HUDM 5126 Linear Models and Regression Analysis Homework 2[*]

Yifei Dong

Sep 9 2020

1 Grade Point Average

2.4 Refer to **Grade point average** Problem 1.19.

(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?

The 99% confidence interval for β_1 is (0.0054, 0.0723). We are 99% confident that for each extra 1 point increase in ACT scores, student's Grade Point Average increase by as little as 0.0054 to as much as 0.0723. It does not include zero. The director of admissions should be interested in whether the confidence interval includes zero because if the confidence interval includes zero, then we should conclude that increase student's ACT scores have no contributions on GPA increase.

(b) Test, using the test statistic t^*, whether or nor 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 .01. State the alternatives, decision rule, and conclusion.

$$H_0\text{: } \beta_1 = 0$$
$$H_a\text{: } \beta_1 \neq 0$$

The decision rule (Two-sided Test):

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

Since t-statistic=3.040 > t(0.995,118)=2.618, we reject H_0 and conclude that β_1 is significant and not equal to 0.

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

The P-value is 0.00292. Since 0.00292 < 0.01, it supports the conclusion reached in part (b).

Extra Exercises[1]: The scatter plot with the regression line, confidence intervals and prediction intervals can be found in Figure 1.

---

[*]This homework is written in LaTeX.
[1]The graph is made using ggplot2 package.

Figure 1: This is a scatter plot with the regression line, confidence intervals and prediction intervals

Confidence Intervals for Mean Values and Prediction Intervals for New Values
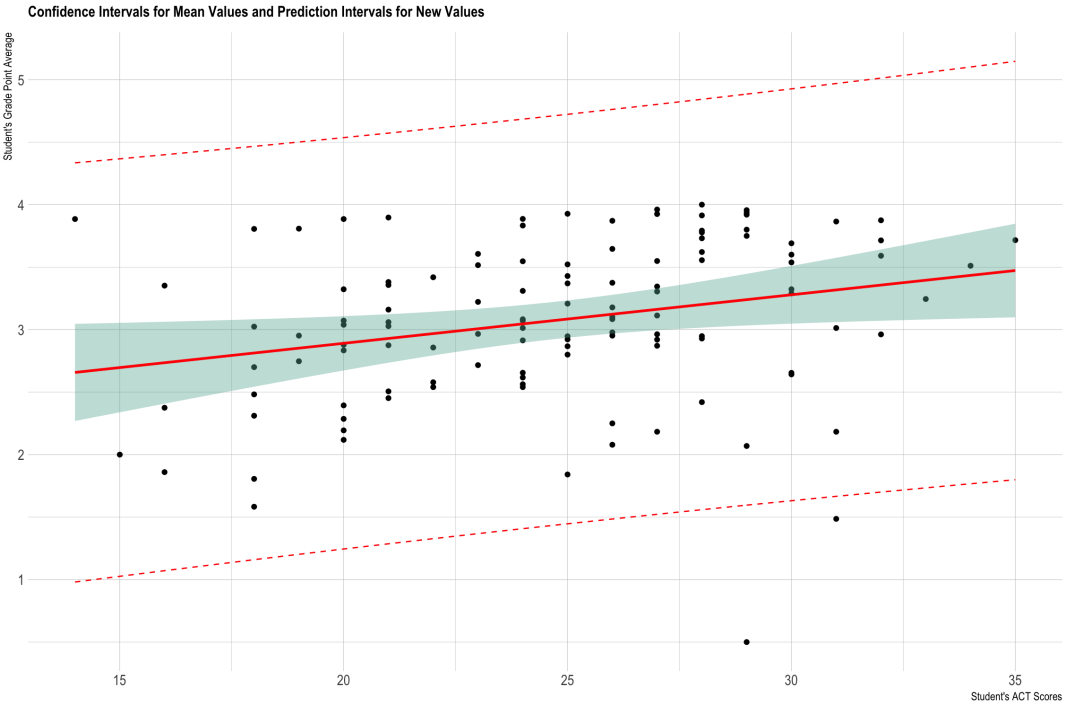
2 Bid Preparation

2.44 A building construction consultant studied the relationship between cost of bid preparation (Y_1) and amount of bid (Y_2) for the consulting firm's clients. In a sample of 103 bids prepared by clients, $r_{12} = .87$. Assume that bivariate normal model (2.74) applies.

(a) Test whether or not $\rho_{12} = 0$; control the risk of Type I error at .10. State the alternatives, decision rule, and conclusion. What would be the implication if $\rho_{12} = 0$?

$$H_0\colon \rho_{12} = 0$$
$$H_a\colon \rho_{12} \neq 0$$

The decision rule:

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

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{.87 \cdot \sqrt{103-2}}{\sqrt{1-.87^2}} = 17.73321 \tag{1}$$

t(.95,101)=1.66. Since $17.733 > 1.66$, we reject the null hypothesis and conclude that $\rho_{12} \neq 0$. If $\rho_{12} = 0$, then we conclude that there is no linear association between cost of bid preparation (Y_1) and amount of bid (Y_2) for the consulting firm's clients. In the case where Y_1 and Y_2 are jointly normally distributed, $\rho_{12} = 0$ implies that Y_1 and Y_2 are independent.

(b) Obtain a 90 percent confidence interval for ρ_{12}. Interpret this interval estimate.

The point estimate of ρ_{12} was r_{12}=.87. To obtain an approximate 90 percent confidence interval estimate, we require:

$$\mathbf{E}\{z'\} = \zeta = \frac{1}{2}\log_e(\frac{1+\rho_{12}}{1-\rho_{12}}) = \frac{1}{2}\log_e(\frac{1+.87}{1-.87}) = \frac{1}{2}\log_e(\frac{1.87}{.13}) = 1.3331 \tag{2}$$

$$\sigma^2\{z'\} = \frac{1}{n-3} = \frac{1}{103-3} = .01 \tag{3}$$

therefore, $\sigma = \sqrt{.01} = .1$. and $z(1 - \alpha/2) = z(1 - 0.1/2) = z(1 - .05) = z(.95) = 1.645$.

Hence, the confidence limits for ζ, are $1.333 \pm 1.645(.1)$, and the approximate 90 percent confidence interval is:

$$1.1685 \leq \zeta \leq 1.4975$$

To transform back to ρ_{12}, using the following equation,:

$$\rho_{12} = \frac{e^{2\zeta} - 1}{1 + e^{2\zeta}} \tag{4}$$

We obtain:

$$0.824 \leq \rho_{12} \leq 0.905$$

(c) Convert the confidence interval in part(b) to a 90 percent confidence for ρ_{12}^2.

$$(0.824)^2 \leq \rho_{12}^2 \leq (0.905)^2 = 0.679 \leq \rho_{12}^2 \leq 0.819$$

3   Proof

Show that the ratio SSR/SSTO is the same whether Y_1 is regressed on Y_2 or Y_2 is regressed on Y_1. [*Hint*: Use (1.10a) and (2.51)].

Recall equation 1.10a and equation 2.51:

$$b_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2} \tag{1.10a}$$

$$SSR = b_1^2 \sum(X_i - \overline{X})^2 \tag{2.51}$$

First, Y_1 is regressed on Y_2, then $\widehat{Y_1} = b_0 + b_1 Y_2$

$$b_1 = \frac{\sum(Y_{i2} - \overline{Y_2})(Y_{i1} - \overline{Y_1})}{\sum(Y_{i2} - \overline{Y_2})^2} \tag{3.11}$$

$$SSR = \left[\frac{\sum(Y_{i2} - \overline{Y_2})(Y_{i1} - \overline{Y_1})}{\sum(Y_{i2} - \overline{Y_2})^2}\right]^2 \left[\sum(Y_{i2} - \overline{Y_2})^2\right] \tag{3.12}$$

$$SSTO = \sum(Y_{i1} - \overline{Y_1})^2 \tag{3.13}$$

therefore,

$$\frac{SSR(Y_2)}{SSTO} = \frac{\left[\sum(Y_{i2} - \overline{Y_2})(Y_{i1} - \overline{Y_1})\right]^2}{\sum(Y_{i1} - \overline{Y_1})^2 \sum(Y_{i2} - \overline{Y_2})^2} \tag{3.14}$$

Second, Y_2 is regressed on Y_1, then $\widehat{Y_2} = b_0 + b_1 Y_1$

$$b_1 = \frac{\sum(Y_{i1} - \overline{Y_1})(Y_{i2} - \overline{Y_2})}{\sum(Y_{i1} - \overline{Y_1})^2} \tag{3.21}$$

$$SSR = \left[\frac{\sum(Y_{i1} - \overline{Y_1})(Y_{i2} - \overline{Y_2})}{\sum(Y_{i1} - \overline{Y_1})^2}\right]^2 \left[\sum(Y_{i1} - \overline{Y_1})^2\right] \tag{3.22}$$

$$SSTO = \sum(Y_{i2} - \overline{Y_2})^2 \tag{3.23}$$

therefore,

$$\frac{SSR(Y_1)}{SSTO} = \frac{\left[\sum(Y_{i1} - \overline{Y_1})(Y_{i2} - \overline{Y_2})\right]^2}{\sum(Y_{i1} - \overline{Y_1})^2 \sum(Y_{i2} - \overline{Y_2})^2} \tag{3.24}$$

Since equation 3.14 equals equation 3.24, we conclude that the ratio SSR/SSTO is the same whether Y_1 is regressed on Y_2 or Y_2 is regressed on Y_1.