STAT 463 Spring, 2020

Group #21: Jack Fink, Izzat Bin Ahmad Adly,Huanyu Liu, Zeping

Project Assignment

The data set we will be using for the report is the World Happiness Report compiled on Kaggle. The rankings and happiness scores come from a Gallup World Poll. The reference for the dataset can be found in the appendix. The scores are based on the self-evaluations of the population samples from each of their respective countries. In the preparation of the data, they used Gallup weights to make the estimates representative. Even though there are 6 columns that make up the score, we will be only looking at the happiness score for each country for our report. We will also be looking at the scores in 2017.

The population of our assignment is the self-evaluation scores of the samples chosen from their specific countries. The single random variable we are looking at is the happiness scores for each country. It seems like the data fits a uniform distribution because the probability of getting any one score is the same as getting any other score. The research question we are trying to answer in this report is what the average happiness score of the population. The underlying characteristics of the research question is how happy each citizen with respect to their countries. The research question is interesting because it helps build a greater understanding of where the happiness score for your country stands in comparisons to other countries around the world. This can also give more information about why your country's happiness scores that way and whether you should take drastic action or not.

The two estimators that we are using are the sample average and the sample midrange. These estimators are good for the data set because both are different ways in getting the average happiness of the population. Based on the data set, our estimates for the sample average is 5.35 and the estimates for our midrange is 5.12.

The bias of estimator by sample midrange is -0.00575695 and the bias of estimator of sample average is -0.2303889 when we repeat the simulation 1000 times. The variance of estimator by sample midrange is 0.0005356781 and the variance of estimator by sample average is 0.01255611. This means that the bias of the sample midrange is closer to zero compared to the sample average. We can also see that the variance of the sample midrange is less than the variance of the sample average. So, in conclusion, the sample midrange is a better estimator than the sample average. Thus, we find the sample midrange is a better estimator of the average happiness of the population.

Appendix

https://www.kaggle.com/unsdsn/world-happiness#2017.csv​

# n = 155   
# m = 1000  
# Uniform distribution: runif  
  
## Midrange estimator #######################################################  
n1 = 155  
l = 2.693  
h = 7.537  
  
estimator\_Midrange <- function(n, lowest, highest)  
{  
 y = runif(n, min = lowest, max = highest) # simulate n samples from Uniform(lowest, highest)  
 output = (max(y) + min(y))/2 # calculate the midrange of the sample  
 return(output)  
}  
  
est\_para1 = 5.12  
  
m = 1000  
  
### run estimator\_uniform(n, 2.693, 7.537) for m times  
estimates\_midrange = replicate(m,estimator\_Midrange(n=n1, l, h))  
  
mean(estimates\_midrange)

## [1] 5.114243

(bias\_midrange = mean(estimates\_midrange)-est\_para1)

## [1] -0.00575695

(var\_midrange = var(estimates\_midrange))

## [1] 0.0005356781

(mse\_midrange = mean((estimates\_midrange-est\_para1)^2))

## [1] 0.0005682849

## average estimator ########################################################  
  
estimator\_avg <- function(n, lowest, highest)  
{  
 y = runif(n,min = lowest, max = highest) # simulate n samples from Uniform(lowest, highest)  
 output = mean(y) # calculate the mean of the sample  
 return(output)  
}  
  
est\_para2 = 5.35  
  
### run estimator\_uniform(n, 2.693, 7.537) for m times  
estimates\_avg = replicate(m, estimator\_avg(n=n1, l, h))  
  
(bias\_avg = mean(estimates\_avg)-est\_para2)

## [1] -0.2303889

(var\_avg = var(estimates\_avg))

## [1] 0.01255611

(mse\_avg = mean((estimates\_avg-est\_para2)^2))

## [1] 0.06562261

## Interpretation ###########################################################  
  
bias\_midrange

## [1] -0.00575695

bias\_avg

## [1] -0.2303889

var\_midrange

## [1] 0.0005356781

var\_avg

## [1] 0.01255611