STA 471 – Regression Analysis
Homework #1

September 14th, 2023

Richard McCormick

RLM443

1. Suppose that you are asked to fit a model to the data (X_i, Y_i), i = 1, 2, ..., n. If the model is $Y = \beta_1 X + \varepsilon$, derive the least squares estimate of β_1.

$$\frac{d}{d(\beta_1)} \sum_{i=1}^{n} (Y_i - \beta_1 X_i)^2 = \sum_{i=1}^{n} 2(Y_i - (\beta_1 X_i)(-X_i))$$

$$0 = \sum_{i=1}^{n} (X_i Y_i) - (\beta_1 X_i^2)$$

$$= \sum_{i=1}^{n} X_i Y_i - \beta_1 \sum_{i=1}^{n} X_i^2$$

$$\beta_1 \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$$

$$\boxed{\beta_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}}$$

2. Show that the point $(\bar{X},\ \bar{Y})$ lies on the fitted line $\hat{Y} = \beta_0 + \beta_1 X$.

$$\hat{y} = b_0 + b_1 X \quad \text{For point } (\bar{X}, \bar{Y})$$

$$\hat{y} = b_0 + b_1 \bar{X} \qquad b_0 = \bar{y} - b_1 \bar{x}$$

$$\hat{y} = (\bar{y} - \cancel{b_1 \bar{x}}) + \cancel{b_1 \bar{X}}$$

$$\boxed{\hat{y} = \bar{y}}$$

3. For the data given in problem F on page 99, consider the model $Y = \beta_0 + \beta_1 X + \varepsilon$ and answer the following questions (without the use of a statistical computer package):

a) Determine the equation of the fitted line.

From Text

X	4.7	5.0	5.2	5.2	5.9	4.7	5.9	5.2	5.3	5.9	5.6	5.0
Y	3	3	4	5	10	2	9	3	7	6	6	4

$\sum_1 X = 63.6 \quad \sum_1 X^2 = 339.18 \quad \sum_1 Y = 62 \quad \sum_1 Y^2 = 390$

$\sum_1 x^2 = 2.10 \quad \bar{X} = 5.3 \quad \sum y^2 = 69.67 \quad \bar{Y} = 5.17$

$\sum_1 XY = 339.1 \quad \sum_1 xy = 10.5$

$\beta_0 = \bar{Y} - \beta_1 \bar{X} \qquad \beta_1 = \frac{S_{xy}}{S_{xx}}$

$S_{xx} = \sum_1 X_i^2 - \frac{(\sum X_i)^2}{n} = 339.18 - \frac{63.6^2}{12}$

$\boxed{S_{xx} = 2.1}$

$S_{xy} = \sum_1 (X_i \cdot Y_i) - \frac{\sum X_i \cdot \sum Y_i}{n} = 339.1 - \frac{63.6 \cdot 62}{12}$

$\boxed{S_{xy} = 10.5}$

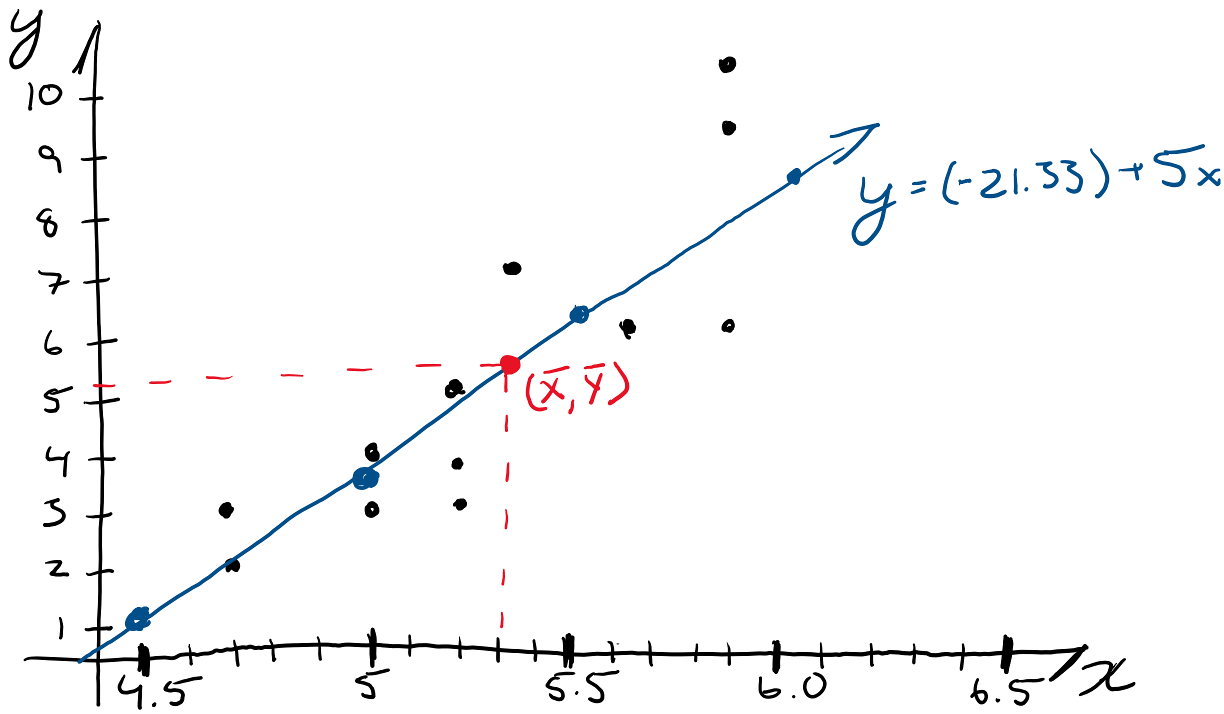
$\Rightarrow \beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{10.5}{2.1} = 5$

$\beta_0 = \bar{Y} - \beta_1 \cdot \bar{X} = 5.17 - (5 \cdot 5.3)$

$\beta_0 = (-21.33)$

$$\boxed{\hat{Y} = (-21.33) + 5X}$$

b) Construct a scatter plot of Y versus X and draw the fitted line on the plot.



c) Find out how much of the variation in Y is explained by the fitted line.

$$S_{xx} = 2.1 \qquad S_{xy} = 10.5$$

$$SS_{reg} = \frac{S_{xy}^2}{S_{xx}} = \frac{(10.5)^2}{2.1} = 52.5$$

$$TSS = S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(y_i)^2}{n} = 390 - \frac{(62)^2}{12} = 69.\overline{6}$$

$$R^2 = \frac{52.5}{69.\overline{6}} = 0.753 \times 100 = 75.3\%$$

$$\boxed{\text{75.3\% of variation is explained by the fitted line.}}$$

4. For the simple linear regression mode $Y = \beta_0 + \beta_1 X + \varepsilon$, show:

a) $SS_{reg} = S^2_{XY} / S_{XX}$

$$SS_{reg} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \qquad S_{xy} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$S_{xx} = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$= \sum (\hat{Y}_i - \bar{Y})^2 \cdot \frac{\sum (X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}$$

$$\boxed{SS_{reg} = \sum \frac{(\hat{Y}_i - \bar{Y})^2 (X_i - \bar{X})^2}{(X_i = \bar{X})^2} = \frac{\sum [(X_i - \bar{X})(Y_i - \bar{Y})]^2}{\sum (X_i - \bar{X})^2} = \frac{S_{xy}^2}{S_{xx}}}$$

b) $R^2 = r^2_{XY}$

$$R^2 = \frac{SS_{reg}}{TSS} = \frac{\frac{S_{xy}^2}{\cancel{S_{xx}}}}{S_{yy}} \cdot \left(\frac{\cancel{S_{xx}}}{S_{xx}} \right)$$

$$\Longrightarrow \boxed{\frac{S_{xy}^2}{S_{yy} S_{xx}} = r_{xy}^2}$$

$$\sqrt{\frac{S_{xy}^2}{S_{yy} S_{xx}}} = \frac{S_{xy}}{\sqrt{S_{yy} S_{xx}}} = r_{xy}$$