TEXTING AND NUMERIC SCORES: HYPOTHESIS AND Z-TEST

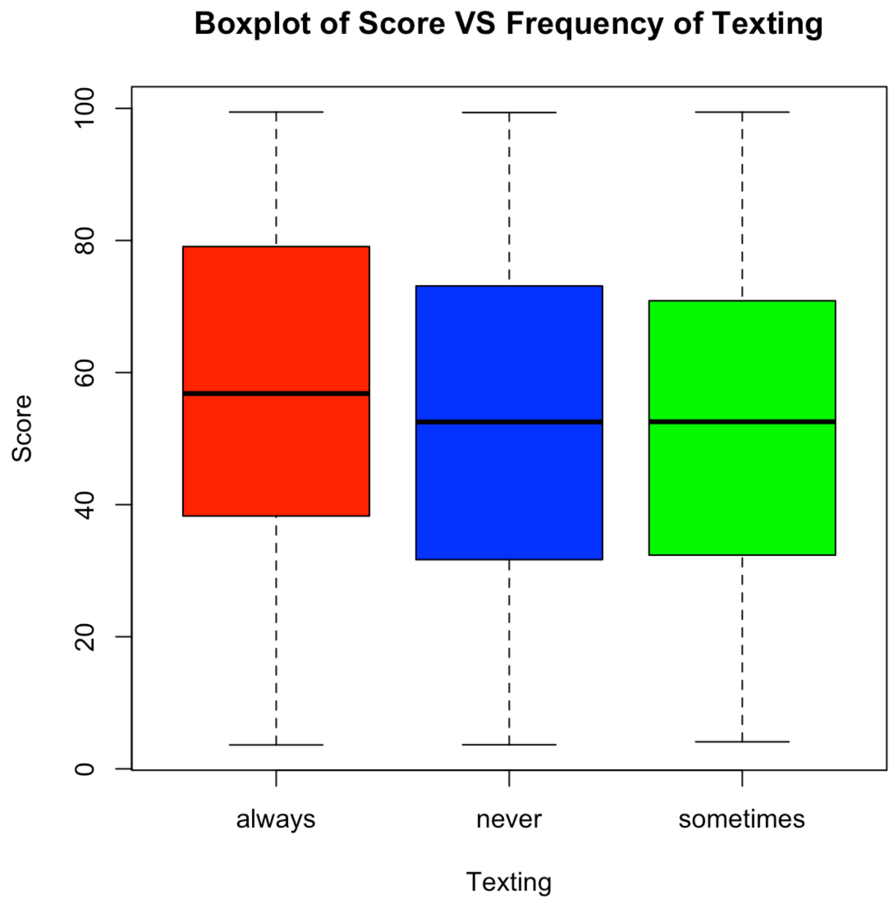
BY FAUZAN AMJAD

Background Information

- We’re given a rich dataset with letter grades, numerical scores, frequency of texting expressed as a categorical, frequency of asking questions expressed as a categorical, and the frequency to participate in class express as a quantitative measure on an index.
- Professor Moody’s grading scheme is rather unique, with numerical scores not being the sole criterion for letter grade. Although higher numerical grades correlate with higher letter grades, the ranges for each letter grade overlap.
- Let’s look at frequency of texting vs numerical grades using a boxplot.

Background Information (cont.)

- A boxplot allows us to visualize categorial vs numerical data.
- As seen through the plot, the range of values are roughly the same, the respective medians are rather close, skew is visually the same, and interquartile range is close.
- Does this mean they're not significantly different?
  - No, it doesn't.



Hypothesis and Null Hypothesis

- Null Hypothesis: The numerical scores for students who always text is not different than the numerical scores for students who sometimes text.
- Alternative Hypothesis: The numerical scores for students who always text is different than the numerical scores for students who sometimes text.

WE ASSUME THE NULL HYPOTHESIS TO BE TRUE

R Code (Read Comments)

```r
## Dataset Import
setwd("~/Desktop")
moody <- read.csv("~/Desktop/moody2020b.csv")
## Colors
colors <- c("red", "blue" , "green")
## Main Boxplot
boxplot(moody$score ~ moody$texting, main = "Boxplot of Score VS Frequency of Texting", xlab = "Texting", ylab = "Score", col = colors)
## Subset our data
alwaysTexting.data <- subset(moody, moody$texting == "always")
somestimesTexting.data <- subset(moody, moody$texting == "sometimes")
## Get our Scores
alwaysTexting.scores <- alwaysTexting.data$score
sometimesTexting.scores <- somestimesTexting.data$score
## Get our Mean Values
mean.alwaysTexting <- mean(alwaysTexting.scores)
mean.sometimesTexting <- mean(sometimesTexting.scores)
## Get our Standard Deviations
sd.alwaysTexting <- sd(alwaysTexting.scores)
sd.sometimesTexting <- sd(sometimesTexting.scores)
## Get our Lengths
len_alwaysTexting <- length(alwaysTexting.scores)
len_sometimesTexting <- length(sometimesTexting.scores)
## Get Combined Standard Deviation
sd.al.st <- sqrt(sd.alwaysTexting^2/len_alwaysTexting + sd.sometimesTexting^2/len_sometimesTexting)
## Calculate our Z-value
zeta.al.st <- (mean.alwaysTexting - mean.sometimesTexting)/sd.al.st
zeta.al.st
##Get our P-value
p.al.st = 1 - pnorm(zeta.al.st)
p.al.st
```

Z-Score

- The z-score is utilized as a measurement to describe how far we are from the mean. Standard deviations is what is used to measure z-scores.
- From our Z-test, we got a z-score of 2.234855.
  - We are 2.234855 standard deviations away from the mean.

```
> zeta.al.st
[1] 2.234855
```

P-Value and Interpretation

- The calculated p-value derived from our z-score will be used as our evidence against the null hypothesis.
- We'll use 0.05 as our threshold to either reject the null hypothesis or fail to reject the null hypothesis.
- The p-value I got was 0.01271343.
    - Given our data, there is a 1.271343% chance that our data occurred under the null hypothesis.

```
> p.al.st
[1] 0.01271343
```

WE REJECT THE NULL HYPOTHESIS

We can conclude with a high degree of certainty that the alternative hypothesis is true.