

## STC 122 - Assignment 5 submission instructions

### Instructions

- Answer the questions that follow and save all your code in a single R script. Name the script **Assignment5.R** where **Assignment5** is the filename and **.R** is the file extension.
- Two submissions are required for this Assignment, namely a code submission and an interpretation submission.

### Submission 1: Code

- Submit your R script on Gradescope before **Tuesday, 24 October 2023, 23:59**.
- Multiple code submissions are allowed and your autograded results will be available shortly after each submission.
- **Ensure all variables are named correctly, as incorrectly named variables will not be awarded any marks.** (Remember variable names are case sensitive.)
- **Ensure your code does not contain any syntax errors.** If your code produces errors when run, the autograder will not be able to mark it.
- Any code commented out will be considered rough work and will not be marked.
- Once you have completed your submission, ensure the file is submitted on Gradescope with the correct file name, **Assignment5.R** where **Assignment5** is the filename and **.R** is the file extension. **The autograder will only be able to grade your submission if you use the correct filename.**
- Students must add **detailed** explanations of their working as comments in their code. This must be done for every question in the assignment. Students should also add citations for any resources they make use of in their assignment. Refer to the document `commenting_and_citing_in_your_code.pdf` on ClickUP.

### Submission 2: Interpretation

- Submission 2 will be an invigilated assessment.
- Full details regarding the date, time and format of this assessment will be communicated as soon as possible on ClickUP.

## Guidelines

- This assignment covers Linear Regression.
- The assignment is based on Lab 8 from the Openintro website. The lab is available on ClickUP. You should work through this lab before attempting this assignment.
- You should work through Chapters 7 and 24 in the prescribed textbook in order to understand the underlying theory to do this assignment.
- This assignment is based on an article published in *Journal of Statistics and Data Science Education*<sup>1</sup>
- Where applicable, answer the questions below by typing the appropriate **code** in the R script template provided on ClickUP. Some questions are theoretical and no coding is needed to answer those questions.
- For this assignment you should use the dataset in the file `lego_sample.csv` that is available on ClickUP.

## Questions

**Consider only the sets with themes "Lego City" and "Lego Friends".**

**Investigate the relationship between the Amazon Price of a set and the number of pieces contained in the set.**

### Question 1

Calculate the correlation coefficient between the Amazon price of a set and the number of pieces contained in the set. Save your response into a variable called Q1

### Question 2

Fit a linear regression model to determine whether the number of pieces contained in the set is a good predictor for the Amazon price of a set.

- a) Save the parameter estimates of the linear model into a variable called Q2a
- b) Predict the Amazon price of a Lego set with 350 pieces. Save your response into a variable called Q2b
- c) Determine the percentage variation in the response variable that is explained by the regression line. Give your answer as a percentage (the value should be between 0 and 100). Save your response into a variable called Q2c

---

<sup>1</sup>Anna D. Peterson & Laura Ziegler (2012) Building a Multiple Linear Regression Model with LEGO Brick Data, *Journal of Statistics and Data Science Education*, 29:3, 297-303, DOI: 10.1080/26939169.2021.1946450

### Question 3

Create a scatterplot comparing the Amazon price of a set and the number of pieces contained in the set. Overlay the least squares regression line calculated in Question 2.

### Question 4

Obtain the residuals of the linear regression model and save your response into a variable called Q4

### Question 5

Test whether the mean Amazon Price increases with R2 for an additional piece added to the set.

- a) Save the point estimate into a variable called Q5a
- b) Calculate the test statistic value and save your response into a variable called Q5b

**Investigate the relationship between the Amazon price of a set and the number of pieces contained in the set for different Lego themes**

### Question 6

Fit a linear regression model for each of the two themes. Save the slope parameter for the linear regression model fitted on the Lego Friends sets as Q6a and the slope parameter for the linear regression model fitted on the Lego City sets as Q6b

### Question 7

Fit a linear regression model using the theme to explain the Amazon price.

- a) Save the parameter estimates into a variable called Q7a
- b) Use an  $F$ -test to test whether the linear regression line is significant. Save the  $p$ -value of the test into a variable called Q7b