

Homework 4

STAT 462 (Fall 2020)

UNAUTHORIZED DISTRIBUTION AND/OR UPLOADING OF THIS DOCUMENT IS STRICTLY PROHIBITED.

Clearly label your answers to each question and each sub-question. Your answers MUST be uploaded to Canvas as a <HWx_Yourfirstname.nb.html> file by the deadline.

Suburban towns often spend a large fraction of their municipal budgets on public safety (police, fire, and ambulance) services. A taxpayers' group felt that very small towns were likely to spend large amounts per person because they have such small financial bases. The group obtained data on the per capita expenditure for public safety of 29 suburban towns in a metropolitan area, as well as the population of each town in units of 10,000 people. The data are given in *expenditure.csv*.

- (a). Identify the explanatory and response variable in the problem.
 - (b). If the taxpayer's group is correct, what sign should the slope of the regression model have?
 - (c). Fit a regression line for the problem. Does the slope in the output confirms the opinion of the group? Explain.
 - (d). Obtain a scatter plot for the problem, and comment on your observations.
 - (e). Does the scatter plot in part (d) suggest that the regression line estimated in part (c) is misleading? Explain.
 - (f). There seems to be an unusual observation (town) in upper right corner of the scatter plot. Remove that unusual observation from the dataset.
 - (g). Quantify the percent of per capita expenditure variation explained by town population, without fitting the regression line. Show your work.
 - (h). Obtain the least squares regression line with the unusual town removed. Write down the equation of the fitted line, define any terms you might have used.
 - (i) Interpret the slope of the regression line in part (h)
 - (j) Test if there is supporting evidence towards taxpayers' group belief (use the data without the unusual point)
 - (k) Test if the slope is significantly different from 1. Show all your work. Use the data without the unusual point.
- (2). Do a quick Google search on "extrapolation in regression". Write what is extrapolation in your words and explain the consequences.
- (3). Suppose we are interested in purchasing a multifunctioning inkjet printer. We want to see how the performance factors related to the price of the printer. To understand the relationship data has been collected from 20 printers. Corresponding data is available in the R package *Lock5Data*. Use the below code lines to install and load the package, and to load the *InkjetPrinters* dataset.

```
install.packages("Lock5Data")  
library(Lock5Data)  
data(InkjetPrinters)
```

Suppose we are interested in predicting the price of a printer from the printing rate.

PPM: Printing rate (pages per minutes) for a set of print jobs

Price: Typical retail price (in dollars) at the time of the review.

- (a). Make a scatterplot of the data and comment on your observations.
- (b). Obtain the least squares regression line for predicting the price using printing rate. Define any terms you used when writing the least squares regression line.
- (c). Interpret the slope parameter in the context of this problem?
- (d). We want to test if printers with higher printing rate are more expensive. Write the appropriate null and alternative hypothesis you would formulate to test this, calculate the test statistic and the p value. Based on the p value you obtained, what can you say about the relationship?
- (e). Suppose you found an inkjet printer with a printing speed of 3 pages per minute for a sale price of \$129. Is this a deal that should not be missed? To answer this question, should you be using 95% confidence interval or a prediction interval? Explain the reason for your choice of interval to answer the above question. Based on the interval you used, what is your decision about the deal?