1 Item 1

1.1 Algorithm

Step 1: Let B be the number of bootstrap samples taken. With $n = 15$, do SRSWR from schools 1 to 15.

$$\begin{aligned} B_1^* &= \{(X_{11}^*, Y_{11}^*), (X_{21}^*, Y_{21}^*), \ldots, (X_{n1}^*, Y_{n1}^*)\} \\ B_2^* &= \{(X_{12}^*, Y_{12}^*), (X_{22}^*, Y_{22}^*), \ldots, (X_{n2}^*, Y_{n2}^*)\} \\ \vdots &= \vdots \\ B_B^* &= \{(X_{1B}^*, Y_{1B}^*), (X_{2B}^*, Y_{2B}^*), \ldots, (X_{nB}^*, Y_{nB}^*)\} \end{aligned} \tag{1.1}$$

Step 2: Let $S_{x_b}^*$ and $S_{y_b}^*$ be the standard deviations of the variables, X_b^* and Y_b^*, respectively, where $b = \{1, 2, \ldots, B\}$. Calculate the pearson product coefficient of correlation, r_b^*

$$r_b^* = \frac{\frac{1}{n-1} \sum \left(X_{ib}^* - \bar{X}_b^*\right) \left(Y_{ib}^* - \bar{Y}_b^*\right)}{S_{x_b}^* S_{y_b}^*} \tag{1.2}$$

to yield

$$r = \{r_1^*, r_2^*, \ldots, r_B^*\} \tag{1.3}$$

Step 3: Calculate $\widehat{se}(r)$ using

$$\widehat{se}(r) = \sqrt{\frac{\sum_{b=1}^{B} \left(r_i^* - \bar{r}^*\right)^2}{B-1}} \tag{1.4}$$

1.2 Algorithm implementation

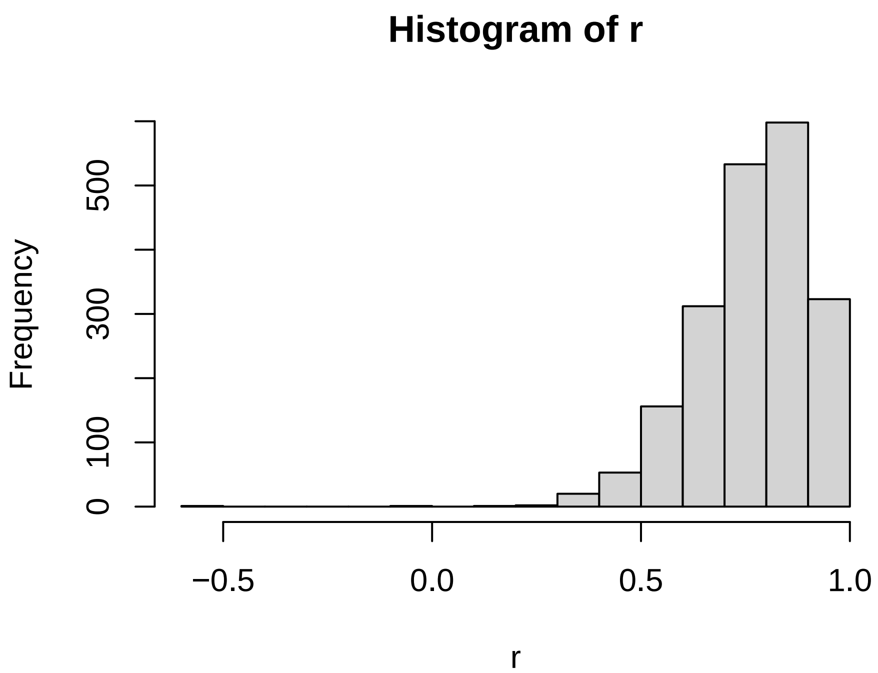
```
# set seed for reproducibility
set.seed(7)

r <-
  # Step 2: take the pearson correlation for each bootstrap sample
  sapply(
    # Step 1: 2000 bootstrap samples (with replacement) of size 15
    lapply(
      1:2000,
      function(x){ law_school_data[sample(law_school_data$School,
                                          15,
                                          replace = TRUE),
                                   c(2,3)] }
    ),
    function(x){cor(x$LSAT,x$GPA, method = "pearson")}
  )

# Step 3: take the bootstrap estimate of the standard error
se_r_boot <- sd(r)

# take the percentile 95% CI for r
ci_r_boot <- c(quantile(r,.025),quantile(r,.975))
```

i. bootstrap estimate of the standard error of r: 0.1369
ii. 95% confidence interval for ρ (the true population correlation): (0.451, 0.9625)
iii. a histogram showing the bootstrap distribution of the correlation r



1.3 Change maximum r_b^*

```
source("../R/s02_i01_bs_sampling.R")

r_max <- max(r)

r_replaced <- replace(r, r==r_max, 100*r_max)
r_replaced_max <- max(r_replaced)
se_r_replaced_boot <- sd(r_replaced)

se_percent_change <- ((se_r_replaced_boot - se_r_boot)/se_r_boot)*100
```

The original maximum of r_b^*'s is 0.9937 while the new one is 99.3656. Calculating the new $\widehat{se}(r)$ gives the value 2.209. This meant an increase of 1514.0138% compared to the original value.

2 Item 2

2.1 Fill in the table

```
    5%     10%     15%     20%     50%     70%     85%     90%     95% 
0.5244 0.5859 0.6224 0.6588 0.7878 0.8517 0.9044 0.9228 0.9472 
[1] 0.9937
[[1]]
   95%
```

```
0.3477

[[2]]
   90%
0.2159

[[3]]
   85%
0.1461

[[1]]
   95%
0.3477

[[2]]
   90%
0.2159

[[3]]
   85%
0.1461

[1] 99.37
```

2.2 Compute $\tilde{se}_\alpha(r)$

2.3 Change maximum r_b^* and recompute

3 Item 3

3.1 Algorithm: $se(\hat{\beta}_2)$ estimation

Step 1: Let $\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2), i = 1, \ldots, 24$. Under the model, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$, estimate

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \tag{3.1}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)^2 \tag{3.2}$$

Step 2:

a. Repeat B times: Let $e_i^* \sim N(0, \hat{\sigma}^2), i = 1, \ldots, n$. Compute $y_i^* = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + e_i^*, i = 1, \ldots, n$
b. Obtain $\hat{\beta}_2$ from the B OLS estimates for each b bootstrap dataset.

$$\hat{\boldsymbol{\beta}}_b^* = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}_b^* \tag{3.3}$$

3.2 Algorithm implementation: $se(\hat{\beta}_2)$ estimation

```
[,1]
```

```
[1,] 17.8469
[2,]  1.1031
[3,]  0.3215
[4,]  1.2889
```

```
[1] 0.03395
```

```
      [,1]
[1,] 3.072
```

```
          [,1]      [,2]      [,3]      [,4]
[1,]  4.007506 -0.312971  0.001358 -0.373510
[2,] -0.312971  0.108619 -0.005144 -0.023493
[3,]  0.001358 -0.005144  0.001377 -0.001364
[4,] -0.373510 -0.023493 -0.001364  0.089090
```

```
[1] 2.00188 0.32957 0.03711 0.29848
```

3.3 Algorithm: $\frac{\hat{\beta}_1}{\hat{\beta}_3}$ 95% CI estimation

Repeat step 1 up to step 2.a. of the first part.

b. Obtain $\frac{\hat{\beta}_1}{\hat{\beta}_3}$ from the B OLS estimates for each b bootstrap dataset.

Step 3: Calculate the quantiles to get the 95% confidence interval.

3.4 Algorithm implementation: $se(\hat{\beta}_2)$ estimation

```
  2.5%  97.5%
0.3492 1.9545
```

We are 95% confident that the true value of the ratio is between 0.3492 and 1.9545

References

Fox, Jean-Paul, and Sukaesi Marianti. 2017. “Person-Fit Statistics for Joint Models for Accuracy and Speed.” *Journal of Educational Measurement*.