STAT 423

WIN 2016

Properties of OLS estimators, Part II

Objective: Infer properties of a {process / population} from data from {an experiment / a survey}

Post-experiment: Data are fixed, known
$\hat{\beta}_0, \hat{\beta}_1$ are fixed, known

Pre-experiment: Data are unknown, random
$\hat{\beta}_0, \hat{\beta}_1$ are unknown, random

Pre-experimental Question: Is the probability dist. of $(\hat{\beta}_0, \hat{\beta}_1)$ centered around a meaningful representation of the process or population?

Probability Model of an experiment

Experiment: An experimental outcome y_i is generated for each experimental condition x_i, $i = 1, \ldots, n$.

Linear Model: $E[Y_i | X_i = x_i] = \beta_0 + \beta_1 x_i$ for some β_0, β_1

equivalently, $\left\{ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \;,\; E[\epsilon_i] = 0 \right\}$ (A1)

Return to the question: Are the values of $(\hat{\beta}_0, \hat{\beta}_1)$ we will get from the data likely to be representative of the population or process?

Result 1: If (A1) holds, then $E[\hat{\beta}_0] = \beta_0$, $E[\hat{\beta}_1] = \beta_1$ ($(\hat{\beta}_0, \hat{\beta}_1)$ are unbiased for (β_0, β_1).)

Pf: $E[\hat{\beta}_1 | x] = E\left[\frac{S_{xy}}{S_{xx}} | x\right] = \frac{1}{S_{xx}} E\left[\sum (x_i - \bar{x})(y_i - \bar{y}) | \underset{\sim}{x}\right]$

$= \frac{1}{S_{xx}} \sum (x_i - \bar{x}) E[(y_i - \bar{y}) | \underset{\sim}{x}]$

$= \frac{1}{S_{xx}} \sum (x_i - \bar{x}) \beta_1 (x_i - \bar{x}) = \beta_1 \frac{S_{xx}}{S_{xx}} = \beta_1 .$

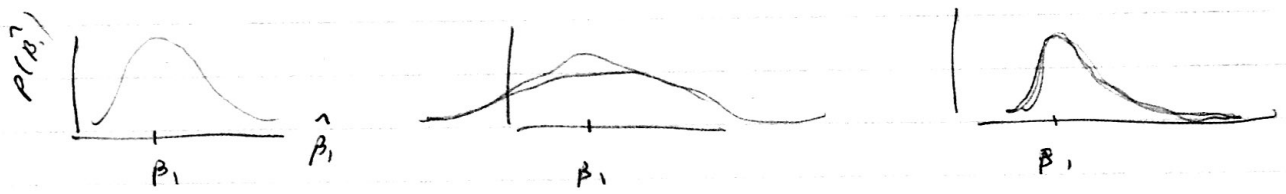
β_0: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$E[\hat{\beta}_0 | \underset{\sim}{x}] = E[\bar{y} | \underset{\sim}{x}] - E[\hat{\beta}_1 | \underset{\sim}{x}]\,\bar{x}$$

$$= \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0 \quad \text{(for all } x\text{)}.$$

Result 1 does not require { independence, normality, constant variance }

but it does not tell us much:



Result #1 doesn't say <u>how concentrated</u> the dist. of $\hat{\beta}_1$ is around β_1

<u>Concentration</u> of $(\hat{\beta}_0, \hat{\beta}_1)$ around (β_0, β_1) can be measured with variance and <u>covariance</u>. But to say something about variance, we need to make additional assumptions:

(A1) $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $E[\epsilon_i | \underset{\sim}{x}] = 0$

(A2) $Var(\epsilon_i | \underset{\sim}{x}) = \sigma^2$ for all i (constant variance)
$Cov(\epsilon_i, \epsilon_j | \underset{\sim}{x}) = 0$ for all i, j (uncorrelated errors)

Result 2: If (A1) + (A2) hold, then

$$E[(\hat{\beta}_0, \hat{\beta}_1)] = (\beta_0, \beta_1) \quad \underline{\text{and}} \quad Var(\hat{\beta}_1 | \underset{\sim}{x}) = \sigma^2 \times \frac{1}{S_{xx}}$$

$$Var(\hat{\beta}_0 | \underset{\sim}{x}) = \sigma^2 \left(\frac{1}{n} + \bar{x}^2 \frac{1}{S_{xx}}\right)$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1 | \underset{\sim}{x}) = -\sigma^2 \left(\frac{\bar{x}}{S_{xx}}\right)$$

} will prove when we do multiple linear reg.

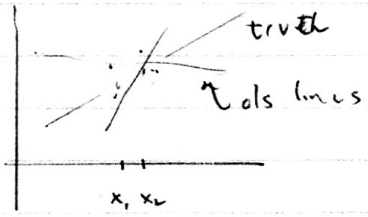
Interpretation:

$$\mathrm{Var}(\hat{\beta}_1 | \underset{\sim}{x}) = \sigma^2 / S_{xx}$$

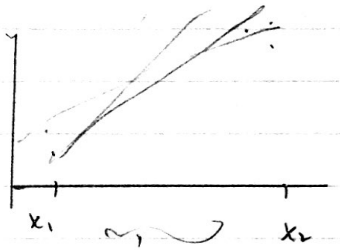
$$= \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2} = \frac{\sigma^2}{n} \frac{1}{s_x^2}$$

- Depends on $\underset{\sim}{x}$
- Small if { experimental error σ^2 is small; n is big; s_x^2 is big } $\Rightarrow$ distribution of $\hat{\beta}_1$ is highly concentrated around β_1.

Picture:



small variance in x_i's

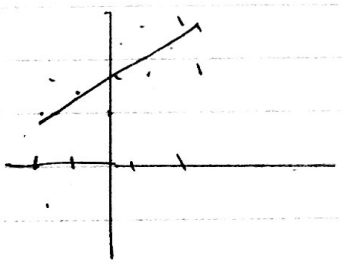


big variance in x_i's

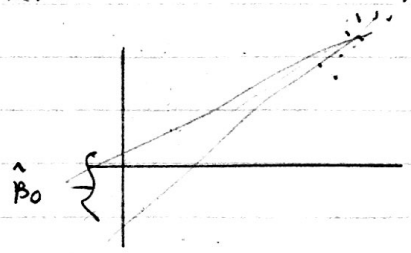
$$\mathrm{Var}(\hat{\beta}_0 | \underset{\sim}{x}) = \sigma^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

$$= \frac{\sigma^2}{n} \left(1 + \frac{\bar{x}^2}{s_x^2} \right) = \frac{\sigma^2}{n} \frac{1}{s_x^2} \left(\Sigma x_i^2 \right)$$

Same interpretation as for $\hat{\beta}_1$, except $\mathrm{Var}(\hat{\beta}_0)$ is larger x_i's are not centered around zero.



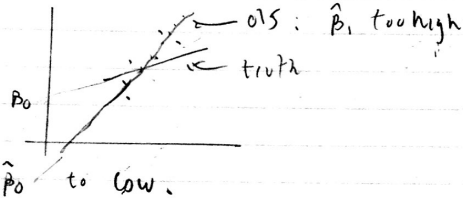
versus



$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1 | \underset{\sim}{X}) = -\sigma^2 \left(\frac{\bar{x}}{S_{xx}}\right)$. Why negative?

Interpretation: if $\hat{\beta}_1$ is higher than β_1, expect $\hat{\beta}_0$ lower than β_0.

Picture:



Math:

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1 | \underset{\sim}{X}) = \text{Cov}(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1 | \underset{\sim}{X})$$

$$= \text{Cov}(\bar{y}, \hat{\beta}_1 | \underset{\sim}{X}) - \text{Cov}(\hat{\beta}_1 \bar{x}, \hat{\beta}_1 | \underset{\sim}{X})$$

$$= 0 - \bar{x}\,\text{Cov}(\hat{\beta}_1, \hat{\beta}_1 | \underset{\sim}{X})$$

$$= -\bar{x}\,\text{Var}(\hat{\beta}_1 | \underset{\sim}{X})$$

$$= -\bar{x}\,\frac{\sigma^2}{S_{xx}}$$

Exercise: Suppose you want to design an experiment to estimate β_0, β_1. How should you choose your x_i's?

Estimating the variance

Q: Is $\hat{\beta}_1$ a precise est. of β_1?

A: $V(\hat{\beta}_1 | \underset{\sim}{X}) = \frac{\sigma^2}{n} \frac{1}{S_x^2}$

Q: S_x^2 is known, n is known, but σ^2 isn't. So, is $\hat{\beta}_1$ a precise est of β_1?

A: Need to know σ^2.

(5)

Idea: Under (A1) + (A2),

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ $\quad Var(\epsilon_i | \underset{\sim}{X}) = \sigma^2$, $Cov(\epsilon_i, \epsilon_j | \underset{\sim}{X}) = 0$

$\epsilon_i = y_i - [\beta_0 + \beta_1 x_i]$

Sample var $\epsilon_i \overset{?}{\approx} \sigma^2$

But dont have ϵ_i, as we don't know β_0, β_1

Idea: $\hat{\beta}_0 \approx \beta_0$, $\hat{\beta}_1 \approx \beta_0$, so $\epsilon_i \approx y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i] = \hat{\epsilon}_i$ = residual

$Var(\hat{\epsilon}_1, \ldots \hat{\epsilon}_n) \approx Var(\epsilon_1 \ldots \epsilon_n) \approx \sigma^2$

More Precisely: Let $RSS = \sum (y_i - [\hat{\beta}_0 + \hat{\beta}_1 x_i])^2 = \sum \hat{\epsilon}_i^2$

then $E[RSS] = \sigma^2 \times (n-2)$

$E[RSS/n-2] = \sigma^2$

so $\hat{\sigma}^2 = \frac{RSS}{n-2}$ is an unbiased estimate of σ^2.

Q: Is $\hat{\beta}_1$ a precise est of β_1?

A: $V(\hat{\beta}_1 | \underset{\sim}{X}) = \frac{\sigma^2}{n} \frac{1}{s_x^2}$ $\quad SD(\hat{\beta}_1 | \underset{\sim}{X}) = \frac{\sigma}{\sqrt{n}} \frac{1}{s_x}$

$\approx \frac{\hat{\sigma}^2}{n} \frac{1}{s_x^2}$ $\quad \approx \frac{\hat{\sigma}}{\sqrt{n}} \frac{1}{s_x} = SE(\hat{\beta}_1 | \underset{\sim}{X})$

The estimated standard deviation is the standard error

Similarly, $SE(\hat{\beta}_0 | \underset{\sim}{X}) = \sqrt{\hat{\sigma}^2 \left(1 + \frac{\bar{x}^2}{s_{xx}}\right)} = \hat{\sigma} \sqrt{1 + \frac{\bar{x}^2}{s_{xx}}}$

$\hat{\sigma}^2$, $SE(\hat{\beta}_0 | \underset{\sim}{X})$, $SE(\hat{\beta}_1 | \underset{\sim}{X})$ will help us with { confidence intervals, hypothesis tests, prediction intervals.

Normal Linear Model

Confidence intervals, hypothesis tests, Probability statements } require we know more about the distribution of $(\hat{\beta}_0, \hat{\beta}_1)$ beyond expectations, variances.

Result 3: If $\epsilon_1 \dots \epsilon_n$ are normally distributed, then $\hat{\beta}_0, \hat{\beta}_1$ " also.

Result 3b: Even if $\epsilon_1 \dots \epsilon_n$ are not normally distributed, $\hat{\beta}_0, \hat{\beta}_1$ are approximately normally distributed for large n.

More specifically:

(A1) $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $E[\epsilon_i] = 0$

(A2) $\mathrm{Var}(\epsilon_i | x) = \sigma^2$, $\mathrm{Cov}(\epsilon_i, \epsilon_j | x) = 0$ all i, j

(A3) ϵ_i is normally distributed

(A23) $\epsilon_1 \dots \epsilon_n \sim$ IID $N(0, \sigma^2)$

Result 3: If A1 + A2 + A3 hold, then (given in result 2.)

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \mathrm{Cov}\left(\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \right) \right)$$

This says: $E\left[\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} | x \right] = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ ✓ already showed this

$\mathrm{Cov}\left(\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} | x \right) =$ prev ✓ "

$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$ are jointly normally distributed. * Need to show this.

Normality of $\hat{\beta}_1$

Recall: (1) $Z \sim N(0,1) \Rightarrow a + bZ \sim N(a, b^2)$

(2) $W_1 \ldots W_n$ indep. normal, $W_i \sim N(\mu_i, V_i)$

then $\underset{\sim}{a} \cdot \underset{\sim}{W} = \sum a_i W_i \sim N\left(\sum a_i \mu_i, \sum a_i^2 V_i\right)$

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{1}{S_{XX}} \cdot (\underset{\sim}{x} - \bar{x}\underset{\sim}{1}) \cdot (\underset{\sim}{Y} - \bar{Y}\underset{\sim}{1})$$

$$(\underset{\sim}{Y} - \bar{Y}\underset{\sim}{1}) = (I - 1\underset{\sim}{1}^T/n)\underset{\sim}{Y} = C\underset{\sim}{Y}$$

$$\Rightarrow \quad (\underset{\sim}{x} - \bar{x}\underset{\sim}{1}) \cdot (\underset{\sim}{Y} - \bar{Y}\underset{\sim}{1}) = (\underset{\sim}{x} - \bar{x}\underset{\sim}{1}) C \underset{\sim}{Y}$$

$$\Rightarrow \quad \hat{\beta}_1 = \frac{1}{S_{XX}} (\underset{\sim}{x} - \bar{x}\underset{\sim}{1})^T C \underset{\sim}{Y}$$

$$= \underset{\sim}{a} \cdot \underset{\sim}{Y}$$

Now by (1), $Y_i = (\beta_0 + \beta_1 X_i) + \epsilon_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$

(2) $Y_1 \ldots Y_n$ indep. normal, so $\underset{\sim}{a} \cdot \underset{\sim}{Y}$ is normal

(already computed the mean and variance)