4.2_Exercise1_Torres_Gloria.R

2023-04-09

```
# Assignment: 4.2 Exercise 1
# Name: Torres, Estrada
# Date: 2023-04-05

# Complete the following exercises by creating an R Script for each.
# Test Scores
# 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.
library(ggplot2)
library(graphics)
setwd('C:/Users/glori/OneDrive/Documents/Gloria GIT/Gloria_Torres_DSC_520')

scores<- read.csv("data/scores.csv")

# Use the appropriate R functions to answer the following questions:

# What are the observational units in this study?
#In this example the observational units will be all the values measured on the unit (Regular and Sport Section).


# Identify the variables mentioned in the narrative paragraph and determine which are categorical and quantitative?
str(scores)

## 'data.frame':    38 obs. of  3 variables:
##  $ Count  : int  10 10 20 10 10 10 10 30 10 10 ...
##  $ Score  : int  200 205 235 240 250 265 275 285 295 300 ...
##  $ Section: chr  "Sports" "Sports" "Sports" "Sports" ...

# variables: Count, Score and Section

class(scores$Count) # quantitative
```

```
## [1] "integer"
```

```
class(scores$Score) # quantitative
```

```
## [1] "integer"
```

```
class(scores$Section) # Categorical
```

```
## [1] "character"
```

```
# Create one variable to hold a subset of your data set that contains only
the Regular Section and one variable for the Sports Section.
library('dplyr')
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

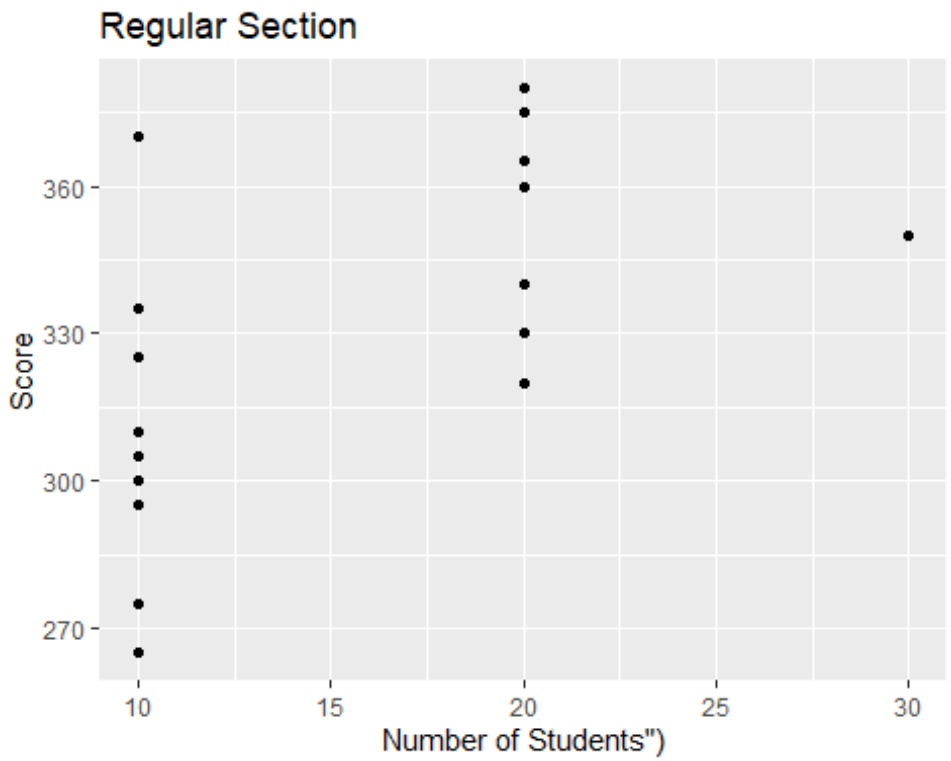
```
only_regular <- subset(scores, Section =="Regular")
only_regular
```

```
##    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
```

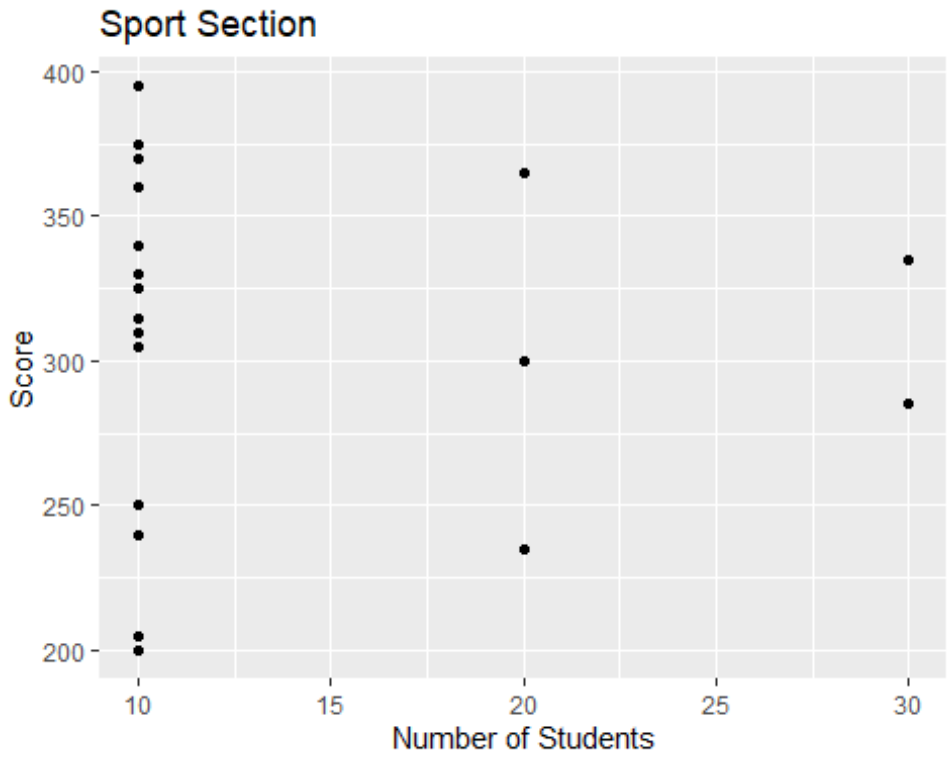
```
only_sport <- subset(scores, Section =="Sports")
only_sport
```

```
##    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
```

```
# 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.
ggplot(only_regular, aes(x=Count, y=Score)) + geom_point() + ggtitle("Regular
Section") + xlab('Number of Students")') + ylab('Score')
```



```
ggplot(only_sport, aes(x=Count, y=Score)) + geom_point() + ggtitle('Sport
Section') + xlab('Number of Students') + ylab('Score')
```




```
# Once you have produced your Plots answer the following questions:
# 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.
# 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.
# The data doesn't display the scores per student; however, it does display
that the scores for the Regular section are higher than the Sport one.
# The statistics tendency: In this example, the tendency can be explained
with the media and mean; the above numbers for the regular section are higher
than the sport one.
# Consistency: The regular section shows more consistency Since the standard
deviation for the regular section is lower than the sport one.

library(pastecs)

##
## Attaching package: 'pastecs'

## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
stat.desc(only_regular, basic = FALSE, norm = TRUE)
```

```
##                      Count         Score Section
## median       10.0000000000  325.00000000      NA
## mean         15.2631578947  327.63157895      NA
## SE.mean       1.4035087719    7.63157895      NA
## CI.mean.0.95  2.9486625126   16.03335241      NA
## var          37.4269005848 1106.57894737      NA
## std.dev       6.1177529032   33.26528141      NA
## coef.var      0.4008182937    0.10153259      NA
## skewness      0.5959903810   -0.07341545      NA
## skew.2SE      0.5689464285   -0.07008412      NA
## kurtosis     -0.7882239006   -1.08697079      NA
## kurt.2SE     -0.3885671578   -0.53583906      NA
## normtest.W    0.7333935101    0.96952149      NA
## normtest.p    0.0001429465    0.76684878      NA
```

```
stat.desc(only_sport, basic = FALSE, norm = TRUE)
```

```
##                     Count        Score Section
## median       1.000000e+01  315.0000000      NA
## mean         1.368421e+01  307.3684211      NA
## SE.mean      1.569171e+00   13.3134085      NA
## CI.mean.0.95 3.296705e+00   27.9704333      NA
## var          4.678363e+01 3367.6900585      NA
## std.dev      6.839856e+00   58.0318021      NA
## coef.var     4.998356e-01    0.1888021      NA
## skewness     1.437922e+00   -0.4193351      NA
## skew.2SE     1.372674e+00   -0.4003071      NA
## kurtosis     5.849792e-01   -1.0607215      NA
## kurt.2SE     2.883746e-01   -0.5228991      NA
## normtest.W   5.928077e-01    0.9445561      NA
## normtest.p   3.759425e-06    0.3179747      NA
```

```
sum(only_regular$Score)
```

```
## [1] 6225
```

```
sum(only_sport$Score)
```

```
## [1] 5840
```

```
# What could be one additional variable that was not mentioned in the
narrative that could be influencing the point distributions between the two
sections?
# Score per student would be helpful to understand the scores of a single
student in both sections (regular/sport).
```