STAT 420
Spring 2010

Name ANSWERS

Version A

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 with notes on it.

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. An instructor wishing to cut down on cheating makes three different exams and distributes them randomly to his students. After collecting the exams, he grades them. The instructor would like to know whether the three exams are equally difficult. He will decide this by investigating whether the scores have equal population means.

Version	Scores				sample mean	sample variance
A	47	53	55	65	55	56
B	43	48	52	61	51	58
C	31	38	46	49	41	66

a) Test $H_0: \mu_A = \mu_B = \mu_C$ at $\alpha = 0.10$ using the ANOVA F test.

$N = n_A + n_B + n_C = 4 + 4 + 4 = 12.$ $\bar{y} = \dfrac{4 \cdot 55 + 4 \cdot 51 + 4 \cdot 41}{12} = 49.$

$\text{SSB} = 4 \cdot (55 - 49)^2 + 4 \cdot (51 - 49)^2 + 4 \cdot (41 - 49)^2 = 416.$

$\text{MSB} = \dfrac{\text{SSB}}{J - 1} = \dfrac{416}{2} = 208.$

$\text{SSW} = 3 \cdot 56 + 3 \cdot 58 + 3 \cdot 66 = 540.$ $\text{MSW} = \dfrac{\text{SSW}}{N - J} = \dfrac{540}{9} = 60.$

$\text{SSTot} = \text{SSB} + \text{SSW} = 416 + 540 = 956.$

$\text{F} = \dfrac{\text{MSB}}{\text{MSW}} = \dfrac{208}{60} \approx \mathbf{3.46667}.$

ANOVA table:

Source	**SS**	**DF**	**MS**	**F**
Between	416	2	208	3.46667
Within	540	9	60	
Total	956	11		

$\text{F}_{0.10}(2, 9) = \mathbf{3.01}.$ **Reject H_0** at $\alpha = 0.10$.

$0.05 < \text{p-value} < 0.10$ (p-value ≈ 0.0765)

b) Use a 90% confidence level and Scheffé's multiple comparison procedure to compare the averages for versions A and B with that for version C.

$$\sum_{j=1}^{J} c_j \, \overline{y}_j \pm \sqrt{F_{\alpha}(J-1, N-J)} \cdot \sqrt{\text{MSW}} \cdot \sqrt{(J-1) \cdot \sum_{j=1}^{J} \frac{c_j^2}{n_j}}$$

$c_{\text{A}} = ½$, $c_{\text{B}} = ½$, $c_{\text{C}} = -1$. $\text{F}_{0.10}(2, 9) = 3.01$.

$$\left(\frac{55+51}{2} - 41\right) \pm \sqrt{3.01} \cdot \sqrt{60} \cdot \sqrt{2 \cdot \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{4}\right)}$$

12 ± 11.6383

c) Use a 95% confidence level and Tukey's pairwise comparison procedure to compare the average for versions A with that for version C.

$$\left(\overline{\text{Y}}_i - \overline{\text{Y}}_j\right) \pm \frac{q_{\gamma, J, N-J}}{\sqrt{2}} \cdot \sqrt{\text{MSW}} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

$q_{0.05,3,9} = 3.95$.

$$55 - 41 \pm \frac{3.95}{\sqrt{2}} \sqrt{60} \sqrt{\frac{1}{4} + \frac{1}{4}}$$

14 ± 15.2983

2. An instructor wishing to cut down on cheating makes three different exams and distributes them randomly to his students. After collecting the exams, he grades them. The instructor would like to know whether the three exams are equally difficult. He will decide this by investigating whether the scores have equal population means.

Version A	47	53	55	65
Version B	43	48	52	61
Version C	31	38	46	49

Test $H_0: \mu_A = \mu_B = \mu_C$ at $\alpha = 0.05$ using the Kruskal-Wallis test.

C	C	B	C	A	B	C	B	A	A	B	A
31	38	43	46	47	48	49	52	53	55	61	65
1	2	3	4	5	6	7	8	9	10	11	12

$\bar{r}_A = 9$ $\bar{r}_B = 7$ $\bar{r}_C = 3.5$ $\bar{r} = 6.5$

Test Statistic:

$$K = \frac{12}{12 \cdot 13}\left[4 \cdot (9-6.5)^2 + 4 \cdot (7-6.5)^2 + 4 \cdot (3.5-6.5)^2\right] = \mathbf{4.76923}.$$

Critical Value:

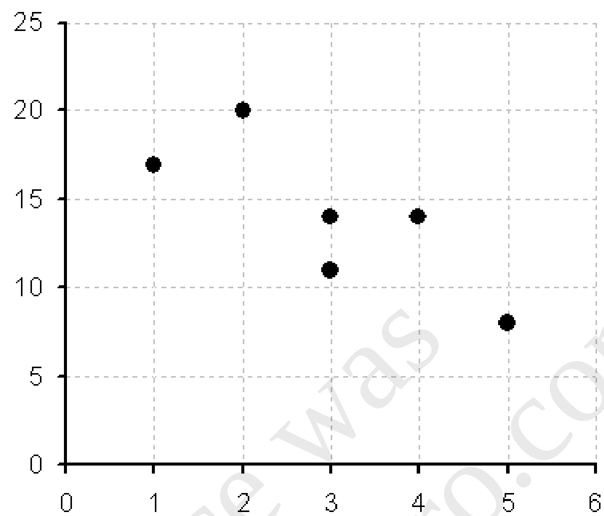
$$\chi^2_\alpha(J-1) = \chi^2_{0.05}(2) = \mathbf{5.991}.$$

K < 5.991. **Do NOT Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

$0.05 < \text{p-value} < 0.10$ $(\text{p-value} \approx 0.092)$

3. We wish to examine the relationship between the age of a vehicle in years (x) and the selling price (y) (in thousands of \$) for a particular brand of minivan at *Honest Harry's Used Car Dealership*. The data are as follows:

x	y
1	17
2	20
3	11
3	14
4	14
5	8



$\sum x = 18$, $\sum y = 84$, $\sum x^2 = 64$, $\sum y^2 = 1266$, $\sum x y = 228$,

$\sum (x - \bar{x})^2 = 10$, $\sum (y - \bar{y})^2 = 90$, $\sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x})\, y = -24$.

Consider the model $Y_i = \alpha + \beta x_i + \varepsilon_i$, where ε_i's are i.i.d. $N(0, \sigma^2)$.

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

$$\hat{\beta} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{-24}{10} = \mathbf{-2.4}.$$

OR

$$\hat{\beta} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} = \frac{6\cdot 228 - 18\cdot 84}{6\cdot 64 - 18^2} = \frac{-144}{60} = \mathbf{-2.4}.$$

$$\bar{x} = \frac{\sum x}{n} = \frac{18}{6} = 3. \qquad \bar{y} = \frac{\sum y}{n} = \frac{84}{6} = 14.$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x} = 14 + 2.4\cdot 3 = \mathbf{21.2}.$$

The least-squares regression line: $\hat{y} = \mathbf{21.2 - 2.4 \cdot x}$.

b) Give an estimate for σ^2, the variance of the observations about the true regression line.

$$\text{SSRegr} = \hat{\beta}^2\,\text{SXX} = (-2.4)^2 \cdot 10 = 57.6.$$

$$\text{SSResid} = \text{SYY} - \text{SSRegr} = 90 - 57.6 = 32.4.$$

OR

x	y	$\hat{y}$	e	e^2
1	17	18.8	– 1.8	3.24
2	20	16.4	3.6	12.96
3	11	14	– 3	9
3	14	14	0	0
4	14	11.6	2.4	5.76
5	8	9.2	– 1.2	1.44

$$\Rightarrow \quad \sum (y - \hat{y})^2 = 32.4.$$

$$s_e^2 = \frac{1}{n-2}\sum\left(y-\hat{y}\right)^2 = \frac{32.4}{4} = \mathbf{8.1}. \qquad s_e = \sqrt{8.1} \approx 2.846.$$

OR

$$\hat{\sigma}^2 = \frac{1}{n}\sum\left(y-\hat{y}\right)^2 = \frac{32.4}{6} = \mathbf{5.4}. \qquad \hat{\sigma} = \sqrt{5.4} \approx 2.3238.$$

c) Test for the significance of the regression at a 5% level of significance.

$H_0: \boldsymbol{\beta = 0}$ vs. $H_1: \boldsymbol{\beta \neq 0}$

Test Statistic:

$$T = \frac{\hat{\beta} - \beta_0}{s_e \big/ \sqrt{\sum\left(x_i - \bar{x}\right)^2}} = \frac{-2.4 - 0}{\sqrt{8.1} \big/ \sqrt{10}} = \mathbf{-2.667}.$$

Rejection Region:

Reject H_0 if $T < -t_{0.025}(6-2=4 \text{ df})$ or $T > t_{0.025}(4 \text{ df})$

$\pm t_{0.025}(4 \text{ df}) = \pm \mathbf{2.776}$ – Critical Values.

Do NOT Reject $\mathbf{H_0}$ $0.05 <$ p-value < 0.10 (p-value = 0.056)

OR

ANOVA table:

Source	SS	DF	MS	F
Regression	57.6	1	57.6	**7.11111**
Residuals	32.4	$n-2=4$	8.1	
Total	90	$n-1=5$		

Rejection Region: Reject H_0 if $F > F_{\alpha}(1, n-2)$

Reject H_0 if $F > F_{0.05}(1, 4)$ $F_{0.05}(1, 4) = \mathbf{7.71}$.

Do NOT Reject $\mathbf{H_0}$ $0.05 <$ p-value < 0.10 (p-value = 0.056)

d) Test $H_0: \alpha = 25$ vs. $H_1: \alpha < 25$ at a 10% level of significance.

Test Statistic:

$$T = \frac{\hat{\alpha} - \alpha_0}{s_e \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{\sum(x_i - \bar{x})^2}}} = \frac{21.2 - 25}{\sqrt{8.1}\sqrt{\frac{1}{6} + \frac{3^2}{10}}} = \mathbf{-1.293}.$$

Rejection Region:

Reject H_0 if $T < -t_{0.10}(6-2=4 \text{ df})$

$-t_{0.10}(4 \text{ df}) = \mathbf{-1.533}$ – Critical Value.

Do NOT Reject $\mathbf{H_0}$ $0.10 <$ p-value < 0.25 (p-value ≈ 0.133)

e) Construct a 90% prediction interval for the selling price of a minivan that is $x = 5$ years old.

$x = 5$ $\quad\quad$ $\hat{y} = 21.2 - 2.4 \cdot x = 21.2 - 2.4 \cdot 5 = 9.2$

$$\hat{y} \pm t_{\gamma/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

$t_{0.05}(6 - 2 = 4 \text{ df}) = 2.132$

$$9.2 \pm 2.132 \cdot \sqrt{8.1} \cdot \sqrt{1 + \frac{1}{6} + \frac{(5-3)^2}{10}} \qquad \mathbf{9.2 \pm 7.595}$$

f) What proportion of observed variation in selling price is explained by a straight-line relationship with the age of a vehicle?

Need $R^2 = ?$ (coefficient of determination)

$$R^2 = \frac{\text{SSRegr}}{\text{SYY}} = \frac{57.6}{90} = 1 - \frac{\text{SSResid}}{\text{SYY}} = 1 - \frac{32.4}{90} = \mathbf{0.64}. \qquad \mathbf{64\%}.$$

4. A agricultural experiment designed to assess differences in yields of corn for four different varieties, using three different fertilizers, produced the results (in bushels per acre) shown in the table below.

Fertilizer	Corn Variety 1	2	3	4	$\overline{Y}_{i\bullet}$
1	86	87	78	81	83
2	91	93	83	89	89
3	78	84	82	76	80
$\overline{Y}_{\bullet j}$	85	88	81	82	84

$Y_{ij} = \mu + \text{Fert}_i + \text{Corn}_j + \varepsilon_{ij}$, $\quad i = 1, 2, 3, \quad j = 1, 2, 3, 4.$

ε_{ij} are independent $N(0, \sigma^2)$ random variables,

$\text{Fert}_1 + \text{Fert}_2 + \text{Fert}_3 = 0, \qquad \text{Corn}_1 + \text{Corn}_2 + \text{Corn}_3 + \text{Corn}_4 = 0.$

a) Complete the ANOVA table.

$$\text{SSA} = J \sum_{i=1}^{I} \left(\overline{y}_{i\bullet} - \overline{y}_{\bullet\bullet}\right)^2 = 4 \cdot [(83-84)^2 + (89-84)^2 + (80-84)^2] = 168.$$

$$\text{SSB} = I \sum_{j=1}^{J} \left(\overline{y}_{\bullet j} - \overline{y}_{\bullet\bullet}\right)^2 = 3 \cdot [(85-84)^2 + (88-84)^2 + (81-84)^2 + (82-84)^2] = 90.$$

SSResid = SSTotal – SSA – SSB = 318 – 168 – 90 = 60.

ANOVA table:

Source	SS	DF	MS	F
Row (Fertilizer)	**168**	$I - 1 = \mathbf{2}$	**84**	**8.4**
Column (Corn)	**90**	$J - 1 = \mathbf{3}$	**30**	**3**
Residuals	**60**	$(I-1)(J-1) = \mathbf{6}$	**10**	
Total	318	$IJ - 1 = \mathbf{11}$		

b) Test for differences in corn varieties. Use a 5% level of significance.

H_0: $\text{Corn}_1 = \text{Corn}_2 = \text{Corn}_3 = \text{Corn}_4 = 0$

Critical Value: $F_{0.05}(3, 6) = \mathbf{4.76}$.

F = 3 < 4.76. Decision: **Do NOT Reject H_0**.