

Due : November 3, 2022

Name:

PUID:

*Instruction: Please submit your R code along with a brief write-up of the solutions (do not submit raw output containing ERRORS). Some of the questions below can be answered with very little or no programming. However, write code that outputs the final answer and does not require any additional paper calculations.*

**Q.N. 1)** The R data frame "HairEyeColor" contains classifications of 592 students by gender, hair color and eye color.

- a) Is hair color independent of eye color for men?
- b) Is hair color independent of eye color for women?

**Q.N. 2)** A clinical dietician wants to compare two different diets, A and B, for diabetic patients. She hypothesizes that diet A (Group 1) will be better than diet B (Group 2), in terms of lower blood glucose. She plans to get a random sample of diabetic patients and randomly assign them to one of the two diets. At the end of the experiment, which lasts 6 weeks, a fasting blood glucose test will be conducted on each patient. She also expects that the average difference in blood glucose measure between the two group will be about 10 mg/dl. Furthermore, she also assumes the standard deviation of blood glucose distribution for diet A to be 15 and the standard deviation for diet B to be 17. How many subjects are needed in each group assuming equal sized groups? (Please use  $\alpha = 0.05$  and Power=0.8).

**Q.N. 3)** Suppose that a fire insurance company wants to relate the amount of fire damage in major residential fires to the distance between the burning house and the nearest fire station. The study is conducted in a suburb of a major metropolitan area. The data collected were the distance in miles between the nearest fire station and the fire and the amount of damage to the house ( in thousands of dollars).

Distance	Damage
3.4	26.2
1.8	17.8
4.6	31.3
2.3	23.1
3.1	27.5
5.5	36.0
0.7	14.1
3.0	22.3
2.6	19.6
4.3	31.3
2.1	24.0
1.1	17.3
6.1	43.2
4.8	36.2
3.8	26.5

- a) Fit a simple linear regression model and analyze the residual plots.
- b) What is the expected Damage if the fire station is 4 miles away?
- c) Use the Box-Cox transformation to choose an appropriate value of  $\lambda$  to improve the model.
- d) Fit a simple linear regression model after transformation.
- e) Compare and contrast models in (a) and (d).

**Q.N. 4)** An author maintains a website on a particular book and using Google Analytics, records the number of visits on this particular website on each day of the year. As expected there are more hits during weekdays then on weekends. Since the book is used as a textbook for a statistics course there are more hits during the time when the classes are in session. Table below provides the data for 35 weeks from April through November 2009. To explore the week by week visit patterns of these

Week	Hits
1	148
2	148
3	157
4	112
5	125
6	155
7	154
8	135
9	140
10	164
11	154
12	138
13	129
14	131
15	113
16	124
17	119
18	110
19	166
20	105
21	132
22	132
23	144
24	152
25	152
26	166
27	161
28	168
29	170
30	179
31	154
32	136
33	147
34	151
35	188

- Display the data using a scatterplot.
- Calculate the correlation coefficient to measure the association between the week and the number of hits on the website. Check whether rank correlation is more appropriate than Pearson correlation
- Test for the significance of the correlation at **0.05** level.

**Q.N. 5)** The data set `cars` is one of the data sets installed with R and is available in base package. The data set contains 50 observations of `speed(mph)` and `dist(stopping distance in feet)`.

- a) Display the data using scatter plot.
- b) Fit a simple regression model using `speed` as a predictor variable.
- c) Add the fitted line to the scatter plot.
- d) Calculate the residuals and fitted values and print only first five observations of the residuals and fitted values.
- e) Create a scatter plot of the residuals and fitted values.
- f) Assuming that no intercept model is appropriate fit a simple linear regression model.
- g) Calculate and compare the coefficient of determination for both the with intercept and no-intercept models.
- h) Using your fitted model predict the stopping distance for a car with an speed of 21 mph.