Lab 9 Assignment

Dr. Purna Gamage

Problem 1

Load the Bangladesh data. We made inference about Arsenic in water in the Lab 9. Let's make inference about Cobalt.

```
bdesh <- read.csv("Bangladesh.csv")
head(bdesh)
```

```
##   Arsenic Chlorine Cobalt
## 1    2400      6.2   0.42
## 2       6    116.0   0.45
## 3     904     14.8   0.63
## 4     321     35.9   0.68
## 5    1280     18.9   0.58
## 6     151      7.8   0.35
```

a. Plot the boxplot of Cobalt. What's look wrong? Do you think a transformation can explain the data better?

b. Make a log transformation of the data (Cobalt) and plot the boxplot. What do you see differently? Comment.

c. Are there an missing values? If so, remove the missing values and draw bootstrap samples(10000) of sample mean. (remeber we are still using the log transformation)

d. What is the bias?

e. Find the standard error of the bootstrap estimate.

Also, calculate the estimate(using the bootstrap sample) of mean squared error(MSE) =var(X)+ (bias)ˆ2

Comment on your results.

f. Make a 95% Percentile Bootstrap CI for mean of log transformed(log10) Cobalt data. Comment on your results.

g. Repeat the bootstrap sampling process to estimate the median of the actual data(Cobalt- no transformation).

Find the bias and the standard error of the bootstrap estimate. Comment on your results.

Problem 2: Hypothesis testing using bootstrap approach.

(Chihara2011)

Import the data set Titanic.csv which contains survival data (0 = death, 1 = survival) and ages of 658 passengers of the Titanic which sank on April 15, 1912 (the day when Americans had to file income tax returns for the first time). Examine the null hypothesis that the mean ages of survivors and of victims are the same against the alternative that these mean ages are different, using a bootstrap approach.

```
Titanic = read.csv("Titanic.csv",stringsAsFactors=FALSE)
head(Titanic)
```

```
##   ID Survived   Age
## 1  1        1  0.92
## 2  2        0 30.00
## 3  3        1 48.00
## 4  4        0 39.00
## 5  5        0 71.00
## 6  6        0 47.00
```

Let's make a hypothesis. Are two population means different?

Null Hypothesis: Mean ages of survivors and of victims are the same. (That means that there is no difference between the two population means)

Let's use the bootstrap approach.

a. First let's seperate the two samples "survivors(=1)" and "victims(=0)".

b. Make many bootstrap samples(100000) of difference of sample means.

c. Make a 90% percentile bootstrap confidence interval for the difference estimate. Using this CI, conclude on your hypothesis.

Null Hypothesis: Mean ages of survivors and of victims are the same. (That means that there is no difference between the two population means)

d. Make a 95% percentile bootstrap confidence interval for the difference estimate and using this CI, conclude on your hypothesis.

Null Hypothesis: Mean ages of survivors and of victims are the same. (That means that there is no difference between the two population means)