Skewness Analysis and Data Visualization

Roland Dela Rosa

2025-02-02

1. Skewness Calculation

Load necessary libraries

```
library(moments)  # For skewness calculation
```

Read the dataset

```
data <- read.table("results.txt", header = TRUE, na.strings = "NA")
```

Compute Pearson's skewness for each subject

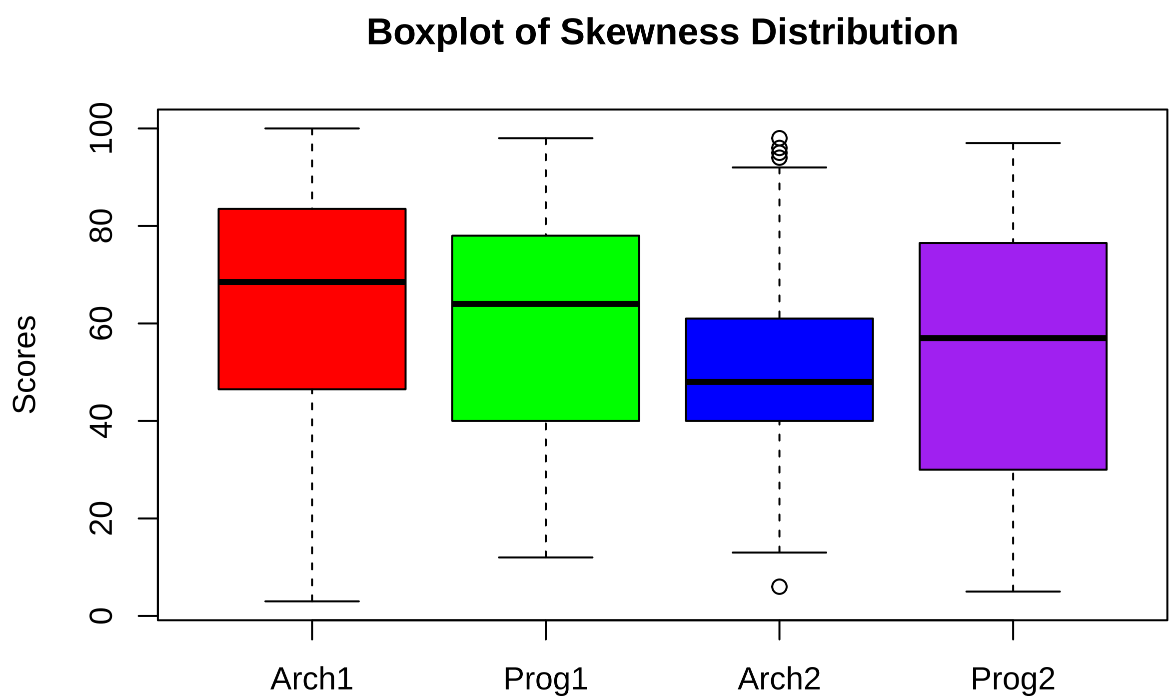
```
skewness_pearson <- function(x) {
  mean_x <- mean(x, na.rm = TRUE)
  median_x <- median(x, na.rm = TRUE)
  sd_x <- sd(x, na.rm = TRUE)

  return(3 * (mean_x - median_x) / sd_x)
}

subjects <- c("Arch1", "Prog1", "Arch2", "Prog2")
skewness_values <- sapply(subjects, function(subj) skewness_pearson(data[[subj]]))
skewness_values <- as.numeric(skewness_values)
```

Compute Normal Skewness

```
skewness_exact <- sapply(subjects, function(subj) skewness(data[[subj]], na.rm = TRUE))
skewness_exact <- as.numeric(skewness_exact)
```

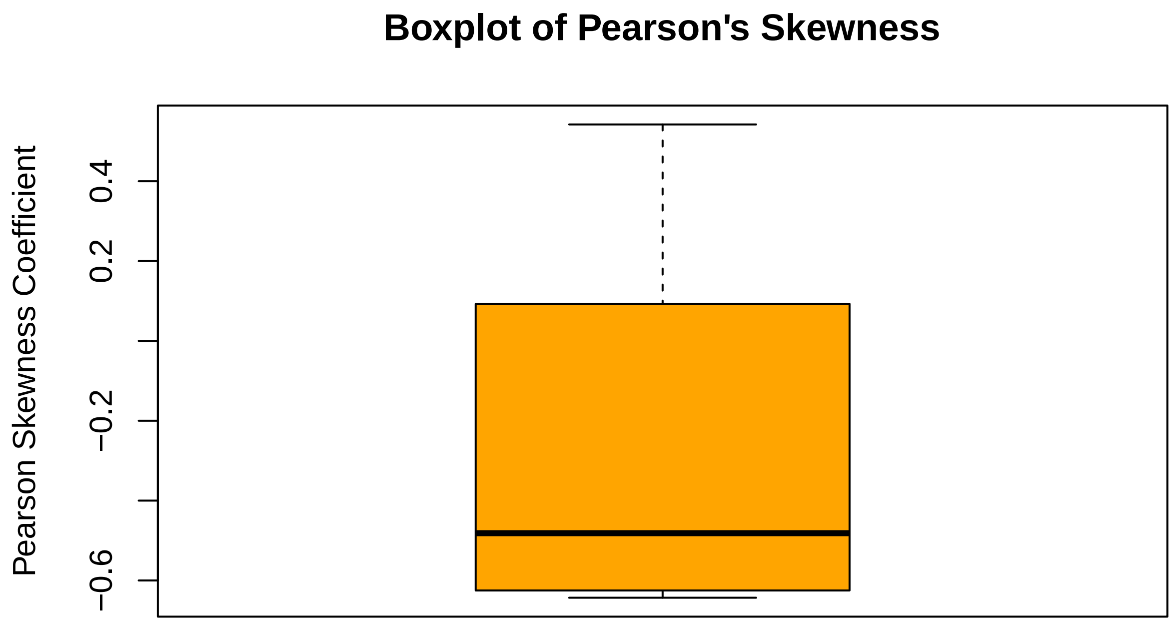
Boxplot for Skewness Visualization

```r
boxplot(data[, subjects], main = "Boxplot of Skewness Distribution", col = c("red", "green", "blue", "pu
        names = subjects, ylab = "Scores")
```



Boxplot for Pearson's Skewness

```r
boxplot(skewness_values, main = "Boxplot of Pearson's Skewness", col = "orange",
        names = "Subjects", ylab = "Pearson Skewness Coefficient")
```



Interpretation

Pearson's Skewness Pearson's skewness is a simpler measure that estimates skewness using mean, median, and standard deviation. It is useful when a quick approximation is needed but may not always capture extreme skewness accurately.

Normal Skewness The exact skewness calculated using the `moments::skewness()` function provides a more precise measure of the asymmetry in the distribution. It is generally more reliable than Pearson's approximation, especially for highly skewed data.

Comparison Comparing both values, Pearson's skewness is generally close to the normal skewness but may deviate when the data distribution is highly skewed or has extreme outliers. If the values are significantly different, it suggests that Pearson's method is less effective in capturing the true nature of the data's asymmetry.

The additional boxplot for Pearson's skewness visually represents its variability across different subjects, helping to compare how consistent the approximation is.

---

2. Data Visualization for Computing Students

Define the data

```
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)
```

(a) Stem-and-Leaf Display

```
cat("\nStem-and-Leaf for Females:\n")
```

```
##
## Stem-and-Leaf for Females:
```

```
stem(females, scale = 2)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   4 | 134
##   4 | 8
##   5 | 1
##   5 | 5679
##   6 | 0
##   6 | 58
##   7 | 1
##   7 | 55889
##   8 | 0133
##   8 | 5
```

```
cat("\nStem-and-Leaf for Males:\n")
```

```
##
## Stem-and-Leaf for Males:
```

```
stem(males, scale = 2)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   3 | 0012
##   3 | 57
##   4 | 1224
##   4 | 899
##   5 | 01113
##   5 | 668
```

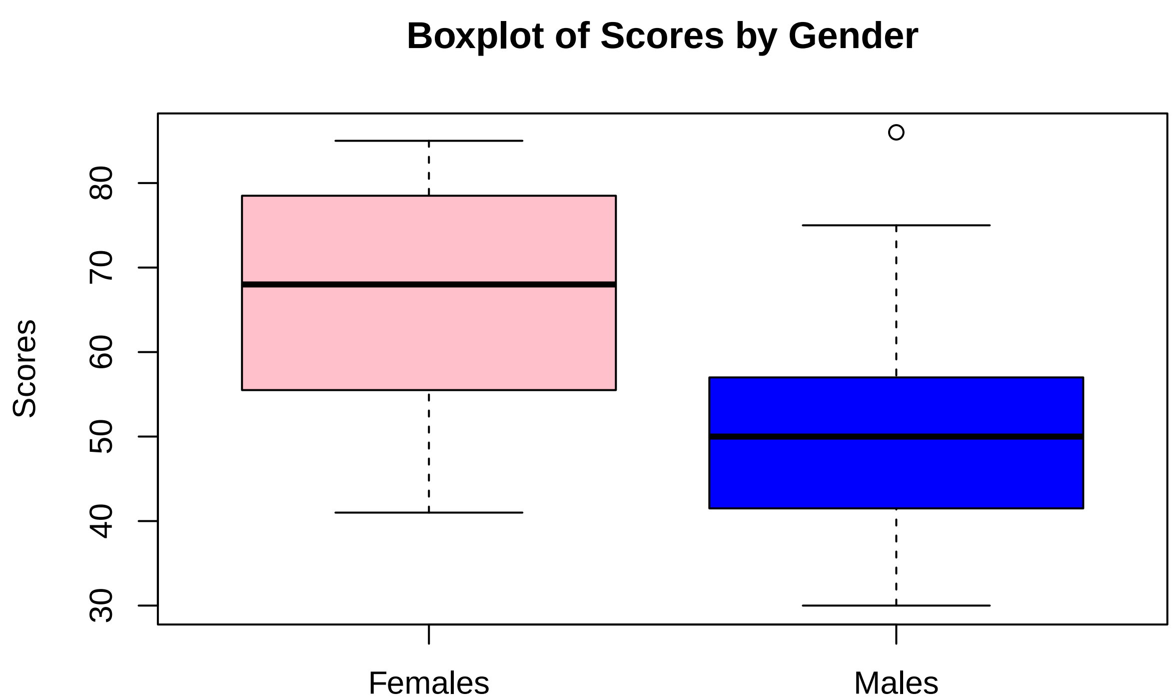
```
##    6 | 44
##    6 | 57
##    7 |
##    7 | 5
##    8 |
##    8 | 6
```

Interpretation of Stem-and-Leaf vs. Histogram

The stem-and-leaf plot preserves individual data points while showing the overall shape of the distribution. Compared to a histogram, it provides a more detailed view of the actual values while still displaying the distribution.

(b) Boxplot Comparison

```r
boxplot(females, males, names = c("Females", "Males"), col = c("pink", "blue"),
        main = "Boxplot of Scores by Gender", ylab = "Scores")
```



Interpretation

The boxplot allows us to visualize the spread, median, and presence of outliers. If the median line is not centered, it indicates skewness. The interquartile range (IQR) shows score variability, and outliers (if present) highlight extreme values.

From the boxplot, we can compare the central tendency and dispersion between male and female students.

Exporting to PDF

To generate a PDF, save this file as an R Markdown (`.Rmd`) and click `Knit`. Ensure **rmarkdown** and **TinyTeX** are installed:

```r
install.packages("rmarkdown")
install.packages("tinytex")
tinytex::install_tinytex()
```