CSCI E-106:Assignment 2

Due Date: September 21, 2020 at 7:20 pm EST

Instructions

Students should submit their reports on Canvas. The report needs to clearly state what question is being solved, step-by-step walk-through solutions, and final answers clearly indicated. Please solve by hand where appropriate.

Please submit two files: (1) a R Markdown file (.Rmd extension) and (2) a PDF document, word, or html generated using knitr for the .Rmd file submitted in (1) where appropriate. Please, use RStudio Cloud for your solutions.

Problem 1

Refer to the regression model $Y_i = \beta_0 + \epsilon_i$. (25pts)

a-) Derive the least squares estimator of β_0 for this model.(10pts)

b-) Prove that the least squares estimator of β_0 is unbiased.(5pts)

c-) Prove that the sum of the Y observations is the same as the sum of the fitted values.(5pts)

d-) Prove that the sum of the residuals weighted by the fitted values is zero.(5pts)

Problem 2

Refer to the Grade point average Data. The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from the ACT test score (X). (30 points, each part is 5 points)

a-) Obtain a 99 percent confidence interval for β_1. Interpret your confidence interval. Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?

b-) Test, using the test statistic t^*, whether or not a linear association exists between student's ACT score (X) and GPA at the end of the freshman year (Y). Use a level of significance of $\alpha = 0.01$. State the alternatives, decision rule, and conclusion.

c-) What is the P-value of your test in part (b)? How does it support the conclusion reached in part (b)?

d-)Obtain a 95 percent interval estimate of the mean freshman GPA for students whose ACT test score is 28. Interpret your confidence interval.

e-) Mary Jones obtained a score of 28 on the entrance test. Predict her freshman GPA-using a 95 percent prediction interval. Interpret your prediction interval.

f-) Is the prediction interval in part (e) wider than the confidence interval in part (d)? Should it be?

g-) Determine the boundary values of the 95 percent confidence band for the regression line when $X_h = 28$. Is your-confidence band wider at this point than the confidence interval in part (d)? Should it be?

Problem 3

Refer to the Crime rate data. A criminologist studying the relationship between level of education-and crime rate in medium-sized U.S. counties collected the following data for a random sample of 84 counties; X is the percentage of individuals in the county having at least a high-school diploma, and Y is the crime rate (crimes reported per 100,000 residents) last year. (45 points, each part is 5 points)

a-)Obtain the estimated regression function. Plot the estimated regression function and the data. Does the linear regression function appear to give a good fit here? Discuss.

b-) Test whether or not there is a linear association between crime rate and percentage of high school graduates, using a t test with $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c-) Estimate β_1, with a 99 percent confidence interval. Interpret your interval estimate.

d-) Set up the ANOVA table.

e-) Carry out the test in part a by means of the F test. Show the numerical equivalence of the two test statistics and decision rules. Is the P-value for the F test the same as that for the t test?

f-) By how much is the total variation in crime rate reduced when percentage of high school graduates is introduced into the analysis? Is this a relatively large or small reduction?

g-) State the full and reduced models.

h-) Obtain (1) SSE(F), (2) SSE(R), (3) dfF. (4) dfR, (5) test statistic F* for the general linear test, (6) decision rule.

i-)Are the test statistic F* and the decision rule for the general linear test numerically equivalent to those in part a?