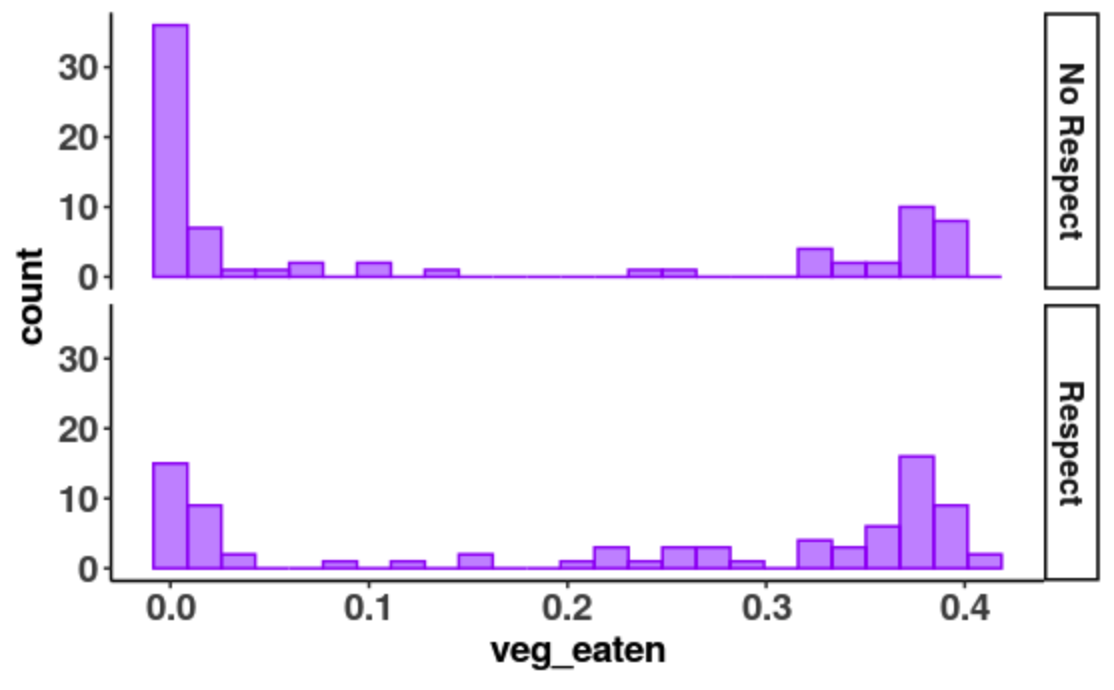
# Name:

# Classwork 17: **Yeager, Hirschi, & Josephs’ Null Model versus Respect Model**

1. Let’s re-visit the distribution triad. [mini-lecture]
2. Now you know how to build models. Find the best fitting empty model and respect model and write it in GLM notation.
3. Draw your two models in different colors (and label them) in the distribution below.



1. Which model explains more variation: null or respect? How do you know?
2. Does that prove that the respect model is better? Why or why not?
3. Okay. Let’s imagine we are working for a hardcore skeptic who says, “Hey, just using a little bit of respectful language is not going to make a big difference in how much vegemite students eat. I think the DGP is just . What’s your best estimate of ?” What would you say?
4. We are doing the best we can. What are our chances of being right? Explain your thinking.
5. If we had a similar but slightly different sample, what kind of variation is this? Explained? Unexplained? Random variation? Sampling variation? Variation caused by respect?
6. Now imagine if we really could pluck 1000 random samples that were roughly similar to ours. Could we use that to learn how much our parameter estimate could vary? Why or why not?
7. One of the superpowers of statistics is that we can simulate random processes. We’ve done a lot with shuffle. Now let’s try resampling (or bootstrapping) – a related but slightly different random process.

Run this code: resample(YeagerData, 3)

What do you think it did? What does this function give you? A single number? A variable? A data frame?

1. Why is this a random process?
2. Here is a new function to try out: b0(veg\_eaten ~ NULL, YeagerData)

\*\*Note: We made **veg\_eaten**, a variable, way back in Classwork 12. You may have named it something slightly different.

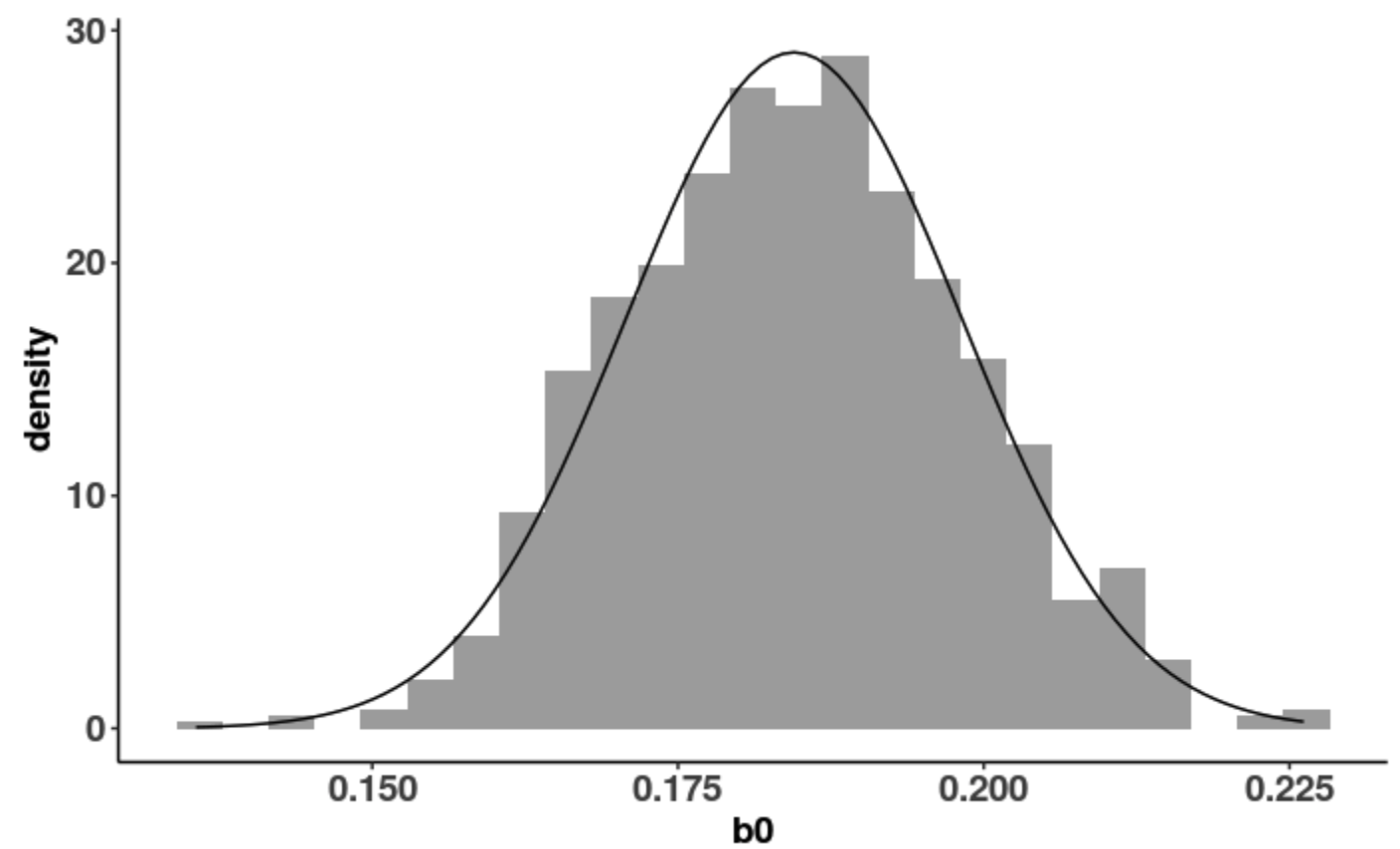
What do you think this function does?

1. Now I want you to imagine that the world, “the population,” is just like our sample. Imagine doing this experiment again but we just randomly select a slightly different group of 160 people. How would you use the function **resample()** to do this? Write the code.
2. If we fit the empty model onto this new resampled data, would we come up with the exact same ? Why or why not?
3. Try to use the two functions **b0()** and **resample()** together somehow to select a new group of 160 people and calculate a new . Write the code here.
4. What is result of running this code? What does that number stand for? Did the grand mean of our sample change?
5. Run that code 10 times. You don’t need to write down the numbers. What are each of these numbers? What does it stand for?
6. What do you notice about these s? How much do you think the s vary?
7. Let’s use the **do()** function to run that code 10 times. What does the **do()** function do? What does this function give you? A single number? A variable? A data frame?
8. If we ran this code 1000 times, using the **do()** function, what kind of distribution do we get: a sample, the DGP, or sampling distribution?
9. Could we visualize this distribution? What do you think the shape, center, and spread of this distribution would be?
10. Go ahead and try to create a visualization of these s. Write the code here.
11. Dr. Ji saved her bootstrapped s in a data frame called **SamplingDistributionOfb0** and ran this code:

gf\_dhistogram(~ b0, data = SamplingDistributionOfb0) %>%

gf\_fitdistr()

What is depicted in this histogram? What is the shape, center, and spread of this distribution?



1. What is the shape of the sample distribution? What is the shape of the sampling distribution of s? Why are the shapes so different?
2. How does what we learned fit back into the distribution triad we started with? [mini-lecture]