STAT 420 – U3, G4
Fall 2013

Name ANSWERS

Version B

Exam 1

Page	Points
1	
2	
3	
4	
5	
6	
7	
8	
Total	

Be sure to show all your work; your partial credit might depend on it.

No credit will be given without supporting work.

The exam is closed book and closed notes.
You are allowed to use a calculator and one 8½" x 11" sheet (both sides) with notes.

Academic Integrity

The University statement on your obligation to maintain academic integrity is:

If you engage in an act of academic dishonesty, you become liable to severe disciplinary action. Such acts include cheating; falsification or invention of information or citation in an academic endeavor; helping or attempting to help others commit academic infractions; plagiarism; offering bribes, favors, or threats; academic interference; computer related infractions; and failure to comply with research regulations.

Rule 33 of the Code of Policies and Regulations Applying to All Students gives complete details of rules governing academic integrity for all students. You are responsible for knowing and abiding by these rules.

1. A substance used in biological and medical research is shipped to users in cartons of 1,000 ampules. The data, involving eight shipments, were collected on the number of times the carton was transferred from one transport to another (including the initial loading) over the shipment route (x) and the number of ampules found to be broken upon arrival (y). Assume the simple linear regression model is appropriate:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \text{where } \varepsilon_i\text{'s are i.i.d. } N(0, \sigma^2).$$

x	y
2	16
1	9
3	17
1	12
4	22
2	13
1	8
2	15

$\sum x = 16,\quad \sum y = 112,\quad \sum x^2 = 40,\quad \sum y^2 = 1712,\quad \sum xy = 256,$

$\sum (x - \bar{x})^2 = 8,\quad \sum (y - \bar{y})^2 = 144,\quad \sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x})\, y = 32.$

a) (8) Find the equation of the least-squares regression line.

$$\bar{x} = \frac{\sum x}{n} = \frac{16}{8} = 2. \qquad \bar{y} = \frac{\sum y}{n} = \frac{112}{8} = 14.$$

$$\hat{\beta}_1 = \frac{\text{SXY}}{\text{SXX}} = \frac{32}{8} = \mathbf{4}. \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 14 - 4 \cdot 2 = \mathbf{6}.$$

Least-squares regression line: $\boldsymbol{\hat{y} = 6 + 4x}$.

b) (4) Give an estimate for σ^2.

$$\text{SSRegr} = \hat{\beta}_1^2 \cdot \text{SXX} = 4^2 \cdot 8 = 128. \qquad \text{RSS} = \text{SYY} - \text{SSRegr} = 144 - 128 = 16.$$

$$s_e^2 = \frac{\text{RSS}}{n-2} = \frac{16}{6} \approx \mathbf{2.6667}. \qquad s_e \approx 1.633.$$

OR

$$\hat{\sigma}^2 = \frac{\text{RSS}}{n} = \frac{16}{8} = \mathbf{2}. \qquad \hat{\sigma} \approx 1.4142.$$

1. (continued)

c) (7) Test $H_0 : \beta_1 = 3$ vs. $H_1 : \beta_1 > 3$ at a 5% level of significance.

Test Statistic: $t = \dfrac{\hat{\beta}_1 - \beta_{10}}{s_e / \sqrt{\text{SXX}}} = \dfrac{4 - 3}{1.633 / \sqrt{8}} \approx \mathbf{1.732}.$

$n - 2 = \mathbf{6}$ df.

Critical Value: $t_{0.05}(6) = \mathbf{1.943}.$

$1.440 = t_{0.10}(6) < t < t_{0.05}(6) = 1.943$

P-value = Area to the **right** of t. **0.05 < p-value < 0.10**.

(p-value ≈ 0.067.)

Do NOT Reject $\mathbf{H_0}$ at $\alpha = 0.05$.

d) (6) Construct a 95% prediction interval for the number of broken ampules if a carton was transferred from one transport to another $x = 5$ times.

$$\hat{y} \pm t_{\alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\text{SXX}}}$$

$\hat{y} = 6 + 4 \cdot 5 = 26.$ $t_{0.025}(6) = 2.447.$

$$26 \pm 2.447 \cdot 1.633 \cdot \sqrt{1 + \frac{1}{8} + \frac{(5 - 2)^2}{8}} \qquad \mathbf{26 \pm 6}$$

1. (continued)

e) (7) Test $H_0 : \beta_0 = 3$ vs. $H_1 : \beta_0 \neq 3$ at the 5% level of significance.

Test Statistic: $$t = \frac{\hat{\beta}_0 - \beta_{00}}{s_e \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{\text{SXX}}}} = \frac{6 - 3}{\sqrt{\frac{16}{6}} \cdot \sqrt{\frac{1}{8} + \frac{(2)^2}{8}}} \approx \mathbf{2.3238}.$$

$n - 2 = \mathbf{6}$ df.

Critical Values: $\pm t_{0.025}(6) = \pm \mathbf{2.447}$.

$1.943 = t_{0.05}(6) < t < t_{0.025}(6) = 2.447$

P-value = Area of **two** tails.

0.025 < one tail < 0.05.

0.05 < p-value < 0.10.

(p-value ≈ 0.059.)

Do NOT Reject H_0 at $\alpha = 0.05$.

2. (13) Suppose the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

was fit to $n = 31$ data points.

```
> sum( lm( y ~ 1 )$residuals^2 )
[1]  270
> sum( lm( y ~ x2 + x4 )$residuals^2 )
[1]  240
> sum( lm( y ~ x1 + x3 + x5 )$residuals^2 )
[1]  180
> sum( lm( y ~ x1 + x2 + x3 + x4 + x5 )$residuals^2 )
[1]  150
```

Test $H_0 : \beta_2 = \beta_4 = 0$ at a 5% level of significance.

Full Model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

$\dim(V) = p = 6.$ $\qquad$ $\text{SSResid}_{\text{Full}} = 150.$

Null Model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_5 x_5 + \varepsilon$$

$\dim(V_0) = q = 4.$ $\qquad$ $\text{SSResid}_{\text{Null}} = 180.$

	SS	*DF*	*MS*	*F*
Diff.	$\text{SSResid}_{\text{Null}} - \text{SSResid}_{\text{Full}}$	$\dim(V) - \dim(V_0)$	…	…
Full	$\text{SSResid}_{\text{Full}}$	$n - \dim(V)$	…	
Null	$\text{SSResid}_{\text{Null}}$	$n - \dim(V_0)$		

	SS	*DF*	*MS*	*F*	
Diff.	30	2	15	**2.5**	← Test Statistic
Full	150	25	6		
Null	180	27			

Critical Value: $F_{0.05}(2, 25) = \mathbf{3.39}$.

$F = 2 < 3.39 = F_{0.05}(2, 25)$ Decision: **Do NOT Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

Null model is preferred.

3. At the Actuarial Science Meet the Firms, a student is waiting in line to talk to a Province Ranch Insurance recruiter. For each of $n = 10$ students waiting in line ahead of him, he notes the number of Actuarial Science exams passed by the student, whether or not the student has already had an internship, and the time the students spends talking to the recruiter. Let

y = time (in minutes),

x_1 = number of exams passed,

x_2 = internship (0 = no, 1 = yes).

x_1	x_2	y
3	0	5.1
4	1	5.6
3	0	3.1
2	1	6
1	0	3
1	1	3.7
1	0	1.5
2	0	3.8
2	1	4
1	1	3.2

Consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \qquad \text{where } \varepsilon\text{'s are i.i.d. } N(0, \sigma^2).$$

Then $$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 10 & 20 & 5 \\ 20 & 50 & 10 \\ 5 & 10 & 5 \end{bmatrix}, \qquad (\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 0.6 & -0.2 & -0.2 \\ -0.2 & 0.1 & 0 \\ -0.2 & 0 & 0.4 \end{bmatrix},$$

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 39 \\ 86 \\ 22.5 \end{bmatrix}, \quad \text{and} \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.7 \\ 0.8 \\ 1.2 \end{bmatrix}, \qquad \sum(y_i - \hat{y}_i)^2 = 6.5, \quad \text{and} \quad \sum(y_i - \overline{y})^2 = 16.5.$$

a) (14) Perform the significance of the regression test at the 10% level of significance.

$n = 10$, $\qquad p = (\text{\# of } \beta\text{'s}) = (\text{\# of columns of matrix } \mathbf{X}) = 3.$

Source	*SS*	*df*	*MS*	*F*
Regression	10.0	$p - 1 = 2$	5.0	**5.385**
Residual	6.5	$n - p = 7$	0.92857	
Total	16.5	$n - 1 = 9$		

$H_0 : \boldsymbol{\beta_1 = \beta_2 = 0}$ vs H_1 : at least one of β_1, β_2 is not zero.

Critical Value: $F_{0.10}(2, 7) = \mathbf{3.26}$. $\qquad$ Decision: **Reject H_0**.

(p-value ≈ 0.03837.)

3. (continued)

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.7 \\ 0.8 \\ 1.2 \end{bmatrix}, \qquad (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 0.6 & -0.2 & -0.2 \\ -0.2 & 0.1 & 0 \\ -0.2 & 0 & 0.4 \end{bmatrix}.$$

b) (7) Test $\mathrm{H}_0 : \beta_1 = 1.5$ vs. $\mathrm{H}_1 : \beta_1 < 1.5$ at the 10% level of significance.

$\hat{\mathrm{Var}}(\hat{\beta}_1) = \mathrm{C}_{11} \times s^2 = 0.1 \times 0.92857 = 0.092857.$

Test Statistic: $\mathrm{T} = \dfrac{0.8 - 1.5}{\sqrt{0.092857}} = -2.297.$

Critical Values: $-t_{0.10}(7\text{ df}) = -1.415.$

Reject H$_0$ at $\alpha = 0.10$.

p-value = left tail. $0.025 < \text{p-value} < 0.05.$

(p-value ≈ 0.0276.)

c) (7) Test $\mathrm{H}_0 : \beta_2 = 0$ vs. $\mathrm{H}_1 : \beta_2 \neq 0$ at the 10% level of significance.

$\hat{\mathrm{Var}}(\hat{\beta}_2) = \mathrm{C}_{22} \times s^2 = 0.4 \times 0.92857 = 0.37143.$

Test Statistic: $\mathrm{T} = \dfrac{1.2 - 0}{\sqrt{0.37143}} = 1.969.$

Critical Values: $\pm t_{0.05}(7\text{ df}) = \pm 1.895.$

Reject H$_0$ at $\alpha = 0.10$.

p-value = 2 tails. $0.025 < 1\text{ tial} < 0.05.$

$0.05 < \text{p-value} < 0.10.$

(p-value ≈ 0.0896.)

3. (continued)

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.7 \\ 0.8 \\ 1.2 \end{bmatrix}, \qquad (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 0.6 & -0.2 & -0.2 \\ -0.2 & 0.1 & 0 \\ -0.2 & 0 & 0.4 \end{bmatrix}.$$

d) (7) Construct a 95% prediction interval for the time to talk to the recruiter for a student who passed 3 exams and has had an internship (that is, $x_2 = 1$).

$$\mathrm{X}_0^{\mathrm{T}} = [\ 1 \quad 3 \quad 1\ ] \qquad \hat{\mathrm{Y}}_0 = 1 \times 1.7 + 3 \times 0.8 + 1 \times 1.2 = 5.3.$$

$$\mathrm{X}_0^{\mathrm{T}}\,\mathbf{C}\,\mathrm{X}_0 = [1 \quad 3 \quad 1] \cdot \begin{bmatrix} 0.6 & -0.2 & -0.2 \\ -0.2 & 0.1 & 0 \\ -0.2 & 0 & 0.4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = [1 \quad 3 \quad 1] \cdot \begin{bmatrix} -0.2 \\ 0.1 \\ 0.2 \end{bmatrix} = 0.3.$$

$$[\,1 + \mathrm{X}_0^{\mathrm{T}}\,\mathbf{C}\,\mathrm{X}_0\,]\,s^2 = (1 + 0.3) \times 0.92857 = 1.207.$$

$$t_{0.025}(7) = 2.365. \qquad 5.3 \pm 2.365 \times \sqrt{1.207} \qquad \mathbf{5.3 \pm 2.6}$$

e) (4) What proportion of the observed variation in the time spent talking to the recruiter has been explained by the linear relationship with the number of exams passed and whether or not a student has had an internship?

$$\mathrm{R}^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{SYY}} = 1 - \frac{6.5}{16.5} \approx \mathbf{0.60606}.$$

4. At a local grocery shop, the weekly sales of eggs (X) and weekly sales of bacon (Y) follow a bivariate normal distribution with

$$\mu_X = \$250, \quad \sigma_X = \$50, \quad \mu_Y = \$200, \quad \sigma_Y = \$20, \quad \rho = 0.60.$$

a) (8) What is the probability that the weekly sales of eggs are below \$350, given that the weekly sales of bacon are \$240? That is, find $P(X < 350 \mid Y = 240)$.

Given $Y = 240$, X has Normal distribution

with mean $\mu_X + \rho \dfrac{\sigma_X}{\sigma_Y}(y - \mu_Y) = 250 + 0.60 \cdot \dfrac{50}{20} \cdot (240 - 200) = 310$

and variance $\left(1 - \rho^2\right) \cdot \sigma_X^2 = \left(1 - 0.6^2\right) \cdot 50^2 = 1600$

(standard deviation = 40).

$$P(X < 350 \mid Y = 240) = P\left(Z < \frac{350 - 310}{40}\right) = P(Z < 1.00) = \mathbf{0.8413}.$$

b) (8) What is the probability that the weekly sales of eggs and bacon exceed \$500? That is, find $P(X + Y > 500)$.

$X + Y$ has Normal distribution,

$E(X + Y) = \mu_X + \mu_Y = 250 + 200 = 450,$

$\mathrm{Var}(X + Y) = \sigma_X^2 + 2\sigma_{XY} + \sigma_Y^2 = \sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2$

$= 50^2 + 2 \cdot 0.6 \cdot 50 \cdot 20 + 20^2 = 4100.$

$SD(X + Y) \approx 64.03.$

$$P(X + Y > 500) = P\left(Z > \frac{500 - 450}{\sqrt{4100}}\right) = P(Z > 0.78) = \mathbf{0.2177}.$$