# Descriptive Statistics

Dr Sreya R

2024-09-21

In this section, we will deal with loading an Excel file named Sales into the R Studio Console and do the descriptive statistics. Importing Excel file to R studio

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.0.5
```

```
Sales <- read_excel("Dataset_MA/Sales.xlsx")
View(Sales)
```

In order to identify the variables in the dataset, we will first use the names() function.

```
names(Sales)
```

```
##  [1] "Supplier"           "Order No."          "Item No."
##  [4] "Item Description"   "Item Cost"          "Quantity"
##  [7] "Cost per order"     "A/P Terms (Months)" "Order Date"
## [10] "Arrival Date"       "Order size"
```

We then calculate the mean, median, minimum, maximum, quantiles, standard deviation, variance, skewness and kurtosis for a particular variable. Skewness and kurtosis functions are found in the moments package and it needs to be installed. Rest of the functions are found in the base package of R.

```
mean(Sales$Quantity)
```

```
## [1] 5857.404
```

```
median(Sales$Quantity)
```

```
## [1] 1915
```

```
min(Sales$Quantity)
```

```
## [1] 90
```

```
max(Sales$Quantity)
```

```
## [1] 25000
```

```
range(Sales$Quantity)
```

```
## [1]    90 25000
```

```
sd(Sales$Quantity)
```

```
## [1] 7252.404
```

```
var(Sales$Quantity)
```

```
## [1] 52597359
```

```
quantile(Sales$Quantity)
```

```
##    0%   25%   50%   75%  100%
##    90   500  1915  9750 25000
```

```
library(moments)
skewness(Sales$Quantity)
```

```
## [1] 1.235493
```

```
kurtosis(Sales$Quantity)
```

```
## [1] 3.181855
```

Next we will calculate the correlation coefficient for two numeric variables.

```
cor(Sales$Quantity,Sales$`Item Cost`)
```

```
## [1] -0.3342939
```