ST512

Fall Quarter, 2002

Second Midterm Exam

Name:______________________________________________________________

Directions: Answer questions as directed. Not that much computation is involved, but show work where appropriate. For true/false questions, circle either true or false.
1. An experiment looks at the amylase specific activity \((y)\) of sprouted maize under 32 treatment conditions that are factorial combinations of three factors:

- A: analysis temperature (8 levels),
- B: growth temperature (2 levels)
- C: variety (2 cultivars).

There were a total of \(N = 96\) experimental units randomized to these 32 treatment combinations. Use \(\alpha = 0.05\) in the following questions.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>(F)</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7</td>
<td>327812</td>
<td>46830</td>
<td>72.94</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>B</td>
<td>1155</td>
<td>1155</td>
<td>1.80</td>
<td>0.1846</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>63808</td>
<td>63808</td>
<td>99.38</td>
<td>&lt; .0001</td>
<td></td>
</tr>
<tr>
<td>A*B</td>
<td>7157</td>
<td>1022</td>
<td>1.59</td>
<td>0.1538</td>
<td></td>
</tr>
<tr>
<td>A*C</td>
<td>1174</td>
<td>168</td>
<td>0.26</td>
<td>0.9666</td>
<td></td>
</tr>
<tr>
<td>B*C</td>
<td>10648</td>
<td>10648</td>
<td>16.58</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>A<em>B</em>C</td>
<td>6258</td>
<td>894</td>
<td>1.39</td>
<td>0.2240</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>41091</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>459101</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) True/False: The third-order interaction \(ABC\) is not significant.

(b) True/False: The second-order interaction involving factors \(A\) and \(B\) is significant.

(c) True/False: There is no evidence of any association between mean amylase specific activity and growth temperature.

(d) True/False: The effect of factor \(A\), analysis temperature, does not depend significantly on either factor \(B\) or factor \(C\).

(e) True/False: In the ANOVA table above, \(MSE = 41091/64 = 642\).
2. A study was performed to look at systolic blood pressure (BP) in three different diet groups:

- strict vegetarians (SV)
- lactovegetarians (LV) (eat dairy products)
- normal - (standard American diet)

Samples of size $n = 5$ from each gender were taken from each dietary group and BP was measured. Total sample size is $N = 30$. Let $y_{ijk}$ denote the BP for subject $k$ in diet group $i$ and gender $j$. Use the SAS output at the end of the problem to answer the questions below:

(a) Estimate the standard deviation of BPs among people in a given dietary-gender classification group, assuming homoscedasticity across the 6 groups.

(b) True/false: The $MS[E]$ is the average of the sample variances for the 6 treatment combinations:

$$MS[E] = \frac{1}{6}(s_{11}^2 + s_{12}^2 + s_{21}^2 + s_{22}^2 + s_{31}^2 + s_{32}^2)$$

where $s_{ij}^2 = \frac{1}{5}\sum_{k=1}^{5}(y_{ijk} - \bar{y}_{ij+})^2$.

(c) Test for a diet×gender interaction effect at level $\alpha = 0.05$.

(d) Report a $p$-value for the null hypothesis that the average BPs in the three diet populations are equal.

(e) Estimate the difference between mean BP for men and women.
(f) True/False: There is significant ($\alpha = 0.05$) evidence to suggest that population differences among diet groups are different for men than for women.

(g) Let $\mu_{NOR}, \mu_{LV}, \mu_{SV}$ denote the mean BPs in the three populations after averaging over gender. Prepare a table of estimates, with estimated standard errors in parentheses:

$$
\hat{\mu}_{NOR} = \quad \quad ( ) \\
\hat{\mu}_{SV} = \quad \quad ( ) \\
\hat{\mu}_{LV} = \quad \quad ( ) 
$$

(h) Suppose you are interested in $c = 4$ contrasts: the 3 pairwise contrasts and the difference $\theta$ between mean BP for normal eaters and the average of the mean BPs for the two vegetarian populations. Use Scheffé’s procedure to obtain simultaneous 95% confidence intervals for these four contrasts. Note that

\[
\sqrt{(3 - 1) F(0.95, 2, 24) MS[E]} \left( \frac{2}{10} \right) = 10.1 \\
\sqrt{(3 - 1) F(0.95, 2, 24) MS[E]} = 22.5 \\
\sqrt{\frac{1}{10} + \frac{1}{4} + \frac{1}{4} = 0.4}
\]

$$
\mu_{NOR} - \mu_{SV} : 11.0 \pm 10.1 \\
\mu_{NOR} - \mu_{LV} : \pm \\
\mu_{LV} - \mu_{SV} : \pm \\
\theta : \pm 
$$
(i) Label each sum of squares below so that the source of variability which it quantifies is clear

\[ SS[\text{diet}], SS[\text{gender}], SS[\text{diet} \times \text{gender}], SS[E], SS[Tot] \]

<table>
<thead>
<tr>
<th>Sum of squares</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (y_{ijk} - \bar{y}_{ij+})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (\bar{y}<em>{i++} - \bar{y}</em>{+++})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (\bar{y}<em>{j+-} - \bar{y}</em>{++-} - \bar{y}<em>{+} + \bar{y}</em>{++} + \bar{y}_{++})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (\bar{y}<em>{++} - \bar{y}</em>{++})^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{5} (y_{ijk} - \bar{y}_{+++})^2 ]</td>
<td></td>
</tr>
</tbody>
</table>
proc glm;
  class DIET GENDER;
  model bp=DITY|GENDER;
  means DIET|GENDER;
run;

The SAS System
Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIET</td>
<td>3</td>
<td>LV NCR SV</td>
</tr>
<tr>
<td>GENDER</td>
<td>2</td>
<td>female male</td>
</tr>
</tbody>
</table>

Sum of
Source                  DF  Squares  Mean Square  F Value  Pr > F
Model                   5  1705.066667  341.013333   4.56    0.0046
Error                   24  1793.600000  74.733333
Corrected Total        29  3498.666667

R-Square  Coeff Var  Root MSE  BP Mean
0.487348   7.539108   8.644844   114.6667

Source                  DF  Type I SS  Mean Square  F Value  Pr > F
DIET                   2  669.066667  334.533333   4.48    0.0223
GENDER                 1  896.533333  896.533333   12.00    0.0020
DIET*GENDER            2  139.466667  69.733333   0.93     0.4071

The GLM Procedure

Level of
 DIET N   Mean  Std Dev
LV  10  112.600000  7.3665913
NOR 10  121.200000 10.336586
SV  10  110.200000 12.3809890

Level of
 GENDER N  Mean  Std Dev
female 15  109.200000 9.87927123
male  15  120.133333 9.39503415

Level of  Level of
 DIET     GENDER N  Mean   Std Dev
LV  female  5  110.000000 7.8740079
LV  male  5  115.200000 6.5726707
NOR female  5  115.200000 8.3186537
NOR male  5  127.200000 9.011043
SV  female  5  102.400000 10.3344085
SV  male  5  118.000000 9.2736185
3. A common method of assessing cardiovascular capacity is through treadmill exercise testing. Maximal oxygen uptake, or $VO_2\text{ max}$, is an index measured using a treadmill with an inclined protocol. A random sample of $n = 12$ adults was randomized to two treatment groups:

- $T_1$: 12-week step-aerobic training program
- $T_2$: 12-week outdoor running regimen on flat-terrain.

$VO_2\text{ max}$ is measured before and after treatment and the change $y$ is used as the response. Ages of subjects are also measured. Data appear below. The point of the study is to see if the first treatment group shows greater increases in $VO_2\text{ max}$ than the 2nd.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Age</th>
<th>$y$</th>
<th>Treatment</th>
<th>Age</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>17.05</td>
<td>2</td>
<td>23</td>
<td>-0.87</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>4.96</td>
<td>2</td>
<td>22</td>
<td>-10.74</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>10.4</td>
<td>2</td>
<td>22</td>
<td>-3.27</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>11.05</td>
<td>2</td>
<td>25</td>
<td>-1.97</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>0.26</td>
<td>2</td>
<td>27</td>
<td>7.5</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>2.51</td>
<td>2</td>
<td>20</td>
<td>-7.25</td>
</tr>
<tr>
<td>avg</td>
<td>25.8</td>
<td>7.7</td>
<td>avg</td>
<td>23.2</td>
<td>-2.8</td>
</tr>
</tbody>
</table>

An ANCOVA model was fit using SAS (output at end of problem.)

(a) Use an independent variable $z_j$ denoting age of subject $j$ and an indicator variable $x_j$ for step-aerobic group to propose a multiple linear regression model for mean change in $VO_2\text{ max}$ for subject $j$. (Don’t fit any models here, just propose one using regression coefficients.)

$$
\mu_j = \quad \text{for } j = 1, 2, \ldots, 12.
$$

$$
x_j = \quad \text{for } j = 1, 2, \ldots, 12.
$$
(b) Use the SAS output to report the estimated mean change in $V O_2$ max separately for each group as a function of age. (Estimate the model you specified in part a)

(c) Use the model to complete the table of adjusted and unadjusted means. Use the fact that the average age among all subjects is $\bar{z} = 24.5$. Refer to the original dataset on the previous page.

<table>
<thead>
<tr>
<th>Trt. group</th>
<th>Unadjusted mean</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step-aerobic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flat-terrain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(d) Is there evidence that mean change in $VO_2$ max depends on age? Explain briefly.

(e) Estimate the treatment effect and its standard error.

(f) Draw a conclusion regarding change in $VO_2$ max brought about by the two exercise regimens.
data one;
  input grp z y @@;
  if grp=1 then x=1;
  if grp=2 then x=0;
cards;
  1 31 17.05 1 23 4.96 1 27 10.4 1 28 11.05 1 22 0.26 1 24 2.51
  2 23 -0.87 2 22 -10.74 2 22 -3.27 2 25 -1.97 2 27 7.5 2 20 -7.25
;run;

proc reg;
  model y=x z;
run;

The SAS System
The REG Procedure
Model: MODEL1
Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>647.87492</td>
<td>323.93746</td>
<td>41.42</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>70.38877</td>
<td>7.82097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>718.26369</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE    2.79660  R-Square   0.9020
Dependent Mean 2.46917  Adj R-Sq   0.8802
Coeff Var    113.26091

| Variable  | DF | Parameter | Standard Error | t Value | Pr > |t|   |
|-----------|----|-----------|----------------|---------|------|-----|
| Intercept | 1  | -46.45650 | 6.93653        | -6.70   | <.0001|
| x         | 1  | 5.44262   | 1.79645        | 3.03    | 0.0143|
| z         | 1  | 1.88589   | 0.29534        | 6.39    | 0.0001|