Lec 4/19

Wednesday, April 19, 2017 14:58

$$T = \frac{\hat{B} - \beta}{\hat{\Sigma}} \sqrt{\frac{(n-2)\, S_{xx}}{n}} \sim t_{n-2}$$

CI for β: $\left(\hat{\beta} \pm t_{\frac{\alpha}{2}, n-2}\, \hat{\sigma} \sqrt{\frac{n}{(n-2) S_{xx}}} \right)$

Ex: construct a 95% CI for β given data from last time.

Sol: $3.471 \pm 2.306 \cdot 4.720 \cdot \sqrt{\frac{10}{8 \cdot 376}} = (2.84, 4.10)$

§14.5 Normal Correlation Analysis (Google: Gaussian Copula)

Last time we required $w(y|x)$ is normal (and x is observed already)

Now we drop assumption x is fixed. $(X, Y) \sim$ Bivariate normal dist:

$$f(x,y) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)} \left(\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1} \right)\left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right) \right]$$

(Note: $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$)

Given paired data set $\{(x_i, y_i) : i = 1, \ldots, n\}$,

$$L = \prod_{i=1}^{n} f(x_i, y_i) = \left(\frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \right)^n \exp\left\{ \frac{-1}{2(1-\rho^2)} \left[\sum_{i=1}^{n} \left(\frac{x_i - \mu_1}{\sigma_1} \right)^2 - 2\rho \sum_{i=1}^{n} \left(\frac{x_i - \mu_1}{\sigma_1} \right)\left(\frac{y_i - \mu_2}{\sigma_2} \right) + \sum_{i=1}^{n} \left(\frac{y_i - \mu_2}{\sigma_2} \right)^2 \right] \right\}$$

$\Downarrow$

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = \bar{y}, \quad \hat{\sigma}_1 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \hat{\sigma}_2 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2}, \quad \hat{\rho} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_1 \hat{\sigma}_2} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$$

also called sample correlation coefficient, usually denoted r).

Remarks:

① ρ measures strength of the "linear" relationship between X and Y.

We often are interested in tests concerning ρ

i.e. $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$.

② When $X = x$ is given, the conditional variance of Y has the following formula:

$$\sigma^2_{Y|x} = \sigma_2^2 (1-\rho^2) \qquad \text{(Thm 6.9)}$$

thus, $\rho = \pm 1 \Rightarrow \sigma^2_{Y|x} = 0 \Rightarrow Y = \alpha + \beta X$ for some $\alpha, \beta \in \mathbb{R}$. → Y completely determined by X.

③ $\sigma^2_{Y|x}$ is a function of σ_2 and ρ

thus by invariance prop. of MLE,

$$\hat{\sigma}^2_{Y|x} = \hat{\sigma}_2^2 (1-\hat{\rho}^2)$$

$$\Rightarrow r^2 = \hat{\rho}^2 = \frac{\hat{\sigma}_2^2 - \hat{\sigma}^2_{Y|x}}{\hat{\sigma}_2^2}$$

← cond. variance of Ys for fixed values of x.

↑ measures total variance of Ys

thus, $100 r^2$ is percentage of total variation of the Ys that is accounted for by the relationship with x.

④ let R be the RV whose value is r

One can show that $\frac{1}{2} \log \left(\frac{1+R}{1-R} \right) \sim N\left(\frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right), \frac{1}{n-3} \right)$

$$\Rightarrow \quad Z = \frac{\sqrt{n-3}}{2} \left[\log \left(\frac{1+R}{1-R} \right) - \log \left(\frac{1-\rho}{1-\rho} \right) \right] \sim N(0,1)$$

Which can be used as a test stat for hyp. test or CI for ρ.