MANOVA

**Description:**

* MANOVA stands for a multivariate analysis of variance. Specifically, MANOVA is most commonly used when you want to examine the relationship of one to many independent variables to two or more dependent variables.
* Regular ANOVA is for one to many independent variables with *one* dependent variable. MANOVA lets you combine dependent variables to look at overall group differences on a combined set of variables. Then you can break down the variables and see which individual dependent variable contributes to group differences the most.
* When you think about ANOVA 🡪 Post hoc tests as the process to break down the individual level/condition means, MANOVA allows you to combine DVs, and then follow that same process. This type of procedure is Type 1 error control because you do not follow up the ANOVAs for non-significant DVs.

**Definitions/Abbreviations:**

* IV – independent variable. This variable *has* to be a categorical variable. You can put people into groups based on any category (gender, handedness) or your experimental manipulation (instructions versus no instructions).
* DV – dependent variable. The dependent variable *needs* to be a continuous variable or another type of analysis might work better (see log regression). Your dependent variable should be the measurement you took in your study or what information you are expecting to see changed over groups.
* DV combinations – Wilk’s Lambda, Roy’s Largest Root, Hotelling’s Trace, Pillia’s – these are all listed in the multivariate test. They are different ways to combine the DVs in such a way that creates large group differences on your IVs. Think of these as *giant means* for your DV, if you could create one mean of all the DVs such that your groups were maximally different. The most commonly used is Wilk’s Lamba.

**Types:**

* Number of IVs: You can use one or more IVs, and the same post hoc procedures as ANOVA apply.
* Number of DVs: You can use two or more DVs, which is the point of MANOVA.
  + Sometimes, instead of analyzing a mixed design, people will use MANOVA to analyze each repeated measures level separately. However, you would not compare the repeated measures levels to each other.
* Types: generally, you will only have between subjects MANOVA because you are treating the repeated part as separate DVs. You can do repeated measures and mixed designs; however, it is more common to use mixed ANOVA for those analyses.
  + If you use mixed or repeated designs, be sure to follow up with the appropriate type of repeated test – the easiest would be to use ezANOVA.

**The process:**

* If the ANOVA level is not significant, do not analyze the post hocs.
* 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?
  + Also, be sure to follow up with the correct test type – do not do dependent t on the between subjects factor. See the previous ANOVA notes about specific post hoc procedures for each ANOVA type.

# Complete Example

# 2X2 Between Subjects MANOVA

Researchers have measured participants on their femininity and masculinity and want to know how those two variables affect a range of dependent measures. They measured self-esteem, attitude about women’s roles, and neuroticism to see if there were differences across femininity and masculinity scores.

**IVS:**

* Femininity scale (low versus high)
* Masculinity scale (low versus high)

**DVS:**

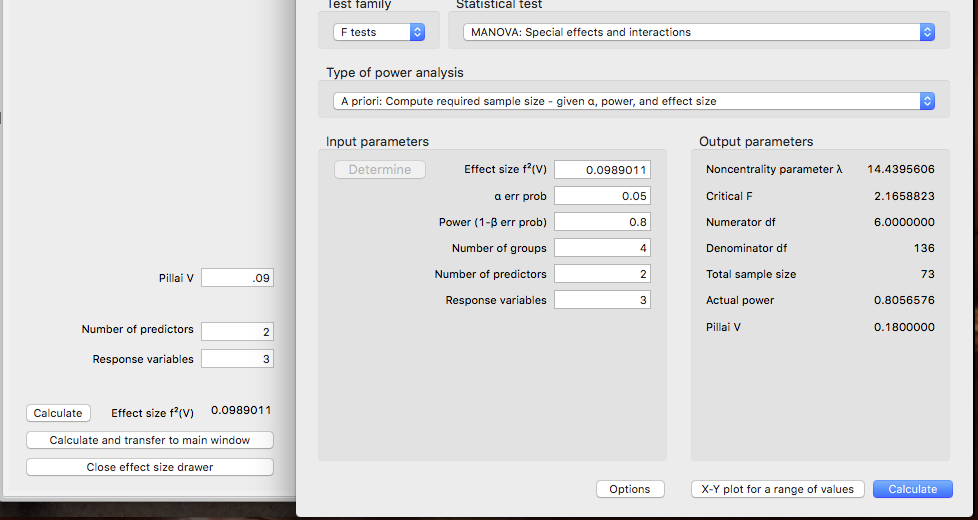
* Self esteem
* Attitude toward the role of women
* Neuroticism

**Research question:**

* We want to see if the different feminine-masculine groups will score differently on a combination of DVs. So we might see if they differ on “personal factors”, which is the combination of several of our DVs.

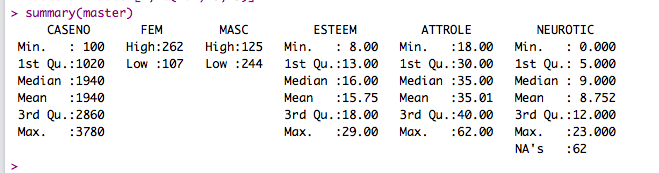
**Power:**

1. Open Gpower!
   1. Test family: F-test
   2. Statistical Test: MANOVA: Special effects and interactions
      1. Usually you are looking for the interaction, so you’ll use this page to estimate the number of people needed for that test.
   3. Estimate an effect size: click determine 🡪 use eta square sizes you think might be accurate, remember small, medium, and large estimates from the notes.
      1. Note: Pillai V will be part of the output (under value for Pillai) but it is still easiest to calculate from eta squared.
      2. Number of predictors: number of IVs.
      3. Response variables: number of DVs.
      4. Calculate and transfer to main.
   4. Alpha = .05
   5. Power (1-beta .20) = .80
   6. Number of groups: number of levels or conditions.
   7. Number of predictors: number of IVs.
   8. Response variables: number of DVs.
2. Let’s estimate the following:
   1. Medium effect size (eta = .09)
   2. Number of groups: 4 (2X2 is four conditions)
   3. Number of predictors: 2
   4. Response variables: 3
3. Says we needed to run 73 people to find a significant effect with a medium 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, not go below zero, and each scale has a slightly different maximum (we are going to assume the max is ok).



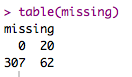
1. Missing:
   1. With the summary function, I can see that I have missing data. So here’s the percent missing by participant:
   2. Code:

percentmissing = function (x){ sum(is.na(x))/length(x) \* 100}

missing = apply(*dataset*, 1, percentmissing)

table(missing)

* 1. 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.
     1. And as you see here, we have 62 people with 20% missing – so we can’t replace any data.

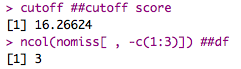


1. Outliers:
   1. Since we have several DVs, we have several columns to work with, which means we can use Mahalanobis values. We want to use this format for data screening because it accounts for the fact that people have more than one measurement. We would not want to ignore that person one is person one for all DVs.
   2. BUT: don’t forget that you cannot use the factored columns in Mahalanobis.
   3. Create the Mahalanobis values:
      1. mahal = mahalanobis(*dataset*,

colMeans(*dataset*, na.rm = T),

cov(*dataset*, use = “pairwise.complete.obs”))

* 1. Create the cut off score:
     1. cutoff = qchisq(1-.001, ncol(*dataset*))
  2. Remember you can use:
     1. cutoff to get the cutoff score
     2. ncol(*dataset*) to get the *df*

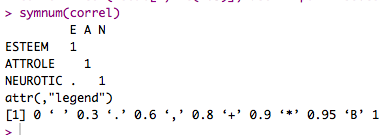


* 1. See how many outliers you have:
     1. summary(mahal < cutoff)

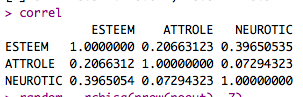


* + 1. Remember FALSE is bad.
    2. I have no outliers!
  1. Exclude outliers:
     1. noout = subset(*dataset*, mahal < cutoff)

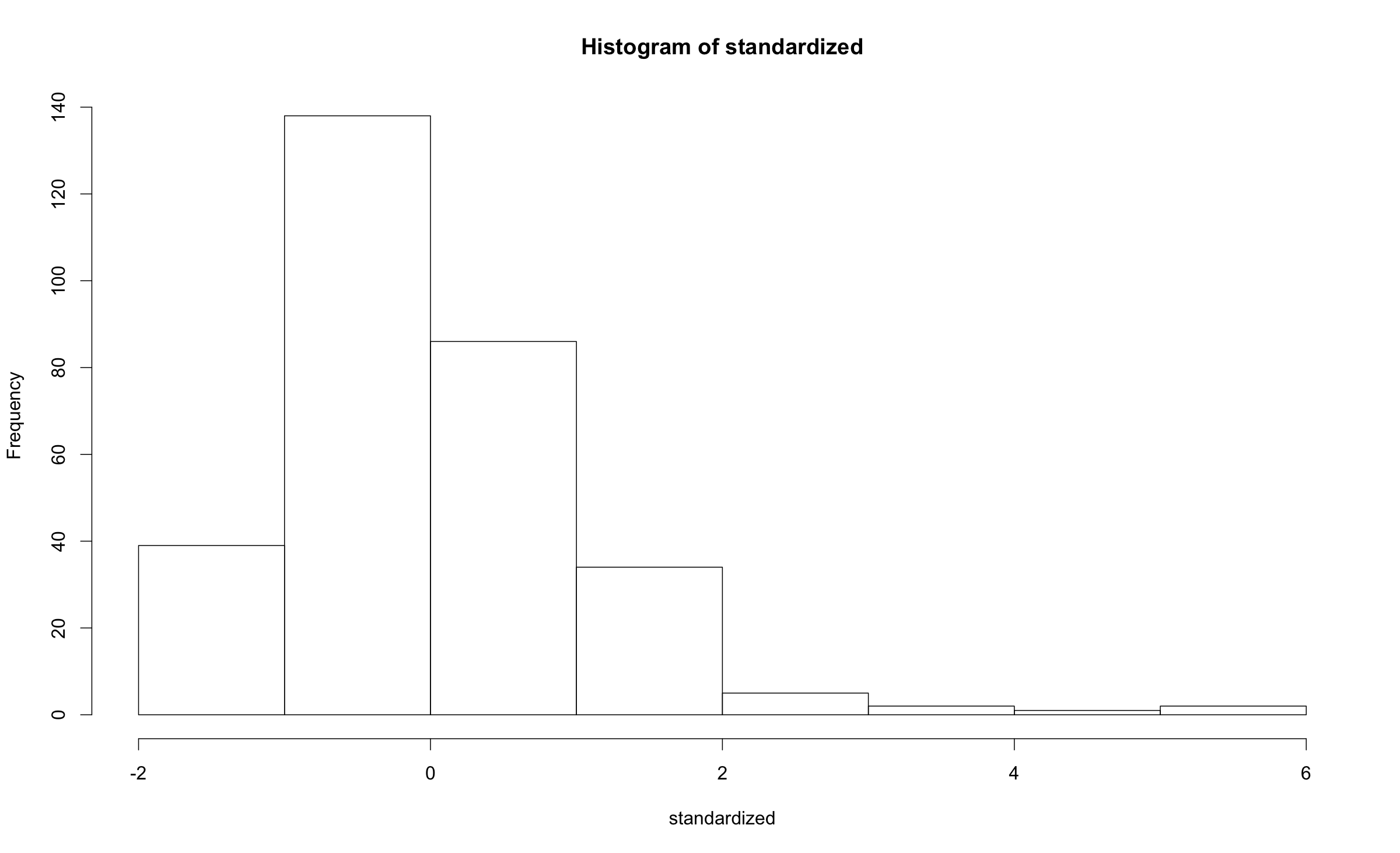
1. Additivity
   1. In general, you *want* the various measurements to be highly correlated – it will give you more power if they are correlated and less if they are not.
   2. However, they cannot be perfectly correlated or the MANOVA will not run.
   3. Mainly we are checking that we don’t get any 1s other than the diagonal in our symbols chart. So, basically, the rule is the *r* < .999.
   4. Get the correlations:
      1. correl = cor(*dataset*, use = “pairwise.complete.obs”)
   5. Get the symbols chart:
      1. symnum(correl)
   6. Look for 1s NOT on the diagonal:



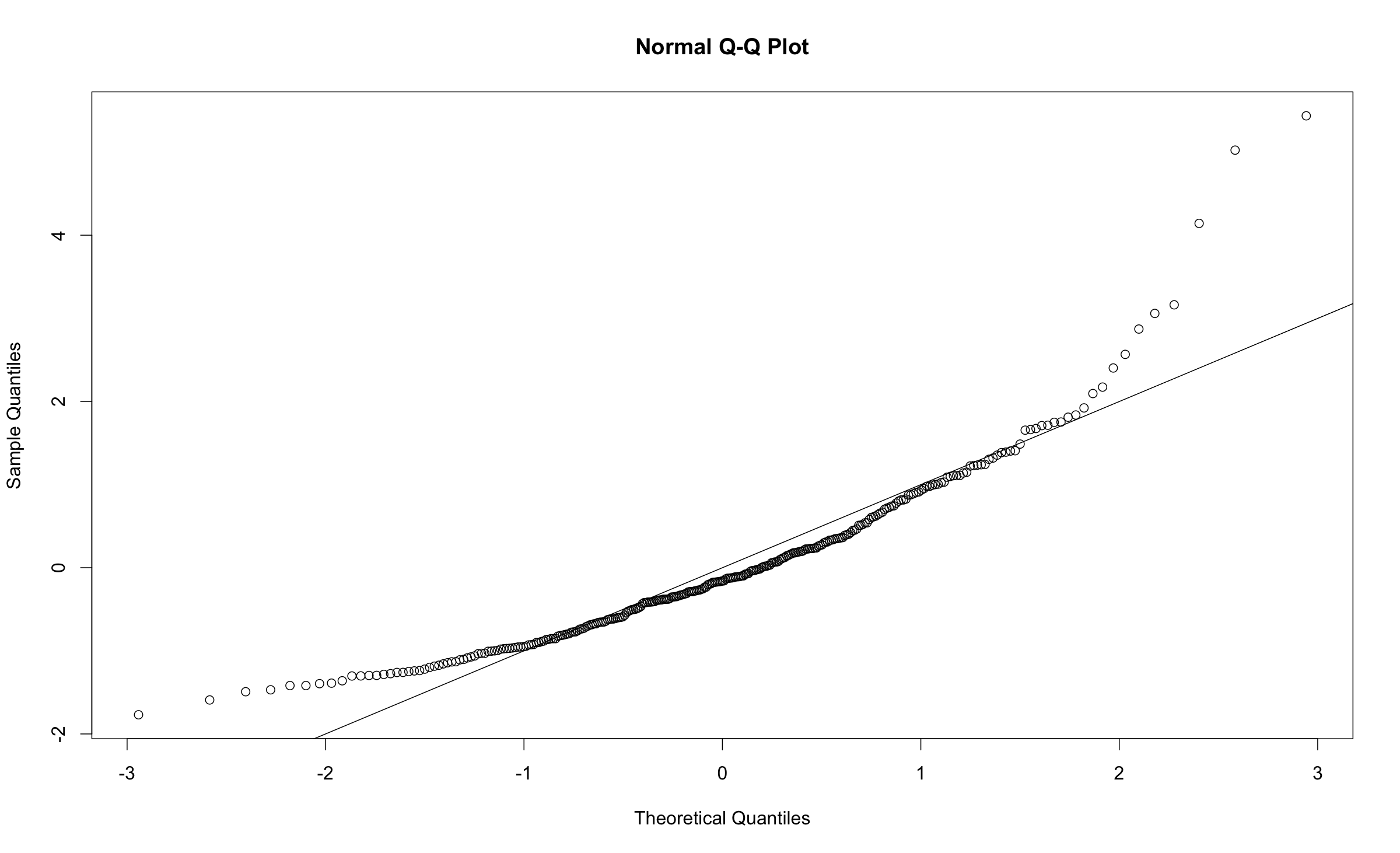
* 1. The ones marked in blue are ok, they are the column correlated with itself (which should be 1).
  2. So, our numbers appear ok.
  3. You can run correl to see the numbers – it appears these might not be the best DVs to combine as they aren’t really correlated.



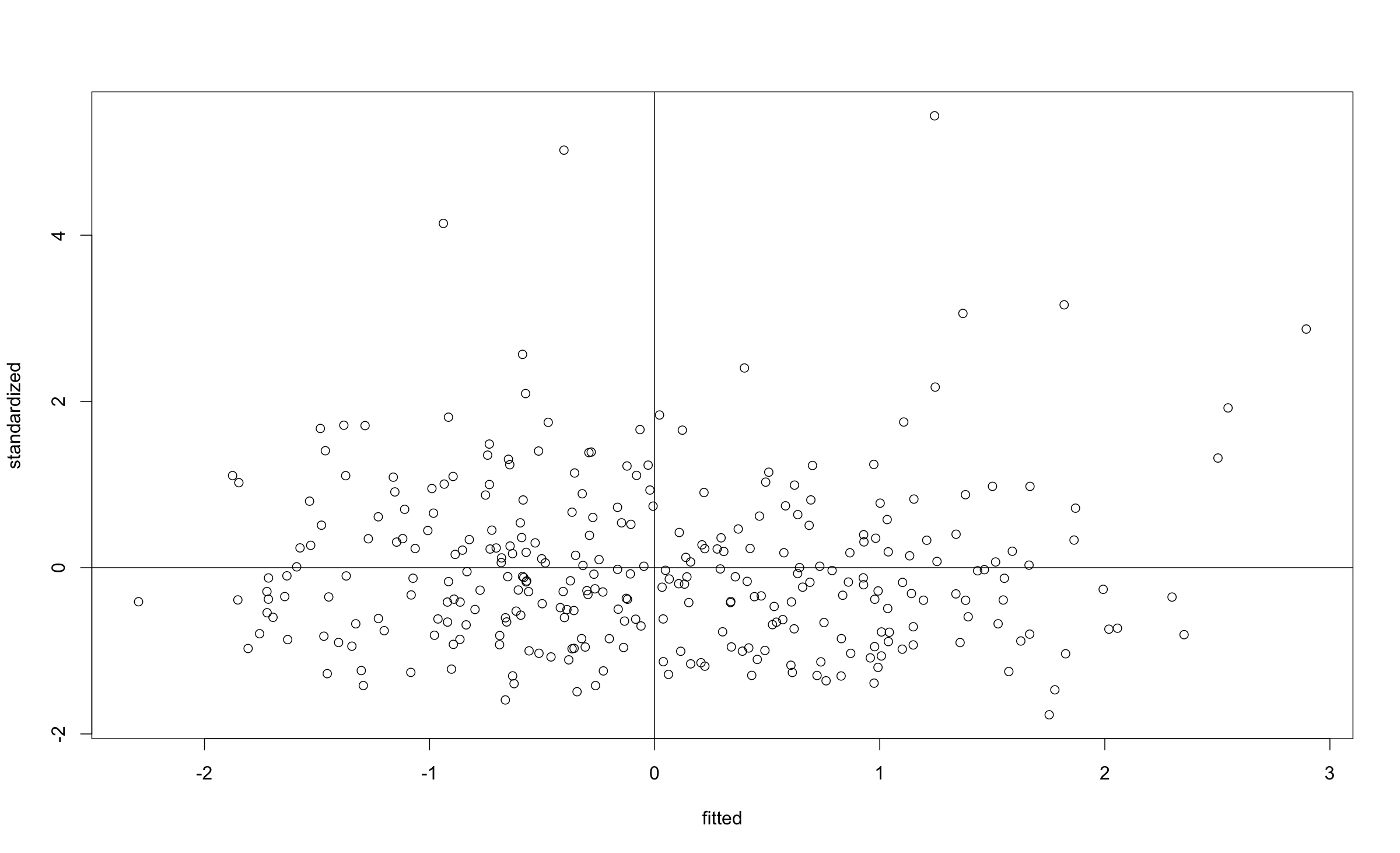
1. Set up the rest of the assumptions:
   1. Make a random variable:
      1. random = rchisq(nrow(*dataset*), 7)
   2. Run a fake regression:
      1. fake = lm(random~., data = *dataset*)
   3. Create the standardized residuals:
      1. standardized = rstudent(fake)
   4. Create the fitted values:
      1. fitted = scale(fake$fitted.values)
2. 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 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 very suspect.
        1. Test yourself – run it a couple times (by rerunning the random assumption set up section), you should get a bad graph every time.
        2. I ran it a couple times and found half-half on bad and sort of ok. I’d say it’s probably ok, it will just lower power trying to use a linear design on this data.
     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 iffy – I ran a couple versions and mostly found between -2 and 4 for vertical, with pretty consistent -2 to 2 on the horizontal. Mostly, it seems ok.
        1. We will also use Levene’s test to determine if it’s a problem, but we will have to check out each DV individually.
     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).

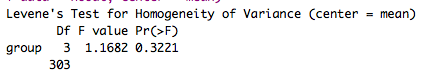


* 1. Homogeneity: Take 2 Levene’s Test
     1. Levene’s is a test for homogeneity between groups, so it looks to see if the variances are equal across your IV levels.
     2. It is notoriously **oversensitive**, but can be a good place to start if you want to check a real number, rather than this scatterplot.
     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.
  2. **HOW TO GET IT:**
     1. Because we are not using ezANOVA just yet, we have to get Levene’s separately.
     2. Load the car library.
        1. library(car)
     3. Run the Levene’s test with only the between subjects factors:
        1. leveneTest(*dv column* ~ *iv column*\**iv column*,

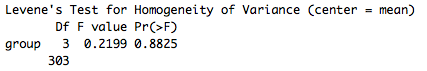
data = *dataset,* center = mean)

* + - 1. If you only have one between subjects factor, just take out the \**iv column* half. If you have more than two, just add more \**iv columns.*
      2. Use center = mean because we are analyzing means in ANOVA.
    1. They are all ok because *p* > .001 on all of them.

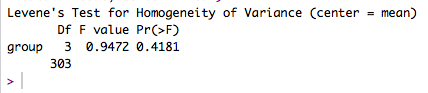
Esteem:



Attitudes:



Neuroticism:



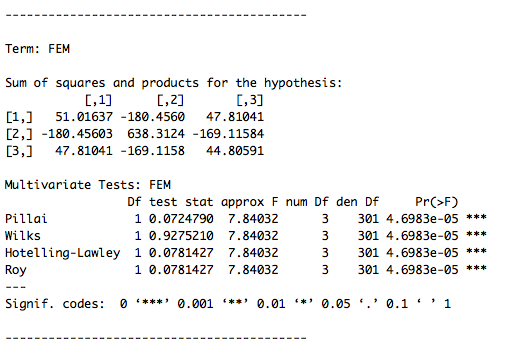
**Running the MANOVA:**

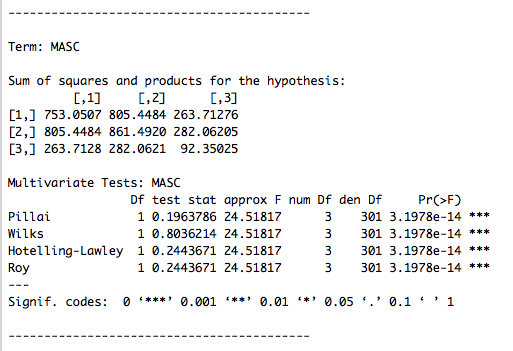
1. Set up the DV:
   1. DV = cbind(*dataset$column, dataset$column*)
   2. This code creates a separate dataset just called DV.
2. Set up the test with lm():
   1. output = lm(DV~ *IV\*IV*, data = *IV* *dataset,*

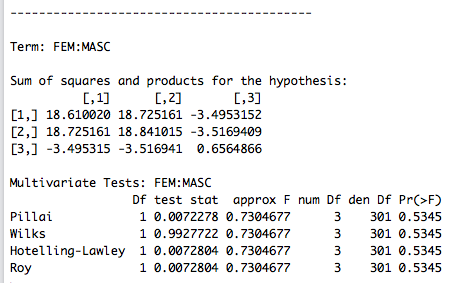
contrasts=list(*IV*=contr.sum, *IV*=contr.sum))

* 1. What’s going on?
     1. We are going to create a formula for our DV…it is ~ (predicted) by the IVs. We multiply the IVs to get interactions.
     2. The contrasts part makes sure this test is type 3 sum of squares, which will match all the other major programs.

1. Get the MANOVA output.
   1. manova\_out = Manova(output, type = “III”)
   2. Note the M being capital is not a mistake.
   3. The car library must still be loaded.
   4. summary(manova\_out, multivariate = T)
   5. Ignore the intercept part!







Interpret the output:

1. First, we want to look at Wilk’s test, as it is the most common DV math to interpret.
   1. What are the numbers?
      1. Ignore the first DF.
      2. Test stat lamba for Wilks.
         1. To get eta squared 1 – lamba.
      3. Approx F = F value.
      4. Num Df = DF model, between, numerator
      5. Den Df = DF error, within, denominator
      6. Pr(>F) = p value.
   2. Each section is separated by -------.
   3. Write that up:
      1. Femininity: *F*(3, 301) = 7.84, *p* <.001, η2 = .07.
      2. Masculinity: *F*(3, 301) = 24.52, *p* <.001, η2 = .20.
      3. Interaction: *F*(3, 301) = 0.73, *p* = .53, η2 = .01.
2. Post Hoc Interpretation/Plan:
   1. Now, we should run ANOVAs for the significant effects only to see which DVs are significant.
   2. We can use ezANOVA for those effects running each DV one at a time.
   3. Load the ez library.
      1. library(ez)
   4. Run the ANOVA (all these lines):
      1. ezANOVA(data = *dataset*,

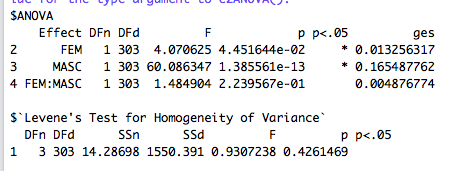
wid = partno,

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

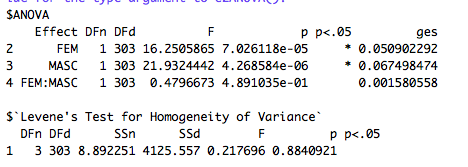
dv = *column of DV*,

type = 3)

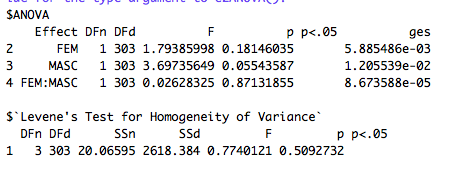
Esteem:



Attitudes of Roles:



Neuroticism:



* 1. What’s significant?
     1. Remember ignore the interactions, as the MANOVA was not significant.
     2. Esteem:
        1. **Fem: *F*(1,303) = 4.07, *p* = .04, ηp2 = .01**
        2. **Masc: *F*(1,303) = 60.09, *p* <.001, ηp2 = .17**
     3. Attitude Roles:
        1. **Fem: *F*(1,303) = 16.25, *p* < .001, ηp2 = .05**
        2. **Masc: *F*(1,303) = 21.93, *p* < .001, ηp2 = .07**
     4. Neuroticism:
        1. Fem: *F*(1,303) = 1.79, *p* = .18, ηp2 = .01
        2. Masc: *F*(1,303) = 3.70, *p* = .06, ηp2 = .01
  2. All of these effects are on two levels, so we can just run means.
     1. If there were more, you can run post hocs just like the ANOVA guides.
     2. If the interaction were significant, you can run those post hocs by splitting one variable, and then using pairwise.t.test on the other half.

Post hoc notes if you have more than two levels or are analyzing the interaction:

* 1. Use the subset function to split the IV if you are analyzing the interaction.
  2. Use the pairwise.t.test() function to run t.test you learned earlier on all groups at once.
     1. Remember, you use paired = T for **dependent** t-tests, which is what we want to use for **repeated-measures** ANOVA.
        1. p.adjust.method is the *correction*.
        2. pairwise.t.test(*dataset*$*DV*, *dataset*$*IV*,

paired = T,

p.adjust.method = "bonferroni")

* + 1. Remember, you use paired = F for **independent** t-tests, which is what we want to use for **between-subjects** ANOVA.
       1. p.adjust.method is the *correction*.
       2. var.equal = T for homogeneity.
       3. pairwise.t.test(*dataset*$*DV*, *dataset*$*IV*,

paired = F,

var.equal = T,

p.adjust.method = "bonferroni")

* 1. Be sure to use the new dataset for interactions! You will have to do the same thing for each dataset that you just created.
  2. 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.
  3. 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.

To interpret post hocs and for only two levels:

* 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.

Copying just the means here:

Fem – Esteem

High Low

15.50909 16.70115

Low fem have higher self esteem.

Masc – Esteem

High Low

13.51000 16.97585

Lower masc have higher self esteem.

Fem – Attitudes

High Low

36.25000 33.05747

Higher fem have higher attitudes toward the role of women.

Masc – Attitudes

High Low

32.85000 36.55072

Lower masc have higher attitudes toward the role of women.

Fem – Neuroticism

High Low

8.504545 9.379310

They are equal in neuroticism (remember F was not significant).

Masc – Neuroticism

High Low

7.840000 9.193237

They are equal in neuroticism (remember F was not significant).

Remember, you can use the earlier notes to figure out post hoc tables, effect sizes, etc.

**Graphs:**

1. The best type of chart for anything analyzing group means is a bar chart with error bars – you are going to make one for each DV!
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. Load the ggplot2 library.
   1. library(ggplot2).
4. We are going to clean up the gray background, the nondiscriminate axes, and the tiny type that always happens with plots.
   1. Separate from the graph code, run this code exactly:

cleanup = 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 cleanup, which then we can add to our graph.

1. 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.
2. 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.
3. Add things to the plot:

bargraph +

stat\_summary(fun.y = mean,

geom = "bar") +

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

geom = "errorbar",

position = position\_dodge(width = 0.90),

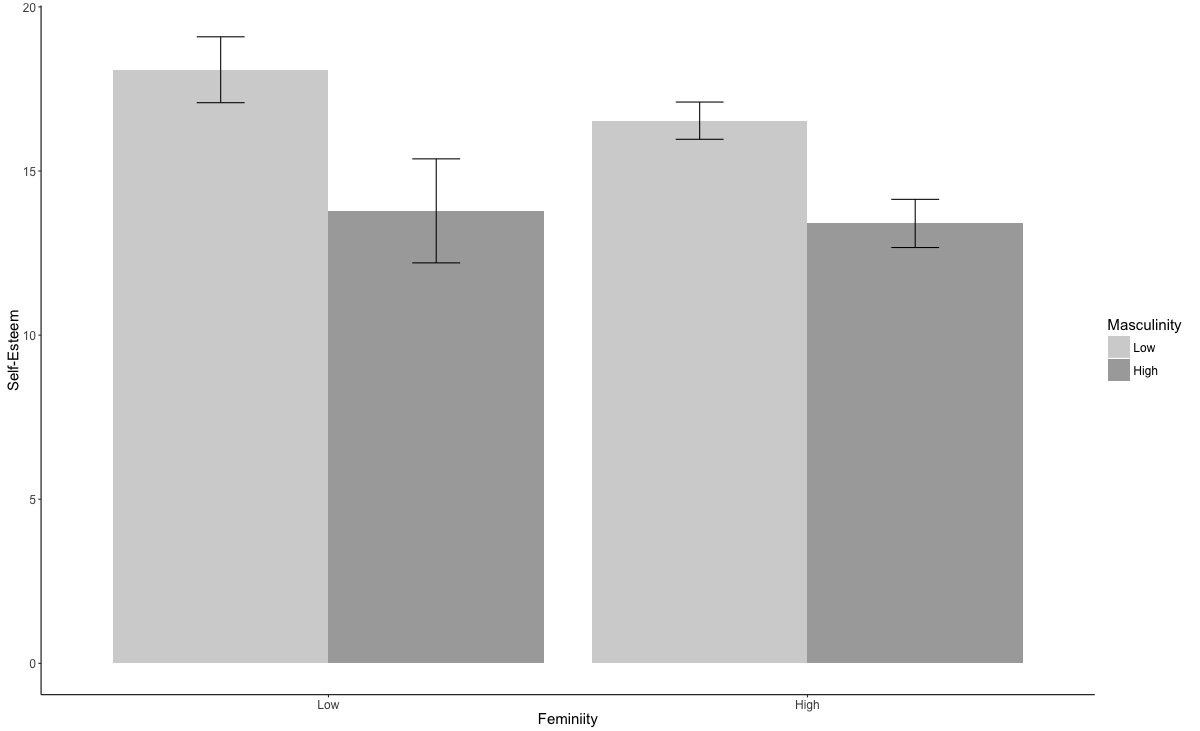
width = 0.2)

* 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

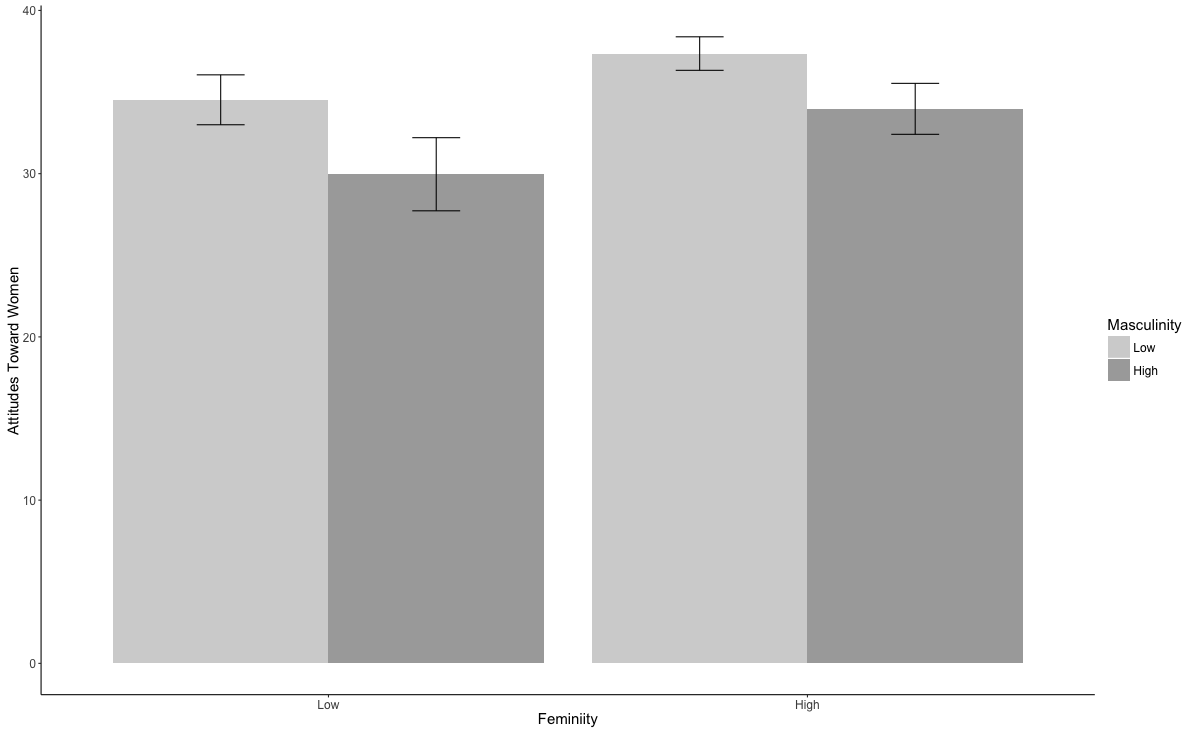
1. Label X and Y:
   1. xlab(“Text that you want”) + ylab(“Text that you want”) will fix the axes labels.
2. Next issue – the bad looking legend and colors:
   1. Notice in the first line we created the graph, we used the word FILL.
   2. We can do scale\_fill\_manual to fix that problem. The name part 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” ,…*),

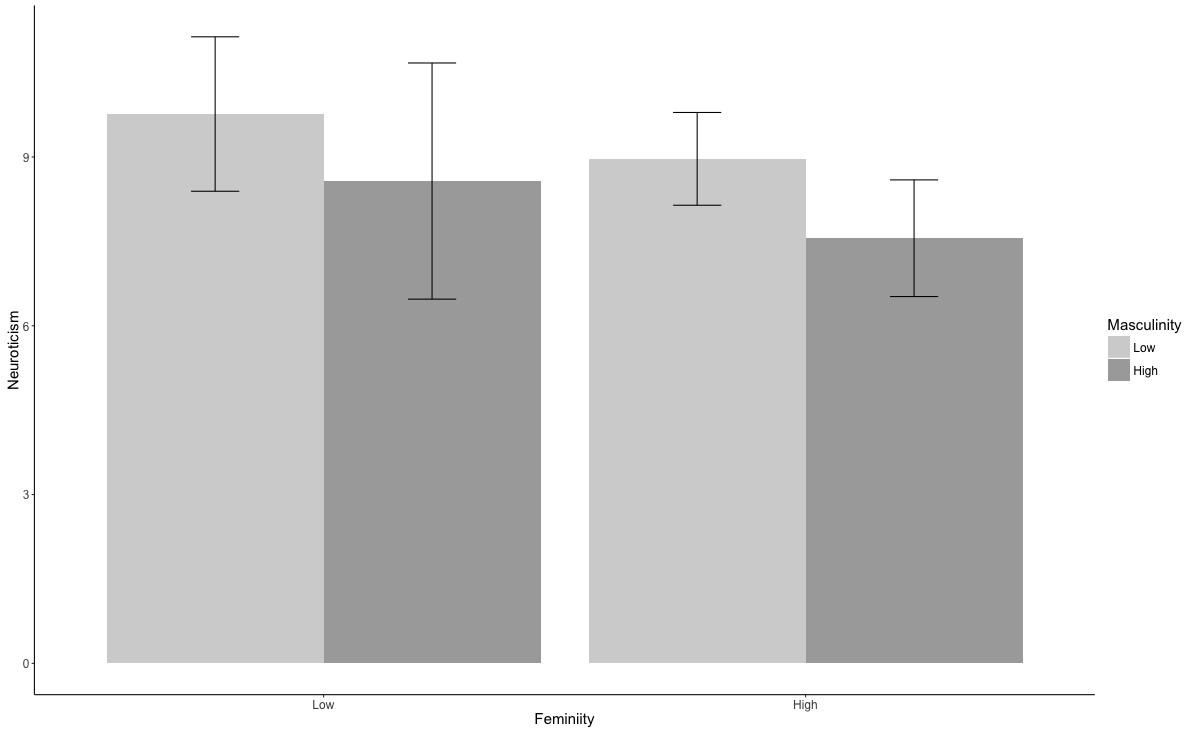
values = c(“*color*”, *“color”, …*))



*Figure 1.*



*Figure 2.*



*Figure 3.*

(Note: this write up’s numbers may not match exactly, but you should include all these numbers).

**Results**

Prior to analysis, data were screened for missing data and outliers. Several participants were identified as univariate outliers, but no cases were multivariate outliers (using Mahalanobis distance). All participants were retained for this analysis. Data were found to be multivariate normal, linear, and homogeneity was met (Levene’s *p*s > .001). A 2 X 2 between subjects MANOVA was analyzed with femininity (low, high) and masculinity (low, high) predicting neuroticism, self-esteem, and attitude toward the role of women. Figures 1, 2, and 3 depict the average scores for the conditions on the dependent variables.

Significant multivariate main effects were found for femininity (*F*(3, 363) = 8.47, *p* < .001, *ηp2* = .07) and masculinity (*F*(3,363) = 31.80, *p* < .001, *ηp2*= .21), but not for the interaction between femininity and masculinity (*F*(3,363) = .71, *p* = .54, *ηp2* = .01). Univariate ANOVAs were used to examine individual dependent variable contributions to main effects. Femininity scores showed a significant difference in self-esteem scores (*F*(1, 365) = 5.18, *p* = .02, *ηp2*= .01), where low femininity participants (*M* = 15.89, *SD* = 2.36) scored higher than high femininity participants (*M* = 14.91, *SD* = 1.23). Attitudes toward the role of women showed the opposite effect (*F*(1, 365) = 16.91, *p* <.001, *ηp2*= .04), where low femininity (*M* = 32.07, *SD* = 2.64) scored lower than high femininity (*M* = 35.18, *SD* = 1.41). Neuroticism scores were not significantly different across femininity groups (*F*(1, 365) = 1.54, *p* = .22, *ηp2*= .004).

Masculinity scores were significantly different on self esteem (*F*(1, 365) = 73.10, *p* < .001, *ηp2*= .17), and low masculinity participants (*M* = 17.23, *SD* = 2.25) had higher self-esteem than high masculinity participants (*M* = 13.57, *SD* = 2.35). Low masculinity (*M =* 35.86, *SD* = 1.44) participants had higher ratings of the role of women (*F*(1, 365) = 35.18, *p* < .001, *ηp2*= .09), than high masculinity participants (*M* = 31.39, *SD* = 1.61). Lastly, neuroticism scores were significantly different for different masculinity groups, *F*(1, 365) = 6.12, *p* = .01, *ηp2*= .02. Low masculinity participants (*M* = 9.37, *SD* = 2.35) scored higher on the neuroticism scale than high masculinity participants (*M* = 7.88, *SD* = 2.49).