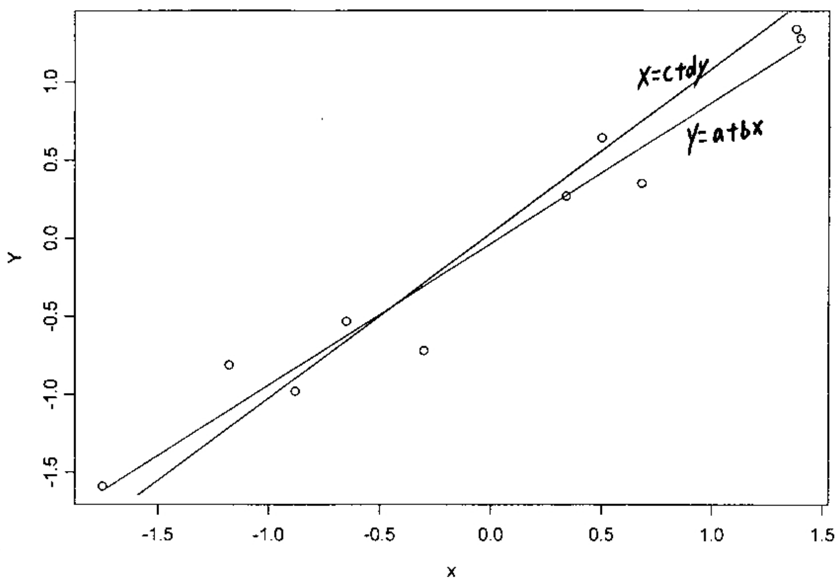
Liang chen

30

2.



a. Fit $y=a+bx$ by the method of least squares.

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

$$=\frac{10.4342\times(-0.75)-(-0.46)\times 9.4522}{10\times 10.4342-(-0.46)^2}$$

$$=-0.0334$$

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

$$=\frac{10\times 9.4522-(-0.46)\times(-0.75)}{10\times 10.4342-(-0.46)^2}$$

$$=0.9044$$

$$\therefore\quad y=-0.0334+0.9044x$$

b. Fit $x=c+dy$ by the method of least squares

$$c=\frac{(\sum_{i=1}^{n} y_i^2)(\sum_{i=1}^{n} x_i)-(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} y_i x_i)}{n\sum_{i=1}^{n} y_i^2-(\sum_{i=1}^{n} y_i)^2}$$

$$=\frac{8.9829\times(-0.46)-(-0.75)\times 9.4522}{10\times 8.9829-(-0.75)^2}$$

$$=0.0331$$

$$d=\frac{n\sum_{i=1}^{n} y_i x_i-(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n\sum_{i=1}^{n} y_i^2-(\sum_{i=1}^{n} y_i)^2}$$

$$=\frac{10\times 9.4522-(-0.46)\times(-0.75)}{10\times 8.9829-(-0.75)^2}$$

$$=1.0550$$

$$\therefore\quad x=0.0331+1.0550y$$

c. They are not the same. For the first line, X is independent. The line minimizes the sum of squared vertical distance. For the second line, Y is independent. The line minimizes the sum of squared horizontal distance. So they are different.

4. a. set up a linear model $y_i = \beta_0 + \beta_1 X_i + e_i$ $e_i \sim Norm(0, \sigma^2)$

Y: the measurement value

X: because the two objects are weighed 3 g. so we can denote

X = 1 just one object is weighed

X = 0 object 1 minus object 2

X = 2 two objects are weighed

So

Y	X
3	1
1	0
7	2
$\bar{Y} = 3.67$	$\bar{X} = 1$

b. the least squares estimates of the slope b and intercept e are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{3}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{3}(X_i-\bar{Y})^2}$$

$$= \frac{0 + (-1)(1-3.67) + 1\times(7-3.67)}{0^2+1^2+1^2}$$

$$= 3$$

$$\cancel{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 3.67 - 3\times 1 = 0.67}$$

$$\cancel{\hat{Y} = \hat{\beta}_1 X + \hat{\beta}_0 = 3X + 0.67}$$ ~~estimate w_1, w_2, thus~~ $X=0$

$\therefore$ $\hat{w}_1 = \hat{w}_2 = 3\times 1 + 0.67 = 3.67$

c. the estimate of σ^2 is $S^2 = \dfrac{\sum_{i=1}^{3}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n-2}$

$$= \frac{\cancel{(3-0.67-3)^2+(1-0.67)^2+(7-0.67-6)^2}}{1}$$

$$\cancel{= 0.6667}$$

d $\hat{\omega}_1 = \hat{\omega}_2 = \hat{\beta}_1 + \hat{\beta}_0$

$Var(\hat{\omega}_1) = Var(\hat{\omega}_2) = Var(\hat{\beta}_0) + Var(\hat{\beta}_1) + 2cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{\frac{\sum_{i=1}^{n} x_i^2}{n}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} + \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \frac{2\sigma^2\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\left[\frac{\sum_{i=1}^{n} x_i^2}{n} + 1 - 2\right.}{\sum_{i=1}^{n}(x_i-\bar{x}}$

$\therefore S_{\hat{\omega}_1} = S_{\hat{\omega}_2} = \sqrt{\frac{\frac{\sum_{i=1}^{n} x_i^2}{n} + 1 - 2\bar{x}}{\sum_{i=1}^{5}(x_i-\bar{x})^2} S^2} = \sqrt{\frac{\frac{5}{3}+1-2}{2} \times 0.6667} = 0.4714$

e. $\widehat{\omega_1-\omega_2} = \hat{\beta}_1 \times 0 + \hat{\beta}_0 = 0.67$

$Var(\widehat{\omega_1-\omega_2}) = Var(\hat{\beta}_0) + Var(0\times\hat{\beta}_1) + 2cov(\hat{\beta}_0, 0\times\hat{\beta}_1) = Var(\hat{\beta}_0) = \frac{\frac{\sum_{i=1}^{n} x_i^2}{n}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$\therefore S_{\widehat{\omega_1-\omega_2}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2/n \; S^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \sqrt{\frac{\frac{5}{3}\times 0.6667}{2}} = 0.7454$

f $\because \omega_1 - \omega_2 = \beta_1 \times 0 + \beta_0 = \beta_0$ $\quad \widehat{\omega_1-\omega_2} = \hat{\beta}_1\times 0 + \hat{\beta}_0 = \hat{\beta}_0$ $\quad S_{\hat{\beta}_0} = S_{\widehat{\omega_1-\omega_2}}$

H_0: $\omega_1 = \omega_2$ thus $\omega_1 - \omega_2 = 0$

$\because \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} \sim t_{n-2}$ thus $\frac{\widehat{\omega_1-\omega_2} - (\omega_1-\omega_2)}{S_{\widehat{\omega_1-\omega_2}}} \sim t_{n-2}$

$t = \frac{\widehat{\omega_1-\omega_2}}{S_{\widehat{\omega_1-\omega_2}}} = \frac{0.67}{0.7454} = 0.8988$

under H_0, $t \sim t_1$, because $t_{0.70} = 0.727$ $t_{0.80} = 1.376$, the p-value should be greater than 0.2 and less than 0.3. It's significant. So we can't reject H_0

11 a $Var(\hat{\mu}_0) = Var(\hat{\beta}_0) + Var(x_0\hat{\beta}_1) + 2cov(\hat{\beta}_0, x_0\hat{\beta}_1)$

$= Var(\hat{\beta}_0) + x_0^2 Var(\hat{\beta}_1) + 2x_0 cov(\hat{\beta}_0, \hat{\beta}_1)$

$\because Var(\hat{\beta}_0) = \frac{\frac{\sum_{i=1}^{n} x_i^2}{n}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$\therefore Var(\hat{\mu}_0) = \frac{\frac{\sum_{i=1}^{n} x_i^2}{n}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} + \frac{x_0^2\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} - \frac{2x_0\sigma^2\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

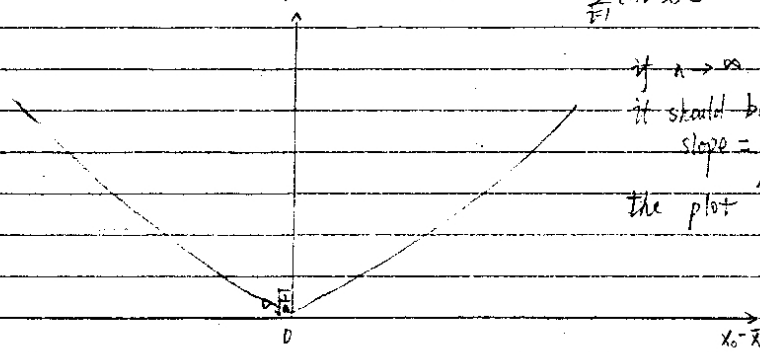
$= \frac{\sigma^2\left(\sum_{i=1}^{n} x_i^2 + n x_0^2 - 2n x_0\bar{x}\right)}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$= \frac{\sigma^2\sum_{i=1}^{n}(x_0-x_i)^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}$

$$= \frac{\sigma^2\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(x_0-\bar{x})^2\right]}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$= \sigma^2\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$$

b the standard deviation of $\hat{\mu}_0$ is $\sigma_{\hat{\mu}_0} = \sigma\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{\frac{1}{2}}$ if $n \to \infty$ $\frac{\sigma|x_0-\bar{x}|}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$



~~if $n \to \infty$ for the $x_0-\bar{x}>0$ regi~~
~~it should be a line with~~
slope = $\frac{\sigma}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$ ~~And~~
the plot is symmetry.

c $$S_{\hat{\mu}_0} = S\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{\frac{1}{2}} \qquad S = \left\{\frac{1}{n-2}\left[\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2\right]\right\}^{\frac{1}{2}}$$

the 95% confidence interval is

$$\hat{\mu}_0 \pm S_{\hat{\mu}_0}\, t_{n-2}(0.025)$$

16. Under these assumptions, the Y_i are independent and normally distributed with means $\beta_0+\beta_1x_i$ and variance σ^2

$$f(y_1,y_2,\dots,y_n|\beta_0,\beta_1) = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left[\frac{y_i-\beta_0-\beta_1x_i}{\sigma}\right]^2\right)$$

the log likelihood is

$$l(\beta_0,\beta_1) = -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i)^2$$

$$\frac{\partial l}{\partial\beta_0} = -\frac{1}{2\sigma^2}\times(-2)\times\sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_i) \qquad \text{let } \frac{\partial l}{\partial\beta_0}=0$$

thus $\sum_{i=1}^{n}\hat{y}_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n}x_i = 0$

$$\frac{\partial l}{\partial \beta_1} = -\frac{1}{2\sigma^2}(-2)\left[\sum_{i=1}^{n} x_i(y_i-\beta_0-\beta_1 x_i)\right] \qquad \text{let } \frac{\partial l}{\partial \beta_1}=0$$

thus $\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$

$$\therefore \hat{\beta}_0 = \frac{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

They are the same as the least squares estimates.

21 let Y denote the final score and X is the midterm score

For $Y = \beta_0 + \beta_1 X$

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \cdot \sqrt{\frac{S_{YY}}{S_{XX}}} = r\sqrt{\frac{S_{YY}}{S_{XX}}} = 0.5\sqrt{\frac{10}{10}} = 0.5$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 75 - 0.5 \times 75 = 37.5$$

$$\therefore \hat{Y} = 37.5 + 0.5X$$

For $X = \gamma_0 + \gamma_1 Y$

$$\hat{\gamma}_1 = \frac{S_{XY}}{S_{XX}} = \hat{\beta}_1 = 0.5$$

$$\hat{\gamma}_0 = \bar{X} - \hat{\gamma}_1 \bar{Y} = 75 - 0.5 \times 75 = 37.5$$

$$\therefore \hat{X} = 37.5 + 0.5Y$$

$\therefore$ for (a) $X=95$ $\hat{Y}=37.5+0.5\times95=85$.

for (b) $Y=85$ $\hat{X}=37.5+0.5\times85=80$.

33 denote $Y=\ln(\text{pressure})$

$X=\frac{1}{T}$

then

$$\therefore \hat{B}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$\doteq \frac{-0.01620}{7.60431\times10^{-7}}$$

$$=-2.13\times10^{4}$$

$$\therefore \hat{A}=\bar{Y}-\hat{B}\bar{X}$$

$$=-4.2715-(-2.13\times10^{4})\times0.00105713$$

$$\doteq 18.18$$

$$\therefore S_{\hat{B}}=S\frac{1}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}} \qquad S=\sqrt{\frac{\sum_{i=1}^{n}(Y_i-\hat{A}-\hat{B}X_i)^2}{n-2}}$$

$$=0.1160$$

$$=\frac{0.1160}{\sqrt{7.60436\times10^{-7}}}$$

$$=1.33\times10^{2}$$

$\therefore$ 95% confidence interval for B is

$$\hat{B}\pm t_{n-2}(0.025)S_{\hat{B}}=\hat{B}\pm t_{30}(0.025)S_{\hat{B}}$$

$$=-2.13\times10^{4}\pm2.042\times1.33\times10^{2}$$

$$=-2.13\times10^{4}\pm2.72\times10^{2}$$

$$\therefore S_{\hat{A}}=S\sqrt{\frac{\sum_{i=1}^{n}X_i^2}{n\sum_{i=1}^{n}(X_i-\bar{X})^2}}$$

$$=0.1160\times\sqrt{\frac{3.61536\times10^{-5}}{32\times7.60436\times10^{-7}}}=0.14$$

$\therefore$ 95% confidence interval for A is

$$\hat{A} \pm t_{30}(1.025)\, s_{\hat{A}}$$

$$= 18.18 \pm 2.042 \times 0.14$$

$$= 18.18 \pm 0.29.$$