CH5_QR_RMD

Fasih Atif

10/26/2020

Introduction to Analysis

I downloaded the GDP per capita and CO2 emissions per capita data from World Bank Indicators website. I took values of 2016 since there was no info on CO2 emissions per capita beyond 2016. I imported both tables and merged them together. There were a total of 264 countries(observations). I then filtered for missing values which left me with 238 observations.I then arranged the data in ascending order of gdp and split the data into 2 groups with the help of the median. All the countries with GDP per capita above and equal 5555.335 were assigned group 2 while the rest were assigned group 1.

h0: There is no difference in average CO2 emissions per capita between the two groups

ha: There is a difference in average the CO2 emissions per capita between the the two groups

Here is what the data looked like:

```
head(gdp_co2_data,3)
```

```
##                    country      gdp        co2  group
## 1                  Burundi 282.1931 0.04720110 group1
## 2                   Malawi 315.7773 0.07544878 group1
## 3 Central African Republic 402.1913 0.06545780 group1
```

```
tail(gdp_co2_data,3)
```

```
##           country       gdp       co2  group
## 236   Switzerland  80172.23  4.117490 group2
## 237    Luxembourg 104278.39 15.442613 group2
## 238 Liechtenstein 165629.19  1.363269 group2
```

I then calculated the means of each group

```
group1_mean <- mean(gdp_co2_data$co2[gdp_co2_data$group == 'group1'])
group2_mean <- mean(gdp_co2_data$co2[gdp_co2_data$group == 'group2'])
```

I calculated the difference in observed mean of the 2 groups

```
mean_diff <- group2_mean - group1_mean
```

I proceeded to create the Bootstrap procedure with 10000 samples

```r
set.seed(13456)
n_group1 <- 119
n_group2 <- 119

# No of boostrap samples
B <- 10000
```

I converted the Boot-sample into a column

```r
boot_group1 <- matrix(sample(gdp_co2_data$co2[gdp_co2_data$group == 'group1'],
size = B*n_group1, replace = TRUE),ncol = B,nrow = n_group1)

boot_group2 <- matrix(sample(gdp_co2_data$co2[gdp_co2_data$group == 'group2'],
size = B*n_group2, replace = TRUE),ncol = B, nrow = n_group2)
```

I calculated the difference in MEANS for each of the bootsamples

```r
Boot_diff_in_means <- colMeans(boot_group1) - colMeans(boot_group2)
```

The "PERCENTILE" bootstrap confidence interval (95%)

```r
quantile(Boot_diff_in_means, prob = 0.025)
```

```
##      2.5%
## -7.270373
```

```r
quantile(Boot_diff_in_means, prob = 0.975)
```

```
##     97.5%
## -4.932408
```

The standard Error

```r
std <- sd(gdp_co2_data$co2)
SE <- std/238^1/2
SE
```

```
## [1] 0.01167965
```

Result

We are 95% confident that the true/population difference in means of CO2 emissions per capita between the two groups is somewhere between -7.27 and -4.93. We reject null hypothesis.