

**STAT 260 202401**  
**R ASSIGNMENT #1**

**Total = 20**

The assignment is due at **6pm, Friday, Jan 26th**. You should try to submit your assignment early, if possible.

- Your assignment should be submitted electronically through the cross-listed Crowdmark page. Please submit in pdf format a typed document, uploading your answer to each question separately. If you don't know how to export a Word or OpenOffice document to a pdf, you can ask the computer helpdesk in Clearihue or email [helpdesk@uvic.ca](mailto:helpdesk@uvic.ca), or "your favourite search engine" is your best friend.
- Please include in the upper left-hand corner of the each page of your assignment:  
    {Last Name, First Name}  
    {Student Number}  
    STAT 260 R Assignment {Number}
- You **MUST** include both R commands and outputs for ALL questions, unless the questions said "R is optional".

1. Do students find math or stats harder? We compared the number of requests for help in an intro stats course and a comparable math course.

**Requests for Stats Help:**

578 183 423 429 188 295 273 232 169 271 251 219 313 571 467 628 402 175

**Requests for Math Help:**

544 87 743 391 676 724 991 801 498 175 392 108 402 305 723 662 599 488 319 321

It is probably wise to copy and paste the numbers into R using the `scan()` function, rather than manually typing in all the observations (see the document *R Assignment 1 Introduction*).

- (a) **[1 mark]** Create two vectors named **stats.help.requests** and **math.help.requests** for the vectors of Requests for Stats Help and Requests for Math Help respectively.
- (b) **[3 marks]** Create one side-by-side boxplot of the two sets of scores (i.e. both boxplots on the same axes). The axes for the boxplots should have appropriate labels. Copy and paste this boxplot into your Word document. The boxplots themselves may be either horizontal or vertical (your choice).

- (c) **[3 marks]** Use R to calculate the mean and standard deviation of the number of requests for each type of help. Copy and paste the relevant commands and output from the R Console Window into your document.
  - (d) **[2 marks]** In your opinion, which class seems more difficult? Write a few sentences justifying your opinion. You should make reference to some of the relevant features of the two data sets (e.g. the mean or median, the spread of the data, minimum/maximum values, etc.)
2. Here we will practice more descriptive statistics procedures using simulated data. Run the following three R commands in the exact following order:

```
set.seed(14253)
simvector = rnorm(600, 100, 24)
summary(simvector)
```

Note: The first command makes sure the random data are the same for each simulation. The second command generate 600 measurements from a population with mean 100 and standard deviation 24.

- (a) **[2 marks]** Produce a histogram and a boxplot for the data. No label is necessary.
  - (b) **[1 mark]** Compute the interquartile range.
  - (c) **[2 marks]** Compute the lower and upper limits for identifying outliers using the formula [lower quartile - 1.5 x IQR, upper quartile + 1.5 x IQR].
  - (d) **[1 mark]** Use the histogram and boxplot to briefly comment on the dataset (symmetry, numbers of peak, etc.)
3. The following table shows the career percentage of 1st serves made ( $x$ ) by 10 professional tennis players and the corresponding percentage of points won after making a 1st serve ( $y$ ) .

$x$	56,	58,	59,	59,	60,	60,	62,	64,	65,	72
$y$	76,	77,	76,	76,	71,	70,	67,	69,	64,	64

- (a) **[2 marks]** Produce a scatter plot, putting the percentage of 1st serves made on the  $x$ -axis and the percentage of points won on the  $y$ -axis. Be sure to include a title and labels.
- (b) **[1 mark]** Use R to compute the sample correlation coefficient  $r$ .
- (c) **[2 marks]** Is it reasonable to model this data as a linear relationship? Justify your answer. If it is linear, would the straight line have a positive or negative slope? You may want to review the notes in Set 3.