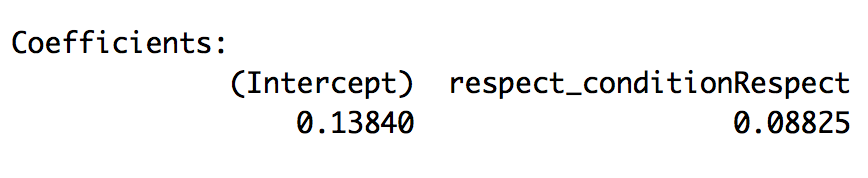
# Name:

# Classwork 12: **Analyzing Yeager, Hirschi, & Josephs’ Hypothesis**

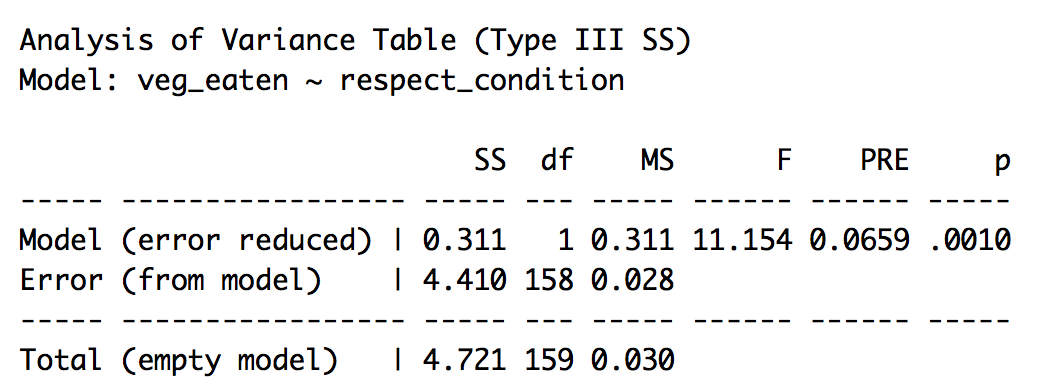
1. Let's assume this was a well conducted experiment. In that case, there are only two reasons why these two groups look different on their vegemite eating. What are those reasons?
2. How would we write the DGP of randomness in GLM notation? If we used the empty model to predict how much vegemite students’ ate, what value would it predict for each student?
3. But enough with the empty model! Let’s create a model that’s more interesting! How would you write a word equation for the more interesting DGP?
4. How would you create the best fitting model of the more interesting data generating process? Go ahead and run the R code and write the resulting model in GLM notation.



1. Interpret the respect model’s estimates here. Also, what does the and stand for?
2. What does the respect model predict: (a) whether or not someone is in the respect group or (b) how much vegemite someone would eat?
3. What would be for someone in the respect group? What would the respect model predict for someone in the respect group?
4. What would be for someone in the no respect group? What would the respect model predict for someone in the no respect group?
5. Save the predictions of the respect model into the data frame. What is the R code to put the predictions of the respect model into the data frame? How would you put those predictions on a jitterplot?
6. Draw the predictions of the respect model in the jitter plot below. Draw some residuals from the respect model. What do those residuals represent: the variation that has been *explained* or *left unexplained* by the respect model? How does GLM represent these residuals?



1. In a different color, draw the empty model into the jitter plot (previous question). Draw some residuals from the empty model. What do those residuals represent: the variation that has been *explained* or *left unexplained* by the empty model? How does GLM represent these residuals?
2. In a different color, draw the part of the residual that has been explained? Which model did the explaining?
3. What is the problem with just trying to add up these residuals (any of these residuals)?
4. To answer this question (“How much variation has been explained?”), it helps to start with this question: “How much total error from the empty model is there to begin with?” Then we can figure out what part of that has been explained. Run **supernova()** on the empty model. How much is the total SS?
5. Now run **supernova()** on the respect model. So which of these numbers tell us *how much* of the error has been explained? What unit are those numbers in?



1. Why didn’t the total SS change in the two supernova outputs: supernova(emptymodel) vs. supernova(respectmodel)?
2. What is the relationship between SS Model, SS Error, and SS Total?
3. Interpret the PRE for the Respect model. Is that a lot? Is that a little? Is it hard to tell?
4. Is it *possible* that we could have gotten this PRE from a random process (where these groups are essentially the same but they are just a little different because of randomness)?