DSC 520: Assignment 2.1 - Test Scores

Demond Love

6/17/2018

**A professor has recently taught two sections of the same course with only one difference between the sections. In one section, he used only examples taken from sports applications, and in the other section, he used examples taken from a variety of application areas. The sports themed section was advertised as such; so students knew which type of section they were enrolling in. The professor has asked you to compare student performance in the two sections using course grades and total points earned in the course. You will need to import the Scores.csv dataset that has been provided for you.**

file = read.csv("scores.csv", stringsAsFactors = FALSE)  
file

## Count Score Section  
## 1 10 200 Sports  
## 2 10 205 Sports  
## 3 20 235 Sports  
## 4 10 240 Sports  
## 5 10 250 Sports  
## 6 10 265 Regular  
## 7 10 275 Regular  
## 8 30 285 Sports  
## 9 10 295 Regular  
## 10 10 300 Regular  
## 11 20 300 Sports  
## 12 10 305 Sports  
## 13 10 305 Regular  
## 14 10 310 Regular  
## 15 10 310 Sports  
## 16 20 320 Regular  
## 17 10 305 Regular  
## 18 10 315 Sports  
## 19 20 320 Regular  
## 20 10 325 Regular  
## 21 10 325 Sports  
## 22 20 330 Regular  
## 23 10 330 Sports  
## 24 30 335 Sports  
## 25 10 335 Regular  
## 26 20 340 Regular  
## 27 10 340 Sports  
## 28 30 350 Regular  
## 29 20 360 Regular  
## 30 10 360 Sports  
## 31 20 365 Regular  
## 32 20 365 Sports  
## 33 10 370 Sports  
## 34 10 370 Regular  
## 35 20 375 Regular  
## 36 10 375 Sports  
## 37 20 380 Regular  
## 38 10 395 Sports

**1. What are the observational units in this study?**

The observational units are 38 total students, split into two groups of Sports and Regular section students.

**2. Identify the variables mentioned in the narrative paragraph and determine which are categorical and quantitative?**

Count: Quantitative

Score: Quantitative

Section: Categorical

**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.**

Reg = subset(file, Section == 'Sports')  
Reg

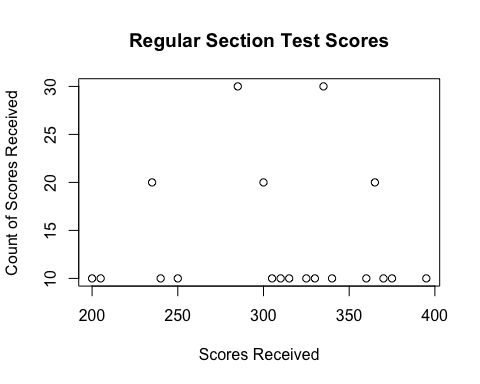
## Count Score Section  
## 1 10 200 Sports  
## 2 10 205 Sports  
## 3 20 235 Sports  
## 4 10 240 Sports  
## 5 10 250 Sports  
## 8 30 285 Sports  
## 11 20 300 Sports  
## 12 10 305 Sports  
## 15 10 310 Sports  
## 18 10 315 Sports  
## 21 10 325 Sports  
## 23 10 330 Sports  
## 24 30 335 Sports  
## 27 10 340 Sports  
## 30 10 360 Sports  
## 32 20 365 Sports  
## 33 10 370 Sports  
## 36 10 375 Sports  
## 38 10 395 Sports

Sport = subset(file, Section == 'Regular')  
Sport

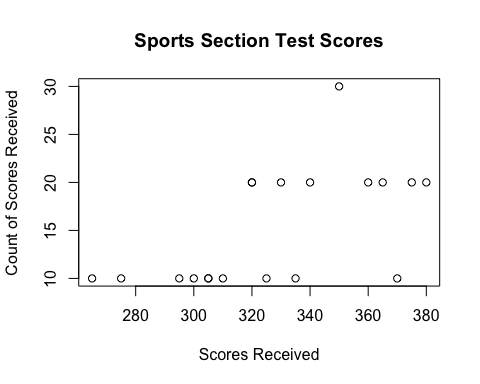
## Count Score Section  
## 6 10 265 Regular  
## 7 10 275 Regular  
## 9 10 295 Regular  
## 10 10 300 Regular  
## 13 10 305 Regular  
## 14 10 310 Regular  
## 16 20 320 Regular  
## 17 10 305 Regular  
## 19 20 320 Regular  
## 20 10 325 Regular  
## 22 20 330 Regular  
## 25 10 335 Regular  
## 26 20 340 Regular  
## 28 30 350 Regular  
## 29 20 360 Regular  
## 31 20 365 Regular  
## 34 10 370 Regular  
## 35 20 375 Regular  
## 37 20 380 Regular

**4. Use the Plot function to plot each Sections scores and the number of students achieving that score. Use additional Plot Arguments to label the graph and give each axis an appropriate label. Once you have produced your Plots answer the following questions:**

plot(Count ~ Score, data = Reg, main = 'Regular Section Test Scores', xlab = 'Scores Received', ylab = 'Count of Scores Received')



plot(Count ~ Score, data = Sport, main = 'Sports Section Test Scores', xlab = 'Scores Received', ylab = 'Count of Scores Received')



**a. 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.**

Looking at the scatterplots, it does appear that the Sports section tended to score more points than the the Regular section. It can be seen that there is a higher frequency of high-scores in the Sports section. This can be validated by looping through the dataset, to calulate if this is actually the case and if the Sports section does in fact score higher than the Regular section.

t = 2  
m = 0  
for (i in Reg[,1]) {  
 u = (Reg[t,1])\*(Reg[t,2])  
 m = u + m  
 if (t < 19) {  
 t = t + 1  
 }  
}  
# Regular section Total Score:  
print (m)

## [1] 81750

t = 2  
k = 0  
for (i in Sport[,1]) {  
 u = (Sport[t,1])\*(Sport[t,2])  
 k = u + k  
 if (t < 19) {  
 t = t + 1  
 }  
}  
# Sports section Total Score:  
print (k)

## [1] 102100

As shown above, with the Sports section did indeed have higher overal test scores.

**b. 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.**

No, not every student in one section scores more points than every student in the other section. The idea of statistical tendency is another name for central tendency, which is the way to measure the “middle” of a set of data. Therefore, not every student in one section scores more points than every student in the other section, the average score of the students from one section is above the average score of students from the other section. This results in the overall higher total test scores shown above.

**c. What could be one additional variable that was not mentioned in the narrative that could be influencing the point distributions between the two sections?**

One additional variable would be the preparedness of students of one section to the next. If the students who enrolled in the Sports sections had, on average, higher grades than the students in the Regular section, then this would have as much, or more impact on the higher test scores as would the categorical examples that the professor used throughout the course.