Lec 4/21

Friday, April 21, 2017 14:57

$$\frac{1}{2}\log\left(\frac{1+R}{1-R}\right) \overset{\text{Approx}}{\sim} N\left(\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n-3}\right)$$

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

<u>Ex:</u>

Morning x	afternoon y
$\vdots$	$\vdots$
$\vdots$	$\vdots$

Test whether there is a relationship.

$H_0: \rho = 0, \quad H_1: \rho \neq 0$

<u>Sol:</u> $S_{xx} = \sum x_i^2 - \frac{1}{n}(\sum x_i)^2 = 19.661, \quad S_{yy} = 30.796, \quad S_{xy} = 23.024.$

$$r = \hat{\rho} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = 0.936.$$

Under H_0, $Z = \left[\frac{1}{2}\log\left(\frac{1+r}{1-r}\right) - \frac{1}{2}\log(1)\right]\sqrt{7} = \frac{\sqrt{7}}{2}\log\left(\frac{1+r}{1-r}\right) \sim N(0,1)$

actual value of Z is $\frac{1}{2}\log\left(\frac{1+0.936}{1-0.936}\right) = 4.5$

CR is $|Z| \geq Z_{\frac{\alpha}{2}}$. $\underset{2.576}{\underset{\shortparallel}{Z_{\frac{0.01}{2}}}} < 4.5$ so we reject H_0. H_0

Conclude there is a linear relationship between the two things.

<u>Sec</u> 14.6, 14.7 Multiple linear regression & Matrix notation.

<u>Recall:</u> $\mu_{Y|x} = \mathbb{E}(Y|x)$

$\Downarrow$ k-dim case, $\vec{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$

$$\mu_{Y|\vec{x}} = \mathbb{E}(Y|x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

As in §14.3, coefficients $\beta_0, \dots, \beta_k$ are usually estimated by LSE. given n data points $\{(x_{i1}, \dots, x_{ik}, y_i) : i = 1, \dots, n\}$, then the least-squares estimates will be:

$$(\hat{\beta}_0^*, \hat{\beta}_1^*, \dots, \hat{\beta}_k^*) = \underset{(\hat{\beta}_0, \dots, \hat{\beta}_k)}{\arg\min} \overbrace{\sum_{i=1}^{n} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})\right)^2}^{q}$$

to solve minimization,

$$\frac{\partial q}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} (-2) \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})\right) = 0$$

$$\frac{\partial q}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} (-2 x_{i1}) \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})\right) = 0$$

$$\vdots$$

$$\frac{\partial q}{\partial \hat{\beta}_k} = \sum_{i=1}^{n} (-2 x_{ik}) \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})\right) = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n} y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \dots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}$$

$$\sum y_i x_{i1} = \hat{\beta}_0 \sum x_{i1} + \hat{\beta}_1 \sum x_{i1}^2 + \dots + \hat{\beta}_k \sum_{i=1}^{k} x_{i1} x_{ik}$$

System of linear equations.

$$\vdots$$

$$\sum y_i x_{ik} = \hat{\beta}_0 \sum x_{ik} + \hat{\beta}_1 \sum x_{i1} x_{ik} + \dots + \hat{\beta}_k \sum x_{ik}^2$$

$$\overbrace{\begin{pmatrix} \sum y_0 \\ \sum y_i x_{i1} \\ \vdots \\ \sum y_i x_{ik} \end{pmatrix}}^{I_1} = \overbrace{\begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1} x_{ik} \\ \vdots & & & \vdots \\ \sum x_{ik} & \sum x_{ik} x_{i1} & \dots & \sum x_{ik}^2 \end{pmatrix}}^{I_2} \cdot \overbrace{\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}}^{\mathbb{B}}$$

$$B = I_2^{-1} I_1$$

Maybe on exam.

$$I_1 = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \vdots \\ \sum y_i x_{ik} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \vdots & & & \vdots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{pmatrix}} \cdot \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbb{Y}}$$

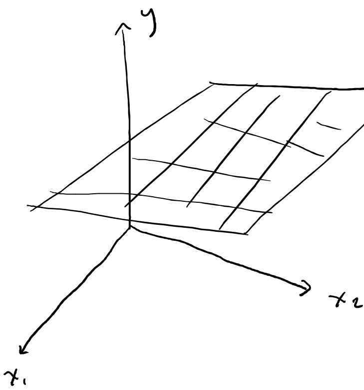
$$I_1 = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \vdots \\ \sum y_i x_{ik} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \vdots & & & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{pmatrix}}_{\mathbb{X}^T} \cdot \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbb{Y}}$$

$$I_2 = \mathbb{X}^T \mathbb{X}$$

Thus we have the LSE are given by $\mathbb{B} = \begin{pmatrix} \hat{\beta}_0^* \\ \vdots \\ \hat{\beta}_k^* \end{pmatrix} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbb{Y}$

Ex

$$\mathbb{Y} = \begin{pmatrix} 212{,}000 \\ \vdots \\ 307{,}500 \end{pmatrix} \quad \mathbb{X} = \begin{pmatrix} 1 & 3 & 2 \\ 1 & 2 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 4 & 2 \end{pmatrix} \quad \mathbb{B} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbb{Y} = \begin{pmatrix} 224{,}929 \\ 15{,}314 \\ 10{,}957 \end{pmatrix}$$



Positive correlation

Normal multiple regression