ST512

Fall Semester, 2005

Exam IV, Monday, Dec 5

Name:____

Directions: Do problem 1. Then select three out of the four remaining questions to solve. Only four will be graded, so please indicate which one you do not want to solve.

Please show work. Don't leave any questions blank. Partial credit may be awarded. For true/false questions, circle either true or false. Each problem has a corresponding printout of SAS code and output at the end of the exam. The printouts are entitled

- PEANUTS PROBLEM
- SLEEPING PROBLEM
- CORNYIELDS PROBLEM
- IRON PROBLEM
- COOKIES PROBLEM

The SAS code used may fit the wrong models, so inspect it carefully. Any time a question asks about "statistical signficance", you may assume a level of significance of $\alpha = 0.05$.

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 (CRSPD) or randomized complete block (RCBSPD)?

   (b) Estimate all variance components from the model.

   (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

ii. Obtain a 95% confidence interval for θ_2

iii. Obtain a 95% confidence interval for θ_3

iv. Obtain the lack-of-fit F-ratio and df for a test of the hypothesis that mean CP is linear in PPFD.

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

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 randomly samples $N = 21$ 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 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. SAS code and output appear as "SLEEPING PROBLEM."

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

   (b) What are the whole plot factors?

   (c) What are the whole plot units?

   (d) What are the split plot factors?

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

(f) Claim: there is no evidence of person-to-person variability in sleep quality. Formulate an appropriate null hypothesis, report any F-ratios, p-values and associated parameter estimates for quantities of interest. Draw a qualitative conclusion. (Regardless of the result, leave the random effect in the model for all other questions.)

(g) Estimate the three simple effects of reading separately for each level of alcohol consumption. In each case, report standard errors and p-values for tests of no reading effect.

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.

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

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

   (d) The p-value for column effects is not significant. Should we drop it from the model? Explain.

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. Gender and residence?

      ii. Student and gender?

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

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

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

   (e) True or false: there is no significant evidence that iron intake varies by gender.

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

5. An experiment is conducted with chocolate chip cookies to see whether or not adding fine pecan fragments enhances the taste. Five batches are prepared. Each batch is separated into two halfs with equal amounts of batter. Pecan fragments are added to one half at random, for each batch. A panel taste-tests each half-batch, resulting in 10 measurements. Let

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

where $i = 1$ for with-pecan cookies, $i = 2$ for without-pecan cookies and $j = 1, 2, 3, 4, 5$ indexes the batches. Pertinent SAS code and output may be found at the end of the exam under the title "COOKIES PROBLEM."

(a) True/false: This is a randomized complete block design.

(b) True/false: This is a split-plot design.

(c) 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.

(d) 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.

(e) For designs such as this, specify the relationship between an F statistic like the one obtained in part (c) and a t-statistic like the one obtained in part (d).

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

Source	DF	Sum of 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 ppfd	Level of supplement	N	CP Mean	CP 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-noP noN-yesP yesN-noP yesN-yesP

                                     Sum of
 Source                     DF      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-noP        129.875000
                         noN-yesP       138.325000
                         yesN-noP       140.175000
                         yesN-yesP      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;
   means 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

                                       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

 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        .       .

            Level of            ----------panelmean----------
            treatment     N             Mean          Std Dev
            no pecans     5       3.80000000       2.80713377
            pecans        5       7.10000000       1.71026314
```

Table of critical values ($t(df, \alpha)$) from t-distributions:

df	$\alpha = 0.2$	$\alpha = 0.15$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$
1	1.37638	1.96261	3.07768	6.31375	12.7062	31.8205	63.6567	318.309
2	1.06066	1.38621	1.88562	2.91999	4.3027	6.9646	9.9248	22.327
3	0.97847	1.24978	1.63774	2.35336	3.1824	4.5407	5.8409	10.215
4	0.94096	1.18957	1.53321	2.13185	2.7764	3.7469	4.6041	7.173
5	0.91954	1.15577	1.47588	2.01505	2.5706	3.3649	4.0321	5.893
6	0.90570	1.13416	1.43976	1.94318	2.4469	3.1427	3.7074	5.208
7	0.89603	1.11916	1.41492	1.89458	2.3646	2.9980	3.4995	4.785
8	0.88889	1.10815	1.39682	1.85955	2.3060	2.8965	3.3554	4.501
9	0.88340	1.09972	1.38303	1.83311	2.2622	2.8214	3.2498	4.297
10	0.87906	1.09306	1.37218	1.81246	2.2281	2.7638	3.1693	4.144
11	0.87553	1.08767	1.36343	1.79588	2.2010	2.7181	3.1058	4.025
12	0.87261	1.08321	1.35622	1.78229	2.1788	2.6810	3.0545	3.930
13	0.87015	1.07947	1.35017	1.77093	2.1604	2.6503	3.0123	3.852
14	0.86805	1.07628	1.34503	1.76131	2.1448	2.6245	2.9768	3.787
15	0.86624	1.07353	1.34061	1.75305	2.1314	2.6025	2.9467	3.733
16	0.86467	1.07114	1.33676	1.74588	2.1199	2.5835	2.9208	3.686
17	0.86328	1.06903	1.33338	1.73961	2.1098	2.5669	2.8982	3.646
18	0.86205	1.06717	1.33039	1.73406	2.1009	2.5524	2.8784	3.610
19	0.86095	1.06551	1.32773	1.72913	2.0930	2.5395	2.8609	3.579
20	0.85996	1.06402	1.32534	1.72472	2.0860	2.5280	2.8453	3.552
21	0.85907	1.06267	1.32319	1.72074	2.0796	2.5176	2.8314	3.527
22	0.85827	1.06145	1.32124	1.71714	2.0739	2.5083	2.8188	3.505
23	0.85753	1.06034	1.31946	1.71387	2.0687	2.4999	2.8073	3.485
24	0.85686	1.05932	1.31784	1.71088	2.0639	2.4922	2.7969	3.467
25	0.85624	1.05838	1.31635	1.70814	2.0595	2.4851	2.7874	3.450
26	0.85567	1.05752	1.31497	1.70562	2.0555	2.4786	2.7787	3.435
27	0.85514	1.05673	1.31370	1.70329	2.0518	2.4727	2.7707	3.421
28	0.85465	1.05599	1.31253	1.70113	2.0484	2.4671	2.7633	3.408
29	0.85419	1.05530	1.31143	1.69913	2.0452	2.4620	2.7564	3.396
30	0.85377	1.05466	1.31042	1.69726	2.0423	2.4573	2.7500	3.385
40	0.85070	1.05005	1.30308	1.68385	2.0211	2.4233	2.7045	3.307
60	0.84765	1.04547	1.29582	1.67065	2.0003	2.3901	2.6603	3.232
80	0.84614	1.04320	1.29222	1.66412	1.9901	2.3739	2.6387	3.195
100	0.84523	1.04184	1.29007	1.66023	1.9840	2.3642	2.6259	3.174
200	0.84342	1.03913	1.28580	1.65251	1.9719	2.3451	2.6006	3.131
100000	0.84162	1.03644	1.28156	1.64487	1.9600	2.3264	2.5759	3.090