"SEC 1-FA1 GROUP 2-QUIJANO, JP"

Julian Philip S. Quijano

2025-01-31

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.table("results.txt", header = TRUE)

skew <- function(x) {
  xbar <- mean(x, na.rm = TRUE)
  sum2 <- sum((x-xbar)**2, na.rm = TRUE)
  sum3 <- sum((x-xbar)**3, na.rm = TRUE)
  skew <- (sqrt(length(x))* sum3)/(sum2**(1.5))
  skew
  }

pearson_skewness <- function(x){
  mean <- mean(x, na.rm = TRUE)
  median <- median(x, na.rm = TRUE)
  sdev <- sd(x, na.rm = TRUE)

  skewness <- (3 * (mean-median)) / sdev
  return(skewness)
}

results_df <- data.frame(
  Arch1 = c(skew(results$Arch1), pearson_skewness(results$Arch1)),
  Prog1 = c(skew(results$Prog1), pearson_skewness(results$Prog1)),
  Arch2 = c(skew(results$Arch2), pearson_skewness(results$Arch2)),
  Prog2 = c(skew(results$Prog2), pearson_skewness(results$Prog2))
)

rownames(results_df) <- c("Normal Skewness", "Pearson Skewness")

print(results_df)
```

```
##                       Arch1      Prog1     Arch2      Prog2
## Normal Skewness  -0.5195368 -0.3362643 0.4558875 -0.3125144
## Pearson Skewness -0.6069042 -0.6432290 0.5421286 -0.3562908
```

What can you say about these data?

The normal skewness indicates that scores are leaned left, meaning more students scoring higher than lower, although only by a small majority.

Is it a reasonable approximation?

Looking at the data, it seems there is a reasonable similarity between the normal skewness and the Pearson skewness, with the latter having slighlty higher or even lower results.

For the class of 50 students of computing detailed in Exercise 1.1, use R to form the stem-and-leaf display for each gender, and discuss the advantages of this representation compared to the traditional histogram.

```
Females <- c(57, 59, 78, 79, 60, 65, 68, 71, 75, 48, 51, 55, 56, 41, 43, 44,
             75, 78, 80, 81, 83, 83, 85)
Males <- c(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)

print("First-Year Java Programming Exam Scores of Females")
```

```
## [1] "First-Year Java Programming Exam Scores of Females"
```

```
print(stem(Females))
```

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

```
print("First-Year Java Programming Exam Scores of Males")
```

```
## [1] "First-Year Java Programming Exam Scores of Males"
```

```
print(stem(Males))
```

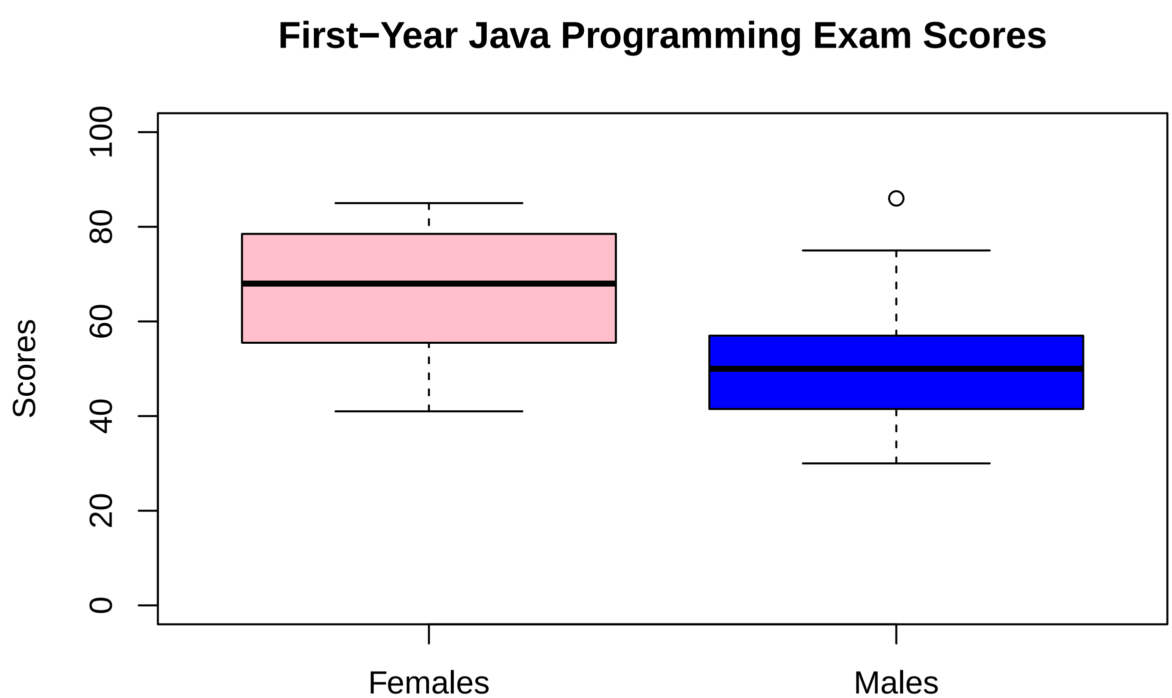
```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   3 | 001257
##   4 | 1224899
##   5 | 01113668
```

```
##   6 | 4457
##   7 | 5
##   8 | 6
##
## NULL
```

Using a stem and leaf display model can easily allow one to visually assess and compare data quickly, as it highlights patterns and points with a simple design. However, one can imagine that it would not be so useful or ideal with large amounts of data.

Construct a box-plot for each gender and discuss the findings.

```
boxplot(Females, Males, names = c("Females", "Males"),
  main = "First-Year Java Programming Exam Scores",
  ylab = "Scores",
  ylim = range(0,100),
  col = c("pink", "blue")
)
```



This graph shows that, generally, the females outperformed the males in the Java Programming Exam by a decent amount. It must also be noticed that the male scores is less varied when compared to the female scores.

Github Link: https://github.com/SylTana/APM1110-QUIJANO-JULIAN_PHILIP/tree/main/FA1