Econ 103: Homework 2

Manu Navjeevan

August 18, 2021

Single Linear Regression Theory Review

1. Recall that we define our parameters of interest β_0 and β_1 as the parameters governing the "line of best fit" between Y and X:
$$\beta_0, \beta_1 = \arg\min_{b_0, b_1} \mathbb{E}[(Y - b_0 - b_1 X)^2]. \tag{1}$$
Once we define these parameters we define the regression error term $\epsilon = Y - \beta_0 - \beta_1 X$ which then generates the linear model
$$Y = \beta_0 + \beta_1 X + \epsilon.$$
   (a) Using the first order conditions for β_0 and β_1 (set the derivatives of the right hand side of (1) with respect to b_0 and b_1 equal to zero at) show why $\mathbb{E}[\epsilon] = \mathbb{E}[\epsilon X] = 0$.

   (b) Using the definition of β_0 and β_1 as line of best fit parameters, give an intuitive explanation for why $\mathbb{E}[\epsilon] = 0$.

Hypothesis Testing and Confidence Intervals

In the following questions, whenver running a hypothesis test, please state the null and alternative hypotheses, show some work, and state the conclusion of the test.

1. In an estimated simple regression model based on $n = 64$, the estimated slope parameter, $\hat{\beta}_1$, is 0.310 and the standard error of $\hat{\beta}_1$ is 0.082.

   (a) What is $\hat{\sigma}^2_{\beta_1}$? Recall σ_{β_1} is the terms such that, approximately for large n,
$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \sim N(0, \sigma_{\beta_1}).$$

   (b) Test the hypothesis that the slope is zero against the alternative that it is not at the 1% level of significance ($\alpha = 0.01$).

   (c) Test the hypothesis that the slope is negative against the alternative that it is positive at the 1% level of significance ($\alpha = 0.01$).

   (d) Test the hypothesis that the slope is positive against the alternative that it is negative at the 5% level of significance. What is the p-value?

   (e) Generate a 99% confidence interval for β_1. How can we use this interval to run the hypothesis test in part (b)?

2. Consider a simple regression of log-income (income is measured thousands of dollars), Y, against years of education, X. After collecting a sample of size $n = 50$ we estimate the following regression equation.
$$\widehat{Y} = \hat{\beta}_0 + 0.0180X.$$
   (a) Using the following information to solve for $\hat{\beta}_0$ as well as the estimated variance $\widehat{\text{Var}}(\hat{\beta}_0)$, which is the square of the standard error.

- The standard error of $\hat{\beta}_0$ is 2.174
- The test statistic, t^*, associated with the hypothesis test for

$$H_0 : \beta_0 = 0 \text{ vs. } H_1 : \beta_0 \neq 0,$$

is equal to 1.257.

(b) Use the following information to solve for the standard error $\hat{\beta}_1$ as well as the estimated variance $\widehat{\text{Var}}(\hat{\beta}_1)$, which is the square of the standard error.

- The test statistic, t^*, associated with the hypothesis test for

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0,$$

is equal to 5.754

(c) Given that Y is a logged variable, $Y = \log(\text{income})$, how do we interpret $\hat{\beta}_1$?

(d) Suppose that we are interested in the average value of log-income for someone with 16 years of education. We want to use the model above to test the hypothesis that the average value of log-income for someone with 16 years of education is less than or equal to 1.85. That is we want to test

$$H_0 : \lambda = \beta_0 + 16\beta_1 \leq 1.85 \text{ vs. } H_1 : \lambda = \beta_0 + 16\beta_1 > 1.85.$$

Use the fact that $\widehat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_1) = 2.84$ to test this hypothesis at level $\alpha = 0.1$.

(e) Use the above to generate a 90% confidence interval for λ.

3. (Challenge) Suppose we find that $\hat{\beta}_1 > 0$. If we reject the null hypothesis that $\beta_1 = 0$ in favor of an alternative hypothesis that $\beta_1 \neq 0$ at level α, up to what level can we be sure that would we reject the null hypothesis that $\beta_1 \leq 0$ against an alternative that $\beta_1 > 0$? (Please give some explanation here as well as your answer, which will be some multiple of α).

R^2 and Goodness of Fit

1. Consider the following estimated regression equation.

$$\widehat{Y} = 6.83 + 0.869X.$$

Write the estimated regression equation that would result if

(a) All values of X were divided by 20 before estimation.

(b) All values of Y were divided by 20 before estimation.

(c) All values of X and Y were divided by 20 before estimation.

2. Given the quantities in the questions below, calculate and interpret R^2:

(a) $\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 631.63$ and $\sum_{i=1}^{n} \hat{\epsilon}_i^2 = 182.85$.

(b) $\sum_{i=1}^{n} Y_i^2 = 5930.94$, $\bar{Y} = 16.035$, $n = 20$, and $\text{SSR} = 666.72$.

3. Suppose $R^2 = 0.7911$, $\text{SST} = 552.36$, and $n = 20$. Find $\hat{\sigma}_\epsilon^2$.