# Name:

# Classwork 19: **Confidence Intervals of**

1. Yeager and colleagues thought respectful language would make a difference in vegemite eating. Let’s see how far we’ve come in trying to create a model around that idea.

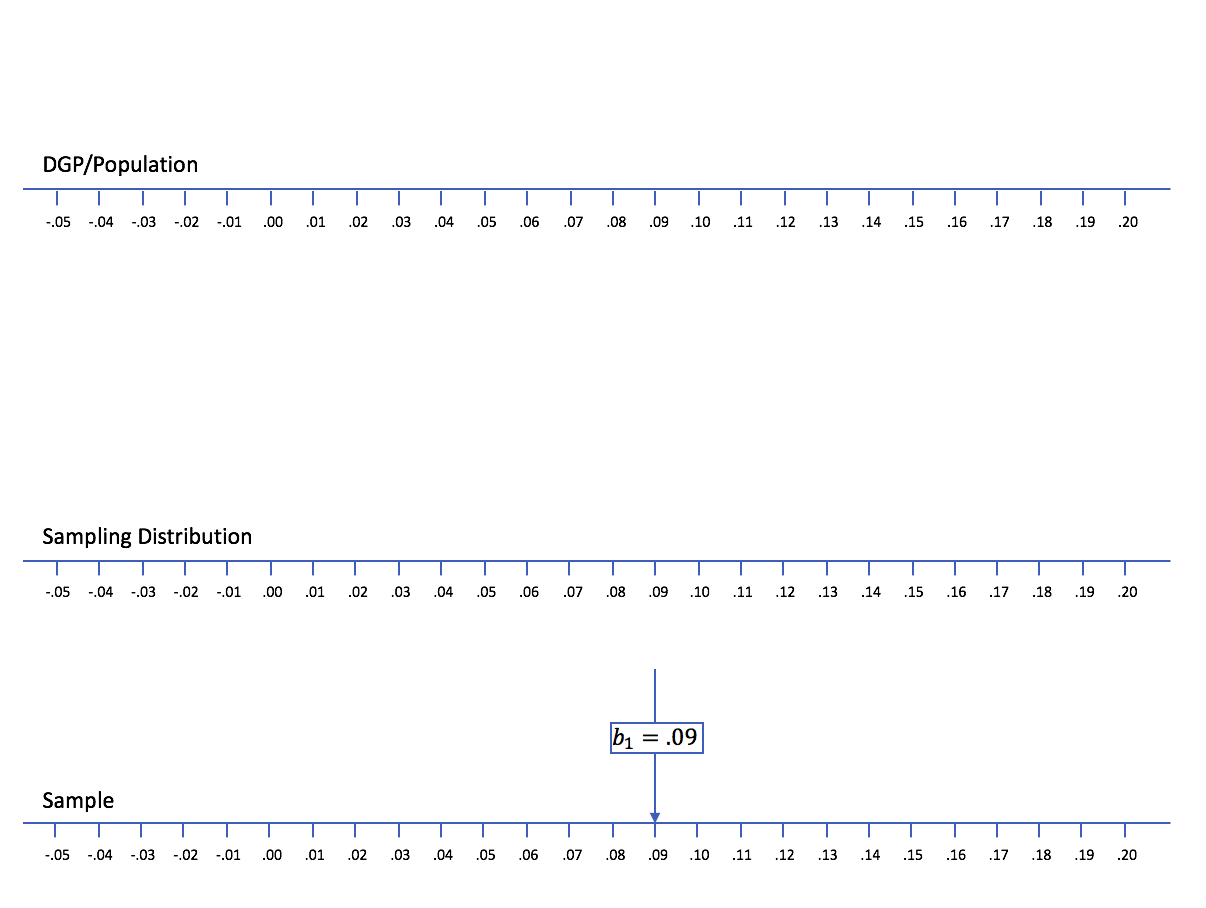
|  |  |
| --- | --- |
| Word equation |  |
| Making visualizations with R code:  faceted histogram  jitter plot |  |
| Specifying the model in GLM (no numbers) |  |
| Finding best fitting parameter estimates with R code |  |
| Writing best fitting GLM model (with numbers) |  |
| How much variation have we explained with this model? |  |
| Our best point estimate of |  |
| Our 95% confidence interval of |  |

1. What is the difference between these two sampling distributions?

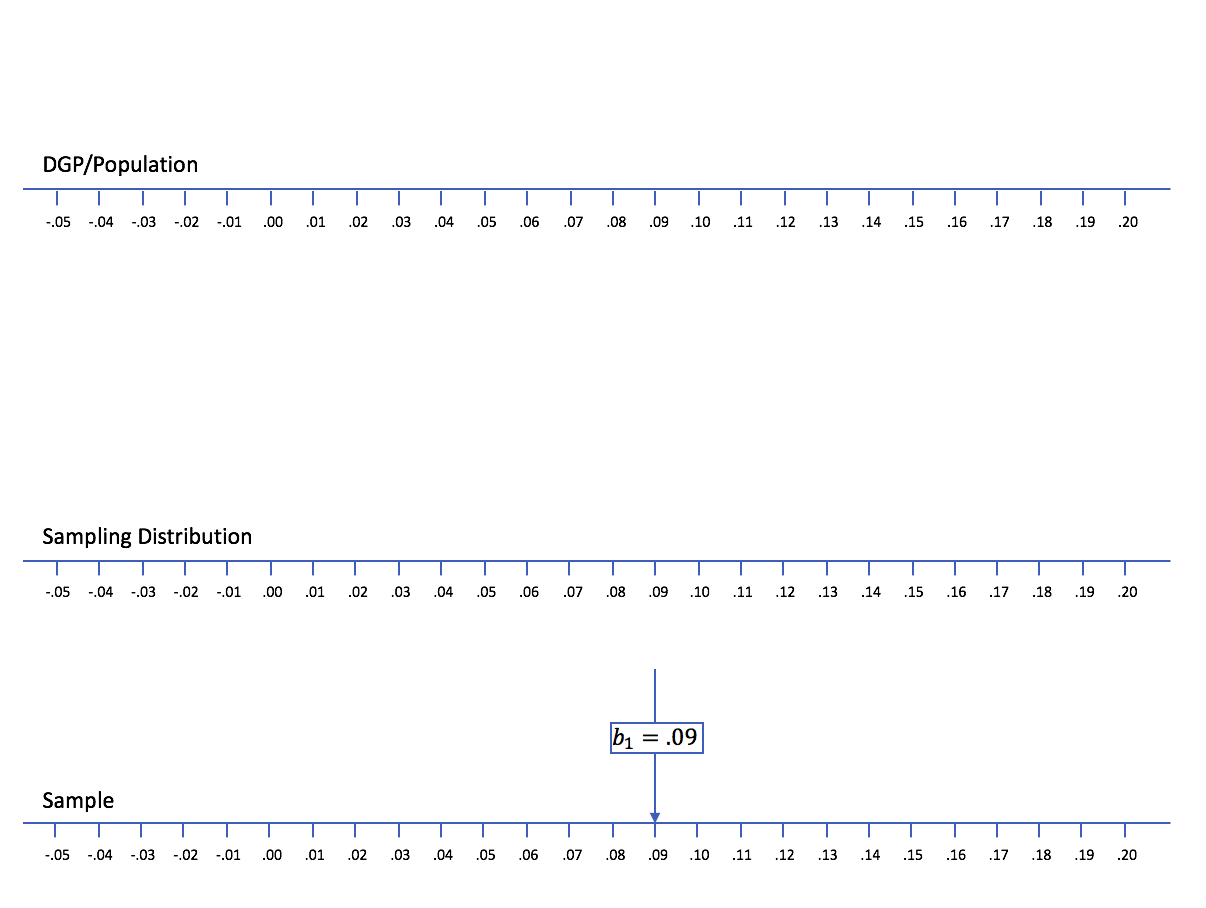
do(1000) \* b1(veg\_eaten ~ respect\_condition, data = resample(YeagerData, 160))

do(1000) \* b1(veg\_eaten ~ shuffle(respect\_condition), data = YeagerData)

1. Use the space below to draw out what we did with the cut out sampling distributions. Using the cut out sampling distributions and your worksheet, let’s figure out how to understand the lowest number in the confidence interval. Draw what you did in the space below and provide an explanation.



1. Now let’s do the same for the highest number in the confidence interval. Draw what you did with the cut out sampling distributions in the space below and provide an explanation.



# Name:

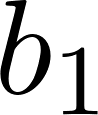
# Classwork 20: **Presidential Tweets**

Let’s start with a data set called **twitter\_presidents** from [FiveThirtyEight](https://fivethirtyeight.com/features/higher-rates-of-hate-crimes-are-tied-to-income-inequality/). If you want to know more information about these variables at any time, google: **twitter\_presidents r documentation**.

1. Go ahead and use R to look at the first six lines of this data frame. What are the cases in this data frame?
2. Just out of curiosity, who had the most favorited tweet in this data set?
3. As data scientists – we want to *explain variation*. In the twitterverse, if you “retweet,” you project a tweet out to your network and it can potentially reach more viewers. Do you think the number of retweets a tweet has might help us predict how many “favorites” (hearts) it receives? Why or why not?
4. Let’s run through all our skills with our retweeting hypothesis!

|  |  |
| --- | --- |
| Word equation |  |
| Making visualizations with R code: |  |
| Specifying the model in GLM (no numbers) |  |
| Finding best fitting parameter estimates with R code |  |
| Writing best fitting GLM model (with numbers) |  |
| How much variation have we explained with this model? |  |

1. Is it possible to have gotten this pattern of data even if there was no relationship between retweets and favorites in the DGP? Is it likely?
2. Is it possible to have gotten this pattern of data if the relationship between retweets and favorites in the DGP was basically like our sample? Is it likely?
3. We have been exploring two random processes of generating data in our class: shuffle and resampling. Which of these processes is like #5? Which one is like #6?
4. Which of these processes is like the “empty model”?
5. If the true DGP was basically like the empty model, what would be the true value of ?
6. Can we ever know the true value of ?
7. The first scatterplot is the real data. Are the rest shuffled or resampled data?

In each box, try to estimate what the [](https://www.codecogs.com/eqnedit.php?latex=b_1%250) is.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |