/75

STAT 2509 B
Assignment #3
SOLUTION

1. A researcher believes that the six-month certificate of deposit interest rate can be predicted from the price of gold (X_1, in dollars per troy ounce) and the money supply (X_2, in billions of dollars). A random sample of months was selected and the following data were obtained.

Deposit Interest rate (Y)	Price of Gold (X_1)	Money Supply (X_2)
8.17	410.11	800.66
6.50	363.33	833.72
4.13	338.50	953.69
3.22	329.35	1021.90
5.17	380.35	1143.66
5.73	386.23	1145.18
5.69	352.06	1067.52
6.07	283.41	1147.99
6.68	273.68	1089.66
2.66	441.76	1393.11
1.92	310.25	1187.34

Consider the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

[5] (a) State all assumptions which are necessary for the statistical inference.

Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, ***n* = 11** 1

Assumptions: **(i)** ***X₁, X₂*** **are observed without error** 1 1

(ii) ***y*'s (or ε's) are independently distributed with mean**

1/2 $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ **(or** $E(\varepsilon) = 0$**)**

(iii) variance of *y*'s (or ε's) is constant, 1/2 σ^2 **for all X_1, X_2**

(iv) ***y*** ~ $N(E(y), \sigma^2)$ **for any value of X_1, X_2 (or** $\varepsilon \sim N(0, \sigma^2)$ **for any value of X_1, X_2)** 1

[5] (b) Use matrices to compute the estimates of the population parameters β_0, β_1, β_2 and hence obtain the fitted least squares prediction line.

$$(X^TX)^{-1}X^TY=\begin{bmatrix}8.2231095529 & -0.012294277 & -0.003554459\\ -0.012294277 & 0.0000374872 & -0.0000008317912\\ -0.003554459 & -0.0000008317912 & 0.0000035909483\end{bmatrix}*\begin{bmatrix}55.94\\19672.7272\\57971.2158\end{bmatrix}=$$

$$=\begin{bmatrix}12.08248\\0.001514\\-0.00703\end{bmatrix}=\begin{bmatrix}\hat\beta_0\\\hat\beta_1\\\hat\beta_2\end{bmatrix}=\hat\beta$$

∴ the least squares fitted regression line is given by: $\hat{Y}=X\hat\beta$**, i.e.**

$\hat{y}$ **= 12.08248 + 0.001514** x_1 **-0.00703** x_2

(c) Set up the ANOVA table and hence test for the significance of the model. Use α = 0.05.

[20]

$$TSS=Y^TY-\frac{\left(\sum_{i=1}^n y_i\right)^2}{n}=\sum_{i=1}^n y_i^2-\frac{\left(\sum_{i=1}^n y_i\right)^2}{n}=320.5914-\frac{(55.94)^2}{11}$$ **= 320.5914 – 284.4803=**

= 36.1111

$$SSR=\hat\beta^T(X^TY)-\frac{\left(\sum_{i=1}^n y_i\right)^2}{n}=[12.08248\quad 0.001514\quad -0.00703]*\begin{bmatrix}55.94\\19672.7272\\57971.2158\end{bmatrix}$$

$-\frac{(55.94)^2}{11}=$ **298.2269 – 284.4803 = 13.74658**

$SSE=TSS-SSR=$ **22.36452**

$MSR=\frac{SSR}{k}=\frac{13.74658}{2}$ **= 6.873292**

$MSE=\frac{SSE}{n-(k+1)}=\frac{22.36452}{8}=$ **2.795565**

$F=\frac{MSR}{MSE}=$ **2.458642**

Source	*d.f.*	*SS*	*MS*	*F*
Regression	2	13.74658	6.873292	2.458642
Error	8	22.36452	2.795565	
Total	10	36.1111		

$H_0 : \beta_1 = \beta_2 = 0$ $\quad \alpha = 0.05$

H_a : *at least one of the* $\beta' s \neq 0$ $\quad$ 1

test-statistics: $F = \dfrac{MSR}{MSE} = \underline{\mathbf{2.458642}}$ 1

R.R: **we reject** H_0 **if** $F > F_{\alpha(k,n-(k+1))} = F_{0.05(2,8)} =$ **4.46** 1

1 1/2

Since F **= 2.458 $\ngtr$ 4.46, we do not reject** H_0 **and conclude that at 5% level of significance there is not enough evidence to say that a linear model is significant (i.e. we say that the model is not significant).** 1/2

[7] (d) Test whether x_1 term (i.e. whether the price of gold) contributes to the given model. Use t-test with α = 0.05.

$H_0 : \beta_1 = 0$ $\quad \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$H_a : \beta_1 \neq 0$ $\quad$ 1

test-statistics: $t = \dfrac{\hat{\beta}_1}{\sqrt{v_{11} MSE}} = \dfrac{0.001514}{\sqrt{(0.0000374872)(2.795565)}} = \underline{\mathbf{0.147894}}$ 1 1

1 (second diagonal element of $(X^TX)^{-1}$)

R.R: **we reject** H_0 **if** $t < -t_{\alpha/2;n-(k+1)} = -t_{0.025;8} =$ **- 2.306**

or $t > t_{\alpha/2;n-(k+1)} = t_{0.025;8} =$ **2.306** $\quad$ 1

1 1/2

Since t **= 0.147894 $\ngtr$ 2.306, we do not reject** H_0 **and conclude that at 5% level of significance there is not enough evidence to say that the** X_1 **term (i.e. the price of gold) contributes to the model.** 1/2

[6] (e) Find the values of the coefficient of determination, r^2, and the adjusted r^2 and interpret their meanings in this problem.

$r^2 = \dfrac{SSR}{TSS} = \dfrac{13.74658}{36.1111} =$ **0.380675** $\cong$ $\underline{\mathbf{38.06\%}}$ 1 1

i.e. approximately 38.06% of the total variation in the data is explained by the regr. line (and 61.94% is due to error). 1

$r^2_{adj} = 1 - \dfrac{SSE/n-(k+1)}{TSS/n-1} = 1 - \dfrac{MSE}{TSS/n-1} = 1 - \dfrac{2.795565}{36.1111/10} = 1 - 0.774157 = \mathbf{0.225843} \cong \underline{\mathbf{22.58\%}}$ 1 1

Since both r^2 and r^2_{adj} are very low (and r^2_{adj} is even lower than r^2) and given our results that the model is not significant and that X_1 term does not contribute to the model, we may conclude that the model is not appropriate, that perhaps higher-order terms are needed. (1)

(f) Run SAS to verify your above results and also use the SAS output to answer part (d) using partial F-test with α = 0.05.

[8]

$$\left.\begin{aligned} H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \end{aligned}\right\} (1) \qquad \alpha = 0.05$$

- **full model**: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ (1)
- **reduced model**: $y = \beta_0 + \beta_2 x_2 + \varepsilon$ (1)

SSR_f = 13.75681 (d.f. = 2)
SSE_f = 22.35427 (d.f. = 8)

SSR_r = 13.69563 (d.f. = 1)
SSE_r = 22.41544 (d.f. = 9)

from SAS output on pp. 5, 6

test-statistics : (1) $F_{part} = \dfrac{[SSR_f - SSR_r]/[df_{SSR_f} - df_{SSR_r}]}{SSE_f/df_{SSE_f}} = \dfrac{(13.75681-13.69563)/(2-1)}{22.35427/8} =$

$= \dfrac{0.06118/1}{2.79428} =$ (1) **0.021895**

or equivalently,

$$F_{drop} = \frac{[SSE_r - SSE_f]/[df_{SSE_r} - df_{SSE_f}]}{SSE_f/df_{SSE_f}} = \frac{(22.41544-22.35427)/(9-8)}{22.35427/8} =$$

$= \dfrac{0.06118/1}{2.79428} =$ **0.021895**

R.R: **we reject** H_0 **if** $F_{part} > F_{\alpha(1,8)} = F_{0.05(1,8)} =$ **5.32** (1)

Since F_{part} = 0.021895 (1) $\not>$ 5.32, we do not reject (1/2) H_0 and conclude that at 5% level of significance there is not enough evidence to say that the X_1 term (i.e. the price of gold) contributes to the model. (1/2)

See SAS output on pages 5 & 6 for results.

The REG Procedure
Model: MODEL1
Dependent Variable: deposit

Number of Observations Read	11
Number of Observations Used	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	13.75681	6.87840	2.46	0.1469
Error	8	22.35427	2.79428		
Corrected Total	10	36.11107			

SSR MSR SSE MSE F-test TSS

Root MSE	1.67161	R-Square	0.3810
Dependent Mean	5.08545	Adj R-Sq	0.2262
Coeff Var	32.87043		

r^2 r^2_{adj}

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12.08246	4.79351	2.52	0.0358
gold	1	0.00151	0.01023	0.15	0.8860
money	1	-0.00703	0.00317	-2.22	0.0573

$\hat{\beta}_0$ $\hat{\beta}_1$ $\hat{\beta}_2$ t-test

3

Full model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

Name, student #

The REG Procedure
Model: MODEL2
Dependent Variable: deposit

Number of Observations Read	11
Number of Observations Used	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
		SSRr			
Model	1	13.69563	13.69563	5.50	0.0437
Error	9	22.41544	2.49060		
Corrected Total	10	36.11107 SSEr			

Root MSE	1.57816	R-Square	0.3793
Dependent Mean	5.08545	Adj R-Sq	0.3103
Coeff Var	31.03292		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12.57909	3.23084	3.89	0.0037
money	1	-0.00699	0.00298	-2.34	0.0437

3

Reduced model: $y = \beta_0 + \beta_2 x_2 + \varepsilon$

Name, student #

```
Footnote 'Name, student #';
Data Interest;
Input deposit gold money @@;
Cards;
    8.17 410.11 800.66 6.50 363.33 833.72 4.13 338.50 953.69
    3.22 329.35 1021.90 5.17 380.35 1143.66 5.73 386.23 1145.18
    5.69 352.06 1067.52 6.07 283.41 1147.99 6.68 273.68 1089.66
    2.66 441.76 1393.11 1.92 310.25 1187.34
Run;
Proc Reg;
      Model deposit=gold money;
      Model deposit=money;
Run;
```

Q.1

2

```
Footnote 'Name, student #';
Data Drug;
Input dose X2 X3 potency;
   X1=log(dose);
   interact12=X1*X2;
   interact13=X1*X3;
Cards;
    0.2  0  0  2.0
    0.4  0  0  4.3
    0.8  0  0  6.5
    1.6  0  0  8.9
    0.2  1  0  1.8
    0.4  1  0  4.1
    0.8  1  0  4.9
    1.6  1  0  5.7
    0.2  0  1  1.3
    0.4  0  1  2.0
    0.8  0  1  2.8
    1.6  0  1  3.4
Run;
Proc Reg;
   Model potency=X1 X2 X3 interact12 interact13;
   Model potency=X1 X2 X3;
Run;
```

Q.2

2

2. Consider the following model:

[8]

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \varepsilon$$

where $x_2 = \begin{cases} 1, & \text{if drug } B \\ 0, & \text{otherwise} \end{cases}$ $\quad x_3 = \begin{cases} 1, & \text{if drug } C \\ 0, & \text{otherwise} \end{cases}$

$x_1 = \ln(dose)$ $\quad y$ = potency of drug

Run **SAS** to test whether the 3 lines are parallel, i.e. test whether the slopes of these 3 lines are the same. Use α = 0.05.

- **full model:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \varepsilon$ (1)

if drug A: $y = \beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(0) + \beta_4 x_1(0) + \beta_5 x_1(0) + \varepsilon$
or $y = \beta_0 + \beta_1 x_1 + \varepsilon$

if drug B: $y = \beta_0 + \beta_1 x_1 + \beta_2(1) + \beta_3(0) + \beta_4 x_1(1) + \beta_5 x_1(0) + \varepsilon$
or $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1 + \varepsilon$

if drug C: $y = \beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(1) + \beta_4 x_1(0) + \beta_5 x_1(1) + \varepsilon$
or $y = (\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1 + \varepsilon$

to test whether the 3 drug lines are parallel is the same as to test whether their slopes are the same, i.e. whether β_4 and β_5 = 0

$$\left.\begin{array}{l} H_0 : \beta_4 = \beta_5 = 0 \\ H_a : \text{at least one of the } \beta' s \neq 0 \end{array}\right\} (1) \qquad \alpha = 0.05$$

- **reduced model:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ (1)

***SSR$_f$* = 55.29350 (d.f. = 5)**
***SSE$_f$* = 0.68900 (d.f. = 6)**

***SSR$_r$* = 48.84417 (d.f. = 3)**
***SSE$_r$* = 7.13833 (d.f. = 8)**

from SAS output on pp. 10, 11

test-statistics :

$$F_{drop} = \frac{[SSE_r - SSE_f]/[df_{SSE_r} - df_{SSE_f}]}{SSE_f / df_{SSE_f}} = \frac{(7.13833 - 0.68900)/(8-6)}{0.68900/6} =$$

$$= \frac{3.224665}{0.114833} = \underline{\mathbf{28.08126}}$$

or equivalently,

$$F_{part} = \frac{[SSR_f - SSR_r]/[df_{SSR_f} - df_{SSR_r}]}{SSE_f / df_{SSE_f}} = \frac{(55.29350 - 48.84417)/(5-3)}{0.68900/6} =$$

$$= \frac{3.224665}{0.114833} = \underline{\mathbf{28.08126}}$$

R.R: **we reject** H_0 **if** $F_{drop} > F_{\alpha(2,6)} = F_{0.05(2,6)} =$ **5.14**

Since F_{drop} = 28.08126 > 5.14, we reject H_0 and conclude that at 5% level of significance there is an evidence to say that the slopes of the 3 drug lines are not parallel.

See SAS output on pages 10 & 11 for results.

The REG Procedure
Model: MODEL1
Dependent Variable: potency

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	55.29350 SSR_f	11.05870	96.30	<.0001
Error	6	0.68900 SSE_f	0.11483		
Corrected Total	11	55.98250			

Root MSE	0.33887	R-Square	0.9877
Dependent Mean	3.97500	Adj R-Sq	0.9774
Coeff Var	8.52505		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.30722	0.21029	34.75	<.0001
X1	1	3.30377	0.21864	15.11	<.0001
X2	1	-2.15481	0.29740	-7.25	0.0004
X3	1	-4.34865	0.29740	-14.62	<.0001
interact12	1	-1.50040	0.30920	-4.85	0.0028
interact13	1	-2.27946	0.30920	-7.37	0.0003

3

Full model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \varepsilon$

Name, student #

The REG Procedure
Model: MODEL2
Dependent Variable: potency

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	48.84417	16.28139	18.25	0.0006
Error	8	7.13833	0.89229		
Corrected Total	11	55.98250			

SSR_r SSE_r

Root MSE	0.94461	R-Square	0.8725
Dependent Mean	3.97500	Adj R-Sq	0.8247
Coeff Var	23.76382		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.58940	0.51309	12.84	<.0001
X1	1	2.04382	0.35187	5.81	0.0004
X2	1	-1.30000	0.66794	-1.95	0.0875
X3	1	-3.05000	0.66794	-4.57	0.0018

3

Reduced model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

Name, student #