January 4, 2024

The results below are generated from an R script.

```
# 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.

# Use the appropriate R functions to answer the following questions:
# 1) What are the observational units in this study?

# This study compares student performance in two sections using the course
# grades and total points earned in the course.

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

# Based on the narrative paragraph, the catagorical variables are the sports
# and regular examples used in the course and the quantitative variables are the
# course grades and total points earned.

# 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.
scores <- read.csv("scores.csv")
regular.scores <- subset(scores, scores$Section == "Regular")
regular.scores
```

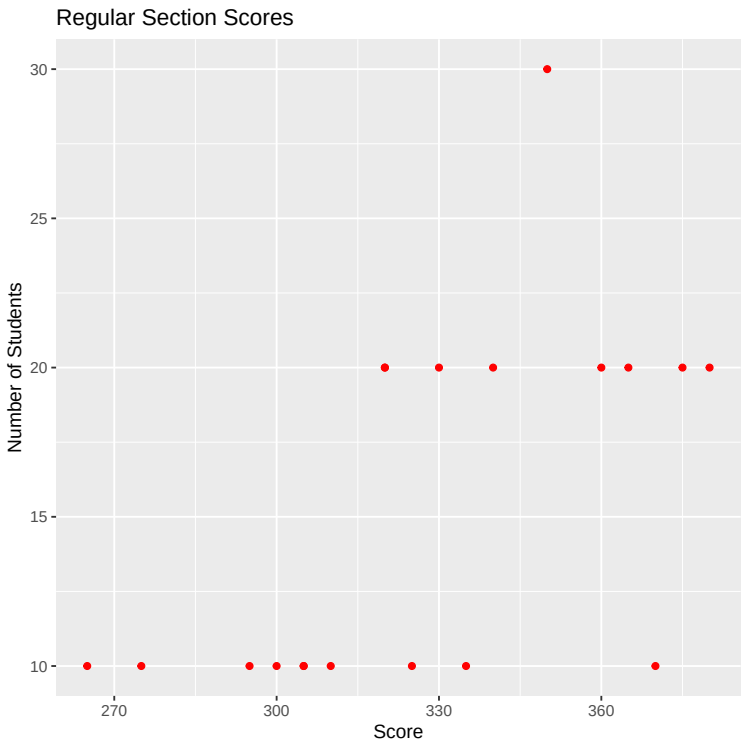
```
##    Count Score Section
## 6     10   265 Regular
## 7     10   275 Regular
## 9     10   295 Regular
## 10    10   300 Regular
## 13    10   305 Regular
## 14    10   310 Regular
## 16    20   320 Regular
## 17    10   305 Regular
## 19    20   320 Regular
## 20    10   325 Regular
## 22    20   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
```

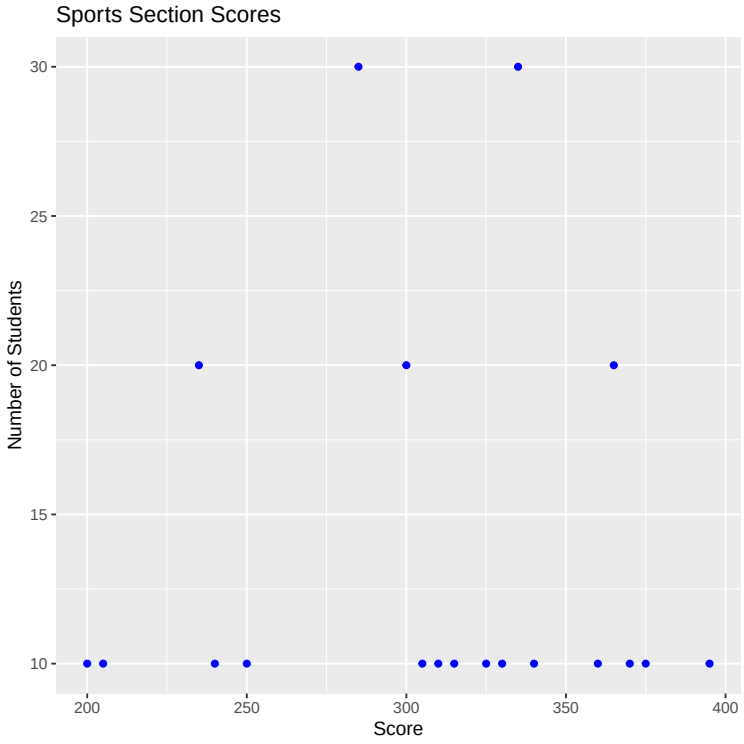
```
sports.scores <- subset(scores, scores$Section == "Sports")
sports.scores
```

```
##    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
```

```
# 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.
library(ggplot2)
ggplot(regular.scores, aes(x = Score, y = Count)) +
  geom_point(color = "red") +
  labs(title = "Regular Section Scores", x = "Score", y = "Number of Students")
```




```
ggplot(sports.scores, aes(x = Score, y = Count)) +
  geom_point(color = "blue") +
  labs(title = "Sports Section Scores", x = "Score", y = "Number of Students")
```



```
# Once you have produced your Plots answer the following questions:
# a. Comparing and contrasting the point distributions between the two sections,
# 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.

# When comparing and contrasting point distributions, it appears that the
# regular section tended to score more points than the sports section.
# Regular - 14 points with scores of 300 or more
# (total of 230 students scored in this range)
# Sports - 13 points with scores of 300 or more
# (total of 150 students scored in this range)
# Regular - 3 points under 300 (total of 30 students scored in this range)
# Sports - 6 points under 300 (total of 90 students scored in this range)
# Though there are more students in the regular section (260) than in the sports
# section (240), 88.5% of students in the regular section scored 300 points or
# more while only 62.5% of students in the sports section attained this.

# 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 scored more points than every student in
# the other section. Both sections had students score high scores as well as low
# scores. Statistical tendency in this context means that, on average, the
# scores in one section were higher than the scores in the other section but not
# that each individual student in one section scored more points than each
# individual student in the other section.

# 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?

# An additional varriable that could be influencing the point distributions
# between sections may be the students' exposure to sports or the overall
# interest that they have in sports which could affect how well they did on the
# examples as opposed to how well they could have done if they were in the
# regular section.
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.3.2 (2023-10-31)
## Platform: x86_64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.5.1
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versio
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-x86_64/Resources/lib/libRlapack.dylib;  LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
```

```
## tzcode source: internal
##
## attached base packages:
## [1] splines   stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] readr_2.1.4        effects_4.2-2      RcmdrMisc_2.9-1    sandwich_3.1-0
##  [5] psych_2.3.12       pastecs_1.3.21     ggplot2_3.4.4      car_3.1-2
##  [9] carData_3.0-5      data.table_1.14.10
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.2.0  dplyr_1.1.4       farver_2.1.1      fastmap_1.1.1
##  [5] digest_0.6.33     rpart_4.1.21      lifecycle_1.0.4   cluster_2.1.4
##  [9] survival_3.5-7    magrittr_2.0.3    compiler_4.3.2    rlang_1.1.2
## [13] Hmisc_5.1-1       tools_4.3.2       utf8_1.2.4        knitr_1.45
## [17] labeling_0.4.3    htmlwidgets_1.6.3 bit_4.0.5         mnormt_2.1.1
## [21] plyr_1.8.9        relimp_1.0-5      abind_1.4-5       tcltk2_1.2-11
## [25] withr_2.5.2       foreign_0.8-85    nnet_7.3-19       grid_4.3.2
## [29] fansi_1.0.5       e1071_1.7-14      colorspace_2.1-0  scales_1.2.1
## [33] MASS_7.3-60       tinytex_0.49      insight_0.19.7    cli_3.6.1
## [37] survey_4.2-1      rmarkdown_2.25    crayon_1.5.2      generics_0.1.3
## [41] rstudioapi_0.15.0 tzdb_0.4.0        readxl_1.4.3      minqa_1.2.6
## [45] DBI_1.1.3         proxy_0.4-27      stringr_1.5.1     parallel_4.3.2
## [49] cellranger_1.1.0  base64enc_0.1-3   mitools_2.4       vctrs_0.6.4
## [53] boot_1.3-28.1     Matrix_1.6-1.1    hms_1.1.3         bit64_4.0.5
## [57] Formula_1.2-5     htmlTable_2.4.2   nortest_1.0-4     glue_1.6.2
## [61] nloptr_2.0.3      stringi_1.8.2     gtable_0.3.4      lme4_1.1-35.1
## [65] munsell_0.5.0     tibble_3.2.1      pillar_1.9.0      htmltools_0.5.7
## [69] R6_2.5.1          tcltk_4.3.2       vroom_1.6.4       evaluate_0.23
## [73] lattice_0.22-5    highr_0.10        haven_2.5.4       backports_1.4.1
## [77] class_7.3-22      Rcpp_1.0.11       gridExtra_2.3     nlme_3.1-163
## [81] checkmate_2.3.1   xfun_0.41         zoo_1.8-12        forcats_1.0.0
## [85] pkgconfig_2.0.3

Sys.time()

## [1] "2024-01-04 22:16:43 EST"
```