UNIVERSITY OF TEXAS AT AUSTIN

HW Assignment 3

Linear Regression.

Please, provide your **complete solutions** to the following problems. Final answers only, even if correct will earn zero points for those problems.

Problem 3.1. (10 points) Solve Problem **3.7.5** from the textbook (pp. 122-123).

Solution: Starting from the given fit, we get

$$\begin{aligned}\hat{y}_i &= x_i\hat{\beta} \\ &= x_i \frac{\sum_{i'=1}^n x_{i'}y_{i'}}{\sum_{j=1}^n x_j^2} \\ &= \frac{\sum_{i'=1}^n x_i x_{i'}y_{i'}}{\sum_{j=1}^n x_j^2} \\ &= \sum_{i'=1}^n \frac{x_i x_{i'}y_{i'}}{\sum_{j=1}^n x_j^2} \\ &= \sum_{i'=1}^n \frac{x_i x_{i'}}{\sum_{j=1}^n x_j^2} y_{i'}\end{aligned}$$

therefore,

$$a_{i'} = \frac{x_i x_{i'}}{\sum_{j=1}^n x_j^2}$$

Problem 3.2. (5 points) *Source: Sample SRM Problem #11.*
You are given the following results from a simple 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).

Solution:

$$(2-4)^2 + (5-3)^2 + (6-9)^2 + (8-3)^2 + (6-4)^2 = 46$$

Problem 3.3. (5 points) *Source: Sample SRM Problem #18.*
For a simple linear regression model the sum of squares of the residuals equals 230 while the coefficient of determination equals 0.64. Calculate the total sum of squares (TSS) for this model.

Solution:

$$TSS = \frac{\sum_{i=1}^n e_i^2}{1-R^2} = \frac{230}{1-0.64} = 638.8889.$$

Problem 3.4. (10 points) *Source: Sample SRM Problem #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.

Solution: There is no reason to use least squares here since the three given price points are all on the same line. The slope of the line is

$$\frac{1.2-1}{16-12} = \frac{0.2}{4} = 0.05$$

We get the intercept from

$$1 - 12(0.05) = 0.4$$

So, the line is

$$y = 0.05x + 0.04$$

The price of a single 24-ounce coffee is

$$0.05(24) + 0.4 = 1.6.$$

The price of a single 48-ounce coffee is

$$0.05(48) + 0.4 = 2.8.$$

Toby and Karen will save $3.2 - 2.8 = 0.4$.

Problem 3.5. (5 points) *Source: An old CAS problem.*
Two variables X and Y exhibit the following relationship:

$$Y = 2 + 1.5X + \varepsilon$$

where ε stands for the standard normal error term independent from X.

Of course, this exact relationship is unknown to the actuary who fits a simple linear regression using ordinary least squares on a data set. In our usual notation, the estimates of the two parameters are $\hat{\beta}_0 = 2.5$ and $\hat{\beta}_1 = 1.3$.

Calculate the *bias* of the estimator.

Solution: This was a trick question on the actual MAS-I exam! I would not give you such a misleading question. There is no bias because the least squares procedure is unbiased in its estimation. The given two estimates of the parameters are red herrings.

Problem 3.6. (15 points) In the context of simple linear regression and using our standard notation, prove that

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$$

Solution: Proving

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$$

is equivalent to proving

$$\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0.$$

As we did in class, let us denote $\varepsilon_i = y_i - \hat{y}_i$ for all $i = 1, \dots, n$. Recall that the sum of residuals is equal to zero, i.e., $\sum \varepsilon_i = 0$. Also, by the least-squares condition, $\sum \varepsilon_i x_i = 0$. Then,

$$\begin{aligned} \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum \varepsilon_i(\beta_0 + \beta_1 x_i - \beta_0 - \beta_1 \bar{x}) \\ &= \sum \varepsilon_i(\beta_1 x_i - \beta_1 \bar{x}) \\ &= \beta_1 \sum \varepsilon_i x_i - \beta_1 \bar{x} \sum \varepsilon_i = 0. \end{aligned}$$