FA - 1

Paul Joaquin M. Delos Santos

2024-02-01

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.

```
##   Subject   Skewness
## 1   arch1 -0.6069042
## 2   prog1 -0.6432290
## 3   arch2  0.5421286
## 4   prog2 -0.3562908
```

```
##       Subject Pearson_Skewness
## arch1   arch1       -0.5063276
## prog1   prog1       -0.3291610
## arch2   arch2        0.4423272
## prog2   prog2       -0.2977574
```

Explanation

~ Pearson Case

- Compared to sample skewness, it is less sensitive to extreme values, less precise, and can be used for quickly computed skewness estimates.
- It works well with data that is regularly disseminated.

~ Normal Skewness case

- It is sensitive to extreme values in the data and is based on moments.
- When working with real-world data samples, it is commonly used.

NO.2

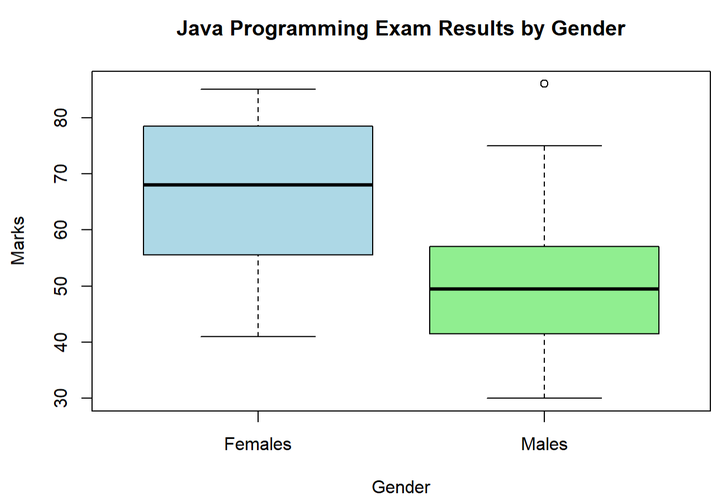
```
##  [1] 57 59 78 79 60 65 68 71 75 48 51 55 56 41 43 44 75 78 80 81 83 83 85
```

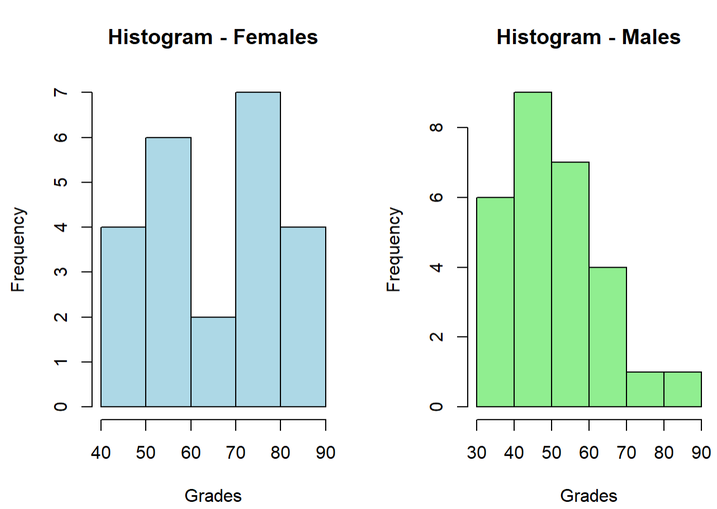
```
##  [1] 48 49 49 30 30 31 32 35 37 41 86 42 51 53 56 42 44 50 51 65 67 51 56 58 64
## [26] 64 75
```

```
##  [1] Female Female Female Female Female Female Female Female Female Female
## [11] Female Female Female Female Female Female Female Female Female Female
## [21] Female Female Female Male   Male   Male   Male   Male   Male   Male
## [31] Male   Male   Male   Male   Male   Male   Male   Male   Male   Male
## [41] Male   Male   Male   Male   Male   Male   Male   Male   Male   Male
## Levels: Female Male
```

```
##
##   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 | 12246899
##   5 | 01113668
##   6 | 4457
##   7 | 5
##   8 | 6
```





Advantages

The stem-and-leaf display and the traditional histogram are both graphical representations used to visualize the distribution of a dataset.

- Individual Data Points: The stem-and-leaf display's ability to preserve individual data points makes it possible to view every observation in the dataset, which is one of its main advantages. A histogram, on the other hand, gathers data into bars. A more in-depth perspective of the distribution is offered by the stem-and-leaf display, which makes it easier to pinpoint particular values and their frequencies.
- Data Integrity: A histogram loses some information because of bundling, but a stem-and-leaf display allows you to piece together the original dataset from the plot. When the original data points are essential for analysis or interpretation, this may be significant.
- Comparison of Groups: The side-by-side stem-and-leaf displays for males and females enable a clear visual comparison of their distributions when comparing groups, as in your example with exam results by gender. This makes it easier to quickly recognize any parallels or discrepancies between the two groups.

BOX PLOT FINDINGS

- Base Tendency:

~ Females: A balanced distribution of scores among females is suggested by the fact that the median (center of the box) for females seems to be approximately in the middle of the interquartile range.
Males: The center part of the male score distribution appears to be skewed towards higher values, as indicated by the male median score appearing to be slightly higher than the female median.
- Distribution of Scores:
Females: The interquartile range (IQR) indicates a moderate spread of scores within the center 50% of the data. It is comparatively constant.

Males: A greater range of scores inside the middle 50% of the data is indicated by the broader IQR for males. This implies that male scores vary more widely.