I. APPLICATION QUESTIONS (25 points)

A Cancer Center researcher is studying the metabolite GHS in the liver of rats. She obtained 24 rats which had received chemotherapy and allocated them randomly to one of four diets (six rats per diet). After one week of being fed this diet, the rats were sacrificed and the metabolite GHS was measured. One rat died after only one day of being fed the diet, and its metabolite level was not measured.

The four diets were:

1: no supplemental folate
2: low level of supplemental folate
3: medium level of supplemental folate
4: high level of supplemental folate.

Scientific Question 1: Does diet affect the level of GHS?

Scientific Question 2: On average, are the GHS levels for those rats with no supplemental folate significantly different from those with some supplemental folate? If so, in what direction and by how much are they different?

Data and output are shown on pages 4–10.
1. Using statistical notation, write down the model which was fit to these data; denote the observations by $Y_{ij}$ where $i = 1, \ldots, t$ and $j = 1, \ldots, r_i$. Include all model assumptions and explain any notation you use. Does your model assume any observations are correlated with each other? If so, which ones?

2. Describe whether or not the model’s assumptions are approximately met. Justify your answer. If some assumptions are not met, describe one remedial measure you could use.

3. Answer the two questions of interest using language a non-statistician could understand. Your answer should describe the data AND the results of a statistical test; you do not need to describe the study design.

4. (a) What null hypothesis is being tested in Question 1? What is the alternative hypothesis?
   (b) What null hypothesis is being tested in Question 2? What is the alternative hypothesis?

5. Using the estimated regression coefficients shown in the output, compute:
   (a) $\hat{\mu}_3$ and $\hat{\mu}_4$
   (b) $\hat{\tau}_3$ and $\hat{\tau}_4$
   (c) a two-sided 95% confidence interval for $\hat{\mu}_4$.

   You may need one of these numbers: $t_{0.025,3} = 3.18$, $t_{0.025,4} = 2.78$, $t_{0.025,19} = 2.09$, $t_{0.025,22} = 2.07$, $t_{0.025,23} = 2.07$, $t_{0.025,24} = 2.06$.

II. METHODS QUESTIONS (25 points)

1. Assume the researcher in Part I had come to talk to you before she began her experiment. How would you have instructed her to carry out the randomization? Your answer should be detailed enough that she could have taken your instructions and done the randomization herself.

2. We learned in class about the Gauss-Markov Theorem, which tells us that the least squares estimates of the $\mu_i$ have two special properties. What are those properties? Your answer must be very specific to get full credit.

MORE QUESTIONS ON THE NEXT PAGE...
3. Consider the formula needed in estimating sample size for a two-sample t-test:

\[ r = \frac{\sigma \left[ z_{\alpha/2} + z_{1-\beta} \right]^2}{\delta^2} \]

What happens to the sample size:
(a) as the desired power gets larger?
(b) as the minimum difference to detect \( \delta \) gets smaller?
(c) as \( \sigma \) doubles?

4. A study is being planned in the Center for Complementary and Alternative Medicine on headache relief. They will compare four groups: a placebo, an herbal remedy, aspirin, and ibuprofin. \( Y_{ij} \) will be the time from taking the medicine until headache relief; assume that \( Y_{ij} \) is approximately normally distributed.

(a) Write down the contrasts for each of the following two comparisons, and define any notation you use:
   i. placebo vs. non-placebo
   ii. herbal vs. pharmaceutical
(b) Write down a third contrast which will provide a complete orthogonal set of contrasts. Demonstrate that your set is orthogonal. For this headache study, what is the interpretation of your contrast?

5. Give an example where a random effects ANOVA would be reasonable. Justify your answer.

EXTRA CREDIT #1 (2 points)

1. What is the difference between an experimental design and an observational study?
2. Give one example where it could be unethical to conduct a designed experiment.

EXTRA CREDIT #2 (2 points)

You have fit a one-way fixed effects ANOVA with 9 treatment groups and would like to test all pairwise comparisons. You could use Tukey-Kramer, Bonferroni, Scheffé, Sidak, or Dunnett. Which is the best procedure to use here and why?
%include 'ex01.readdata.sas';

proc means data=dat n mean std;
   class diet;
   var ghs;
run;

proc glm data=dat;
   class diet;
   model ghs = diet / solution;
   output out=out student=etilde rstudent=estar p=fitted;
   contrast 'contrast' diet 3 -1 -1 -1;
   estimate 'contrast' diet 3 -1 -1 -1 / divisor=3;
run;

data _null_
   set out;
   file 'ex01.diag.dat';
   if _N_=1 then put "diet ghs etilde estar fitted";
   put diet ghs etilde estar fitted;
run;
ods select Moments Quantiles;
run;
proc univariate data=out;
   var estar;
run;

The SAS System 15:21 Friday, March 3, 2006 1

Obs     diet     ghs
1       1         23.71
2       1         28.14
3       1         27.85
4       1         62.73
5       1         80.05
6       1         46.00
7       2         45.55
8       2         42.57
9       2         48.13
10      2         49.25
11      2         61.96
12      2         41.04
13      3         83.89
14      3         78.62
15      3         93.98
16      3         60.57
17      3         44.30
18      3         117.50
19      4         70.41
20      4         58.45
21      4         92.67
22      4         47.45
23      4         71.02

The SAS System 15:21 Friday, March 3, 2006 2

The MEANS Procedure

   Analysis Variable : ghs

<table>
<thead>
<tr>
<th>diet</th>
<th>N</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>44.7466667</td>
<td>22.6941249</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>6</td>
<td>48.0833333</td>
<td>7.4879681</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>6</td>
<td>79.8100000</td>
<td>25.5754773</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>68.0000000</td>
<td>16.8600297</td>
</tr>
</tbody>
</table>

The SAS System 15:21 Friday, March 3, 2006 3

The GLM Procedure
Class Level Information

Class    Levels  Values

diet  4        1  2  3  4

Number of Observations Read  23
Number of Observations Used   23

The SAS System  15:21 Friday, March 3, 2006  4

The GLM Procedure

Dependent Variable: ghs

Sum of
Source       DF  Squares   Mean Square  F Value  Pr > F
Model        3   4921.89092  1640.63031  4.29     0.0180
Error       19  7263.03247   382.26487               
Corrected Total  22  12184.92338

R-Square  Coeff Var  Root MSE  ghs Mean
0.403933  32.68452  19.55159  59.81913

Source       DF  Type I SS     Mean Square  F Value  Pr > F
diet        3   4921.890916  1640.630305  4.29   0.0180

Source       DF  Type III SS    Mean Square  F Value  Pr > F
diet        3   4921.890916  1640.630305  4.29   0.0180

Contrast DF  Contrast SS     Mean Square  F Value  Pr > F
contrast    1   1869.409923  1869.409923  4.89   0.0395

Parameter  Estimate     Standard  t Value  Pr > |t|
contrast  -20.5511111  9.29319880  -2.21   0.0395

Parameter  Estimate     Standard  t Value  Pr > |t|
Intercept   68.00000000  8.74373909    7.78   <.0001
diet  1  -23.25333333  11.83907870   -1.96   0.0643
diet  2  -19.91666667  11.83907870   -1.68   0.1089
diet  3   11.81000000  11.83907870    1.00   0.3310
diet  4   0.00000000  .            .          

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

The SAS System  15:21 Friday, March 3, 2006  5

The UNIVARIATE Procedure

Variable: estar

Moments

N  23  Sum Weights  23
Mean  0.01079533  Sum Observations  0.24829266
Std Deviation  1.08001913  Variance  1.16644131
Skewness  0.41242391  Kurtosis  0.38301894
Uncorrected SS  25.6643893  Corrected SS  25.6617089
<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>2.34951671</td>
</tr>
<tr>
<td>99%</td>
<td>2.34951671</td>
</tr>
<tr>
<td>95%</td>
<td>2.16048358</td>
</tr>
<tr>
<td>90%</td>
<td>1.45120199</td>
</tr>
<tr>
<td>75% Q3</td>
<td>0.76908420</td>
</tr>
<tr>
<td>50% Median</td>
<td>0.00254492</td>
</tr>
<tr>
<td>25% Q1</td>
<td>-0.92699437</td>
</tr>
<tr>
<td>10%</td>
<td>-1.18776168</td>
</tr>
<tr>
<td>5%</td>
<td>-1.19160496</td>
</tr>
<tr>
<td>1%</td>
<td>-2.17644724</td>
</tr>
<tr>
<td>0% Min</td>
<td>-2.17644724</td>
</tr>
</tbody>
</table>
Oncology Study: Liver Metabolite GHS

Studentized residuals vs. fitted values
Oncology Study: Liver Metabolite GHS

Studentized residuals

Theoretical Quantiles

Sample Quantiles

exam1.res.hist.ps
Oncology Study: Liver Metabolite GHS

Observed vs. fitted values

Fitted values vs. GHS Level

exam1.obs.fits.ps