

STAT 260 202501
R ASSIGNMENT 1

Total = 20

The assignment is due at **11:59pm, Tuesday, Jan 21th**. 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, 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. Here's a sample of historical data regarding the number of boys and girls in an engineering department over a few years.

Numbers of Boys:

80 82 85 90 88 92 95 98 100 105 110 115 120 125

Numbers of Girls:

20 22 25 30 35 40 26 30 20 35 19 25 20 15

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) [**2 mark**] Create two vectors named **number.of.boys** and **number.of.girls** for the number of boys and girls in the engineering department.
- (b) [**3 marks**] Create one side-by-side boxplot of the two sets of numbers (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). Provide some comments on comparing the boxplots, for example the center and variability.

2. Use the data from example 14 in the Lecture Notes to answer the following questions:
 - (a) **[3 marks]** Use R to calculate the mean and standard deviation of the two data sets. Copy and paste the relevant commands and output from the R Console into your document.
 - (b) **[2 marks]** Use R to calculate coefficient of variation of the the two data sets. Copy and paste the relevant commands and the output from the R Console into your document.
3. Here we will practice more descriptive statistics procedures using simulated data. Run the following three R commands in the exact following order:

```
set.seed(2025)
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) **[2 mark]** Compute the interquartile range.
 - (c) **[1 mark]** (R is optional) Use the histogram and boxplot to briefly comment on the dataset (symmetry, numbers of peak, etc.)
4. The following table shows data on students' study hours (x) and their corresponding exam scores (y).

x	2	3	4	5	6	7	8	9	10	11
y	60	61	67	72	78	81	82	89	90	95

- (a) **[2 marks]** Produce a scatterplot, putting students' study hours on the x -axis and exam scores 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]** (R is optional) 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.