Spring 2021 MATH 476 Homework #6

Due: May 8, 11:59PM, submit in blackboard

Homework solution is not required to be typed, but must be legible.

Problem 1 Exercise 13.1(a),(c) from the TEXT.

Problem 2 Exercise 14.3 from the TEXT.

Problem 3 Exercise 11.15 from the TEXT.

Problem 4 Refer to the dataset given in Exercise 11.3 on page 573.

1) Find a 95% confidence interval for $E(Y)$ when $x = 2$.
2) Find a 95% prediction interval for Y when $x = 2$.

Problem 5 Consider the dataset from Exercise 11.31 on page 588. We are interested in testing if there is sufficient evidence to indicate that peak current (y) changes as nickel concentrations (x) increase. So we test the null hypothesis $H_0 : \beta_1 = 0$ vs the alternative hypothesis $H_a : \beta_1 \neq 0$.

1) Use an F-test.
2) Use a t-test.
3) Compare the value of F to the value of t, and verify if $F = t^2$.

Problem 6 Consider the dataset from Exercise 11.69 on page 615. Fit the model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$.

1) Carry out an F-test of $H_0 : \beta_2 = 0$ vs $H_a : \beta_2 \neq 0$ at level $\alpha = 0.05$.
2) Test $H_0 : \beta_1 = \beta_2 = 0$ vs H_a: *at least one of the equalities does not hold* at level $\alpha = 0.05$.

Problem 1

13.1 The reaction times for two different stimuli in a psychological word-association experiment were compared by using each stimulus on independent random samples of size 8. Thus, a total of 16 people were used in the experiment. Do the following data present sufficient evidence to indicate that there is a difference in the mean reaction times for the two stimuli?

Stimulus 1	1	3	2	1	2	1	3	2
Stimulus 2	4	2	3	3	1	2	3	3

a Use the ANOVA approach to test the appropriate hypotheses. Test at the $\alpha = .05$ level of significance.

b **Applet Exercise** Use the applet *F-Ratio Probabilities and Quantiles* to determine the exact p-value for the test in part (a).

c Test the appropriate hypotheses by using the two-sample t test for comparing population means, which we developed in Section 10.8. Compare the value of the t statistic to the value of the F statistic calculated in part (a).

d What assumptions are necessary for the tests implemented in the preceding parts?

(a)

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_a: \mu_1 \neq \mu_2 \end{cases}$$

$k = 2$

$n_1 = n_2 = 8 \rightarrow \sum_{i=1}^{k} n_i = 16$

$RR = \{F : F > |F_{\alpha/2}|\}$

$$F = \frac{MST}{MSE}$$

$$MST = \frac{SST}{k-1} = \left[\frac{n_1 n_2}{n}(\bar{Y}_1 - \bar{Y}_2)^2\right] \div (k-1)$$

$$MSE = \frac{SSE}{n-k} = \left[\sum_{i=1}^{k}(n_i - 1)\cdot S_i^2\right] \div (n-k)$$

$$S_1 = \frac{\sum_{i=1}^{n_1}(y_i - \bar{y})^2}{n_1 - 1}$$

$$= \frac{1}{7}\cdot\left[(1-1.875)^2 + (3-1.875)^2 + (2-1.875)^2 + (1-1.875)^2 + (2-1.875)^2 + (1-1.875)^2 + (3-1.875)^2 + (2-1.875)^2\right]$$

$$= 0.696$$

$$S_2 = \frac{\sum_{i=1}^{n_2}(y_i - \bar{y})^2}{n_2 - 1}$$

$$= \frac{1}{7}\cdot\left[(4-2.625)^2 + (2-2.625)^2 + (3-2.625)^2 + (3-2.625)^2 + (1-2.625)^2 + (2-2.625)^2 + (3-2.625)^2 + (3-2.625)^2\right]$$

$$= 0.839$$

$$MST = \frac{8\cdot 8}{16}(1.875 - 2.625)^2 \div (2-1)$$

$$= 2.25$$

$$MSE = \left[(7)\cdot 0.696 + (7)\cdot 0.839\right] \div (16-2)$$

$$= 0.7675$$

$$F = \frac{MST}{MSE} = \frac{2.25}{0.7675} = 2.932$$

$$F_{\alpha/2,\,k-1,\,n-k} = F_{0.025,\,1,\,14} = 6.30$$

F not in RR $\rightarrow$ $\boxed{\text{Fail to reject } H_0 \text{ that } \mu_1 = \mu_2.}$

(c) $\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_a: \mu_1 - \mu_2 \neq 0 \end{cases}$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$= \frac{(7)(0.696) + (7)(0.839)}{14}$$

$$= 0.7675$$

$$\hookrightarrow S_p = \sqrt{0.7675} = 0.876$$

RR: $\{t : |t| > t_{\alpha/2}\}$

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - 0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{4}} = 0.5$$

$$t = \frac{1.875 - 2.625}{0.876 \cdot 0.5}$$

$$= -1.712$$

$$|t| = 1.712$$

$$t_{\alpha/2, df} = t_{0.025, 14} = 2.145$$

t not in RR $\longrightarrow$ $\boxed{\text{Fail to reject } H_0 \text{ that } \mu_1 = \mu_2.}$

We can observe that $t^2 = 1.712^2 = 2.924 \approx F = 2.932.$

$\llcorner\ t^2 \approx F$

We can prove that the true values of t^2 and F (i.e., in the absence of rounding) are equivalent.

proof: $t^2 = F$ when $k = 2$

$$F = \frac{MST}{MSE}$$

$$MSE = \frac{SSE}{n-k}$$

$$= \left[\sum_{i=1}^{k} (n_i - 1) \cdot S_i^2\right] \div (n-k)$$

$$= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n-k}$$

$$= S_p^2$$ — when $k = 2$, as it does here

$$MST = \frac{SSE}{k-1}$$

$$= SSE$$ — when $k=2$, as it does here

$$= \sum_{i=1}^{k} n_i (y_i - \bar{y})^2$$

$$= n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2$$

$$= n_1\left(\bar{y}_1 - \frac{n_1\bar{y}_1 + n_1\bar{y}_2}{n_1 + n_1}\right)^2 + n_2\left(\bar{y}_2 - \frac{n_2\bar{y}_1 + n_2\bar{y}_2}{n_2 + n_2}\right)^2$$

$$= n_1\left(\frac{n\bar{y}_1 - n_1\bar{y}_1 - n_2\bar{y}_2}{n}\right)^2 + n_2\left(\frac{n\bar{y}_1 - n_2\bar{y}_1 - n_1\bar{y}_2}{n}\right)^2$$

$$= n_1\left(\frac{(n - n_1)\bar{y}_1 - n_2\bar{y}_2}{n}\right)^2 + n_2\left(\frac{(n - n_2)\bar{y}_1 - n_1\bar{y}_2}{n}\right)^2$$

$$= \frac{n_1 n_2^2}{n^2}(\bar{y}_1 - \bar{y}_2)^2 + \frac{n_1^2 n_2}{n^2}(\bar{y}_1 - \bar{y}_2)^2$$

$$= \left(\frac{n_1 n_2}{n^2}\right)\left(\underbrace{(n_2 + n_1)}_{=n}(\bar{y}_1 - \bar{y}_2)^2\right)$$

$$= \frac{n_1 n_2}{n}(\bar{y}_1 - \bar{y}_2)^2$$

$$= \frac{n_1 n_2}{n_1 + n_2}(\bar{y}_1 - \bar{y}_2)^2$$

$$= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}(\bar{y}_1 - \bar{y}_2)^2$$

$$F = \frac{MST}{MSE}$$

$$= \frac{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}(\bar{y}_1 - \bar{y}_2)^2}{S_p^2}$$

$$= \frac{(\bar{y}_1 - \bar{y}_2)^2}{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\therefore F = t^2$$

Problem 2

14.3 A city expressway with four lanes in each direction was studied to see whether drivers preferred to drive on the inside lanes. A total of 1000 automobiles were observed during the heavy early-morning traffic, and their respective lanes were recorded. The results are shown in the accompanying table. Do the data present sufficient evidence to indicate that some lanes are preferred over others? (Test the hypothesis that $p_1 = p_2 = p_3 = p_4 = 1/4$, using $\alpha = .05$.) Give bounds for the associated p-value.

Lane	1	2	3	4
Count	294	276	238	192

$$\begin{cases} H_0: p_1 = p_2 = p_3 = p_4 = 0.25 \\ H_a: \text{not all equal} \end{cases}$$

$df = k - 1 = 3$

$E(n_i) = np_i$
$= 1000 \times 0.25$
$= 250$

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E(n_i))^2}{E(n_i)}$$

$$= \frac{(294-250)^2}{250} + \frac{(276-250)^2}{250} + \frac{(238-250)^2}{250} + \frac{(192-250)^2}{250}$$

$$= 24.48$$

$$\chi^2_{0.05,3} = 7.815$$

the largest value the table shows

$P(\chi^2 > 12.8381) = 0.0005.$

using excel: chisq.dist.rt(24.48, 3) = 0.0000

reject if $P(\chi^2 > X^2) \leq 0.05$.

The probability of $X^2 > 24.48$ is too small to put in the table.

$\longrightarrow$ reject H_0 that $p_1 = p_2 = p_3 = p_4$.

Problem 3

11.15 **a** Derive the following identity:

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$
$$= \sum_{i=1}^{n}(y_i - \overline{y})^2 - \hat{\beta}_1 \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = S_{yy} - \hat{\beta}_1 S_{xy}.$$

Notice that this provides an easier computational method of finding SSE.

b Use the computational formula for SSE derived in part (a) to prove that SSE $\leq S_{yy}$. [*Hint:* $\hat{\beta}_1 = S_{xy}/S_{xx}$.]

(a)

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

$$= \sum_{i=1}^{n}\left(y_i - (\bar{y} - \hat{\beta}_1\bar{x}) - \hat{\beta}_1 x_i\right)^2 \longrightarrow \text{least-squares estimator for } \hat{\beta}_0: \ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$= \sum_{i=1}^{n}\left(y_i - \bar{y} + \hat{\beta}_1(\bar{x} - x_i)\right)^2$$

$$= \sum_{i=1}^{n}\left[(y_i - \bar{y})^2 - 2\hat{\beta}_1(y_i - \bar{y})(x_i - \bar{x}) + \hat{\beta}_1^2(\bar{x} - x_i)^2\right]$$

$$= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\hat{\beta}_1 \cdot \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) + \hat{\beta}_1 \cdot \hat{\beta}_1 \cdot \sum_{i=1}^{n}(\bar{x} - x_i)^2$$

$$= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\hat{\beta}_1 \cdot \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) + \hat{\beta}_1 \cdot \underbrace{\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}_{\text{least-squares estimator for } \hat{\beta}_1} \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\hat{\beta}_1 \cdot \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) + \hat{\beta}_1 \cdot \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$$

$$= \underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{S_{yy}} - \hat{\beta}_1 \cdot \underbrace{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}_{S_{xy}}$$

$$= S_{yy} - \hat{\beta}_1 S_{xy} \quad \checkmark\checkmark$$

(b)

$$SSE = S_{yy} - \hat{\beta}_1 \cdot S_{xy}$$

$$= S_{yy} - \left[\frac{S_{xy}}{S_{yy}}\right] \cdot S_{xy}$$

$$= S_{yy} - \frac{S_{xy}^2}{S_{yy}}$$

$S_{xy}^2 \longrightarrow$ a squared value $\Rightarrow$ 0 or greater

$S_{yy} \longrightarrow$ the sum of squared values $\Rightarrow$ 0 or greater

$$\therefore \frac{S_{xy}^2}{S_{yy}} \geq 0$$

and $S_{yy} - \dfrac{S_{xy}^2}{S_{yy}} \leq S_{yy}$

i.e., $SSE \leq S_{yy}$ ✓✓

Problem 4

$\bar{x} = 0 \quad \bar{y} = 1.5$

y	3.0	2.0	1.0	1.0	0.5
x	−2.0	−1.0	0.0	1.0	2.0

1) Find a 95% confidence interval for $E(Y)$ when $x = 2$.
2) Find a 95% prediction interval for Y when $x = 2$.

C.I. for $E(Y \mid x = x^*)$: $\hat{y} \pm t_{\alpha/2} \cdot \sqrt{\dfrac{SSE}{n-2}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{(x^* - \bar{x})^2}{S_{xx}}}$

P.I. for Y when $x = x^*$: $\hat{y} \pm t_{\alpha/2} \cdot \sqrt{\dfrac{SSE}{n-2}} \cdot \sqrt{1 + \dfrac{1}{n} + \dfrac{(x^* - \bar{x})^2}{S_{xx}}}$

for large samples:

$\sqrt{\dfrac{SSE}{n-2}} \Rightarrow \sigma$

$t \Rightarrow z$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$
$$= (-2-0)(3-1.5) + (-1-0)(2-1.5) + (0-0)(1-1.5) + (1-0)(1-1.5) + (2-0)(0.5-1.5)$$
$$= -6$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$
$$= (-2-0)^2 + (-1-0)^2 + (0-0)^2 + (1-0)^2 + (2-0)^2$$
$$= 10$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$
$$= (3-1.5)^2 + (2-1.5)^2 + (1-1.5)^2 + (1-1.5)^2 + (0.5-1.5)$$
$$= 4$$

From table:

$t_{\alpha/2, n-2} = t_{0.025, 3} = 3.182$

$$SSE = S_{yy} - \frac{S_{xy}^2}{S_{yy}} = 4 - \frac{(-6)^2}{10} = 4 - 3.6 = 0.4$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = (\bar{Y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 x = \bar{Y} + \hat{\beta}_1 (x - \bar{x}) = 1.5 - \frac{6}{10}(2-0) = 0.3$$

(a)

$$C.I. = 0.3 \pm 3.182 \times \sqrt{\frac{0.4}{5-2}} \times \sqrt{\frac{1}{5} + \frac{(2-0)^2}{10}} \longrightarrow \boxed{-0.60 \le E(Y \mid x=2) \le 1.20}$$

(b)

$$P.I. = 0.3 \pm 3.182 \times \sqrt{\frac{0.4}{5-2}} \times \sqrt{1 + \frac{1}{5} + \frac{(2-0)^2}{10}} \longrightarrow \boxed{-1.075 \le Y \text{ (given } x=2) \le 1.675}$$

Problem 5

Problem 5 Consider the dataset from Exercise 11.31 on page 588. We are interested in testing if there is sufficient evidence to indicate that peak current (y) changes as nickel concentrations (x) increase. So we test the null hypothesis $H_0 : \beta_1 = 0$ vs the alternative hypothesis $H_a : \beta_1 \neq 0$.

1) Use an F-test.
2) Use a t-test.
3) Compare the value of F to the value of t, and verify if $F = t^2$.

x = Ni (ppb)	y = Peak Current (μA)
19.1	.095
38.2	.174
57.3	.256
76.2	.348
95	.429
114	.500
131	.580
150	.651
170	.722

For test $\begin{cases} H_0: \beta_j = 0 \\ H_a: \beta_j \neq 0 \end{cases}$

Two choices:

(1) F-test $\longrightarrow$ $F = \dfrac{SSR/k}{SSE/(n-k-1)} \sim F_{(k,n-k-1)}$

(2) t-test $\longrightarrow$ $t = \dfrac{\hat{\beta}_1 - 0}{\sqrt{\frac{SSE}{n-2} \cdot \frac{1}{S_{xx}}}} \sim t_{n-2}$

1.

$$F = \frac{SSR/k}{SSE/(n-k-1)} \sim F_{(k,n-k-1)}$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{yy}}$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$= \sum_{i=1}^{n} \left((\hat{\beta}_0 + \hat{\beta}_1 x_i) - (\hat{\beta}_0 + \hat{\beta}_1 \bar{x}) \right)^2$$

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

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

$$= \frac{S_{xy}^2}{S_{xx}}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = 21068.82$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 0.3748$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 88.800$$

$SSE = 0.0005$

$SSR = 0.3743$

$F = 5334.843$

$F_{\alpha, k, n-k-1} = F_{0.05, 1, 7} = 5.59$

$F > F_{\alpha, k, n-k-1}$

$\boxed{\text{We can reject } H_0 \text{ at level } \alpha = 0.05.}$

2.

$$t = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{SSE}{n-2} \cdot \frac{1}{S_{xx}}}}$$

$$= \frac{S_{xy}/S_{xx}}{\sqrt{\frac{SSE}{n-2} \cdot \frac{1}{S_{xx}}}}$$

$$= 73.04$$

$$t_{\alpha/2, n-2} = t_{0.025, 7} = 2.365$$

$$|t| > t_{\alpha/2, n-2}$$

We can reject H_0 at level $\alpha = 0.05$.

3. $t = 73.04$

$F = 5334.843 \approx 73.04^2$

Yes, $F = t^2$.

Problem 6

Problem 6 Consider the dataset from Exercise 11.69 on page 615. Fit the model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$.

1) Carry out an F-test of $H_0 : \beta_2 = 0$ vs $H_a : \beta_2 \neq 0$ at level $\alpha = 0.05$.
2) Test $H_0 : \beta_1 = \beta_2 = 0$ vs H_a: *at least one of the equalities does not hold* at level $\alpha = 0.05$.

x	y
1996	18.5
1997	22.6
1998	27.2
1999	31.2
2000	33.0
2001	44.9
2002	49.4
2003	35.0

(1)

complete model

$$Y = \begin{bmatrix} 18.5 \\ 22.6 \\ 27.2 \\ 31.2 \\ 33.0 \\ 44.9 \\ 49.4 \\ 35.0 \end{bmatrix} \qquad X = \begin{array}{c} \begin{array}{ccc} x^0 & x^1 & x^2 \end{array} \\ \begin{bmatrix} 1 & 1996 & 3984016 \\ 1 & 1997 & 3988009 \\ 1 & 1998 & 3992004 \\ 1 & 1999 & 3996001 \\ 1 & 2000 & 4000000 \\ 1 & 2001 & 4004001 \\ 1 & 2002 & 4008004 \\ 1 & 2003 & 4012009 \end{bmatrix} \end{array}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1996 & 1997 & 1998 & 1999 & 2000 & 2001 & 2002 & 2003 \\ 3984016 & 3988009 & 3992004 & 3996001 & 4000000 & 4004001 & 4008004 & 4012009 \end{bmatrix} \begin{bmatrix} 1 & 1996 & 3984016 \\ 1 & 1997 & 3988009 \\ 1 & 1998 & 3992004 \\ 1 & 1999 & 3996001 \\ 1 & 2000 & 4000000 \\ 1 & 2001 & 4004001 \\ 1 & 2002 & 4008004 \\ 1 & 2003 & 4012009 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^n x_i^0 \cdot x_i^0 & \sum_{i=1}^n x_i^0 x_i^1 & \sum_{i=1}^n x_i^0 x_i^2 \\ \sum_{i=1}^n x_i^1 \cdot x_i^0 & \sum_{i=1}^n x_i^1 x_i^1 & \sum_{i=1}^n x_i^1 x_i^2 \\ \sum_{i=1}^n x_i^2 \cdot x_i^0 & \sum_{i=1}^n x_i^2 x_i^1 & \sum_{i=1}^n x_i^2 x_i^2 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & 15996 & 31984044 \\ 15996 & 31984044 & 6.3952 \cdot 10^{10} \\ 31984044 & 6.3952 \cdot 10^{10} & 1.27873 \cdot 10^{14} \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \rightarrow \text{calculator} \rightarrow$$

$a_1 = 9.51427 \cdot 10^{10}$

$a_2 = -9.51666 \cdot 10^7$

$a_3 = 23797.6$

$b_1 = -9.51666 \cdot 10^7$

$b_2 = 95190.5$

$b_3 = -23.8036$

$c_1 = 23797.6$

$c_2 = -23.8036$

$c_3 = 0.005952$

$$B = (X^T X)^{-1} X^T Y = \begin{bmatrix} -2.16803 \cdot 10^6 \\ 2164.99 \\ -0.540476 \end{bmatrix}$$

reduced model

$$Y = \begin{bmatrix} 18.5 \\ 22.6 \\ 27.2 \\ 31.2 \\ 33.0 \\ 44.9 \\ 49.4 \\ 35.0 \end{bmatrix} \quad X = \begin{array}{c} \begin{array}{cc} x^0 & x^1 \end{array} \\ \begin{bmatrix} 1 & 1996 \\ 1 & 1997 \\ 1 & 1998 \\ 1 & 1999 \\ 1 & 2000 \\ 1 & 2001 \\ 1 & 2002 \\ 1 & 2003 \end{bmatrix} \end{array}$$

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1996 & 1997 & 1998 & 1999 & 2000 & 2001 & 2002 & 2003 \end{bmatrix} \begin{bmatrix} 1 & 1996 \\ 1 & 1997 \\ 1 & 1998 \\ 1 & 1999 \\ 1 & 2000 \\ 1 & 2001 \\ 1 & 2002 \\ 1 & 2003 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i^0 \cdot x_i^0 & \sum_{i=1}^{n} x_i^0 x_i^1 \\ \sum_{i=1}^{n} x_i^1 \cdot x_i^0 & \sum_{i=1}^{n} x_i^1 x_i^1 \end{bmatrix} = \begin{bmatrix} 8 & 15996 \\ 15996 & 31984044 \end{bmatrix}$$

$$(X^TX)^{-1} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = \begin{bmatrix} 95190.6 & -47.6071 \\ -47.6071 & 0.0238 \end{bmatrix}$$

$$B = (X^TX)^{-1} X^TY = \begin{bmatrix} -7213.08 \\ 3.6238 \end{bmatrix}$$

$$SSE = Y^TY - B^TX^TY$$

$$SSE_c = 168.636$$

$$SSE_R = 217.711$$

k = # independent variables in complete model = 2

g = # independent variables in reduced model = 1

$$F = \frac{(SSE_R - SSE_c)/(k-g)}{(SSE_c)(n-k-1)}$$

$$= 1.45506$$

$$F_{\alpha, k-g, n-k-1} = F_{0.05, 1, 5} = 6.61$$

F is not greater than $F_{0.05,1,5}$ $\longrightarrow$ we do not reject the null hypothesis that $\beta_2 = 0$.

(2) $\begin{cases} H_0: \beta_1 = \beta_2 = 0 \\ H_a: \text{at least one not zero/equal} \end{cases}$ $\alpha = 0.05$ $\longrightarrow$ ANOVA F-test

$$S_{YY} = \text{sum}\begin{pmatrix} (18.5 - 32.725)^2 \\ (22.6 - 32.725)^2 \\ (27.2 - 32.725)^2 \\ (31.2 - 32.725)^2 \\ (33.0 - 32.725)^2 \\ (44.9 - 32.725)^2 \\ (49.4 - 32.725)^2 \\ (35.0 - 32.725)^2 \end{pmatrix} = 769.255$$

From a previous problem:

$$SSE = S_{YY} - \frac{S_{XY}^2}{S_{YY}}$$

$$= 739.142$$

$$SSR = \frac{S_{XY}^2}{S_{XX}}$$

$$= 551.544$$

$$S_{XY} = \text{sum}\begin{pmatrix} (1996 - 1999.5)(18.5 - 32.725) \\ (1997 - 1999.5)(22.6 - 32.725) \\ (1998 - 1999.5)(27.2 - 32.725) \\ (1999 - 1999.5)(31.2 - 32.725) \\ (2000 - 1999.5)(33.0 - 32.725) \\ (2001 - 1999.5)(44.9 - 32.725) \\ (2002 - 1999.5)(49.4 - 32.725) \\ (2003 - 1999.5)(35.0 - 32.725) \end{pmatrix} = 152.2$$

$k = 2$

$n - k - 1 = 8 - 2 - 1 = 5$

$$F = \frac{SSR/k}{SSE/(n-k-1)}$$

$$= 1.86549$$

$$F_{\alpha, k-g, n-k-1} = F_{0.05, 2, 5} = 5.79$$

F is not greater than $F_{0.05,2,5}$ $\longrightarrow$ we do not reject the null hypothesis that $\beta_2 = 0$.