1. Bootstrap (Parametric/Nonparametric)

SM Shafqat Shafiq

2023-08-05

1. Parametric Bootstrap

```
set.seed(123)
```

1.1 Generate data from Normal Distribution with mu = 5, sigma_square = 1

```
myData <- rnorm(10, mean = 5, sd = 1)
theta <- mean(myData)
```

1.2 Generate a bootstrap matrix to hold the values of the bootstrap estimates

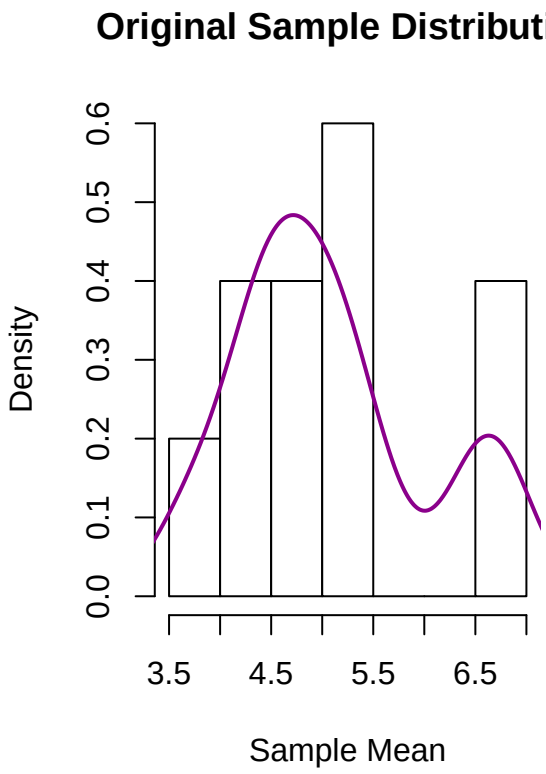
```
boot_matrix <- matrix(NA, 10, 500)
boot_data <- c()
boot_column <- c()
```

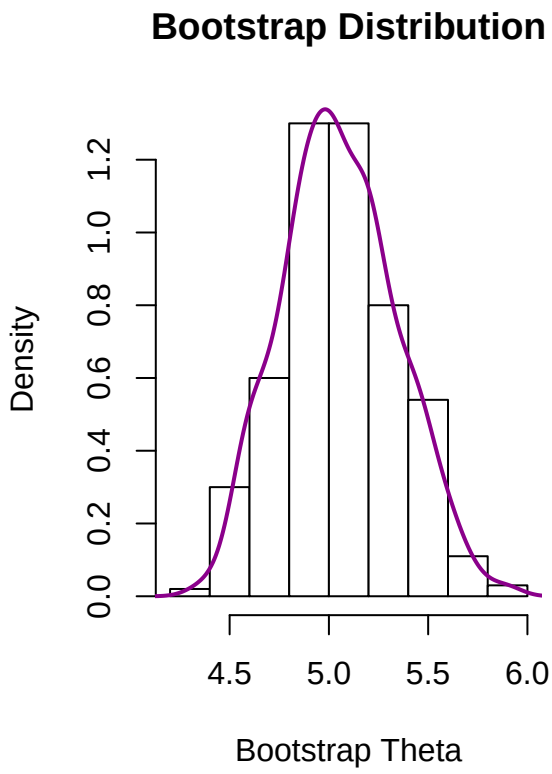
1.3 Use a for-loop to calculate the thetas for 500 re-samples

```
for (m in 1:500) {

  i <- sample(1:length(myData), length(myData), replace = TRUE)
  boot_data <- myData[i]
  boot_column[m] <- mean(boot_data)

}
```

1.4 Plot histogram

```
par(mfrow = c(1, 2))

hist(myData, xlab = "Sample Mean", col = "white",
     main = "Original Sample Distribution", prob = TRUE)
lines(density(myData), lwd = 2, col = "darkmagenta",
      main = "Density of Original Sample")

hist(boot_column, xlab = "Bootstrap Theta", col = "white",
     main = "Bootstrap Distribution", prob = TRUE)
lines(density(boot_column), lwd = 2, col = "darkmagenta",
      main = "Density of Theta")
```





1.5 Calculate Bootstrap Theta

```
theta_hat_mean <- mean(boot_column)
```

1.6 Calculate Bias = Bootstrap Theta - Mean

```
theta_hat_mean - theta
```

```
## [1] -0.02083852
```

2. Nonparametric Bootstrap

2.1 Load dataset

```
library(bootstrap)
```

2.2 Calculate true sample statistics for LSAT and GPA for the 15 schools

```
print(c(mean(law$LSAT), mean(law$GPA)))
```

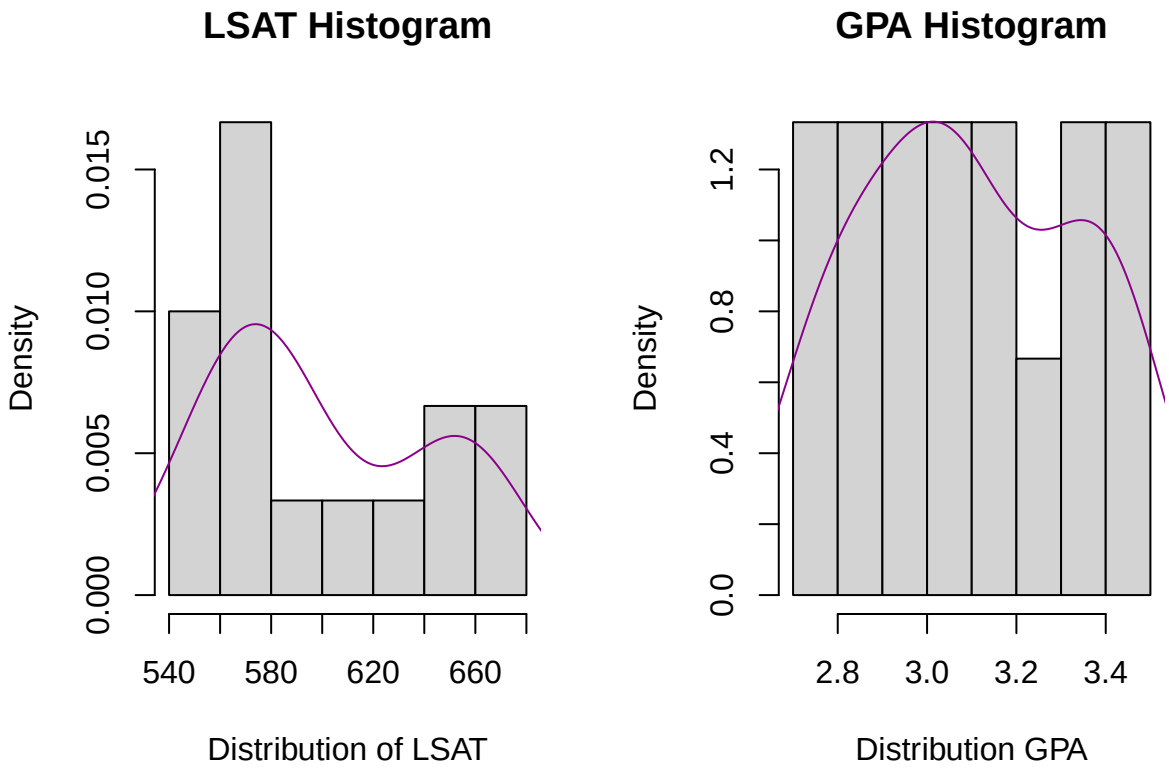
```
## [1] 600.266667   3.094667
```

```
print(cor(law$LSAT, law$GPA))
```

```
## [1] 0.7763745
```

2.3 Plot LSAT, GPA Histograms

```
par(mfrow = c(1, 2))
hist(law$LSAT, prob = TRUE, main = "LSAT Histogram",
     xlab = "Distribution of LSAT")
lines(density(law$LSAT), col = "darkmagenta")
hist(law$GPA, prob = TRUE, main = "GPA Histogram",
     xlab = "Distribution GPA")
lines(density(law$GPA), col = "darkmagenta")
```



2.4 Set up bootstrap for the following:

```
# 1. Bootstrap means for LSAT and GPA
# 2. Bootstrap standard error
```

```
M <- 200                  # Bootstrap replicates
n <- nrow(law)            # Sample size
```

2.5 Vectors to hold the bootstrap values

```
R <- numeric(M)
mean_LSAT <- numeric(M)
mean_GPA <- numeric(M)
```

2.6 For loop to calculate bootstrap means of LSAT and GPA scores

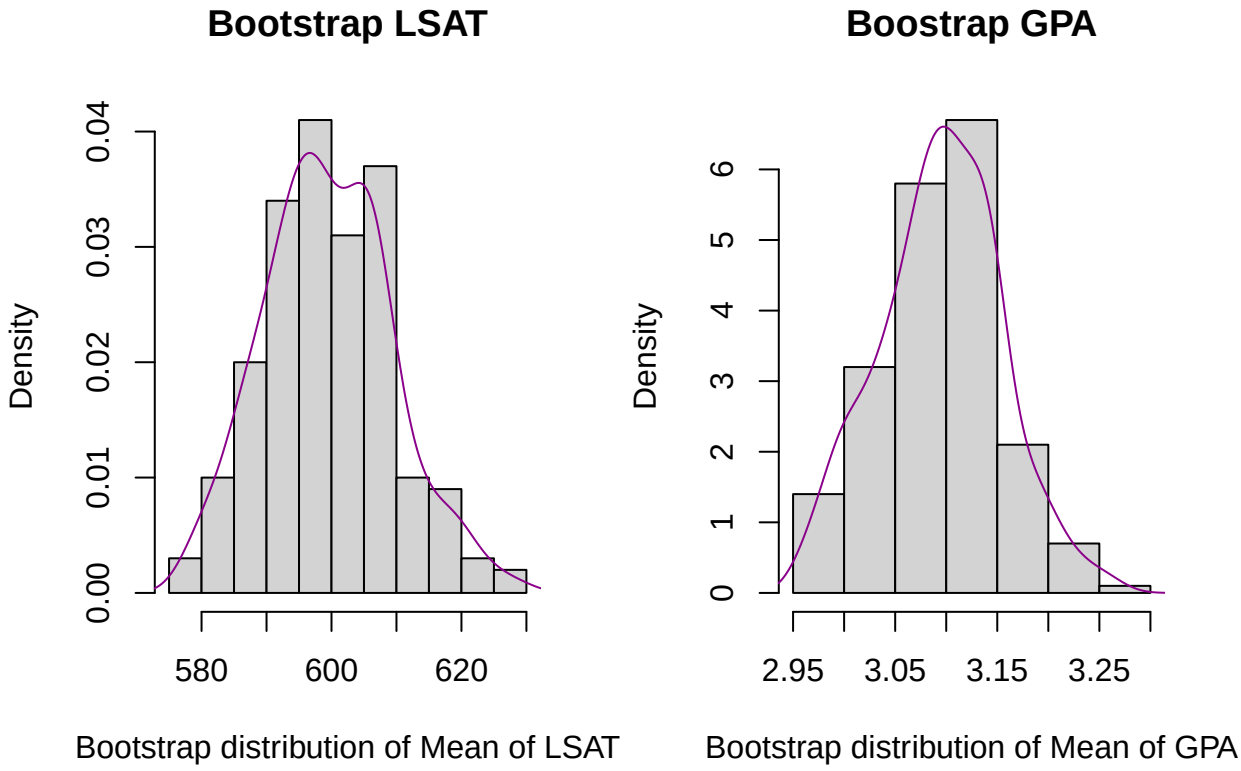
```
for (m in 1:M) {
  i <- sample(1:n, n, replace = TRUE)
  LSAT <- law$LSAT[i]
```

```
  GPA <- law$GPA[i]

  mean_LSAT[m] <- mean(LSAT)
  mean_GPA[m] <- mean(GPA)
}

print(c(mean(mean_LSAT), mean(mean_GPA)))
```

```
## [1] 599.480333   3.094617
```

2.7 Plot

```
par(mfrow = c(1, 2))
hist(mean_LSAT, prob = TRUE, main = "Bootstrap LSAT",
     xlab = "Bootstrap distribution of Mean of LSAT")
lines(density(mean_LSAT), col = "darkmagenta")
hist(mean_GPA, prob = TRUE, main = "Boostrap GPA",
     xlab = "Bootstrap distribution of Mean of GPA")
lines(density(mean_GPA), col = "darkmagenta")
```



2.8 For loop to calculate standard error of Bootstrap correlation

```
for (m in 1:M) {

  i <- sample(1:n, size = n, replace = TRUE)
  LSAT <- law$LSAT[i]
  GPA <- law$GPA[i]
```

```
  R[m] <- cor(LSAT, GPA)

}

print(sd(R))
```

```
## [1] 0.1279363
```

2.9 Plot

```
hist(R, prob = TRUE)
lines(density(R), col = "darkmagenta")
```