

## Problem set 8

Your name here

Due 11/19/2021 at 5pm

NOTE: Start with the file `ps8_2021.Rmd` (available from the github repository at <https://github.com/UChicago-pol-methods/IntroQSS-F21/tree/main/assignments>). Modify that file to include your answers. Make sure you can “knit” the file (e.g. in RStudio by clicking on the **Knit** button). Submit both the Rmd file and the knitted Pbb via Canvas.

In this assignment we will return to data from an experiment that measured the effect of constituent names in emails on legislator replies. The published paper is:

Butler, D. M., & Broockman, D. E. (2011). *Do politicians racially discriminate against constituents? A field experiment on state legislators*. AJPS.

The data file is `Butler_Broockman_AJPS_2011_public_csv.csv` and it is found in the `data/legislators_email` directory of the course github repository.

To load the data you can either download and read in the local file, or you can read in the url from github. Note that reading in by the url will only work when you have an internet connection:

```
file <- 'https://raw.githubusercontent.com/UChicago-pol-methods/IntroQSS-F21/main/data/legislators_email.csv'
bb <- read_csv(url(file))
```

### Question 1: Inference from a single random variable

(1a) Create an object called `theta_hat` which is the mean of the `reply_atall` variable in the data set.

```
# your code here
```

(1b) Create an object called `se_hat` which is the estimate of the standard error of the mean of the `reply_atall` variable in the data set, using the formula based on the unbiased sample variance.

```
# your code here
```

(1c) The formula for the normal approximation-based confidence intervals is below

$$CI_n = \left( \hat{\theta}_n - z_{1-\alpha/2} \times \hat{se}, \hat{\theta}_n + z_{1-\alpha/2} \times \hat{se} \right)$$

$z_c$  describes the  $c$ -th quantile of the standard normal distribution. For 95% confidence intervals,  $\alpha = 0.05$ , so we want to find  $z_{1-\alpha/2} = z_{0.975}$ . Using `qnorm`, get the 97.5-th quantile of the standard normal distribution.

```
# your code here
```

(1d) Using `theta_hat`, `se_hat`, and your answer to the previous question, report the 95% normal approximation-based confidence intervals for the estimate of `theta_hat`

```
# your code here
```

(1e) Interpret what the 95% confidence interval means.

(1f) To get the 90% confidence intervals, we will set  $\alpha$  as 0.10. So we want to find  $z_{1-\alpha/2} = z_{0.95}$ . Using `qnorm`, get the 95-th quantile of the standard normal distribution.

```
# your code here
```

(1g) Using your answer from the question above, report the 90% normal approximation-based confidence intervals for the estimate of `theta_hat`.

```
# your code here
```

(1h) Create a vector of 1000 bootstrapped estimates of the sample mean of `reply_atall`. Save this vector as an object. Report the standard deviation across the estimates. The standard deviation of your bootstrapped estimates should be similar to your answer to 1b above.

*Note: This should look very much like your solution to (2e) on hw 7, but you should be sampling with replacement from `bb$reply_atall`.*

```
# your code here
```

(1i) We can compare the distribution of the estimator under the bootstrap procedure and under the normal approximation. Using the `quantile()` function and your saved vector of 1000 bootstrapped estimates of the sample mean, report the 2.5th and 97.5th quantiles of the estimates under the bootstrap. These cover 95% of the empirical distribution of the bootstrap. How do they compare to your 95% normal approximation-based confidence intervals in your answer to 1d above?

```
# your code here
```

## Question 2: Inference from linear models

(2a) Using `lm_robust`, regress `reply_atall` on `treat_deshawn` interacted with `leg_republican`. Print the model object. Save the vector of coefficients as `theta_hats`.

```
# your code here
```

(2b) From the model object above, report and interpret the standard errors and 95% confidence intervals on `treat_deshawn` and `treat_deshawn:leg_republican`. Do the confidence intervals include zero? If so/if not, what does that imply?

(2c) Using `map()` and `slice_sample()`, `replace = TRUE`, take 1000 bootstrap re-samples with replacement of the same size as the original data from the `bb` dataset. Save your bootstrapped samples as an object.

```
# your code here
```

(2d) Using `map()` again, run the same regression as above on *each* of your bootstrapped samples; extract coefficient estimates; and use `bind_rows()` to create a matrix where each row represents estimates from one of your bootstrap samples, and each column is one of the coefficients.

```
# your code here
```

(2e) Report the bootstrapped estimates of the standard errors of each of the coefficients. To do this, get the standard deviations of each of the columns.

```
# your code here
```

(2f) Produce normal approximation-based confidence intervals for each of the coefficients using the bootstrapped standard errors, inserted into the same formula for confidence intervals as presented in 1c. Compare these to the standard errors from your original `lm_robust()` model object in question 2a.

```
# your code here
```