Stat 563
5-14-19
①

To do inferences in regression,
we need the following:

Let $\hat{\varepsilon}_i = y_i - \hat{y}_i$ (the i^{th} residual)

Note: $\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$

Find $Cov(\hat{\beta}_1, \hat{\varepsilon}_j) = Cov\left(\sum_{i=1}^{n} c_i Y_i, \sum_{i=1}^{n} d_i Y_i\right)$

where $c_i = \frac{1}{S_{xx}}(x_i - \bar{x})$

but what are the d_i's ?

$= \sum_{i=1}^{n} c_i d_i \underbrace{V[y_i]}_{\sigma^2}$

$= \sigma^2 \sum_{i=1}^{n} c_i d_i$

②

$$\hat{\varepsilon}_j = Y_j - \hat{Y}_j = Y_j - (\hat{\beta}_0 + \hat{\beta}_1 X_j)$$

$$= Y_j - (\bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_j)$$

$$\hat{\varepsilon}_j = Y_j - \bar{Y} - \hat{\beta}_1 (X_j - \bar{X})$$

$$= Y_j - \frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{(X_j - \bar{X})}{S_{xx}} \underbrace{\sum_{i=1}^{n}(X_i - \bar{X}) Y_i}_{S_{xy}}$$

$$= \sum_{i=1}^{n} \left(\delta_{ij} - \frac{1}{n} - \frac{(X_j - \bar{X})(X_i - \bar{X})}{S_{xx}}\right) Y_i$$

where $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

So $d_i = \delta_{ij} - \frac{1}{n} - \frac{(x_j-\bar{x})(x_i-\bar{x})}{S_{xx}}$ ③

Now, find $\sum_{i=1}^{n} c_i d_i$

$$= \sum_{i=1}^{n} \frac{1}{S_{xx}}(x_i-\bar{x})\left[\delta_{ij} - \frac{1}{n} - \frac{(x_j-\bar{x})(x_i-\bar{x})}{S_{xx}}\right]$$

$$= \sum_{\substack{i=1 \\ i\neq j}}^{n} \frac{1}{S_{xx}}(x_i-\bar{x})\left[-\frac{1}{n} - \frac{(x_j-\bar{x})(x_i-\bar{x})}{S_{xx}}\right]$$

$$+ \frac{1}{S_{xx}}(x_j-\bar{x})\left[1 - \frac{1}{n} - \frac{(x_j-\bar{x})^2}{S_{xx}}\right]$$

④

$$\cancel{-\frac{1}{nS_{xx}} \underbrace{\sum_{\substack{i=1 \\ i\neq j}}^{n}(x_i-\bar{x})}_{①}} - \cancel{\frac{1}{S_{xx}^2}(x_j-\bar{x}) \underbrace{\sum_{\substack{i=1 \\ i\neq j}}^{n}(x_i-\bar{x})^2}_{②}}$$

$$+ \cancel{\frac{(x_j-\bar{x})}{S_{xx}}} - \cancel{\frac{1}{nS_{xx}}(x_j-\bar{x})} - \cancel{\frac{1}{S_{xx}^2}(x_j-\bar{x})^3} = 0$$

①: $0-(x_j-\bar{x})$

②: $S_{xx}-(x_j-\bar{x})^2$

$\therefore \operatorname{Cov}(\hat{\beta}_1, \hat{\varepsilon}_j) = 0$

$\Rightarrow \hat{\beta}_1$ and $\hat{\varepsilon}_j$ are independent, since they are jointly normal

⑤

$\Rightarrow \hat{\beta}_1$ is indep of $\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = SSE$

$\Rightarrow \hat{\beta}_1$ is indep of $s^2 = \frac{SSE}{n-2}$

Also, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} \Rightarrow \hat{\beta}_0$ is indep. of s^2

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right) \Rightarrow \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}} \sim N(0,1)$$

⑥

Similar to the derivation in Fall,

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2_{n-2}$$

So
$$\frac{\dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}}}{\sqrt{\dfrac{\cancel{(n-2)} s^2}{\cancel{\sigma^2}} \Big/ \cancel{(n-2)}}} \sim t_{n-2}$$

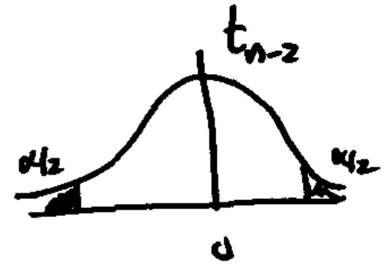
$$\Rightarrow \frac{\hat{\beta}_1 - \beta_1}{\sqrt{s^2 / S_{xx}}} \sim t_{n-2}$$

⑦

This allows us to do hypothesis tests &
Confidence intervals for the slope.

Example: $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

Test stat is $\dfrac{\hat{\beta}_1 - 0}{\sqrt{s^2 / S_{xx}}}$

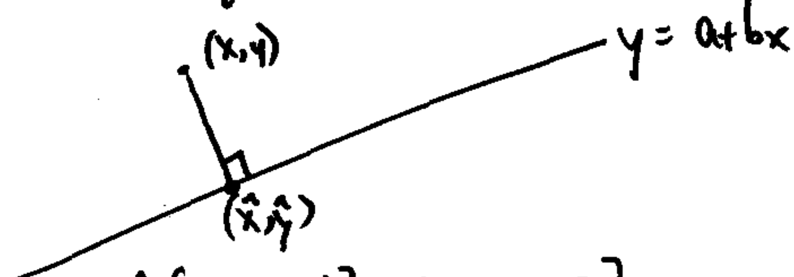


Reject H_0 if the test stat
lies in either tail

⑧

Similarly, $$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t_{n-2}$$

Orthogonal least squares



Minimize $\sum_{i=1}^{n}\left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2\right]$

⑨

Given a point (x,y) not on the line $y = a+bx$,
Find the point $(\hat{x},\hat{y})$ on the line, closest to (x,y)

Let D = squared distance from (x,y) to $(\hat{x},\hat{y})$

$$= (x-\hat{x})^2 + (y-\hat{y})^2$$

$$= (x-\hat{x})^2 + (y-(a+b\hat{x}))^2$$

$$\frac{dD}{d\hat{x}} = \cancel{2}(x-\hat{x})(\cancel{-}1) + \cancel{2}(y-a-b\hat{x})(\cancel{-}b) \stackrel{\text{set}}{=} 0$$

$$x - \hat{x} + by - ab - b^2\hat{x} = 0$$

$$\hat{x} = \frac{x+by-ab}{(1+b^2)}$$

⑩

So $\hat{x} = \frac{x+by-ab}{(1+b^2)}$ and

$$\hat{y} = a + b\left(\frac{x+by-ab}{1+b^2}\right)$$

Our target function that we need to minimize is

$$\sum_{i=1}^{n}\left[\left(x_i - \frac{x_i+by_i-ab}{1+b^2}\right)^2 + \left(y_i - a - b\left(\frac{x_i+by_i-ab}{1+b^2}\right)\right)^2\right]$$

$$= \frac{1}{1+b^2}\sum_{i=1}^{n}(y_i - a - bx_i)^2$$

Fix b. Then $\hat{a} = \bar{y} - b\bar{x}$ (11)

Substitute:

$$\frac{1}{1+b^2}\sum_{i=1}^{n}\left(y_i - (\bar{y} - b\bar{x}) - bx_i\right)^2$$

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

$$= \frac{1}{1+b^2}\left[\sum(y_i-\bar{y})^2 + b^2\sum(x_i-\bar{x})^2 - 2b\sum(x_i-\bar{x})(y_i-\bar{y})\right]$$

$$= \frac{1}{1+b^2}\left[S_{yy} + b^2 S_{xx} - 2b S_{xy}\right]$$

(12)

Take the derivative w.r.t. b:

$$\frac{(1+b^2)\left[2bS_{xx} - 2S_{xy}\right] - \left[S_{yy} + b^2 S_{xx} - 2bS_{xy}\right](2b)}{(1+b^2)^2} \overset{\text{set}}{=} 0$$

$$2bS_{xx} - 2S_{xy} + \cancel{2b^3S_{xx}} - \underline{2b^2S_{xy}}$$

$$- 2bS_{yy} - \cancel{2b^3S_{xx}} + \underline{4b^2S_{xy}} = 0$$

$$b^2 S_{xy} + b(S_{xx} - S_{yy}) - S_{xy} = 0$$

(13)

$$\hat{b} = \frac{S_{yy} - S_{xx} \pm \sqrt{(S_{xx} - S_{yy})^2 + 4 S_{xy}^2}}{2 S_{xy}}$$

and $\hat{a} = \bar{y} - \hat{b}\,\bar{x}$

Next up: measurement error model