BE601 - DATA ANALYTICS I

Describing Data: Numerical

Uyen Vu

29/08/2022

# Measures of central tendency and location

1. Mean, Median, and Mode (describe central tendency)

* arithmetic mean:
* median: odd n vs even n
* mode: most regular observation \*\*
  + one mode: unimodal distribution
  + two modes: bimodal distribution
  + two modes: multimodal distribution

1. Example

#exp 2.1 (p65)  
demand <- c(60,84,65,67,75,72,80,85,63,82,70,75)  
mean(demand)

## [1] 73.16667

median(demand)

## [1] 73.5

df <- as.data.frame(table(demand))  
colnames(df) <- c("value","freq")  
df <- df[order(-df[, "freq"]), ]  
df #mode is 75

## value freq  
## 7 75 2  
## 1 60 1  
## 2 63 1  
## 3 65 1  
## 4 67 1  
## 5 70 1  
## 6 72 1  
## 8 80 1  
## 9 82 1  
## 10 84 1  
## 11 85 1

#exp 2.2 (p65)  
eps <- c(0.05,0.05,0.081,0.136,0.232,0.207,0.12,0.142)  
mean(eps)

## [1] 0.12725

median(eps)

## [1] 0.128

df <- as.data.frame(table(eps))  
colnames(df) <- c("value","freq")  
df <- df[order(-df[, "freq"]), ]  
df

## value freq  
## 1 0.05 2  
## 2 0.081 1  
## 3 0.12 1  
## 4 0.136 1  
## 5 0.142 1  
## 6 0.207 1  
## 7 0.232 1

# Shape of distribution

1. **skewness**

* Skewness is positive: skewed-right (mean > median)
* skewness is negative: skewed-left (mean < median)

1. R

* Base R does not have the function to calculate skewness
* Package “moment” with the command **skewness()**