Properties of $b_0$ & $b_1$ ①

452

5-11

Model was $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

Assume $E[\varepsilon_i] = 0 \quad \forall i$

$V[\varepsilon_i] = \sigma^2 \quad \forall i$

---

$$E[b_1] = E\left[\frac{S_{xy}}{S_{xx}}\right] = \frac{1}{S_{xx}} E[S_{xy}]$$

$$= \frac{1}{S_{xx}} E\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]$$

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

$$= \frac{1}{S_{xx}} \sum_{i=1}^{n}(x_i - \bar{x}) E(\beta_0 + \beta_1 x_i + \varepsilon_i - \beta_0 - \beta_1 \bar{x} - \bar{\varepsilon})$$

②

$$= \frac{1}{S_{xx}} \sum_{i=1}^{n}(x_i - \bar{x})\left[\beta_1 (x_i - \bar{x}) + 0\right]$$

$$= \frac{1}{S_{xx}} \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})^2 = \beta_1$$

$\therefore$ $b_1$ is an unbiased estimator of $\beta_1$

---

$$E(b_0) = E(\bar{y} - b_1 \bar{x})$$

$$= E(\beta_0 + \beta_1 \bar{x} + \bar{\varepsilon} - b_1 \bar{x})$$

$$= \beta_0 + \beta_1 \bar{x} + 0 - \beta_1 \bar{x}$$

$$= \beta_0$$

$\therefore$ $b_0$ is an unbiased estimator of $\beta_0$

③

$$V(b_1) = V\left[\frac{S_{xy}}{S_{xx}}\right] = \frac{1}{S_{xx}^2} V[S_{xy}]$$

$$= \frac{1}{S_{xx}^2} V\left[\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right]$$

$$= \frac{1}{S_{xx}^2} V\left[\sum_{i=1}^{n} (x_i - \bar{x}) y_i - \underbrace{\sum_{i=1}^{n} (x_i - \bar{x})}_{\sum x_i - n\bar{x} = 0} \bar{y}\right]$$

$$= \frac{1}{S_{xx}^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 V(y_i) + \underbrace{\text{covariance terms}}_{}$$

These will be 0 if we assume $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i, j$

④

$$= \frac{1}{S_{xx}^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 V(\beta_0 + \beta_1 x_i + \varepsilon_i)$$

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

$$= \frac{\sigma^2}{(n-1) S_x^2}$$

Note that this $\to 0$ as $n \to \infty$

---

If we also assume that the $\varepsilon_i$'s are normally distributed, then

$$\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \sim N(0,1)$$

5

Also, $\sigma^2$ can be estimated by MSE.

$s = \sqrt{MSE}$ is called the "standard error of the estimate"

$$\frac{b_1 - \beta_1}{s/\sqrt{S_{xx}}} \sim t_{n-2}$$

A confidence interval for $\beta_1$:

$$b_1 \pm t_{\alpha/2}\frac{s}{\sqrt{S_{xx}}}$$

6

Let $\bar{Y}_K$ be the average y-coordinate of k new observations, all with the same x-coordinate, $x_0$

The predicted y-value at $x = x_0$ will be $\hat{y} = b_0 + b_1 x_0$.

So the predicted value of $\bar{Y}_K$ is also $b_0 + b_1 x_0$

Fact 1: $E[\hat{y} - \bar{Y}_K] = 0$

Fact 2: $V[\hat{y} - \bar{Y}_K] = \sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} + \frac{1}{k}\right]$

(7)

Result:

$$\frac{\hat{y}-\bar{Y}_k}{s\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}+\frac{1}{k}}} \sim t_{n-2}$$

(1) Prediction interval for $\bar{Y}_k$:

$$\hat{y} \pm t_{\alpha/2}\, s\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}+\frac{1}{k}}$$

Special case: $k=1$, i.e. a prediction interval for a single new value

$$\hat{y} \pm t_{\alpha/2}\, s\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}$$

(8)

(2) Confidence interval for the mean y-coordinate of all observations whose x-coordinate is $x_0$:

$$\hat{y} \pm t_{\alpha/2}\, s\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}$$

(3) Confidence interval for $\beta_0$:

$$b_0 \pm t_{\alpha/2}\, s\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}}$$

All of $t_{\alpha/2}$ values have $n-2$ df