Formative Assessment 1

Vera Aguila

2024-02-01

1. Write the skewness program, and use it to calculate the skewness coefficient of the four examination subjects in results.txt (results.csv). What can you say about these data?

Pearson has given an approximate formula for the skewness that is easier to calculate than the exact formula given in Equation 2.1.

```
results <- read.csv("C:/Users/Dell/Downloads/FA1/results.csv",header = T)
results$gender <- as.factor(results$gender)
summary(results)
```

```
##  gender      arch1            prog1           arch2           prog2      
##  f: 19   Min.   :  3.00   Min.   :12.00   Min.   : 6.00   Min.   : 5.00  
##  m:100   1st Qu.: 46.75   1st Qu.:40.00   1st Qu.:40.00   1st Qu.:30.00  
##          Median : 68.50   Median :64.00   Median :48.00   Median :57.00  
##          Mean   : 63.57   Mean   :59.02   Mean   :51.97   Mean   :53.78  
##          3rd Qu.: 83.25   3rd Qu.:78.00   3rd Qu.:61.00   3rd Qu.:76.50  
##          Max.   :100.00   Max.   :98.00   Max.   :98.00   Max.   :97.00  
##          NA's   :3        NA's   :2       NA's   :4       NA's   :8
```

SKEWNESS (first method) Using the skewness program from the book:

```
skew <- function(x) {
  xbar <- mean(x,na.rm = T)
  sum2 <- sum((x-xbar)^2, na.rm = T)
  sum3 <- sum((x-xbar)^3, na.rm = T)
  skew <- (sqrt(length(x))*sum3)/(sum2^(1.5))
  skew
}
skew(results$arch1)
```

```
## [1] -0.5195368
```

```
skew(results$prog1)
```

```
## [1] -0.3362643
```

```
skew(results$arch2)
```

```
## [1] 0.4558875
```

```
skew(results$prog2)
```

```
## [1] -0.3125144
```

```
## Skewness for Arch1: -0.5195368
```

```
## Skewness for Prog1: -0.3362643
```

```
## Skewness for Arch2: 0.4558875
```

```
## Skewness for Prog2: -0.3125144
```

SKEWNESS (second method) Using the Pearson's coefficient of skewness:

```
skew2 <- function(x) {
  mn <- mean(x, na.rm = T)
  md <- median(x, na.rm = T)
  standev <- sd(x, na.rm = T)
  skewness <- (3*(mn-md))/standev
  skewness
}
skew2(results$arch1)
```

```
## [1] -0.6069042
```

```
skew2(results$prog1)
```

```
## [1] -0.643229
```

```
skew2(results$arch2)
```

```
## [1] 0.5421286
```

```
skew2(results$prog2)
```

```
## [1] -0.3562908
```

2. For the class of 50 students of computing detailed in Exercise 1.1, use R to

a. form the stem-and-leaf display for each gender, and discuss the advantages of this representation compared to the traditional histogram;

Data: Females: 57, 59, 78, 79, 60, 65, 68, 71, 75, 48, 51, 55, 56, 41, 43, 44, 75, 78, 80, 81, 83, 83, 85 Males: 48, 49, 49, 30, 30, 31, 32, 35, 37, 41, 86, 42, 51, 53, 56, 42, 44, 50, 51, 65, 67, 51, 56, 58, 64, 64, 75

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   4 | 1348
##   5 | 15679
##   6 | 058
##   7 | 155889
##   8 | 01335
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   3 | 001257
##   4 | 1224899
##   5 | 01113668
##   6 | 4457
##   7 | 5
##   8 | 6
```