

Homework 1

STAT 425 - Yu

Due: Sept 18, 2019

Notice: Answer everything! The questions are not rhetorical. The graders and I will take no pity on unanswered questions. (e.g. “What does ____ represent?”)

Please download the file ‘cship.dat’ from Compass, and load it into an R object called **cship**.

You should be able to do this with **read.table()**.

This dataset includes measurements of ship size, capacity, crew, and age for 158 cruise ships.

Question 1 – Warmup (4 points)

- Use the *summary* function to compute summary statistics of the the number of passengers and crew members. Be sure to interpret the output. What do these numbers represent?
- Plot the relationship between the number of passengers (x-axis) and the number of crew members (y-axis) (include appropriate axis labels). Describe the relationship between the variables.
- Use the *cor* function to compute the correlation between the number of passengers and crew members. Save the correlation in an object and use an inline code chunk to report the correlation in a sentence. Please note that the square of the correlation is R^2 .
- Using the *lm* function, fit a SLR model of the number of passengers and crew. Store this lm function into an object named *model1*. Print the summary. Interpret the results in a couple of sentences.

Question 2 – SLR (9 points)

Using the same dataset, (passengers = x, crew = y) calculate the following and output the results. Also, show your work by displaying the code you used. Do not use any extra functions or packages besides *predict()* and *resid()*

a-c) Calculate the following, and store it into an object by the same name:

- SXX
- SXY
- SYY
- Describe the model matrix used for fitting this linear regression. (i.e. what are its dimensions, what is in each column). Please use the actual numbers (not n and p).
- Using only the model matrix, X, and the vector of observations Y, calculate the LS estimates:

$$\hat{\beta}_0, \hat{\beta}_1$$

- Now, instead, use the *cor* function, and the results from Question 2 to calculate an estimate of

$$\hat{\beta}_1$$

- Calculate the RSS and use it to calculate an estimate of the error variance

$$\hat{\sigma}^2$$

- Using the estimate of the error variance, and the model matrix, calculate the variance-covariance matrix

$$Var[\hat{\beta}].$$

also h) What does the element in row 2 column 2 element represent? What does the element in row 2, column 1 represent?

- Calculate the R^2 of this SLR using the following:

$$y_i, \hat{y}_i, \bar{y}$$

Question 3 – MLR: Simulated data (7 points)

Important: First, set the seed to 217.

```
set.seed(217) #seriously, include this line or future you will be sorry
```

- Create an R vector named *x1* by generating 30 values from a random exponential distribution with **mean** = 5. Keep in mind that the mean of an exponential distribution is equal to 1/rate for a Poisson distribution.

#Extra hints:

#Hint 1: To generate from an exponential distribution, you can use the function: rexp()

#hint 2: Typing ?rexp will generate a helpful read-me file. Type that into your console, not the markdown document. Don't need to include this

#Hint 3: Read the help file carefully and double check to make sure your answer is reasonable

b. Create an R vector named `x2` by generating 30 values from a random normal distribution with *mean* = 5, and *sd* = 3.

c. Create a function that takes `x1` and `x2` as inputs, and outputs:

$$y = 2 * x1 - 6 * x2 + \epsilon,$$

where $\epsilon \sim N(0,4)$

d. Run a MLR on this data using `y` as the response, and `x1` and `x2` as predictors.

e. Show (in R) that $X^T \hat{\epsilon} = 0$, where `X` is the model matrix.

f. Which variables are significant at $p=0.05$? Comment on why you think these variables may have been/not been significant.

g. What is (calculate) the hat matrix for this particular model? Briefly describe(in general) what a hat matrix is/does.

“Don’t bring a knife to a gun fight.” -idiom

“Don’t bring a `set.seed(≠ 217)` simulation to a `set.seed(217)` grader” -Albert