

Week 1 - AYU - Individual

Son Nguyen

2023-04-03

Practice Problems

Type 1: Calculating parameter estimators from the sums.

Problem 1.

Determine which of the following statements is NOT true about the equation $Y = \beta_0 + \beta_1 X + \epsilon$

- (A) β_0 is the expected value of Y.
- (B) β_1 is the average increase in Y associated with a one-unit increase in X.
- (C) The error term, ϵ , is typically assumed to be independent of X.
- (D) The equation defines the population regression line.
- (E) The method of least squares is commonly used to estimate the coefficients β_0 and β_1 .

Problem 2.

x	2	3	5	6	1	9
y	1	4	6	4	4	3

Write the equation of the best fitted line.

Problem 3. (SRM - Sample Question 17)

The regression model is $y = \beta_0 + \beta_1 x + \epsilon$. There are six observations. The summary statistics are:

$$\sum y_i = 8.5, \bar{x} = 6, \sum x_i^2 = 16, \sum x_i y_i = 15.5, \sum y_i^2 = 17.25 \quad (1)$$

Calculate the least squares estimate of β_1 .

- (A) 0.1
- (B) 0.3
- (C) 0.5
- (D) 0.7
- (E) 0.9

Problem 4.

The regression model is $y = \beta_0 + \beta_1 x + \epsilon$. There are 8 observations. The summary statistics are:

$$\bar{y} = 12.625, \bar{x} = 6, \sum x_i^2 = 390, \sum x_i y_i = 753, \sum y_i^2 = 1557 \quad (2)$$

Write the equation of the best fitted line.

Problem 5

The regression model is $y = \beta_0 + \beta_1 x + \epsilon$. You are given the follows.

$$n = 10, \bar{y} = 12.625, \bar{x} = 6, \sum x_i^2 = 390, \sum x_i y_i = 753, \sum y_i^2 = 1557 \quad (3)$$

Predict y_{11} given that $x_{11} = 20$

Problem 6 (SRM - Sample Question 11) You are given the following results from a regression model.

Observation number (i)	y_i	$\hat{f}(x_i)$
1	2	4
2	5	3
3	6	9
4	8	3
5	4	6

Calculate the sum of squared errors (SSE).

- (A) -35
- (B) -5
- (C) 5
- (D) 35
- (E) 46

Problem 7 (SRM - Sample Question 18) For a simple linear regression model the sum of squares of the residuals is

$$\sum_{i=1}^{25} e_i^2 = 230$$

and the R^2 statistic is 0.64. Calculate the total sum of squares (TSS) for this model.

- (A) 605.94
- (B) 638.89
- (C) 690.77
- (D) 701.59
- (E) 750.87

Problem 8 (SRM - Sample Question 47)

You are given the following summary statistics:

$$\bar{x} = 3.500 \quad (4)$$

$$\bar{y} = 2.840 \quad (5)$$

$$\sum (x_i - \bar{x})^2 = 10.820 \quad (6)$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = 2.677 \quad (7)$$

$$\sum (y_i - \bar{y})^2 = 1.125 \quad (8)$$

Determine the equation of the regression line, using the least squares method.

- (A) $y = 1.97 + 0.25x$
- (B) $y = 0.78 + 0.59x$
- (C) $y = 0.57 + 0.65x$
- (D) $y = 0.39 + 0.70x$
- (E) The correct answer is not given by (A), (B), (C), or (D).

Type 2: Statistical Inference: Hypothesis Testing

Problem 9 (SRM - Sample Question 44)

Two actuaries are analyzing dental claims for a group of $n = 100$ participants. The predictor variable is sex, with 0 and 1 as possible values.

Actuary 1 uses the following regression model:

$$Y = \beta + \epsilon$$

Actuary 2 uses the following regression model:

$$Y = \beta_0 + \beta_1 \times Sex + \epsilon$$

The residual sum of squares for the regression of Actuary 2 is 250,000 and the total sum of squares is 490,000.

Calculate the F-statistic to test whether the model of Actuary 2 is a significant improvement over the model of Actuary 1.

- (A) 92
- (B) 93
- (C) 94
- (D) 95
- (E) 96

Type 3: Statistical Inference: Confidence Intervals and prediction Interval

Problem 10 (SRM - Sample Question 13) Determine which of the following statements is/are true for a simple linear relationship, $y = \beta_0 + \beta_1 x + \epsilon$

I. If $\epsilon = 0$, the 95% confidence interval is equal to the 95% prediction interval

II. The prediction interval is always at least as wide as the confidence interval.

III. The prediction interval quantifies the possible range for $E(y|x)$

- (A) I only
- (B) II only
- (C) III only
- (D) I, II, and III
- (E) The correct answer is not given by (A), (B), (C), or (D)

Problem 11 (SRM - Sample Question 13)

In a simple linear regression model based on over 100 observations, you are given the following estimates.

- i) The estimated slope is -1.03 .
- ii) The standard error of the estimated slope is 0.06 .

Calculate the 95% confidence interval for the slope.

- (A) $(-1.15, -0.91)$
- (B) $(-1.13, -0.93)$
- (C) $(-1.11, -0.95)$
- (D) $(-1.09, -0.97)$
- (E) $(-1.07, -0.99)$

Problem 12 (SRM - Sample Question 56)

Determine which of the following statements about prediction is true.

- (A) Each of several candidate regression models must produce the same prediction.
- (B) When making predictions, it is assumed that the new observation follows the same model as the one used in the sample.
- (C) A point prediction is more reliable than an interval prediction.
- (D) A wider prediction interval is more informative than a narrower prediction interval.
- (E) A prediction interval should not contain the single point prediction

Type 4: Application of Linear Model.

Problem 13 (SRM - Sample Question 23)

Toby observes the following coffee prices in his company cafeteria:

- 12 ounces for 1.00
- 16 ounces for 1.20
- 20 ounces for 1.40

The cafeteria announces that they will begin to sell any amount of coffee for a price that is the value predicted by a simple linear regression using least squares of the current prices on size.

Toby and his co-worker Karen want to determine how much they would save each day, using the new pricing, if, instead of each buying a 24-ounce coffee, they bought a 48- ounce coffee and shared it.

Calculate the amount they would save.

- (A) It would cost them 0.40 more.
- (B) It would cost the same.
- (C) They would save 0.40.
- (D) They would save 0.80.
- (E) They would save 1.20.