**Descriptive Statistics**

Exercises for central tendency and dispersion

Open the Height and Weight Data.csv file in R.

> file = read.csv("Height and Weight Data.csv") # reads file

> HtWt = with(file, data.frame(height, weight)) # data frames are useful for many tests

> HtWt = na.omit(HtWt) # removes missing data

1. Calculate descriptive statistics (mean, median, range) for two variables, height and weight.

You could use something like this:

> mean(HtWt$weight)

[1] 65.8

and do the same with median() and range(). And then do it all again for HtWt$height.

This is quicker:

> library(psych)

> describe(HtWt)

 vars n mean sd median trimmed mad min max range skew kurtosis

height 1 200 170.02 12.01 169.5 170.32 9.64 57 197 140 -4.00 37.07

weight 2 200 65.80 15.10 63.0 64.21 11.86 39 166 127 2.01 8.96

 se

height 0.85

weight 1.07

Or

> summary(HtWt)

 height weight

 Min. : 57.0 Min. : 39.0

 1st Qu.:164.0 1st Qu.: 55.0

 Median :169.5 Median : 63.0

 Mean :170.0 Mean : 65.8

 3rd Qu.:177.2 3rd Qu.: 74.0

 Max. :197.0 Max. :166.0

1. Split the data by 'Sex' and calculate the same descriptive statistics. It is suggested that you look up the Psych library's describeBy function.

> HtWtSex = with(file, data.frame(height, weight, sex)) # includes 'sex' variable

> library(psych)

> describeBy(HtWtSx, HtWtSex$sex)



1. Produce histograms for the height and weight variables and decide if the distribution shapes resemble that of the Gaussian distribution. It is suggested that you use the hist() function.

> hist(HtWtSx$height)

> hist(HtWtSx$weight)





These do not resemble a Gaussian distribution.