STATISTICAL COMPUTATION

WEEK 13 – LINEAR REGRESSION (1/2)

Annisa Auliya
I Melda Puspita

GET TO KNOW US

ANNISA AULIYA R.

082334174749

I MELDA PUSPITA L.

085257113961

https://intip.in/KomstatC2023

MATERIALS

Simple Linear Regression

- Parameter Estimation
- Significance Test (Simultan & Partial

01

PARAMETER ESTIMATION

ESTIMATION

The **method of least squares** is used to estimate β_0 and β_1. That is, we estimate β_0 and β_1 so that the sum of the squares of the differences between the observations y_i and the straight line is a minimum. From Eq. (2.1) we may write

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{2.3}$$

Equation (2.1) maybe viewed as a **population regression model** while Eq. (2.3) is a **sample regression model**, written in terms of the n pairs of data (y_i, x_i) $(i = 1, 2, \ldots, n)$. Thus, the least-squares criterion is

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2 \tag{2.4}$$

ESTIMATION

The least-squares estimators of β_0 and β_1, say $\hat{\beta}_0$ and $\hat{\beta}_1$, must satisfy

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

and

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i = 0$$

ESTIMATION

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \qquad (2.5)$$

Equations (2.5) are called the **least-squares normal equations**. The solution to the normal equations is

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (2.6)$$

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad (2.7)$$

ESTIMATION

Since the denominator of Eq. (2.7) is the corrected sum of squares of the x_i and the numerator is the corrected sum of cross products of x_i and y_i, we may write these quantities in a more compact notation as

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (2.9)$$

and

$$S_{xy} = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} = \sum_{i=1}^{n} y_i (x_i - \bar{x}) \qquad (2.10)$$

Thus, a convenient way to write Eq. (2.7) is

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad (2.11)$$

02

SIGNIFICANCE TEST

SIGNIFICANCE TEST

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	$SS_R = \hat{\beta}_1 S_{xy}$	1	MS_R	MS_R/MS_{Res}
Residual	$SS_{Res} = SS_T - \hat{\beta}_1 S_{xy}$	$n-2$	MS_{Res}	
Total	SS_T	$n-1$		

$$SS_{Res} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 \equiv SS_T$$

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n-2} = MS_{Res}$$

$$SS_{Res} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy}$$

$$SS_{Res} = SS_T - \hat{\beta}_1 S_{xy}$$

SIGNIFICANCE TEST

The test procedure for H_0: $\beta_1 = 0$ may be developed from two approaches. The first approach simply makes use of the t statistic in Eq. (2.27) with $\beta_{10} = 0$, or

$$t_0 = \frac{\hat{\beta}_1}{\operatorname{se}\left(\hat{\beta}_1\right)}$$

The null hypothesis of significance of regression would be rejected if $|t_0| > t_{\alpha/2, n-2}$.

$$\operatorname{se}\left(\hat{\beta}_1\right) = \sqrt{\frac{MS_{\text{Res}}}{S_{xx}}}$$

SIGNIFICANCE TEST

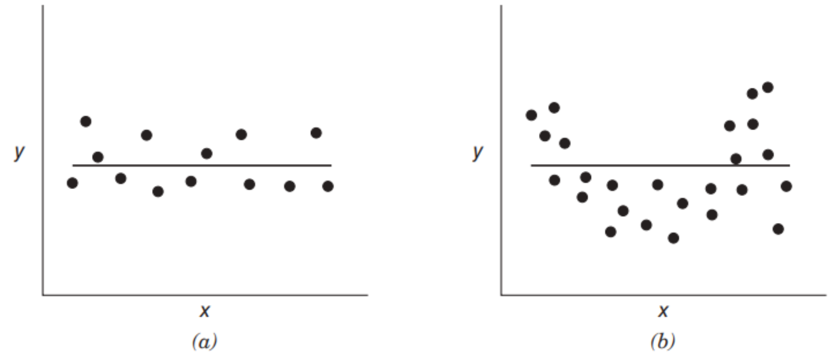

Figure 2.2 Situations where the hypothesis H_0: $\beta_1 = 0$ is not rejected.

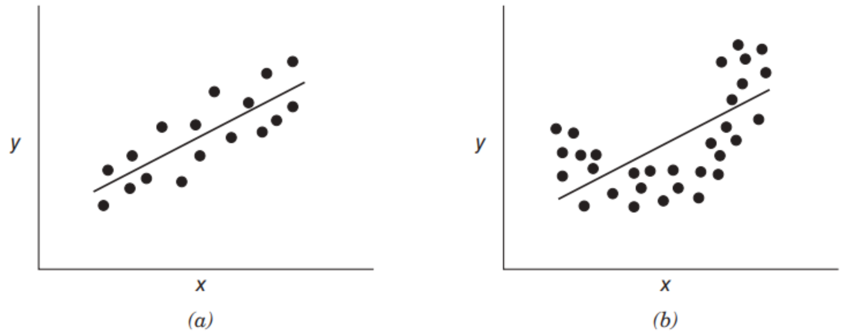

Figure 2.3 Situations where the hypothesis H_0: $\beta_1 = 0$ is rejected.

THANKS

https://intip.in/KomstatC2023