STAT 420 Fall 2012

Homework #8

(due Friday, November 2, by 3:00 p.m.)

1. Do NOT use a computer for this problem.

The data below represent the attendance for STAT 408, STAT 420 – N1, and STAT 420 – D1 for a random sample of 5 days for each class during Spring 2012 semester.

Section	Attendance					$\overline{y}_j$	s_j^2
408	39	40	47	49	50	45	26.5
420 – N1	49	53	56	57	60	55	17.5
420 – D1	48	49	53	55	60	53	23.5

a) Test $H_0: \mu_1 = \mu_2 = \mu_3$ at $\alpha = 0.10$ using the ANOVA F test. Construct an ANOVA table and state your conclusion (Reject H_0 or Do NOT Reject H_0). What is the p-value for this test (you may give a range) ?

$$N = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15. \qquad \overline{y} = \frac{5 \cdot 45 + 5 \cdot 55 + 5 \cdot 53}{15} = 51.$$

$$\text{SSB} = 5 \cdot (45 - 51)^2 + 5 \cdot (55 - 51)^2 + 5 \cdot (53 - 51)^2 = 280.$$

$$\text{MSB} = \frac{\text{SSB}}{J - 1} = \frac{280}{2} = 140.$$

$$\text{SSW} = 4 \cdot 26.5 + 4 \cdot 17.5 + 4 \cdot 23.5 = 270.$$

$$\text{MSW} = \frac{\text{SSW}}{N - J} = \frac{270}{12} = 22.5.$$

$$\text{SSTot} = \text{SSB} + \text{SSW} = 280 + 270 = 550.$$

$$\text{F} = \frac{\text{MSB}}{\text{MSW}} = \frac{140}{22.5} \approx \mathbf{6.2222}.$$

ANOVA table:

Source	SS	DF	MS	F
Between	280	2	140	**6.2222**
Within	270	12	22.5	
Total	550	14		

$F_{0.10}(2, 12) = 2.81.$ **Reject $\mathbf{H_0}$** at $\alpha = 0.10.$

0.01 $<$ p-value $<$ **0.05** (p-value ≈ 0.014)

b) Use a 90% confidence level and Scheffé's multiple comparison procedure to compare the average attendance for both sections of STAT 420 versus the average attendance for STAT 408. (Construct a 90% interval for an appropriate contrast.)

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

$c_1 = -1,\qquad c_2 = \tfrac{1}{2},\qquad c_3 = \tfrac{1}{2}.\qquad\qquad F_{0.10}(2, 12) = 2.81.$

$$\left(\frac{55+53}{2} - 45\right) \pm \sqrt{2.81} \cdot \sqrt{22.5} \cdot \sqrt{2 \cdot \left(\frac{1}{5} + \frac{1}{20} + \frac{1}{20}\right)} \qquad \mathbf{9 \pm 6.159}$$

c) Test $H_0: \mu_1 = \mu_2 = \mu_3$ at $\alpha = 0.10$ using the Kruskal-Wallis test. What is the p-value for this test (you may give a range)?

Section	Attendance											$\overline{y}_j$	s_j^2
408	39	40	47		49	50						45	26.5
ranks	1	2	3		6	8							
420 – N1					49		53		56	57	60	55	17.5
ranks					6		9.5		12	13	14.5		
420 – D1				48	49		53	55			60	53	23.5
ranks				4	6		9.5	11			14.5		

$\bar{r}_1 = \frac{1+2+3+6+8}{5} = 4.$ $\bar{r}_2 = \frac{6+9.5+12+13+14.5}{5} = 11.$

$\bar{r}_3 = \frac{4+6+9.5+11+14.5}{5} = 9.$ $\bar{r} = \frac{N+1}{2} = 8.$

Test Statistic:

$$K = \frac{12}{15 \cdot 16}\left[5 \cdot (4-8)^2 + 5 \cdot (11-8)^2 + 5 \cdot (9-8)^2\right] = \mathbf{6.5}.$$

Critical Value:

$$\chi^2_{\alpha}(J-1) = \chi^2_{0.10}(2) = 4.605.$$

$K > 4.605.$ **Reject $\mathbf{H_0}$** at $\alpha = 0.10.$

$\mathbf{0.025} <$ p-value $< \mathbf{0.05}$ (p-value ≈ 0.039)

2. System Design Inc. produces electronic modules and is keeping records on the hours to failure, Y, of units tested. In addition, for each module tested, data are kept on the frequency of exposure to power surges of more than 8 volts, X1, the power level at which each part is used in the machine, X2, and the number of quality control steps during manufacture, X3. The records for 14 electronic modules tested are given below. Consider the model

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \beta_3 X3_i + \varepsilon_i, \qquad i = 1, 2, \dots, 14,$$

where ε_i's are i.i.d. $N(0, \sigma^2)$.

```
> hourstofailure
     Y  X1    X2  X3
1   55   8   700  13
2   65   8  1100  17
3   90   4  1600  18
4   40   9   400  10
5   75   5  1300  19
6   60   7   800  11
7  100   3  1500  23
8   80   5  1400  20
9   45   8   500  14
10  95   3  1700  24
11  50   9   600  12
12  70   6   900  15
13  65   6  1000  16
14  85   4  1800  21
```

R command `drop1` was applied, and the following results were obtained:

```
> drop1( lm( Y ~ X1 + X2 + X3 ) )
Single term deletions

Model:
Y ~ X1 + X2 + X3
       Df  Sum of Sq     RSS    AIC

<none>                 216.67  46.35

X1      1     188.69   405.36  53.12

X2      1      61.85   278.52  47.87

X3      1       6.85   223.51  44.79 *
```

a) Fill in the missing AIC values. If the AIC model selection criteria is used, can the model be improved. If so, how? ***Justify your answer***. *No credit will be given without proper justification.*

<none> None of the variables have been dropped. `Y ~ X1 + X2 + X3`

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$p = 4$.

$$\text{AIC} = 14 + 14 \ln(2\pi) + 14 \ln\left({}^{216.67}/_{14}\right) + 2 \times 4 = \mathbf{86.08073}.$$

OR

$$\text{AIC} = 14 \ln\left({}^{216.67}/_{14}\right) + 2 \times 4 = \mathbf{46.35045}.$$

X1 X1 has been dropped. Y ~ X2 + X3

$$Y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$p = 3.$

$$\text{AIC} = 14 + 14 \ln(2\pi) + 14 \ln\left(405.36/14\right) + 2 \times 3 = \mathbf{92.85033}.$$

OR $$\text{AIC} = 14 \ln\left(405.36/14\right) + 2 \times 3 = \mathbf{53.12006}.$$

X2 X2 has been dropped. Y ~ X1 + X3

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

$p = 3.$

$$\text{AIC} = 14 + 14 \ln(2\pi) + 14 \ln\left(278.52/14\right) + 2 \times 3 = \mathbf{87.59633}.$$

OR $$\text{AIC} = 14 \ln\left(278.52/14\right) + 2 \times 3 = \mathbf{47.86606}.$$

X3 X3 has been dropped. Y ~ X1 + X2

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$p = 3.$

$$\text{AIC} = 14 + 14 \ln(2\pi) + 14 \ln\left(223.51/14\right) + 2 \times 3 = \mathbf{84.51586}.$$

OR $$\text{AIC} = 14 \ln\left(223.51/14\right) + 2 \times 3 = \mathbf{44.78558}.$$

Want a model with **lowest** AIC value. Therefore, we can improve the model by **dropping X3**: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$

b) Test $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ at $\alpha = 0.10$.
What is the p-value for this test (you may give a range) ?

Full model: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$

$\dim(V) = 4.$ $SSResid_{full} = 216.67.$

Null model: $Y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$

$\dim(V_0) = 3.$ $\text{SSResid}_{\text{full}} = 405.36.$

	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.	188.69	1	188.69	**8.7086**
Full	216.67	10	21.667	
Null	405.36	11		

← Test Statistic

Critical Value: $F_{0.10}(1, 10) = 3.29.$ Decision: **Reject H_0.**

0.01 < p-value < **0.05** (p-value ≈ 0.0145)

c) Test $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$ at $\alpha = 0.05.$
What is the p-value for this test (you may give a range) ?

Full model: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$

$\dim(V) = 4.$ $\text{SSResid}_{\text{full}} = 216.67.$

Null model: $Y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon.$

$\dim(V_0) = 3.$ $\text{SSResid}_{\text{full}} = 278.52.$

	SS	DF	MS	F
Diff.	$\text{SSResid}_{\text{null}} - \text{SSResid}_{\text{full}}$	$\dim(\text{V}) - \dim(\text{V}_0)$	…	…
Full	$\text{SSResid}_{\text{full}}$	$n - \dim(\text{V})$	…	
Null	$\text{SSResid}_{\text{null}}$	$n - \dim(\text{V}_0)$		

	SS	DF	MS	F
Diff.	61.85	1	61.85	**2.8546** ← Test Statistic
Full	216.67	10	21.667	
Null	278.52	11		

Critical Value: $F_{0.05}(1, 10) = \mathbf{4.96}$. Decision: **Do NOT Reject H_0**.

p-value > 0.10 (p-value ≈ 0.122)

d) $\sum (y - \overline{y})^2 = 4573.214$. Find Adjusted R^2 for the full model.

$$\text{Multiple } R\text{-squared} = 1 - \frac{\text{SSResidual}}{\text{SSTotal}} = 1 - \frac{216.67}{4573.214} \approx 0.9526.$$

$$\text{Adjusted } R\text{-squared} = 1 - \frac{n-1}{n-p} \cdot \left(1 - R^2\right) = 1 - \frac{13}{10} \cdot (1 - 0.9526) \approx \mathbf{0.9384}.$$

3. Do NOT use a computer for this problem.

Each of three cars is driven with each of four different brands of gasoline. The number of miles per gallon driven for each of the $IJ = (3)(4) = 12$ different combinations is recorded in the table below.

Car	Gasoline 1	2	3	4	$\overline{Y}_{i\bullet}$
1	31	32	23	26	28
2	36	38	28	34	34
3	23	29	27	21	25
$\overline{Y}_{\bullet j}$	30	33	26	27	29

$\mathrm{Y}_{ij} = \mu + \mathrm{Car}_i + \mathrm{Gas}_j + \varepsilon_{ij}$, $\quad i = 1, 2, 3, \quad j = 1, 2, 3, 4.$

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

$\mathrm{Car}_1 + \mathrm{Car}_2 + \mathrm{Car}_3 = 0,$ $\quad \mathrm{Gas}_1 + \mathrm{Gas}_2 + \mathrm{Gas}_3 + \mathrm{Gas}_4 = 0.$

a) Complete the ANOVA table.

$$\mathrm{SSA} = J \sum_{i=1}^{I} \left(\overline{y}_{i\bullet} - \overline{y}_{\bullet\bullet}\right)^2 = 4 \cdot [(28-29)^2 + (34-29)^2 + (25-29)^2] = 168.$$

$$\mathrm{SSB} = I \sum_{j=1}^{J} \left(\overline{y}_{\bullet j} - \overline{y}_{\bullet\bullet}\right)^2 = 3 \cdot [(30-29)^2 + (33-29)^2 + (26-29)^2 + (27-29)^2] = 90.$$

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

ANOVA table:

Source	**SS**	**DF**	**MS**	**F**
Row (Car)	**168**	$I-1=\mathbf{2}$	**84**	**8.4**
Column (Gas)	**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 cars. Use a 5% level of significance.

H_0: $Car_1 = Car_2 = Car_3 = 0$

Critical Value: $F_{0.05}(2, 6) = 5.14$.

$F = 8.4 > 5.14$. Decision: **Reject H_0.**

c) Test for differences in brands of gasoline. Use a 5% level of significance.

H_0: $Gas_1 = Gas_2 = Gas_3 = Gas_4 = 0$

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

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