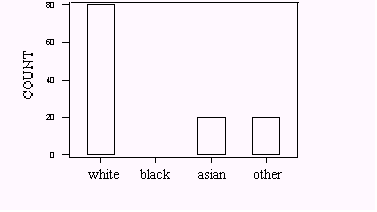
**DS 510 – Week 3**

**Assignment**

Mir Ahmed

1. A sample of **160** workers in the downtown area classified each worker by race. A bar graph of the results is given below, but the bar for blacks in the graph below has been omitted.

Using the information provided, the *proportion* of black workers in the sample must be:



Solution:

From the graph, White workers (W) = 80

Asian workers (A) = 20

Other workers (O) = 20

We know, total workers (T) = 160

So, black workers (B) = T – W – A – O

= 160 – 80 – 20 – 20

= 40

So, proportion of black workers = B/T = 40/160 = 0.25 = 25%

1. The following is a sample of the weights of 12 boxers from various weight classes:

121 173 157 165 170 161 142 171 184 100 145 196

What is the median value of this data set?

Solution:

Let’s sort the values in an ascending order:

100 121 142 145 157 161 165 170 171 173 184 196

Total number of elements = 12

So, the median = (6th element + 7th element)/2

= (161 + 165)/2

= 163

In R:

arr = c(121, 173, 157, 165, 170, 161, 142, 171, 184, 100, 145, 196)

median(arr)

1. The following is a sample of the weights of 12 boxers from various weight classes:

121 173 157 165 170 161 142 171 184 100 145 196

How many values are going to be considered outliers?

Solution:

Let’s sort the values in an ascending order:

100 121 142 145 157 161 165 170 171 173 184 196

Q1 Median Q3

From previous question, The median (M) = 163

So,

First quartile, Q1 = (142+145)/2 = 143.5

Third quartile, Q3 = (171+173)/2 = 172

IQR = Q3 – Q1 = 172 – 143.5 = 28.5

Lower bound = Q1 – 1.5\*IQR = 143.5 – 1.5\*28.5 = 100.75

Upper bound = Q3 + 1.5\*IQR = 172 + 1.5\*28.5 = 214.75

According to the lower bound value, 100 is an outlier. So only one value will be considered outlier.

In R:

arr = c(121, 173, 157, 165, 170, 161, 142, 171, 184, 100, 145, 196)

boxplot.stats(arr)$out

length(boxplot.stats(arr)$out)

1. Too much cholesterol in the blood increases the risk of heart disease. The cholesterol levels of young women aged 20 to 34 vary approximately Normally with mean 185 milligrams per deciliter (mg/dl) and standard deviation 39 mg/dl. About what percent of young women in this age group will have cholesterol levels less than 150 mg/dl?

Solution:

Here, mean ( = 185

Standard deviation (σ) = 39

X = 150

We know, σ

So, Z = (150 – 185)/39 = -0.8974359 = - 0.9 (rounded up to 2nd place)

Now, using the Z-score table, P(Z) = .1841

So, 18.41% of young women in this age group will have cholesterol levels less than 150 mg/dl

In R:

pnorm((150-185)/39)

0.1847432

So, ~18.47% of young women in this age group will have cholesterol levels less than 150 mg/dl