04/08

Warm up

Exercise: Using the exercise 1.27 of the book (muscle mass) and the 6 observations:

Observation	1	2	3	4	5	6
Xi	43	39	41	86	72	76
Yi	106	106	97	60	70	80

a. Obtain a Table in order to compute the least square estimates. Similar to Table 1.1 (Tocula company example) page 19.

b. Compute bo and b1 using the equations (1.10a) and (1.10b).

c. Consider and explain a, b and c in 2.27 (Book) for this exercise.

Figure 1: Midterm bonus problem.

d. Given $X_h = 40$, compute the point estimator of the mean response $E[Y_h]$ and the corresponding 95% confidence interval.

i	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	43	106	-16.5	19.5			
2	39	106	-20.5	19.5			
3	41	97	-18.5	10.5			
4	86	60	26.5	-26.5			
5	72	70	12.5	-16.5			
6	76	80	16.5	-6.5			
Total	357	519					
Mean	59.5	86.5					

a

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

b

c

d

$$\hat{Y}_h = b_0 + b_1 X_h$$

$$(\hat{Y}_h - t(1-\alpha/2; n-2) * s(\hat{Y}_h), \hat{Y}_h + t(1-\alpha/2; n-2) * s(\hat{Y}_h))$$

1.2

$Y = 2X + 300$. Functional relation.

1.8

Yes. Not necessarily

$$E[Y] = \beta_0 + \beta_1 * X$$

1.12

a

b

c

d

1.27

a

b

1.30

04/15

Warm up

Given $X_h = 40$, compute the 90% prediction interval for the mean of the $m = 3$ new Y observations. See Page 60 in the textbook.

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{\text{predmean}\},$$

where

$$s^2\{\text{predmean}\} = \frac{MSE}{m} + s^2(\hat{Y}_h) = MSE\left[\frac{1}{m} + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right].$$

In the example provided by the textbook (Page 61), we have $b_0 = 62.37$, $b_1 = 3.5702$, $\bar{X} = 70$, $\sum(X_i - \bar{X})^2 = 19800$, and $MSE = 2384$. Thus

$$\hat{Y}_h = b_0 + b_1 X_h = 62.37 + 3.5702 \times 40 = 205.2,$$

and

$$s^2(\hat{Y}_h) = 2384\left[\frac{1}{25} + \frac{(40-70)^2}{19800}\right] = 203.72.$$

Therefore

$$s^2\{\text{predmean}\} = \frac{2384}{3} + 203.72 = 998.4.$$

Note that $t(0.95; 23) = 1.714$. We obtain

$$205.2 - 1.714 \times 31.60 \le \bar{Y}_{h(\text{new})} \le 205.2 + 1.714 \times 31.60$$
$$151.0 \le \bar{Y}_{h(\text{new})} \le 259.4.$$

2.9

Formula (2.30). $s(\hat{Y}_h)$ depends on X_h, which needs to be specified.

2.10

(a)

(b)

(c)

2.12

(a) (2.37)

$$\sigma^2\{\text{pred}\} = \sigma^2\{Y_{h(\text{new})} - \hat{Y}_h\} = \sigma^2\{Y_{h(\text{new})}\} + \sigma^2(\hat{Y}_h) = \sigma^2 + \sigma^2(\hat{Y}_h),$$

where $\hat{Y}_h = b_0 + b_1 X_h$.

(b) (2.29b)

$$\sigma^2(\hat{Y}_h) = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right].$$

2.27

(a)

$$H_0 : \beta_1 \geq 0 \qquad \text{vs} \qquad H_a : \beta_1 < 0.$$

$$t^* < t(\alpha; n-2).$$

$$p(T < t^*) < \alpha,$$

where T is a random variable which is t distributed with degrees of freedom $n-2$.

(b)

(c)

$$b_0 + b_1 X; b_0 + b_1(X+1)$$

$$b_1 \pm t(0.975; 58)s(b_1)$$

04/29

Exercise: Using the exercise 1.27 of the book (muscle mass) and the 6 observations:

Observation	1	2	3	4	5	6
Xi	43	39	41	86	72	76
Yi	106	106	97	60	70	80

a. Estimate a 95% confidence interval for the parameters β_0 and β_1

b. Use the p-value for testing $H_0: \beta_1 = 0$, with $\alpha = 0.01$.

c. Compute a 99% confidence interval for the mean response $E\{Y_h\}$ using $X_h = 30$.

d. Test whether $\beta_1 = 0$ using an F test with $\alpha = 0.01$. State the alternatives, decision rule, and conclusion.

Figure 2

What we need from the data:

$$n = 6$$

$$t(1 - 0.05/2; n-2) = t(0.975, n-2) = 2.776$$

$$t(1 - 0.01/2; n-2) = t(0.995, n-2) = 4.604$$

$$\bar{X} = \frac{43 + 39 + 41 + 86 + 72 + 76}{6} = 59.5$$

$$\bar{Y} = \frac{106 + 106 + 97 + 60 + 70 + 80}{6} = 86.5$$

$$\sum (X_i - \bar{X})^2 = 2165.5$$

$$SSTO = \sum (Y_i - \bar{Y})^2 = 1887.5$$

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = -1931.5$$

What we can derive:

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{-1931.5}{2165.5} = -0.892$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 86.5 - (-0.892) \times 59.5 = 139.57$$

$$SSR = b_1^2 \sum (X_i - \bar{X})^2 = 1722.79$$

$$MSR = SSR/1 = 1722.79$$

$$SSE = SSTO - SSR = 164.71$$

$$MSE = SSE/(n-2) = 41.18$$

$$s^2(b_0) = MSE \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right] = 74.19$$

$$s^2(b_1) = \frac{MSE}{\sum(X_i - \bar{X})^2} = 0.019$$

$$t^* = \frac{b_1}{s(b_1)} = \frac{-0.892}{\sqrt{0.019}} = -6.471$$

$$F^* = \frac{MSR}{MSE} = \frac{1722.79}{41.18} = 41.84$$

$$F^* = (t^*)^2$$

a P45 and P49

$$b_0 \pm t(1-\alpha/2; n-2)s(b_0)$$

$$139.57 \pm 2.776 \times \sqrt{74.19} = [115.66, 163.48]$$

$$b_1 \pm t(1-\alpha/2; n-2)s(b_1)$$

$$-0.892 \pm 2.776 \times \sqrt{0.019} = [-1.127, -0.509]$$

b

$$H_a : \beta_1 \geq 0 \quad P(t(4) > t^*)$$

$$H_a : \beta_1 \leq 0 \quad P(t(4) < t^*)$$

p-value:

$$2P(t(4) > |t^*|) = 2P(t(4) > |-6.471|) = 0.00294$$

R code: `2*(1-pt(6.471, 4))` Since the p-value is less than $\alpha = 0.01$, we conclude H_a.

c

$$\hat{Y}_h = b_0 + b_1 X_h = 139.57 + (-0.892) * 30 = 112.81$$

$$s^2(\hat{Y}_h) = MSE\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right] = 23.41$$

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s(\hat{Y}_h)$$

$$112.81 \pm 4.604 \times \sqrt{23.41} = [90.53, 135.09]$$

d

$$H_a : \beta_1 \neq 0$$

p-value: $P(F(1, n-2) > F^*) = 0.0029 < \alpha = 0.01$
R code: `1-pf(41.84, 1, 4)`
quantile: $F(1-\alpha; 1, n-2) = 21.2 < F^*$
R code: `qf(0.99, 1, 4)`
Decision rule: If $F^* \leq F(1-\alpha; 1, n-2)$, conclude H_0. if $F^* > F(1-\alpha; 1, n-2)$, conclude H_a.
Conclusion: H_a

Midterm II Bonus (Due May 21 at 4pm)

Consider the following regression model with two predictors X_1 and X_2 ($p = 3$ in the general linear regression model)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, i = 1, 2, \cdots, n.$$

In matrix form,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

- Compute $(\mathbf{X}'\mathbf{X})^{-1}$. Show all the steps and formulas. Hint: Equation 5.23.
- Show $\mathbf{HH} = \mathbf{H}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the hat matrix. Show all the steps and formulas.
- Compute the ANOVA table, formulas. Write the model. Hint: the second column of the ANOVA table (Table 6.1 SS) should be expressed using $Y_i, X_{ij}, i = 1, 2, \cdots, n, j = 1, 2, 3$. Equations 6.74 and 6.75 may be useful.