 0 – but there definitely is a skew to the distribution, even after taking out outliers.
   3. Because we have more than 30 transactions (“people”), we do not have to worry because of the central limit theorem.



* 1. Linearity:
     1. qqnorm(standardized)
     2. abline(0,1)
     3. Oh boy – this graph is bad.
        1. Test yourself – run it a couple times (by rerunning the random assumption set up section), you should get a bad graph every time.
     4. When you have non-linearity problems, you should switch to a non-parametric test, such as Freidman’s, Mann-Whitney U, or Kruskal-Wallis.



* 1. Homogeneity:
     1. plot(fitted,standardized)
     2. abline(0,0)
     3. abline(v = 0)
     4. Here the data is not homogeneous ... see how the top has a wider spread than the bottom (if you use vertical zero as the ruler). Horizontally, it’s ok … mostly between -2 and 2.
        1. Run the random set up section a couple times here to see a couple versions of this graph.
        2. We will also use Levene’s test to determine if it’s a problem.
     5. Now, most people do not talk about homoscedasticity for ANOVA, because homogeneity sort of equals homoscedasticity when one variable is categorical, and the other is continuous (aka the ANOVA set up).



* 1. Homogeneity: Take 2 Levene’s Test
     1. Levene’s is a test for homogeneity between groups, so it looks to see if the variances are equal across your IV levels.
     2. It is notoriously **oversensitive**, but can be a good place to start if you want to check a real number, rather than this scatterplot.
     3. With large sample sizes, it is often significant (remembering the big important rule, p<.001), and with large sample sizes it matters less. Ergo, if you have big *n* in each group, then don’t worry about it so much.
     4. You will have to run the ANOVA to get Levene’s Test, see below.

**Running the ANOVA:**

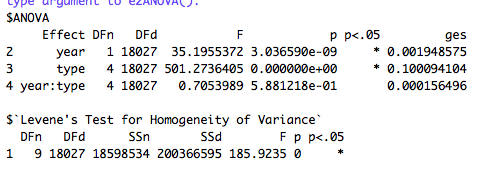
1. First, we must add a participant number to your data if it does not have one.
   1. The ez package requires a participant number, so we will have to add one.
   2. *dataset$partno* = 1:nrow(*dataset*)
2. Load the ez library.
   1. library(ez)
3. Run the ANOVA (all these lines):
   1. PS: This ANOVA is huge and will take some time to run.
   2. ezANOVA(data = *dataset*,

wid = partno,

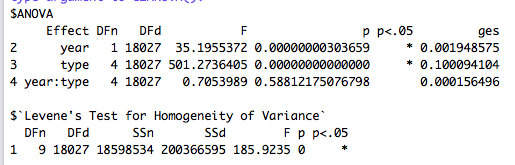
between = .(*column of IV1, column of IV2),*

dv = *column of DV*,

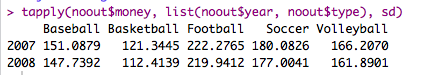
type = 3)



with scientific notation turned off: option(scipen = 999)



1. Interpret the output:
   1. Check Levene’s for Homogeneity – especially if your residual plot was not evenly distributed. You want p > .001.
   2. Look at the last number under *p*, it says 0, which is less than .001, so homogeneity is definitely bad.
      1. We could check out the standard deviations for each condition to see what the problem might be.
      2. tapply(*dataset$DV*, list(*dataset*$*IV, dataset$IV*), sd)



* + 1. The likely problem is that some conditions have low 100s for SD, while other’s have 200s.
       1. These variations are a good example of the problem with Levene’s test being oversensitive. We like to have lots of participants because it gives us more power to find an effect if it exists. However, it also gives Levene’s test more power to say the variances are different (so, it’s a catch-22). Generally, with large samples we tend not to worry much.
  1. Check the Omnibus (overall) test for your IVs:
     1. We have three of them! What happened?!
        1. You will get one *F* test for each IV and then also the interaction.
        2. You will interpret each one separately.
        3. Remember, if the interaction is significant, only do post hocs for the interaction.
     2. The DFn = df numerator or model.
     3. The DFd = df denominator or error.
     4. F = F
     5. p = p value.
     6. p < .05 helpfully tells you if it’s significant at *p* < .05, which is what we want to find.
     7. ges = generalized least squares or a form of η2.
     8. Write that up:
        1. Year: *F*(1, 18027) = 35.20, *p* <.001, η2 = .002.
        2. Type of Sport: *F*(4, 18027) = 501.27, *p* <.001, η2 = .10.
        3. Interaction: *F*(4, 18027) = .71, *p* = .59, η2 < .001.

1. Post Hoc Interpretation/Plan:
   1. To get the means and SDs, we can use tapply.
      1. tapply(*dataset$DV*, list(*dataset*$*IV, dataset$IV*), mean)
      2. tapply(*dataset$DV*, list(*dataset*$*IV, dataset$IV*), sd)
      3. tapply(*dataset$DV*, list(*dataset*$*IV, dataset$IV*), length)
      4. Remember, you can take out one of the IVs to just get main effects.
   2. Year is significant – remember it only has two levels, so I could just examine the means for those years to see what happened:

MEAN

2007 2008

281.0855 266.2025

SD

2007 2008

183.4575 179.2207

N

2007 2008

9437 8600

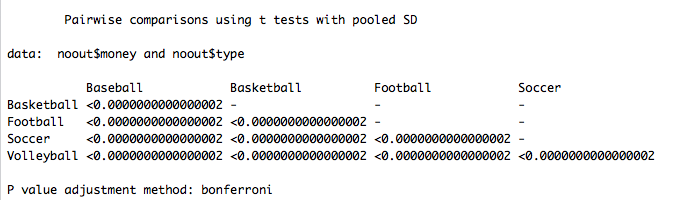
* + 1. What happened? It appears that spending was decreased per transaction for 2008 when compared to 2007.
  1. Sport is also significant:
     1. To analyze this section, I would have to run a post hoc test.
     2. Use the pairwise.t.test() function to run t.test you learned earlier on all groups at once.
        1. Remember, you use paired = F for **independent** t-tests, which is what we want to use for **between-subjects** ANOVA.
        2. We could use var.equal = F to correct for homogeneity, but with large samples, it’s not usually a big deal.
        3. p.adjust.method is the *correction*.
        4. pairwise.t.test(*dataset*$*DV*, *dataset*$*IV*,

paired = F,

var.equal = T,

p.adjust.method = "bonferroni")

* + - 1. Remember that Bonferroni changes the p values biased on the number of tests you are running. That’s good for us, because then we can use p<.05 again to determine if it is significant.
    1. I’ve given an example of how this analysis would work in the R script. You would create the same post hoc table we’ve used before to help you understand what’s going on.
       1. You should find that there are **TEN** post hoc tests, that are all significant.
       2. Football > Soccer > Volleyball > Baseball > Basketball when it comes to individual transactions.



MEANS

Baseball Basketball Football Soccer Volleyball

229.0071 191.7812 350.0400 312.3379 270.3164

* 1. Interaction: in this specific example, it is not significant, so I would not analyze it. However, this guide is to teach you how to analyze these things, so I have example of how it would be analyzed.
     1. We have 2X5 ANOVA – so we have ten boxes to consider:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Type of Sport | | | | |
|  |  | Baseball | Basketball | Football | Soccer | Volleyball |
| Year | 2007 |  |  |  |  |  |
| 2008 |  |  |  |  |  |

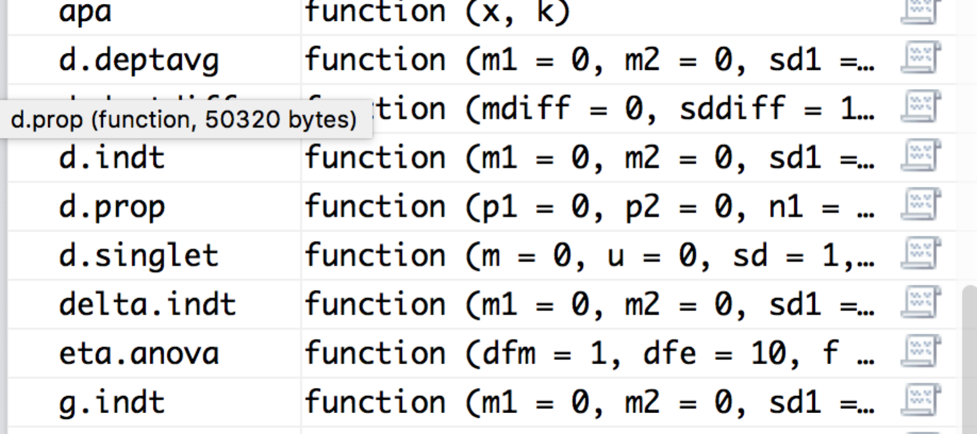
* + 1. How do we want to compare them?
    2. The rule usually is across or down but not both.
       1. If we went across, we would get something like this:
          1. For 2007 ONLY, baseball v basketball, baseball v football, baseball v soccer, etc.
          2. For 2008 ONLY, baseball v basketball, baseball v football, baseball v soccer, etc.
          3. That would be 20 tests! Ouch!
       2. If we went down, we would get:
          1. Baseball ONLY, 2007 v 2008.
          2. Basketball ONLY, 2007 v 2008.
          3. Etc.
          4. That’s five tests! Whew!
       3. Which way?
          1. First, pick based on a hypothesis – if you wanted to know who was spending the most each year, you’d have to do the first way.
          2. Second, go with the lesser number of tests to save type 1 error rate.
    3. Once you pick a direction, you will need to SPLIT the dataset into chunks to analyze each piece separately.
       1. Use the subset function!
       2. Be sure to look at your N values to make sure they changed – it should total up to the full number of rows in the last dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean 1 | Mean 2 | P-value | Explain? | Effect size |
| Baseball  2007  M = 234.76  SD = 151.09  N = 1617 | Baseball  2008  M = 222.73  SD = 147.74  N = 1483 |  |  |  |
| Basketball  2007  M = 197.58  SD = 121.34  N = 1866 | Basketball  2008  M = 185.33  SD = 112.41  N = 1677 |  |  |  |
| Football  2007  M = 360.77  SD = 222.28  N = 2190 | Football  2008  M = 338.72  SD = 219.94  N = 2075 |  |  |  |
| Soccer  2007  M = 320.98  SD = 180.08  N = 1676 | Soccer  2008  M = 302.54  SD = 177.00  N = 1478 |  |  |  |
| Volleyball  2007  M = 275.99  SD = 166.21  N = 2088 | Volleyball  2008  M = 264.03  SD = 161.89  N = 1887 |  |  |  |

* 1. Now, we have to calculate the *post hoc test* and *post hoc correction* to find out what’s going on.
  2. Use the Bonferroni output to fill in your p-values.
     1. Notice how they all show the same pattern – that’s a sign why the interaction is not significant. But we can check out the effect sizes to show that they are roughly equal … you can have interactions with the same pattern (i.e. all increasing or decreasing) but with very different effect sizes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean 1 | Mean 2 | P-value | Explain? | Effect size |
| Baseball  2007  M = 234.76  SD = 151.09  N = 1617 | Baseball  2008  M = 222.73  SD = 147.74  N = 1483 | .025 | Significant, spending decreased across years |  |
| Basketball  2007  M = 197.58  SD = 121.34  N = 1866 | Basketball  2008  M = 185.33  SD = 112.41  N = 1677 | .002 | Significant, spending decreased across years |  |
| Football  2007  M = 360.77  SD = 222.28  N = 2190 | Football  2008  M = 338.72  SD = 219.94  N = 2075 | .001 | Significant, spending decreased across years |  |
| Soccer  2007  M = 320.98  SD = 180.08  N = 1676 | Soccer  2008  M = 302.54  SD = 177.00  N = 1478 | .004 | Significant, spending decreased across years |  |
| Volleyball  2007  M = 275.99  SD = 166.21  N = 2088 | Volleyball  2008  M = 264.03  SD = 161.89  N = 1887 | .02 | Significant, spending decreased across years |  |

* 1. You can use MOTE to calculate the effect sizes OR the R script Dr. B just wrote!
     1. Load the effsize.R script and run the whole thing, so that you get new functions listed in the Environment window.



* 1. We will use d.indt for these calculations because it uses the numbers we have (m, n, sd) for each group.
     1. d.indt(m1 = #, m2 = #, sd1 = #, sd2 = #, n1 = #, n2 = #, a = .05, k = 2)
     2. Here’s the output you should get (remember you might get a warning sometimes – especially if N is large):

M1 = 234.76, SD = 151.09, SE = 3.76, 95%CI[227.39 - 242.13]

M2 = 222.73, SD = 147.74, SE = 3.84, 95%CI[215.20 - 230.26]

t(3098) = 2.24, p = 0.03, d = 0.08, 95%CI[0.01 - 0.15]

M1 = 197.58, SD = 121.34, SE = 2.81, 95%CI[192.07 - 203.09]

M2 = 185.33, SD = 112.41, SE = 2.74, 95%CI[179.95 - 190.71]

t(3541) = 3.11, p < .01, d = 0.10, 95%CI[0.04 - 0.17]

M1 = 360.77, SD = 222.28, SE = 4.75, 95%CI[351.46 - 370.08]

M2 = 338.72, SD = 219.94, SE = 4.83, 95%CI[329.25 - 348.19]

t(4263) = 3.25, p < .01, d = 0.10, 95%CI[0.04 - 0.16]

M1 = 320.98, SD = 180.08, SE = 4.40, 95%CI[312.35 - 329.61]

M2 = 302.54, SD = 177.00, SE = 4.60, 95%CI[293.51 - 311.57]

t(3152) = 2.89, p < .01, d = 0.10, 95%CI[0.03 - 0.17]

M1 = 275.99, SD = 166.21, SE = 3.64, 95%CI[268.86 - 283.12]

M2 = 264.03, SD = 161.89, SE = 3.73, 95%CI[256.72 - 271.34]

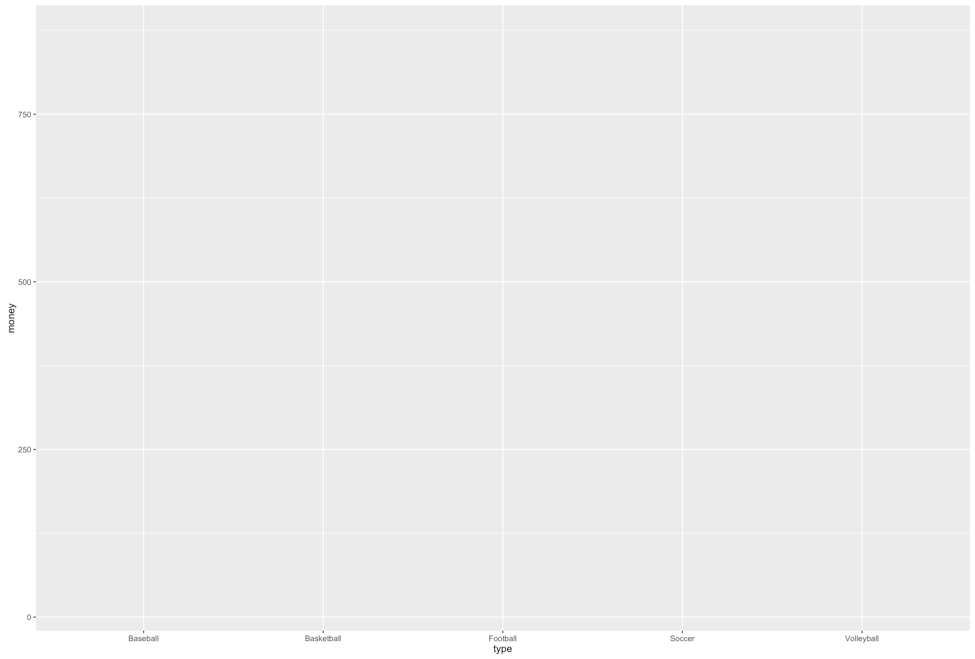
t(3973) = 2.29, p = 0.02, d = 0.07, 95%CI[0.01 - 0.14]

* 1. Make sure each M, SD, and N look correct.
  2. Enter *d* only into your table.
  3. You can make *d* values positive or negative – I tend to report them as always positive because the negative just indicates that you subtracted the smaller mean first, not anything about the actual effect size.
     1. These effects are all small and relatively equal (hence, no interaction).
     2. I would say that the large sample is what is driving the significant effects, but they aren’t really anything super exciting to talk about.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mean 1 | Mean 2 | P-value | Explain? | Effect size |
| Baseball  2007  M = 234.76  SD = 151.09  N = 1617 | Baseball  2008  M = 222.73  SD = 147.74  N = 1483 | .025 | Significant, spending decreased across years | 0.08 |
| Basketball  2007  M = 197.58  SD = 121.34  N = 1866 | Basketball  2008  M = 185.33  SD = 112.41  N = 1677 | .002 | Significant, spending decreased across years | 0.10 |
| Football  2007  M = 360.77  SD = 222.28  N = 2190 | Football  2008  M = 338.72  SD = 219.94  N = 2075 | .001 | Significant, spending decreased across years | 0.10 |
| Soccer  2007  M = 320.98  SD = 180.08  N = 1676 | Soccer  2008  M = 302.54  SD = 177.00  N = 1478 | .004 | Significant, spending decreased across years | 0.10 |
| Volleyball  2007  M = 275.99  SD = 166.21  N = 2088 | Volleyball  2008  M = 264.03  SD = 161.89  N = 1887 | .02 | Significant, spending decreased across years | 0.07 |

**Graphs:**

1. The best type of chart for anything analyzing group means is a bar chart with error bars.
2. We are going to use ggplot2 to build all our graphs.
   1. The package works like a transparency machine – you build layers and add them to the graph. You will really want to learn to stack your code, so that it’s easy to troubleshoot any problems you have.
3. First, load the ggplot2 library.
   1. library(ggplot2).
4. Create a blank graph with the right variables.
   1. X = IV, Y = DV.
   2. bargraph = ggplot(*datasetname,* aes(*Xcolumn, Ycolumn,* fill = *IVcolumn*))
   3. Note: fill has to be a factored variable. This variable will be put into a legend.
   4. Check that it worked – try running just bargraph. You should get a blank plot like this:



1. Which one should be the legend versus X axis?
   1. I put my split variable for interactions on the X axis, so the post hoc tests match the bars that are paired together.
2. Add things to the plot – note this code has changed a little bit:

bargraph +

stat\_summary(fun.y = mean,

geom = "bar",

position = "dodge") +

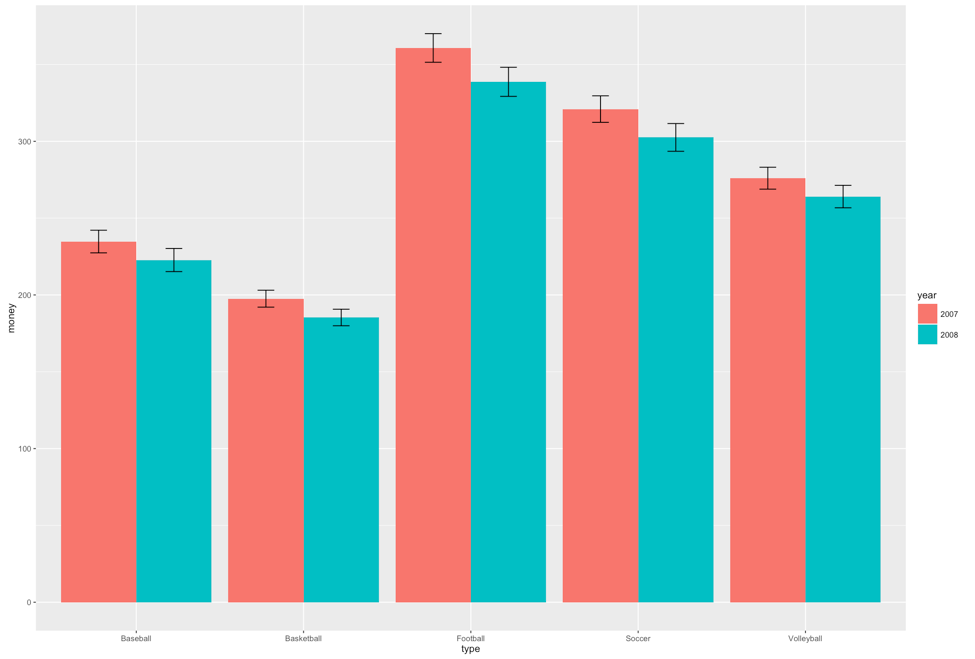
stat\_summary(fun.data = mean\_cl\_normal,

geom = "errorbar",

position = position\_dodge(width = 0.90),

width = 0.2)

* 1. Please note:
     1. That code above stays exactly the same, but remember that “” doesn’t copy correctly sometimes.
     2. What does it do?
        1. The first stat\_summary adds the bars to the graph by graphing the mean for each group.
        2. The second stat\_summary adds the error bars of the confidence interval (approximately 2\*SE). These bars help you see how much the variance is spread around each group.
     3. You should have this now:



* 1. That is the right graph, but it is **hideous.**
  2. First, we are going to clean up the gray background, the nondiscriminate axes, and the tiny type.
  3. Separate from the graph code, run this code exactly:

theme = theme(panel.grid.major = element\_blank(),

panel.grid.minor = element\_blank(),

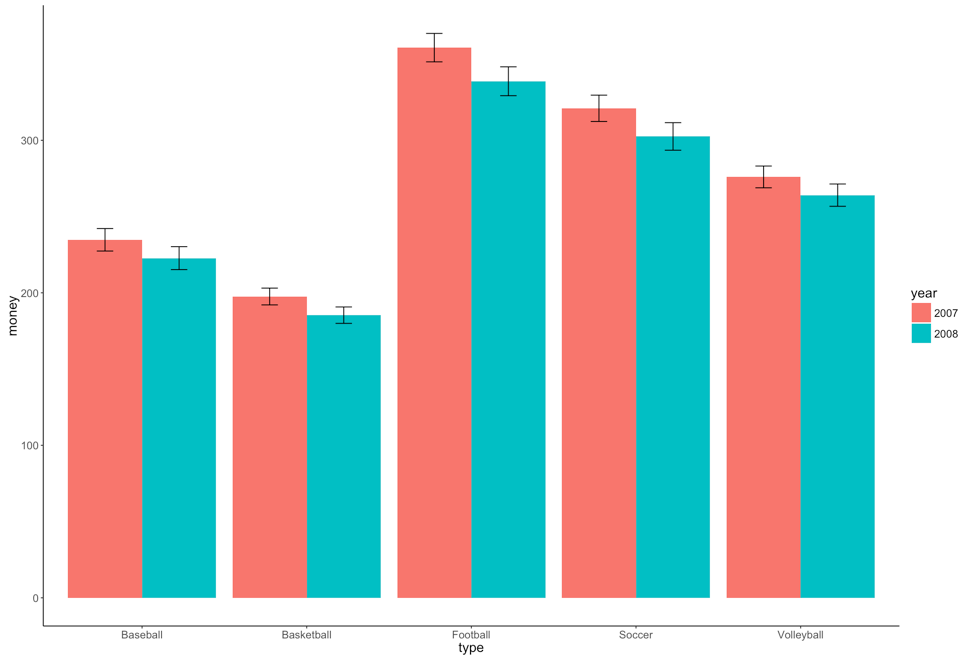
panel.background = element\_blank(),

axis.line = element\_line(colour = "black"),

legend.key = element\_rect(fill = "white"),

text = element\_text(size = 15))

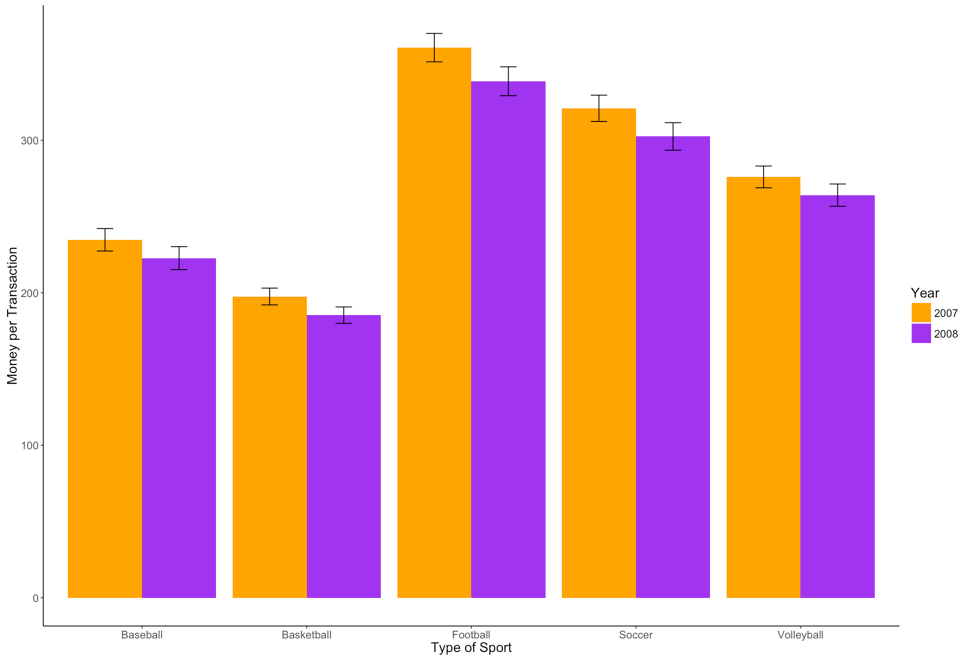
* 1. This code saves a whole bunch of settings as theme, which then we can add to our graph.
  2. NOTE: In this demo, we are walking through one part at a time, but you will run the entire graph code again to recreate the graph. It isn’t quite cool enough to remember what you did a minute ago.



* 1. Ok, that’s better, but now we have two issues:
     1. The x and y axes labels are terrible – what do they even mean?
  2. How to fix that:
     1. xlab(“Text that you want”) + ylab(“Text that you want”) will fix the axes labels.
  3. Next issue – the bad looking legend and colors:
     1. Notice in the first line we created the graph, we used the word FILL.
     2. We can do scale\_fill\_manual to fix that problem. The name part will change the overall label, and you can use labels if you want to fix the level labels.
     3. You can also make it black / gray / white / green / purple by using the values command.
     4. scale\_fill\_manual(name = c(“*Name of IV*”),

labels = c(“*level”, “level” ,…*),

values = c(“*color*”, *“color”, …*))



Note: I don’t really recommend an orange and purple graph … just for fun make sure you understand how to manipulate the colors.

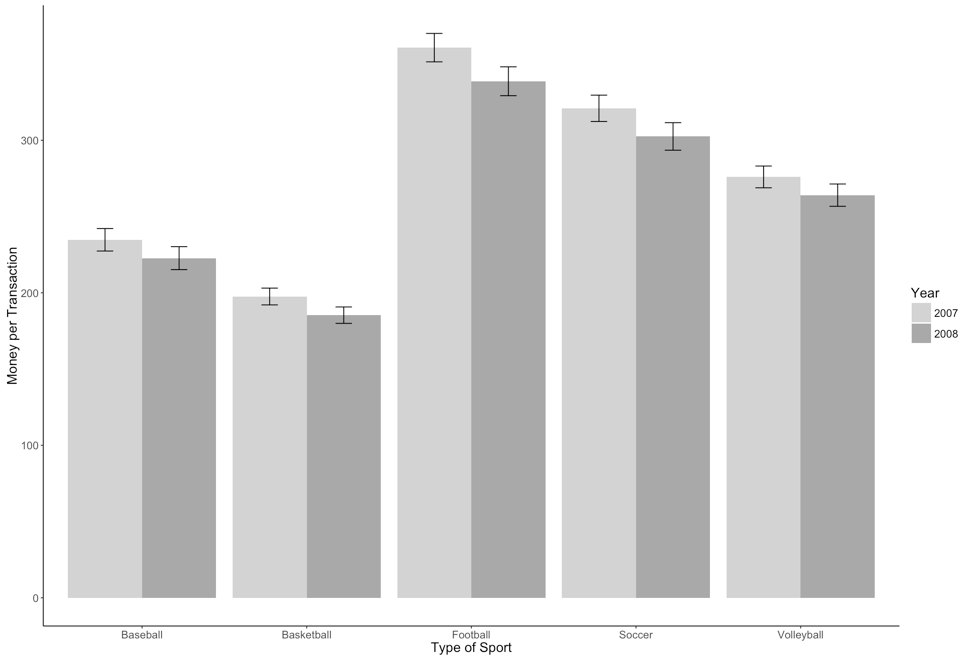
Write Up Example:

**Results**

Monthly budget transactions were examined to determine if type of sport spending (Baseball, Basketball, Football, Soccer, and Volleyball) changed over a two year period (2007, 2008). Data were screened for assumptions (linearity, homogeneity, normality, outliers), and several problems were found with linearity, homogeneity, and outliers. While Levene’s test indicated a potential problem with homogeneity (*F*(9, 18027) = 185.92, *p* <.001) the large sample size may have influenced this factor and a residual plot showed homogeneous groups. Several outliers were found with high standardized z-scores (*n* = 223), which were excluded for analyses after determining the data was correct.

A 2X5 factorial ANOVA was analyzed on year and transaction type. Both main effects of year (*F*(1, 18027) = 35.20, *p* <.001, η2 = .002) and type of sport (*F*(4, 18027) = 501.27, *p* <.001, η2 = .10) were significant. The interaction between year and transaction type was not significant, *F*(4, 18027) = .71, *p* = .59, η2 < .001. Figure 1 shows the interaction. (\*\*normally here you’d talk about the significant main effects post hoc only\*\*). Independent t-tests with a Bonferroni correction were performed to examine if average budgets had decreased across time. Baseball decreased spending by about $12 dollars from 2007 to 2008 (*p* < .001, *d* = 0.08), while Football decreased by about $22 dollars from 2007 to 2008 (*p* < .001, *d* = 0.10).

(\*\*\* you would go on and talk about all the different comparisons\*\*\*)



*Figure 1.*