统计线性模型2023秋期中回忆版

1. （同2022年秋hw1.1）(40 marks) Suppose an appliance store conducts a 5-month experiment to determine the effect of advertising on sales revenue. The results are shown below.

Advertising Expenditure x (hundreds of dollars)	**1**	**2**	**3**	**4**	**5**
Sales Revenue y (thousands of dollars)	1	1	2	2	4

(a) Draw a scatterplot of the data and comment the relationship between y and x.

(b) What is your linear regression model? State the necessary assumptions.

(c) Find the least squares line from the data and plot it on your scatterplot.

(d) Test the hypothesis that the Advertising Expenditure has no effect of the Sales Revenue when a linear model is used (use $\alpha = 0.05$). State the null and alternative hypotheses. Draw the appropriate test conclusions.

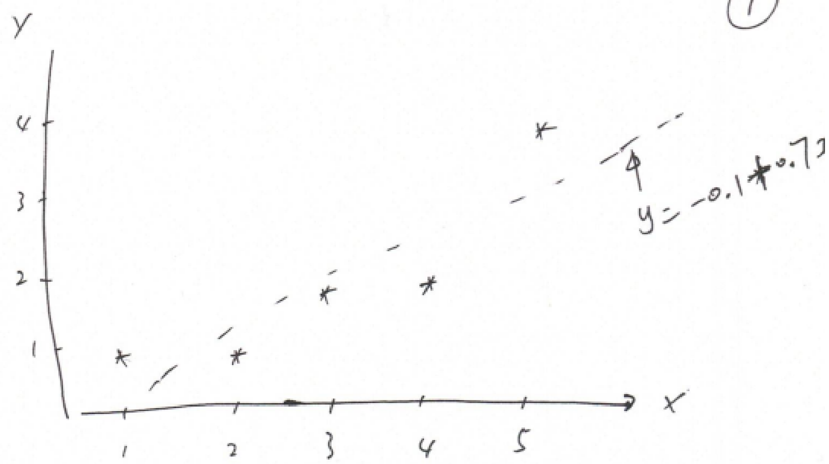
(e) Find a 95% confidence interval for β_1 (slope of the linear regression model). Interpret your results.

(f) Eind the coefficient of determination for the linear regression model. Interpret your result.

(g) Find a prediction for the mean Sales Revenue when 4 hundreds dollars are spent on advertising and its 95% interval. What is the 95% interval for the Sales Revenue?

Assignment 1
Solutions

1. (a)



It shows a rough linear relationship between y and x, but the correlation is ~~not~~ quite strong.

(b) A linear regression model can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

main assumptions are:

(i) the relationship between y_i and x_i is linear

(ii) the variance of y is a constant

(iii) the observations are independent

(iv) (optional) Y is distributed normally.

(c) $\bar{x} = 3 \quad \bar{y} = 2$

$$S_{xx} = \Sigma(x_i - \bar{x})^2 = 10$$

$$S_{xy} = \Sigma(x_i - \bar{x})(y_i - \bar{y}) = 7$$

$$S_{yy} = \Sigma(y_i - \bar{y})^2 = 6$$

$$\hat{\beta}_1 = S_{xy}/S_{xx} = 7/10 = 0.7$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2 - 0.7 * 3 = -0.1$$

Thus the least squares line is $y = -0.1 + 0.7x$, which has been added to the scatterplot

(d) $H_0: \beta_1 = 0$ v.s. $H_1: \beta_1 \neq 0$ $n=5$

$$SSE = \Sigma (y_i - \hat{y}_i)^2 = 1.1$$

$$S^2 = \frac{SSE}{n-2} = 0.3667$$

$$t = \frac{\hat{\beta}_1}{(S/S_{xx})^{1/2}} = \frac{0.7}{(0.3667/10)^{1/2}} = 3.6556$$

$$\geq t_{0.025,3} = 3.1824$$

Thus, reject H_0, meaning the Advertising Expenditure has effect of the Sales Revenue.

(e) $t = \frac{\hat{\beta}_1 - \beta_1}{S/S_{xx}^{1/2}} \sim t_3$, the 95% C.I. of β_1 is

$$\hat{\beta}_1 \pm t_{0.025,3} \cdot S/S_{xx}^{1/2}$$

$$= 0.7 \pm 3.1824 * \left(\frac{0.3667}{10}\right)^{1/2}$$

$$= (0.0906, 1.3094)$$

There is 95% chance that β_1 would take values between 0.0906 and 1.3094.

(f) $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{1.1}{6} = 0.8167$

meaning 81.67% percent of the variations of the Sales Revenue can be explained by the model.

(g) $x_h = 4$, $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 x_h = 2.7$

C. I. of $E(Y_h)$ is

$$\hat{Y}_h \pm t_{0.025,3} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_h - \bar{x})^2}{Sxx}}$$

$$= (1.6445,\ 3.7559)$$

Predictive interval of Y_h at $x_h = 4$ is

$$\hat{Y}_h \pm t_{0.025,3} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{Sxx}}$$

$$= (0.5028,\ 4.8972).$$

2. (20 marks) List at least two generalized inverse of A, verify it.

$$A = \begin{bmatrix} 2 & 2 & 6 \\ 1 & 0 & 2 \\ 3 & 2 & 8 \end{bmatrix}$$

3. （同2022年秋hw3.2）(20 marks) Let $\boldsymbol{y} = (y_1, y_2, y_3)'$ be distributed as $N_3(\boldsymbol{\mu}, \Sigma)$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}.$$

(a) Find the distribution of $\begin{pmatrix} y_1 - y_2 + y_3 \\ 2y_1 + y_2 - y_3 \end{pmatrix}$;

(b) The conditional distribution of (y_1, y_2) given y_3;

(c) The partial correlation between y_1 and y_2 given y_3.

Q2

(a)

$$\begin{pmatrix} y_1 - y_2 + y_3 \\ 2y_1 + y_2 - y_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \sim N(\underset{\sim}{\mu}_1, \underset{\sim}{\Sigma}_1).$$

$$\underset{\sim}{\mu}_1 = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$\underset{\sim}{\Sigma}_1 = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 1 & -1 \end{pmatrix} \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 2 \\ 9 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 4 \\ 4 & 23 \end{pmatrix}.$$

(b) $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mid y_3 \sim N(\underset{\sim}{\mu}_2, \underset{\sim}{\Sigma}_2)$.

$\begin{pmatrix} y_1 \\ y_2 \\ \hline y_3 \end{pmatrix} \begin{matrix} \underset{\sim}{y} \\ \hline x \end{matrix}$ ②

$\underset{\sim}{\Sigma} = \left(\begin{array}{cc|c} 4 & 1 & 0 \\ 1 & 2 & 1 \\ \hline 0 & 1 & 3 \end{array}\right) = \begin{pmatrix} \underset{\sim}{\Sigma}_{yy} & \underset{\sim}{\Sigma}_{yx} \\ \underset{\sim}{\Sigma}_{xy} & \underset{\sim}{\Sigma}_{xx} \end{pmatrix}$

$$\underset{\sim}{\mu}_2 = \mu_y + \underset{\sim}{\Sigma}_{yx} \underset{\sim}{\Sigma}_{xx}^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}_x)$$

$$= \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} 3^{-1} (y_3 - 3)$$

$$= \begin{pmatrix} 2 \\ \frac{y_3}{3} - 2 \end{pmatrix}$$

$$\underset{\sim}{\Sigma}_2 = \underset{\sim}{\Sigma}_{yy} - \underset{\sim}{\Sigma}_{yx} \underset{\sim}{\Sigma}_{xx}^{-1} \underset{\sim}{\Sigma}_{xy}$$

$$= \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} 3^{-1} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & \frac{5}{3} \end{pmatrix}$$

(c) The partial correlation between y_1 and y_2 given y_3 is:

$$\frac{1}{\sqrt{4 * 5/3}} = 0.3873$$

4. (20 marks) $y \sim N_p(\mu, \sigma^2\Sigma),\ \mu, \sigma^2, \sum$ is knows.

(a) $E(y'\Sigma^{-1}y)$;

(b) Distribution of $\frac{1}{\sigma^2} y'\Sigma^{-1}y$.

4.1 Random vector and matrix

- Let $\boldsymbol{Y}$ be a random vector with mean $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{Y})$ and $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{Y})$, then

$$\mathrm{E}(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$$

  where $\boldsymbol{A}$ is a symmetric matrix.

Ch 5.2 Non-Central χ^2, F and t distributions

Theorem 5.1 Let $\boldsymbol{x}_{p\times 1} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{A}$ be symmetric, then $q = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \sim \chi^2_{(r,\,\lambda)}$ where r denoting the rank of $\boldsymbol{A}$ and $\lambda = \frac{\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}}{2}$ if and only if $\boldsymbol{A}\boldsymbol{\Sigma}$ is idempotent.