Solution: Page 5 in https://github.com/yidongzhou/STA-108B-SQ-2021/blob/main/STA108BDiscussion.pdf

University of California, Davis

Department of Statistics

Midterm 1– Spring 2021

Course Title: Applied Statistical Methods: Regression Analysis
Date of Examination: 4/30/2021
Teacher's name: Jairo Fúquene-Patiño
Student's name: ____________________
Course Code: STA 108

Sections: B01 and B02.
Time duration: 1 hour for Q.1. The deadline for Q.2 is 5/3/2021, 5:00 p.m.
Total marks: 100

Q.1) Using the following observations You can use the software R to compute ONLY the probabilities/quantiles for the specific problem. You need to show the steps to find the solutions.):

Observation	1	2	3	4	5	6
X_i	43	39	41	86	72	76
Y_i	106	106	97	60	70	80

- Compute the estimates b_0 and b_1 for the parameters β_0 and β_1 in the simple linear regression model (20 points):

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

  - Y_i: is the value of the response variable in the $i-th$ trial.
  - β_0 and β_1 are the parameters.
  - X is a known constant, namely, the value of the predictor variable in the $i-th$ trial.
  - ϵ_i is a random error term with $E(\epsilon_i) = 0$ and $V(\epsilon_i) = \sigma^2$ and follows a normal distribution $N(\epsilon_i; 0, \sigma^2)$
  - ϵ_i and ϵ_j are uncorrelated so that their covariance is zero (i.e., $cov(\epsilon_i, \epsilon_j) = 0$ for all i, j; with $i \neq j$ and $i = 1, ..., n$
- Estimate a 90% confidence interval for the parameters β_0 and β_1 (20 points).
- Using $\alpha = 0.01$ compute the p-value for the two alternatives in formula

$$H_0 : \beta_1 = 0$$

  versus

$$H_a : \beta_1 \neq 0,$$

  state the decision rule and conclusion (20 points).
- Compute a 95% confidence interval for the mean response $E\{Y_h\}$ using $X_h = 30$ (20 points).
- Test whether or not $\beta_1 = 0$ using an F test with $\alpha = 0.05$. State the alternatives, decision rule, and conclusion (20 points).

80 marks

t test

Two-Sided Test A cost analyst in the Toluca Company is interested in testing, using regression model (2.1), whether or not there is a linear association between work hours and lot size, i.e., whether or not $\beta_1 = 0$. The two alternatives then are:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0 \tag{2.16}$$

The analyst wishes to control the risk of a Type I error at $\alpha = .05$. The conclusion H_a could be reached at once by referring to the 95 percent confidence interval for β_1 constructed earlier, since this interval does not include 0.

An explicit test of the alternatives (2.16) is based on the test statistic:

$$t^* = \frac{b_1}{s\{b_1\}} \tag{2.17}$$

The decision rule with this test statistic for controlling the level of significance at α is:

$$\text{If } |t^*| \leq t(1 - \alpha/2; n - 2), \text{ conclude } H_0$$
$$\text{If } |t^*| > t(1 - \alpha/2; n - 2), \text{ conclude } H_a \tag{2.18}$$

t(1-0.01/2; 4)=4.6

For the Toluca Company example, where $\alpha = .05$, $b_1 = 3.5702$, and $s\{b_1\} = .3470$, we require $t(.975; 23) = 2.069$. Thus, the decision rule for testing alternatives (2.16) is:

$$\text{If } |t^*| \leq 2.069, \text{ conclude } H_0$$
$$\text{If } |t^*| > 2.069, \text{ conclude } H_a$$

Since $|t^*| = |3.5702/.3470| = 10.29 > 2.069$, we conclude H_a, that $\beta_1 \neq 0$ or that there is a linear association between work hours and lot size. The value of the test statistic, F test F^*=MSR/MSE

Construction of Decision Rule. Since the test is upper-tail and F^* is distributed as $F(1, n-2)$ when H_0 holds, the decision rule is as follows when the risk of a Type I error is to be controlled at α:

$$\begin{aligned} &\text{If } F^* \leq F(1-\alpha; 1, n-2), \text{ conclude } H_0 \\ &\text{If } F^* > F(1-\alpha; 1, n-2), \text{ conclude } H_a \end{aligned} \tag{2.62}$$

where $F(1-\alpha; 1, n-2)$ is the $(1-\alpha)100$ percentile of the appropriate F distribution.

Here is \alpha, rather than \alpha/2

For the Toluca Company example, we shall repeat the earlier test on β_1, this time using the F test. The alternative conclusions are:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

As before, let $\alpha = .05$. Since $n = 25$, we require $F(.95; 1, 23) = 4.28$. The decision rule is:

$$\text{If } F^* \leq 4.28, \text{ conclude } H_0$$
$$\text{If } F^* > 4.28, \text{ conclude } H_a$$

We have from earlier that $MSR = 252{,}378$ and $MSE = 2{,}384$. Hence, F^* is:

$$F^* = \frac{252{,}378}{2{,}384} = 105.9$$

Since $F^* = 105.9 > 4.28$, we conclude H_a, that $\beta_1 \neq 0$, or that there is a linear association between work hours and lot size. This is the same result as when the t test was employed, as it must be according to our discussion below.

Q.2) Muscle mass. A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each 10-year age group, beginning with age 40 and ending with age 79. The results follow; X is age, and Y is a measure of muscle mass (20 points) . Use the software R or R-Studio and the attached data set.

1. Set up the ANOVA table. anova()
2. Test whether or not $\beta_1 = 0$ using an F test with $\alpha = 0.05$. State the alternatives, decision rule, and conclusion.
3. Obtain R^2 and r. The sign of r depend on \beta_1
4. Compute the diagnostic plots. State the conclusions.

20 marks