ST512
Exam IV, Monday, Dec. 5, 2005, Key

1. An experiment investigates the effects of sunlight and fertilizer on forage quality of peanuts. In each of three locations in Georgia, four plots are randomized to four different lighting conditions (quantified by something called percent incident photosynthetic photon flux density, PPFD) Each of these twelve plots are divided in two, and a nutritional soil supplement is added to one half, chosen at random and a measurement of leaf crude protein (CP) is made on each, for a total of 24 measurements. Pertinent SAS code and output appear as "PEANUTS PROBLEM" at the end of the exam.

   (a) Which design does this split-plot experiment use: completely randomized (CR-SPD) or $\boxed{\text{randomized complete block (RCBSPD)}}$?

   (b) Estimate all variance components from the model.

$$
\begin{aligned}
\hat{\sigma}^2 &= MSE \\
&= 4.125 \\
\hat{\sigma}^2_{l*p} &= \frac{MS[l*p] - MSE}{2} \\
&= 2.7 \\
\hat{\sigma}^2_{l} &= \frac{MS[l] - MS[l*p]}{8} \\
&= 9.9
\end{aligned}
$$

   (c) Regarding mean CP, consider the following contrasts of interest:

Contrast	Effect
θ_1	100% PPFD - 34% PPFD, averaging over supplement
θ_2	100% PPFD - 34% PPFD when the supplement is used
θ_3	the supplement effect at 100% PPFD
θ_4	linear contrast for PPFD

   If any *df* approximations are required for the remaining problems, use the value from the table closest to $\widehat{df}$.

   i. Obtain a 95% confidence interval for θ_1

$$
\begin{aligned}
\hat{\theta}_1 &= 266.5 - 228.5 = 38 \\
SE &= \sqrt{\frac{2}{6} MS(l*p)}(\text{ on } df = 6) \\
&= 1.78 \\
t(.025, 6) &= 2.45 \\
c.i. &= 38 \pm 4.35
\end{aligned}
$$

ii. Obtain a 95% confidence interval for θ_2

$$\begin{aligned}
\hat{\theta}_2 &= 37 \\
SE &= \sqrt{\frac{2}{3}(\sigma^2_{l*p} + \sigma^2)} \\
\widehat{SE} &= \sqrt{\frac{2}{3}(\frac{MSl*p^2}{2} + \frac{MSE}{2}} \\
&= 2.13 \\
\widehat{df} &= 10.8 \\
qt(0.025, 11) &= 2.2 \\
c.i. &= 37 \pm 4.7
\end{aligned}$$

iii. Obtain a 95% confidence interval for θ_3

$$\begin{aligned}
\hat{\theta}_3 &= -3 \\
\widehat{SE} &= \sqrt{\frac{2}{3}MSE}(df = 8) \\
c.i. &= -3 \pm 3.8
\end{aligned}$$

iv. Does a model in which CP is linear in PPFD exhibit lack of fit?

$$\hat{\theta}_4 = -\bar{y}_{1++} - \bar{y}_{2++} + \bar{y}_{3++} + 3\bar{y}_{4++} = 128.5$$

$$\begin{aligned}
SS(\theta_4) &= \frac{\hat{\theta}_4^2}{20/6} = 4954 \\
SS(LOF) &= 5008 - 4954 \\
&= 54(on df = 2) \\
F_{LOF} &= MS(LOF)/MS(l*p) \\
&= 2.8
\end{aligned}$$

2. An experiment is carried out to evaluate the effects of bedtime reading and alcohol consumption on quality of sleep. The experimental factors and their levels follow:

   - nightly alcohol consumption (3 levels): none, 1 drink, 2 drinks
   - reading (2 levels) : reading at bedtime, not reading at bedtime.

   (A drink has .6 oz ethanol.) The experiment recruits $N = 21$ subjects and allows a "burn-in" period of two weeks to allow them time to learn to judge the quality of their sleep by filling out brief reports the next morning. The subjects are randomized to the three alcohol treatment groups ($n = 7$ each). Each subject in a given drinking group then spends, in random order, one week on the reading regimen, and another week on the non-reading regimen and makes a report of sleep quality after each. Thus, there are a total of 42 measurements. Be sure to take subjects as random effects in your model.

   (a) (3 pts) Name this design. Is it a randomized complete block split-plot design (RCBSPD), or a completely randomized split-plot design (CRSPD)? (Circle one.)

   (b) (3 pts) What are the whole plot factors? *Alcohol consumption*

   (c) (3 pts) What are the whole plot units? *Subjects*

   (d) (3 pts) What are the split plot factors? *Reading*

   (e) (4 pts) Report an F-ratio and p-value for a test of no interaction between the drinking and reading factors.

   $$F = 4.83, p = 0.0210$$

   There is evidence of interaction.

   (f) (5 pts) Claim: there is no evidence of person-to-person variability in sleep quality. Formulate an appropriate null hypothesis ...

   $$H_0 : \sigma_s^2 = 0$$

   $$F = 1.34, p = 0.27, \hat{\sigma}_s^2 = (MS(Subj(D)) - MS(E))/2 = 2.6$$

   No evidence of subject effect.

   (g) (4 pts) Estimate the three simple effects of reading separately for each level of alcohol consumption ...

Drinking Level	Reading Effect	Standard Error	p-value
none	3.57	0.65	< 0.0001
one drink	1.85		0.0109
two drinks	0.7		0.2896

3. An experiment on corn yields divides a plot up into four rows and four columns so that a grid of $N = 16$ subplots is obtained. Four treatments, corresponding to presence/absence of nitrogen and presence/absence of phosphate, are randomized to the subplots in such a way that each treatment appears exactly once in each row and in each column.

   (a) Report the F-ratio and p-value for a test of the null hypothesis that neither nitrogen nor phosphate has any effect on yield.

   $$F = 5.27, p = 0.0406$$

   (b) Is the phosphate effect plausibly the same in the presence of nitrogen as in the absence of nitrogen?

   $$\hat{\theta}_{int} = 31.9, SE = 19.7, t = 1.62 = \sqrt{2.6}$$

   (c) Report a 95% confidence interval for the main effect of nitrogen.

   $$\widehat{\text{effect}} = \frac{140.175 + 180.5}{2} - \frac{138.3 + 129.9}{2} = 26.15$$

   $$SE = \sqrt{2MSE/8} = 9.86$$

   $$t = 26.15/9.86 = 2.65(\text{significant})$$

   (d) The p-value for column effects is not significant. Should we drop it from the model? Explain. Don't test blocks. Leave blocks in model if they're in the design.

4. In an NCSU experiment to study iron intake, ten students are randomly sampled from each combination of residence (on-campus or off-campus) and gender (man or woman). Each student is asked to randomly sample $n = 5$ days during the semester, and iron intakes are measured, in mg, for each day. There are a total of 200 measurements.

   (a) Are these factors crossed or nested? If nested, which is factor is nested in which?

      i. (2 pts) Gender and residence? *crossed.*

      ii. (2 pts) Student and gender? *Student is nested in gender.*

   (b) (3 pts) How much variability is there from day to day for a given student? Estimate an appropriate parameter.

$$MSE = 3.6mg^2$$

   (c) (3 pts) How much variability is there from student to student for a given gender-by-residence combination? Estimate an appropriate parameter.

$$\hat{\sigma}_s^2 = \frac{MS[Student(G \times R)] - MSE}{5} = 6.5mg^2$$

   (d) (5 pts) Report an F-ratio and associated degrees of freedom for a test that all four gender-by-residence mean daily iron intakes are equal.

$$F = \frac{MS[\text{“}Model\text{”}]}{MS[Student(G \times R)]} = \frac{1/3(36.5 + 236.7 + 311.5)}{39.7} = 4.91$$

   Critical value is $F(0.05, 3, 36) = 2.87$, *so there is significant evidence of effects.*

   (e) (4 pts) True or $\boxed{\text{false}}$: there is no significant evidence that iron intake varies by gender.

   (f) (6 pts) Estimate the simple gender effect for people who reside on campus. Report a standard error. Is the difference between men and women significant?

$$\bar{y}_{12+} - \bar{y}_{11+} = 10.288 - 8.646 = 1.642(SE = \sqrt{\frac{2}{50}MS(Stud(G \times R))} = 1.26mg)$$

   This simple effect is not significant.

5. An experiment is conducted with chocolate chip cookies to see whether or not adding fine pecan fragments enhances the taste. Batches of twenty cookies are baked at a time, with 10 from each batch coming from regular batter and the other 10 from batter with the pecan fragments added. Each member of a 10-member panel gets two cookies from each batch, one with pecans, one without. The 10 responses from the panel are averaged for each treatment in each batch. Five batches of cookies are produced, for a total of 10 measurements to be used in the analysis. Let $i = 1$ for with-pecan cookies, $i = 2$ for without-pecan cookies and let $j = 1, 2, 3, 4, 5$ index the batches. Then

$$Y_{ij} = \text{ average taste for treatment } i, \text{ batch } j.$$

(a) (3pts) $\boxed{\text{True}}$/false: This is a randomized complete block design.

(b) (3pts) True/$\boxed{\text{false}}$: This is a split-plot design.

(c) (6pts) In the model you use to analyze the data, assume that any effect of pecans on taste is constant across batches. Report the F-ratio and associated degrees of freedom for a test of no pecan effect on taste.

$$F = MS[treatment]/MS[E] = 27.225/3.01 = 9.04 \text{ on } df = 1, 4$$

(d) (9 pts) Obtain a test statistic for a paired t-test of the form

$$t = \frac{\bar{d}}{s_D/\sqrt{n}}$$

where d_j denotes the difference between with and without pecans for batch j:

$$d_j = y_{2j} - y_{1j} \text{ for } j = 1, 2, 3, 4, 5$$

and $s_d = 2.45$ denotes the observed standard deviation of the $n = 5$ differences. Indicate the associated degrees of freedom. Draw a conclusion about the significance of the observed taste difference.

$$t = \frac{3.3}{2.45/\sqrt{5}} = 3.01$$

The pecan treatment enhances the taste significantly.

(e) (4 pts) Specify the relationship between the F statistic in part (d) and the t-statistic in part (e), thus demonstrating the equivalence of the F-test for this design and the paired t-test.

$$F = T^2$$

PEANUTS PROBLEM

```
proc glm ;
   class ppfd location supplement ;
   model CP=ppfd|supplement location|ppfd;
   random location location*ppfd;
   means ppfd*supplement;
```

```
                              The SAS System                              1
                             The GLM Procedure

                  Class            Levels    Values
                  ppfd                  4    34 56 78 100
                  location              3    1 2 3
                  supplement            2    no yes

                                       Sum of
 Source                      DF       Squares    Mean Square  F Value  Pr > F
 Model                       15   5259.625000     350.641667    85.00  <.0001
 Error                        8     33.000000       4.125000
 Corrected Total             23   5292.625000

 Source                      DF     Type I SS    Mean Square  F Value  Pr > F
 ppfd                         3   5008.125000    1669.375000   404.70  <.0001
 supplement                   1      3.375000       3.375000     0.82  0.3921
 ppfd*supplement              3     13.125000       4.375000     1.06  0.4180
 location                     2    177.750000      88.875000    21.55  0.0006
 ppfd*location                6     57.250000       9.541667     2.31  0.1352

Source                  Type III Expected Mean Square
ppfd                    Var(Error) + 2 Var(ppfd*location) +
                        Q(ppfd,ppfd*supplement)
supplement              Var(Error) + Q(supplement,ppfd*supplement)
ppfd*supplement         Var(Error) + Q(ppfd*supplement)
location                Var(Error) + 2 Var(ppfd*location) + 8 Var(location)
ppfd*location           Var(Error) + 2 Var(ppfd*location)

      Level of      Level of                 --------------CP-------------
      ppfd          supplement        N              Mean          Std Dev
      34            no                3        229.000000       5.19615242
      34            yes               3        228.000000       5.56776436
      56            no                3        243.000000       6.08276253
      56            yes               3        243.000000       1.73205081
      78            no                3        257.000000       2.64575131
      78            yes               3        258.000000       1.00000000
      100           no                3        268.000000       4.35889894
      100           yes               3        265.000000       3.00000000
```

SLEEPING PROBLEM

```
proc glm data=one;
   class subject drinking reading;
   model sleepqual=drinking|reading subject(drinking);
   lsmeans drinking*reading/slice=drinking;
run;
```

The SAS System
The GLM Procedure
Class Level Information

Class	Levels	Values
subject	21	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
drinking	3	none one drink two drinks
reading	2	no reading reading

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	127.9761905	5.5641822	3.71	0.0031
Error	18	27.0000000	1.5000000		
Corrected Total	41	154.9761905			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
drinking	2	33.33333333	16.66666667	11.11	0.0007
reading	1	44.02380952	44.02380952	29.35	<.0001
drinking*reading	2	14.47619048	7.23809524	4.83	0.0210
subject(drinking)	18	36.14285714	2.00793651	1.34	0.2712

Least Squares Means

drinking	reading	sleepqual LSMEAN
none	no reading	4.00000000
none	reading	7.57142857
one drink	no reading	6.28571429
one drink	reading	8.14285714
two drinks	no reading	4.71428571
two drinks	reading	5.42857143

drinking*reading Effect Sliced by drinking for sleepqual

drinking	DF	Sum of Squares	Mean Square	F Value	Pr > F
none	1	44.642857	44.642857	29.76	<.0001
one drink	1	12.071429	12.071429	8.05	0.0109
two drinks	1	1.785714	1.785714	1.19	0.2896

CORNYIELDS PROBLEM

```
proc glm;
   class row col trtcombo;
   model yield = row col trtcombo;
   lsmeans trtcombo;
run;
```

The SAS System
The GLM Procedure
Class Level Information

Class	Levels	Values
row	4	1 2 3 4
col	4	1 2 3 4
trtcombo	4	noN-noK noN-yesK yesN-noK yesN-yesK

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	26059.65563	2895.51729	7.44	0.0119
Error	6	2334.76875	389.12813		
Corrected Total	15	28394.42438			

R-Square	Coeff Var	Root MSE	yield Mean
0.917774	13.39933	19.72633	147.2188

Source	DF	Type III SS	Mean Square	F Value	Pr > F
row	3	17505.06687	5835.02229	15.00	0.0034
col	3	2405.94688	801.98229	2.06	0.2069
trtcombo	3	6148.64188	2049.54729	5.27	0.0406

trtcombo	yield LSMEAN
noN-noK	129.875000
noN-yesK	138.325000
yesN-noK	140.175000
yesN-yesK	180.500000

IRON PROBLEM

```
proc mixed;  class gender residence student;
model intake=gender|residence;
random student(gender*residence);
lsmeans gender*residence;
```

The SAS System

Source	DF	Sum of Squares	Mean Square
gender	1	36.465800	36.465800
residence	1	236.748800	236.748800
gender*residence	1	311.500800	311.500800
stude(gender*reside)	36	1429.208400	39.700233
Residual	160	577.108000	3.606925

Type 3 Analysis of Variance

Source	Expected Mean Square	Error Term
gender	Var(Residual) + 5 Var(stude(gender*reside)) + Q(gender,gender*residence)	MS(stude(gender*reside))
residence	Var(Residual) + 5 Var(stude(gender*reside)) + Q(residence,gender*residence)	MS(stude(gender*reside))
gender*residence	Var(Residual) + 5 Var(stude(gender*reside)) + Q(gender*residence)	MS(stude(gender*reside))
stude(gender*reside)	Var(Residual) + 5 Var(stude(gender*reside))	MS(Residual)
Residual	Var(Residual)	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
gender	1	36	0.92	0.3443
residence	1	36	5.96	0.0196
gender*residence	1	36	7.85	0.0081

Least Squares Means

Effect	gender	residence	Estimate	Standard Error	DF	t Value	Pr > |t|
gender*residence	man	offcampus	8.9660	0.8911	36	10.06	<.0001
gender*residence	man	oncampus	8.6460	0.8911	36	9.70	<.0001
gender*residence	woman	offcampus	5.6160	0.8911	36	6.30	<.0001
gender*residence	woman	oncampus	10.2880	0.8911	36	11.55	<.0001

COOKIES PROBLEM

```
proc print noobs data=cookies;
run;
proc glm;
   class batch treatment;
   model taste = batch|treatment;
run;
```

```
                                The SAS System                              1

                      batch    treatment    panelmean

                        1      pecans          8.1
                        1      no pecans       6.8
                        2      pecans          4.5
                        2      no pecans       0.1
                        3      pecans          6.2
                        3      no pecans       5.7
                        4      pecans          8.4
                        4      no pecans       4.7
                        5      pecans          8.3
                        5      no pecans       1.7

                               The GLM Procedure

                Class          Levels    Values

                batch               5    1 2 3 4 5
                treatment           2    no pecans pecans

Dependent Variable: panelmean

                                        Sum of
 Source                      DF        Squares    Mean Square   F Value   Pr > F
 Model                        9    70.44500000     7.82722222         .        .
 Error                        0     0.00000000              .
 Corrected Total              9    70.44500000

          R-Square     Coeff Var      Root MSE    panelmean Mean
          1.000000             .             .          5.450000

 Source                      DF      Type I SS    Mean Square   F Value   Pr > F
 batch                        4    31.17000000     7.79250000         .        .
 treatment                    1    27.22500000    27.22500000         .        .
 batch*treatment              4    12.05000000     3.01250000         .        .
```