

STAT 260 Summer 2024: R Assignment 1

Due: Tuesday May 28 BEFORE 6pm (PT) on Crowdmark.

Introduction to R: Before attempting this assignment, read and work through the *Introduction to R - Assignment 1* file posted on Crowdmark. This file contains a list of all the R commands needed to complete this assignment.

Submission: Since you will need to copy your R code and the code output, your answers must be typed. The best way to complete this assignment is by using a word processor such as MS Word, Google Docs, or Open Office. I have a link posted on our Brightspace page in the “Useful Links” module where you can download Microsoft 365 for free. ***Hand-written work will NOT be accepted for any R assignment.***

Complete the following questions below, copying and pasting the required R commands, R outputs, and analysis into your Word documents. *Unless otherwise instructed, if you use any R Code, you should copy/paste it in your document.* Save your assignment in the PDF file format. On Crowdmark you will be asked to upload your solutions to some parts separately, so the best way to save your work would be to put your solutions to each part requiring a separate submission upload on separate pages in your Word file. To convert your Word document to a PDF use the “Save As” feature - PDF is one of the output options there. When uploading your work to Crowdmark put your submission in the first upload area, then drag and drop the pages for the other parts into the proper submission areas.

Your solution to Question 1 must be uploaded in the location for Question 1, your solution to Question 2 must be uploaded in the location for Question 2, etc. If uploaded to the wrong location the marker will not be able to grade your work.

Your submissions must be readable. If the screenshot qualities are too low resolution, too blurry, too big/too small, they will not be graded. Be sure to crop excess margins / whitespace, as they are included in Crowdmark’s fixed-width viewing for graders.

Cellphone photos of a laptop/tablet screen will not be graded. Submit the actual files.

Upload your files for submission to the assignment on Crowdmark, before Tuesday May 26 at 6pm (PT). A late penalty of 5% per hour will be applied. The penalty is applied in such a way so that assignments submitted 6pm to 6:59pm will have 5% deducted, assignments submitted 7pm-7:59pm will have 10% deducted, etc. You should keep an electronic copy of your assignment for your records until the end of the semester.

Note: For each of the following, carry out your calculations **only** using R or RStudio. Copy and paste your command(s) and the output into your document as indicated. You should write your solutions as if you were writing a report as part of a statistical analysis (that is, use somewhat formal language and complete sentences).

Part 1 Background: Echinoderms are considered useful bioindicators of water quality and marine sediments. If a region becomes polluted, sea urchins, with their minimal migration patterns, will be among the first species to show signs of the contamination.

Scenario: Suppose that we want to compare the cadmium content of black sea urchins (*Diadema antillarum*) found in two neighbouring regions: North Bay and South Bay. Samples of adult urchins are collected from both regions, and the cadmium content of the urchins' internal tissues are measured as below, in μg of cadmium per kg of dry tissue weight.

Cadmium in North Bay urchin tissue ($\mu\text{g}/\text{kg}$ dry weight):

```
356 329 348 394 324 247 255 320 278 219
329 318 269 395 273 320 219 292 263 274
320 280 407 348 329 319 217 244 308 319
422 419 395 398 313 306 315 459 328 358
319 240 190 214 337 255 273 343 300 337
```

Cadmium in South Bay urchin tissue ($\mu\text{g}/\text{kg}$ dry weight):

```
189 285 276 207 339 189 357 164 310 318
173 419 251 157 462 311 337 248 246 345
316 359 287 191 184 171 105 308 256 147
243 154 409 226 250 229 382 304 269 385
274 276 86 243 259
```

Note: These observations are also available in a text file on Crowdmark, if you wish to make use of the `scan()` function along with the copy and paste capabilities on your computer.

- [2 marks]** Create a histogram of the cadmium content of the North Bay sea urchins. The title and x -axis for the histogram should have appropriate labels, including units where appropriate. Copy and paste the relevant command from the *R Console Window* and the resulting histogram into your Word document. (Do not include the code for how you stored the data into R, just include the one line of code for the histogram.)
- [2 marks]** Repeat the process in (a) for the South Bay urchins.
- [2 marks]** Create one side-by-side boxplot of the two sets of cadmium contents. The graph should be of a readable size, have an appropriate title, its x - and y -axes should have appropriate labels, and the two groups should be labelled. Copy and paste this boxplot and the line of code used to create it into your Word document.
- [3 marks]** Use R to calculate the means and standard deviations of the cadmium contents of the North Bay and South Bay urchins; for this question, use the dedicated commands for mean and standard deviation, not `summary`. Copy and paste the relevant commands and output from the R Console Window into your document. *In plain English, write a short statement summarizing the values of your R output. Explicitly state the values and their units in your statement.*
- [2 marks]** Answer the following question:

Do North Bay sea urchins and South Bay urchins appear to have a different tissue cadmium content? If so, which population appears to have a greater cadmium content?

Write a few plain English sentences explaining your opinion, making explicit references to the relevant features of the two data sets (including the mean, the median, the spread of the data, and the minimum and maximum values). Use your results from both parts (c) and (d) to support your statement. You may wish to run the `summary` command in R for each sample to gather information about the maximum and minimum values (if you do so, include your copy/pasted R code here too).

Part 2 Scenario: We will examine the relationship between soil depth (in cm) and nitrogen content (as percentage of dry mass) in permafrost samples taken from the Canadian subarctic. Use the bivariate dataset of 70 permafrost samples found in the `permaFrost.csv`¹ file on Crowdmark.

- (a) **[2 marks]** Create a scatterplot to compare the permafrost sample depth to nitrogen content. (Hint: here the depth is the value that can be controlled, so it is the x variable.) Your plot should have an appropriate title and the x - and y -axes should be labelled appropriately. Copy and paste the relevant commands and the plot into your Word document. (You do not need to include the code for how you stored the data into R, just include the line of code for the plot.)
- (b) **[1 mark]** In one or two sentences, describe the relationship (if any) between permafrost sample depth and nitrogen content that this scatterplot shows. (e.g. Is it linear or not? Positive or negative?)
- (c) **[2 marks]** Use the `cor` function to compute the correlation coefficient. (Copy and paste the relevant commands and output from the R Console Window into your Word document.) Does this value agree with your answer about the relationship permafrost sample depth and nitrogen content? Explain in a short sentence.

¹csv files can be opened in Excel or Google Sheets