Week4 Assignment

Surenther

2024-09-21

Questions

1. What are the observational units in this study?
   The observational units in this study are the individual students enrolled in the two different sections of the course.Each student's performance, including their total points earned, is being compared between the sections with different teaching examples.

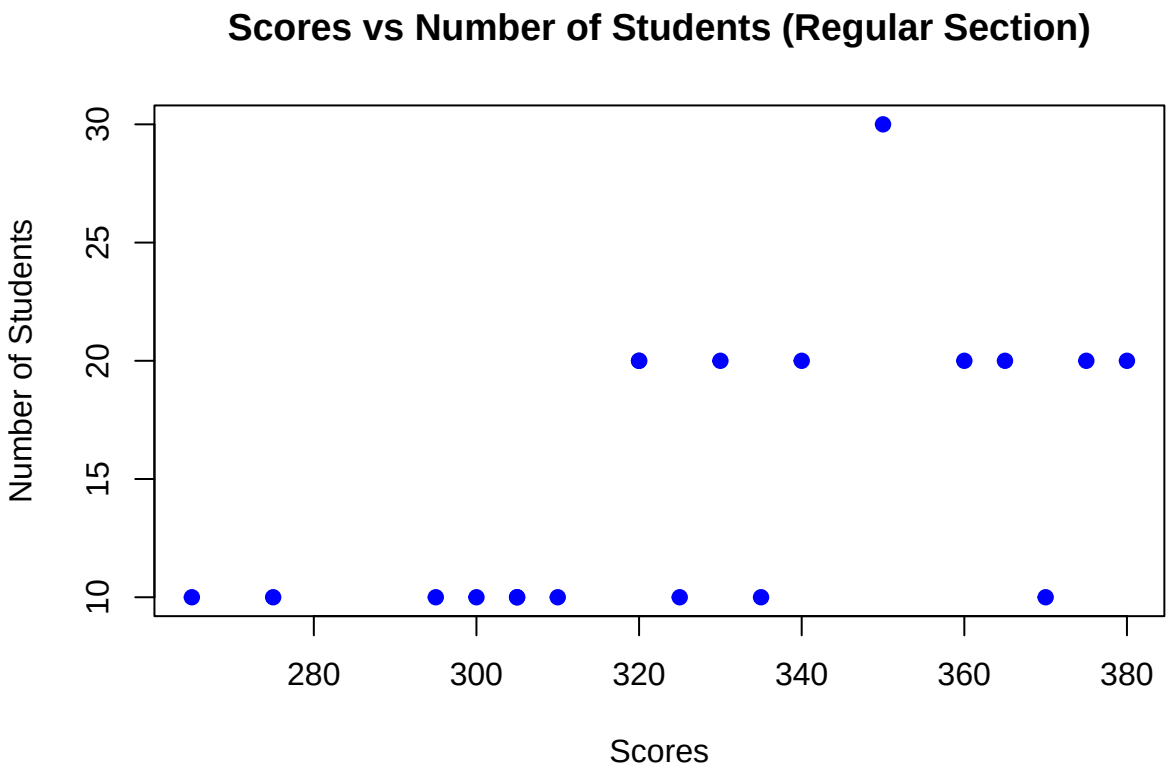
2. Identify the variables mentioned in the narrative paragraph and determine which are categorical and quantitative?
   Section Type is a categorical variable, representing the different types of examples used in the course (sports-themed or a regular). Number of Students and Scores are quantitative variables, providing numerical measurements of the number of students enrolled in each section and the performance outcomes, respectively.

3. Create one variable to hold a subset of your data set that contains only the Regular Section and one variable for the Sports Section.

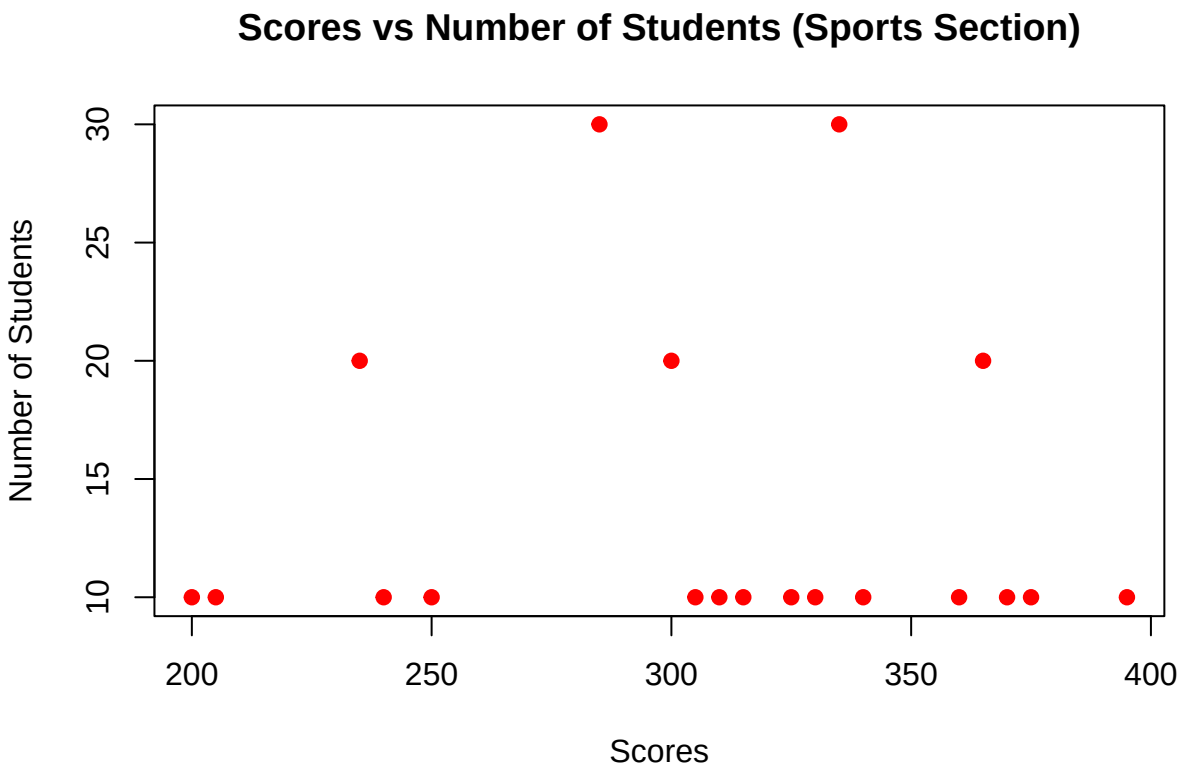
```r
#Import CSV
scores <- read.table(file = "scores.csv", header = TRUE, sep = ",")
# Regular Section
regular_section <- subset(scores, Section == "Regular")
# Sports Section
sports_section <- subset(scores, Section == "Sports")
```

4. Use the Plot function to plot each Sections scores and the number of students achieving that score.

```r
# Plot for Regular Section
plot(regular_section$Score, regular_section$Count,
     main = "Scores vs Number of Students (Regular Section)",
     xlab = "Scores", ylab = "Number of Students",
     col = "blue", pch = 19)
```




```r
# Plot for Sports Section
plot(sports_section$Score, sports_section$Count,
     main = "Scores vs Number of Students (Sports Section)",
     xlab = "Scores", ylab = "Number of Students",
     col = "red", pch = 19)
```



- Comparing and contrasting the point distributions between the two section, looking at both tendency and consistency: Can you say that one section tended to score more points than the other? Justify and explain your answer.

Tendency (Central Tendency) can be assessed using measures like the mean and median, while consistency (Spread) is evaluated through the standard deviation. Both can be computed using the stat.desc function in R.

```
# Stat dec for Regular
library(pastecs)
stat.desc(regular_section$Score, basic=FALSE, desc=TRUE, norm=TRUE)
```

```
##        median          mean       SE.mean  CI.mean.0.95           var
##  325.00000000  327.63157895    7.63157895   16.03335241 1106.57894737
##       std.dev      coef.var      skewness      skew.2SE      kurtosis
##   33.26528141    0.10153259   -0.07341545   -0.07008412   -1.08697079
##      kurt.2SE    normtest.W    normtest.p
##   -0.53583906    0.96952149    0.76684878
```

```
# Stat dec for Sports
stat.desc(sports_section$Score, basic=FALSE, desc=TRUE, norm=TRUE)
```

```
##       median         mean      SE.mean CI.mean.0.95          var      std.dev
```

```
##  315.0000000  307.3684211   13.3134085   27.9704333 3367.6900585   58.0318021
##     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE   normtest.W
##    0.1888021   -0.4193351   -0.4003071   -1.0607215   -0.5228991    0.9445561
##   normtest.p
##    0.3179747
```

Findings
In comparing the two sections, the Regular section shows a higher mean (327.63) compared to the Sports section (307.37), and the median for the Regular section (325) is also higher than that of the Sports section (315). **This indicates that students in the Regular section generally scored higher.** Additionally, the Regular section's standard deviation of 33 suggests that student scores in this group were more consistent compared to those in the Sports section.

- Did every student in one section score more points than every student in the other section? If not, explain what a statistical tendency means in this context.
  Not every student in one section scored more points than every student in the other section. Statistical tendencies, such as mean and median, represent the overall pattern of scores, not individual outcomes. The Regular section has a higher mean score than the Sports section, it means that, on average, students in the regular section scored more points. However, this does not mean that every individual in the regular section outscored every individual in the sports section.

- What could be one additional variable that was not mentioned in the narrative that could be influencing the point distributions between the two sections?
  In the Sports section, students might have enrolled due to a stronger interest in sports, which could lead to higher engagement with the material and potentially better performance.Conversely, students in the Regular section may have chosen that section to avoid sports-related examples, indicating different levels of interest or engagement.**This interest or familiarity with the subject matter** could impact how well students understand and apply the course concepts, thus influencing their overall performance.