

Pstat120C Homewrok 1

Yu Tian

2022-08-09

Reading

The purpose of this portion of the assignment is to guide your reading and help you generate concise reading notes that list the key concepts – generally, terminology, definitions, and theorems. For the submission, treat each bullet point as an exercise and submit your ‘answers’ as you would a problem set.

1. Define deterministic and probabilistic mathematical models. Give an example of each.

Answer:

- Deterministic mathematical model: It suppose that y and x are related according to the equation $y = \beta_0 + \beta_1 x$, where β_0 and β_1 are unknown parameters. This model is called a deterministic mathematical model because it does not allow for any error in predicting y as a function of x . The example is that $y = \beta_0 + \beta_1(5.5)$, y takes the value $\beta_0 + \beta_1(5.5)$ whenever $x = 5.5$.
 - Probabilistic mathematical model: It suppose that y and x are related according to the equation $Y = \beta_0 + \beta_1 x + \epsilon$ (or $E(Y) = \beta_0 + \beta_1 x$), where β_0 and β_1 are unknown parameters and ϵ is a random variable possessing a specified probability distribution with mean 0. This model provides a more accurate description of reality than the deterministic model. The example for $Y = \beta_0 + \beta_1 x + \epsilon$ is that when $x = 5.5$, there is a population of possible values of Y . The distribution is centered on the line $E(Y) = \beta_0 + \beta_1 x$ at the point $x = 5.5$. This population has a distribution with mean $\beta_0 + \beta_1(5.5)$ and variance σ^2 .
2. Write the general equation for a simple linear regression model.

Answer:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

3. Describe, in your own words, the overall concept of the method of least squares.

Answer: The method of least square is finding and fitting a straight line to a set of data points that the sum of squares of the vertical deviations from the fitted line is minimized. Then we can take the partial derivatives of SSE (sum of squares for error) with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and set them equal to zero for estimating the parameters of a line.

4. State the least-squares estimators for the simple linear regression model.

Answer:

- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, where $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ and $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$.
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

5. State the means and variances of the least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ in simple linear regression.

Answer:

$$E(\hat{\beta}_0) = \beta_0$$

$$V(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{nS_{xx}}$$

$$E(\hat{\beta}_1) = \beta_1$$

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

6. State a pair of null and alternative hypotheses for making inferences about single regression parameters and linear functions of the parameters.

Answer:

- single regression parameters: (Hypotheses for β_i)

$$H_0 : \beta_i = \beta_{i0}.$$

$$H_a : \begin{cases} \beta_i > \beta_{i0} & \text{(upper-tail rejection region),} \\ \beta_i < \beta_{i0} & \text{(lower-tail rejection region),} \\ \beta_i \neq \beta_{i0} & \text{(two-tailed rejection region).} \end{cases}$$

- linear function of the parameters: (Hypotheses for $\theta = \alpha_0\beta_0 + \alpha_1\beta_1$)

$$H_0 : \theta = \theta_0$$

$$H_a : \begin{cases} \theta > \theta_0, \\ \theta < \theta_0, \\ \theta \neq \theta_0. \end{cases}$$

Practice

1. (a) Fit the model $Y = \beta_0 + \beta_1x + \epsilon$ to these data, using least squares.

Answer: From the table, we can get:

$$\bar{x} = \frac{10 + 12 + 9 + 27 + 47 + 112 + 36 + 241 + 59 + 167}{10} = 72$$

$$\bar{y} = \frac{9 + 14 + 7 + 29 + 45 + 109 + 40 + 238 + 60 + 170}{10} = 72.1$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$= (10 - 72)(9 - 72.1) + (12 - 72)(14 - 72.1) + (9 - 72)(7 - 72.1) + (27 - 72)(29 - 72.1) \\ + (47 - 72)(45 - 72.1) + (112 - 72)(109 - 72.1) + (36 - 72)(40 - 72.1) + (241 - 72)(238 - 72.1) \\ + (59 - 72)(60 - 72.1) + (167 - 72)(170 - 72.1)$$

$$S_{xy} = 54243$$

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

$$\begin{aligned}
&= (10 - 72)^2 + (12 - 72)^2 + (9 - 72)^2 + (27 - 72)^2 + (47 - 72)^2 + (112 - 72)^2 + (36 - 72)^2 \\
&\quad + (241 - 72)^2 + (59 - 72)^2 + (167 - 72)^2 \\
&S_{xx} = 54714
\end{aligned}$$

Thus,

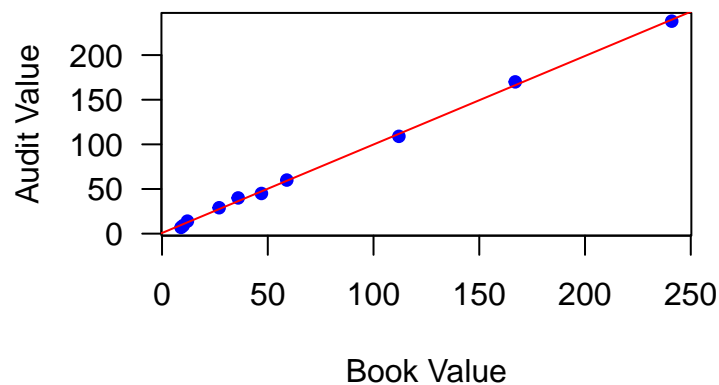
$$\begin{aligned}
\hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} = \frac{54243}{54714} = 0.9914 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = 72.1 - (0.9914)(72) = 0.7192
\end{aligned}$$

Therefore,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 0.7192 + 0.9914x$$

(b) Plot the 10 data points and graph the line representing the model.

Answer:



(c) Calculate SSE and S^2 .

Answer:

$$\begin{aligned}
SSE &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \\
&= (9 - (0.7192 + 0.9914 \cdot 10))^2 + (14 - (0.7192 + 0.9914 \cdot 12))^2 + (7 - (0.7192 + 0.9914 \cdot 9))^2 \\
&\quad + (29 - (0.7192 + 0.9914 \cdot 27))^2 + (45 - (0.7192 + 0.9914 \cdot 47))^2 + (109 - (0.7192 + 0.9914 \cdot 112))^2 \\
&\quad + (40 - (0.7192 + 0.9914 \cdot 36))^2 + (238 - (0.7192 + 0.9914 \cdot 241))^2 + (60 - (0.7192 + 0.9914 \cdot 59))^2 \\
&\quad + (170 - (0.7192 + 0.9914 \cdot 167))^2 \\
&SSE = 56.8454 \\
S^2 &= \left(\frac{1}{n-2}\right) SSE = \frac{56.8454}{10-2} = 7.1057
\end{aligned}$$

(d) Do the data present sufficient evidence to indicate that the slope β_1 differs from zero? Conduct a hypothesis test at the 5% significance level.

Answer: Test of Hypothesis for β_1 :

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

From the question above, we get $\hat{\beta}_1 = 0.9914$, $S_{xx} = 54714$, $s = \sqrt{S^2} = 2.6657$.

Because we interested in the parameter β_1 , we need the value

$$c_{11} = \frac{1}{S_{xx}} = \frac{1}{54714}$$

Then,

$$t = \frac{\hat{\beta}_1 - 0}{s\sqrt{c_{11}}} = \frac{0.9914 - 0}{2.6657 \cdot \sqrt{\frac{1}{54714}}} = 86.99$$

SSE is based on $df = 10 - 2 = 8$. If we take $\alpha = 0.05$, the value of $t_{\alpha/2} = t_{0.025}$ for 2 df is $t_{0.025} = 2.2625$, and the rejection region is

$$\text{reject if } |t| \geq 2.2625.$$

Since 86.99 is larger than 2.2625, we reject the null hypothesis that $\beta_1 = 0$.

Therefore, the data present sufficient evidence to indicate that the slope β_1 differs from zero.

- (e) What is your model's estimate for the expected change in audited value per one-unit change in book value?

Answer: My model's estimate for the expected change in audited value per one-unit change in book value is 0.9914.

- (f) What does your model predict the audited value to be for an item with a book value of \$100?

Answer: My model predict the audited value to be for an item with a book value of 100 is $0.7192 + 0.9914 \cdot 100 = 99.8592$.

2. Let β_0 and β_1 be the least-squares estimates for the intercept and slope in a simple linear regression model. Show that the least-squares equation $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ will always go through the point (\bar{x}, \bar{y}) .

Answer: Since $\beta_0 = \bar{y} - \beta_1 \bar{x}$ and $\beta_1 = \frac{S_{xy}}{S_{xx}}$, so the least squares equation can be

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = \bar{y} - \hat{\beta}_1 \bar{x} + \left(\frac{S_{xy}}{S_{xx}}\right)x = \bar{y} - \left(\frac{S_{xy}}{S_{xx}}\right)\bar{x} + \left(\frac{S_{xy}}{S_{xx}}\right)x = \bar{y} + \left(\frac{S_{xy}}{S_{xx}}\right)(x - \bar{x})$$

Therefore, from this equation above, we can get that the the least-squares equation $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ will always go through the point (\bar{x}, \bar{y}) .

3. Suppose that the model $y = \beta_0 + \beta_1 x + \epsilon$ is fit to the n data points $(y_1, x_1), \dots, (y_n, x_n)$. At what value of x will the length of the prediction interval for y be minimized?

Answer: Since a $100(1 - \alpha)\%$ prediction interval for Y when $x = x^*$ is

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S \left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right)^{\frac{1}{2}}$$

so when $x = \bar{x}$, the length of the prediction interval for y be minimized.