Lec 4/14

Friday, April 14, 2017 15:00

Exam problem: profit/ overbooking.

$$e_i = y_i - (\alpha + \beta x_i)$$

Least squares estimates $\hat{\alpha}^*$ and $\hat{\beta}^*$ minimize $\sum_{i=1}^{n} e_i^2$

Why e_i^2? it is easier that $|e_i|$.

Formula: Theorem 14.2

$$\hat{\alpha}^* = \bar{y} - \hat{\beta}^* \bar{x}, \quad \hat{\beta}^* = \frac{S_{xy}}{S_{xx}}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

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

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

Proof: let $q = \sum_{i=1}^{n} (y_i - (\hat{\alpha} + \hat{\beta} x_i))^2$

$$\frac{\partial q}{\partial \hat{\alpha}} = \sum_{i=1}^{n} (-2)(y_i - (\hat{\alpha} + \hat{\beta} x_i)) = 0$$

$$\frac{\partial q}{\partial \hat{\beta}} = \sum_{i=1}^{n} (-2 x_i)(y_i - (\hat{\alpha} + \hat{\beta} x_i)) = 0$$

solving system for $\hat{\alpha}, \hat{\beta}$, we get desired result.

Ex: hours studied vs testscore:

$\bar{x} = 10 \quad \bar{y} = 56.4$

$S_{xx} = 376, \quad S_{xy} = 1305$

$\hat{\beta}^* = 3.471 \qquad \hat{\alpha}^* = 21.69$

So Least-squares line is $y = 21.69 + 3.471x$

if you study 14 hours you should get $21.69 + 3.471 \cdot 14$
$= 70.284$

Remark: x could be multidimensional. i.e. linear function of several variables.

§14.4: Normal Regression Analysis.

recall that $\mu_{Y|x} = \mathbb{E}(Y|x) = \int_{-\infty}^{\infty} y\, w(y|x)\, dy$

In particular, we assume

$$w(y|x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma^2}\left(y - (\alpha + \beta x)\right)^2\right]$$

(normal density with $\mu = \alpha + \beta x$ and variance σ^2

Thus given $\{(x_i, y_i) : i = 1, \dots, n\}$ we'd like to

1. the estimation of σ, α, β.
2. Test of hypotheses concerning these
3. predictions based on estimate $\mu_{Y|x} = \hat{\alpha} + \hat{\beta}x$

1. MLE of α, β, σ

① MLE of α, β, σ

Likelihood $\prod_{i=1}^{n} \underbrace{w(y_i|x_i) \cdot m(x_i)}_{f(x_i, y_i)}$ ← marginal density of x doesn't contain α β σ.

So for simplicity let $L = \prod_{i=1}^{n} w(y_i|x_i)$

$$\log L = \sum_{i=1}^{n} \log(w(y_i|x_i))$$

$$= \sum_{i=1}^{n} \left(-\log(\sigma) - \frac{1}{2}\log(2\pi) - \frac{1}{2\sigma^2}(y_i - (\alpha + \beta x_i))^2\right)$$

$$= -n\log\sigma - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2$$

Notice that $\forall \sigma$ MLE estimate of α, β are the $\underset{\alpha,\beta}{\operatorname{argmax}}$ (this)

$$= \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2$$

$\Downarrow$

least squares estimate $\hat{\alpha}^*$ $\hat{\beta}^*$

$\hat{\beta}^* \to \frac{S_{XY}}{S_{XX}}$

$\hat{\alpha}^* \to \bar{y} - \hat{\beta}^* \bar{x}$

So MLE = LSE

$$\frac{\partial}{\partial\sigma}\log L = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - (\hat{\alpha} + \hat{\beta}x_i))^2$$

$$\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - (\hat{\alpha} + \hat{\beta}x_i))^2$$

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

$$= \frac{1}{n} \sum_{i=1} (y_i - (\bar{y} - \hat{\beta}\bar{x} + \hat{\beta} x_i))^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y} - \hat{\beta}(x_i - \bar{x}))^2$$

$$= \frac{1}{n} \left[\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{S_{yy}} - 2\hat{\beta} \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}_{S_{xy}} + \hat{\beta}^2 \underbrace{\sum_{i=1}^{n} (x_i - \bar{x})^2}_{S_{xx}} \right]$$

$$= \frac{1}{n} \left[S_{yy} - 2\hat{\beta} S_{xy} + \hat{\beta} S_{xx} \right]$$

$$= \frac{1}{n} \left[S_{yy} - \hat{\beta} S_{xy} \right.$$

$$\Rightarrow \quad \hat{\sigma} = \sqrt{\frac{1}{n}(S_{yy} - \hat{\beta} S_{xy})}$$

Remarks:

1: for the Normal regression,

$$w(y|x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma^2} (y - (\alpha + \beta x))^2 \right]$$

The regression eqn. $\mu_{Y|x} = \alpha + \beta x$ is linear.

hence by 14.1 $\alpha = \mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1$, $\beta = \rho \frac{\sigma_2}{\sigma_1}$

Compare w/ MLE.

$\alpha = \mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1$ $\qquad \beta = \rho \frac{\sigma_2}{\sigma_1} = \frac{\text{Cov}(X,Y)}{\sigma_1^2}$

$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$ ✓ $\qquad \hat{\beta} = \frac{S_{xy}}{S_{xx}}$ ✓