ST512
Fall Semester, 2006
Quiz 3-solutions

1. (Taken from Ott and Longnecker, Ch. 17) An experiment measures tablet hardness by randomly selecting three `batches` of a formulation for a drug, then randomly selecting three 1-kg `samples` from each of these batches. Then 7 tablets are sampled from each of the 9 `samples` and hardness readings (y) are taken on each with the following results:

	Batch 1			Batch 2			Batch 3		
Sample	1	2	3	1	2	3	1	2	3
	85	75	94	109	117	99	71	80	71
	96	88	98	101	108	108	83	70	67
	91	91	94	104	103	98	78	84	81
	97	91	95	109	109	99	69	83	73
	86	88	101	104	101	117	85	72	76
	95	95	100	102	104	109	66	81	80
	93	94	94	107	101	107	76	75	73

The SAS code and output at the end of the quiz, entitled "TABLET PROBLEM", fits the wrong model, but the correct sums of squares required for inference may be recovered easily, using the following algebraic fact.

$$\sum_i\sum_j\sum_k(\bar{y}_{+j+}-\bar{y}_{+++})^2+\sum_i\sum_j\sum_k(\bar{y}_{ij+}-\bar{y}_{i++}-\bar{y}_{+j+}+\bar{y}_{+++})^2=\sum_i\sum_j\sum_k(\bar{y}_{ij+}-\bar{y}_{i++})^2$$

(a) Using a model with random `batch` and random (nested) `sample` effects,

i. report the F-ratio and associated df for a test of the `sample` effect (numerator $df=6$), Let A denote `batch` and B denote `sample`.

$$F=\frac{MS[B(A)]}{MS(E)}=\frac{(SS[B]+SS[B(A)])/(a(b-1))}{MS[E]}$$

$$=\frac{(19.9+216.6)/6}{31.4}=\frac{39.4}{31.4}=1.26,\quad df=6,54$$

ii. report the F-ratio and associated df for a test of equality of `batch` means.

$$F=\frac{MS[A]}{MS[B(A)]}=\frac{4653.8}{39.4}=118.1\quad df=2,6$$

2. Consider two experimental designs for an investigation of the growth of grass using four different nitrogen additives: N_1, N_2, N_3 and N_4. A plot divided into 16 sections (4×4) is to be used. In the tables below, the top of this page is north, the left is west and randomizations are within row for Design 1 and within column for Design 2.

	Design 1			
Block 1	N_4	N_1	N_3	N_2
Block 2	N_2	N_4	N_3	N_1
Block 3	N_4	N_1	N_2	N_3
Block 4	N_1	N_2	N_3	N_4

Design 2

Block 1	Block 2	Block 3	Block 4
N_2	N_3	N_1	N_1
N_1	N_2	N_3	N_3
N_4	N_4	N_4	N_2
N_3	N_1	N_2	N_4

(a) Suppose that prior to treatment with the additives, there is a pronounced fertility gradient in the north-south direction, but not so much in the east-west direction, which design is better? Why?

RCB designs aim to reduce unexplained variability by forming groups of homogeneous experimental units then randomizing within groups. In this case, columns consist of heterogeneous units due to the north-south fertility gradient, so it would be better to block by row, as in Design 1.

(b) Provide an experimental design using this plot that would allow for control of fertility effects in both directions; north-south and east-west. I used these two permutations of $(1, 2, 3, 4)$: 4213 and 2134 to get the randomized latin square on the right

$$\left|\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{array}\right| \xrightarrow{(4213)} \left|\begin{array}{cccc} 2 & 3 & 4 & 1 \\ 4 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{array}\right| \xrightarrow{(2134)} \boxed{\begin{array}{cccc} 3 & 2 & 4 & 1 \\ 1 & 4 & 2 & 3 \\ 2 & 1 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{array}}$$

(c) Sketch an ANOVA table for an analysis that is appropriate for the design you proposed in part (b), with a column for source and a column for *df*.

Source	*df*
Rows	3
Columns	3
Nitrogen additives	3
Error	6
Total	15

3. A researcher studies the effects of four treatments on chicken susceptibility to infection after exposure to a particular bacteria. The experiment uses three different houses. In each house, four cages are randomized to treatments. Infection rates are measured for all twelve cages. He considers a model appropriate for a completely randomized design (CRD). Means and an ANOVA table follow. For $i = 1, 2, 3, 4$ and $j = 1, 2, 3$,

$$Y_{ij} = \mu + \alpha_i + E_{ij}.$$

Treatment number (i)	Treatment description	Treatment label	Treatment mean ($\bar{y}_{i+}$)	sample variance s_i^2
1	not vaccinated, no food additive	$V0A0$	45	37
2	not vaccinated, food additive	$V0A1$	42	172
3	vaccinated, no food additive	$V1A0$	40	64
4	vaccinated, food additive	$V1A1$	17	103

Source	*df*	SS	MS	F	p-value
Treatment	3	1382	494	5.26	.027
Error	8	752	94		
Total	11	2234			

(a) He analyzed his data using the model above and found that the p-value for interaction of vaccination and additive was greater than 0.05. Verify this calculation, complete with an appropriate critical value and sum of squares for interaction.

$$\begin{aligned}
\hat{\theta}_I &= \bar{y}_{1+} - \bar{y}_{2+} - (\bar{y}_{3+} - \bar{y}_{4+}) \\
&= 20 \\
SS(\theta_I) &= \frac{\hat{\theta}^2}{\sum c_i/b^2} \\
&= \frac{400}{4/3} \\
&= 300 \\
t &= \sqrt{\frac{SS(\theta)}{MS[E]}} \\
&= 1.79 \\
t(.025, 8) &= 2.31
\end{aligned}$$

(Since observed t-statistic does not exceed critical value for $\alpha = .05$ test, $p > .05$.)

(b) A statistical consultant learned that the house means were $\bar{y}_{+1} = 38, \bar{y}_{+2} = 44$ and $\bar{y}_{+3} = 26$. Using the sum of squares obtained in part (a), redo the test for interaction, so that it is appropriate for this randomized complete block (RCB) design with house as block. Report a test statistic, critical value and whether or not $p < .05$. Briefly characterize the nature of the interaction.

$$\begin{aligned}
SS(\text{blocks}) &= \sum_i \sum_j (\bar{y}_{+j} - \bar{y}_{++})^2 \\
&= 4(2^2 + 8^2 + (-10)^2) \\
&= 672 \\
SS(E)_{RBD} &= SS(E)_{CRD} - SS(\text{blocks}) \\
&= 752 - 672 \\
&= 80 \\
MS(E)_{RBD} &= \frac{80}{6} \\
F_{inter} &= \frac{SS(\theta_I)}{MS[E]} \\
&= 22.5 \\
t &= \sqrt{22.5} \\
&= 4.7 \\
t(.025, 6) &= 2.45
\end{aligned}$$

(Since observed t-statistic does exceed critical value for $\alpha = .05$ test, $p < .05$.)

4. An experiment measures the weightgain of rainbow trout (*Oncorhynchus mykiss*) fed using three methods of feed delivery: `demand`, `hand` and `pellet`. Six offspring from each of six experimentally bred families are randomized to three diets, for a total of $abn = 36$ observations. Consider a mixed model for these data with fixed method effects and random family (variance component σ_B^2) and interaction (variance component σ_{AB}^2) effects. SAS code and output pertaining to this problem are given at the end of the quiz under the title "TROUT PROBLEM."

   (a) Specify the correlation between weightgains for siblings (fish from the same family)

   - receiving the same feed delivery method,

   $$\frac{\sigma_B^2 + \sigma_{AB}^2}{\sigma_B^2 + \sigma_{AB}^2 + \sigma^2}$$

   - receiving different feed delivery methods.

   $$\frac{\sigma_B^2}{\sigma_B^2 + \sigma_{AB}^2 + \sigma^2}$$

   (b) Estimate all variance components in the model.

   $$\begin{aligned} \hat{\sigma}^2 &= MSE \\ &= 23.2 \\ \hat{\sigma}_{AB}^2 &= \frac{MS[AB] - MSE}{n} \\ &= \frac{35.8 - 23.2}{2} \\ &= 6.3 \\ \hat{\sigma}_B^2 &= \frac{MS[B] - MS[AB]}{na} \\ &= \frac{974 - 35.8}{6} \\ &= 156 \end{aligned}$$

   (c) Pick a variance component other than the error variance σ^2. Give an expression for an approximate 95% confidence interval for that variance component. If the Satterthwaite approximation for *df* is required, provide an expression for $\widehat{df}$.

   Both other variance components estimators are linear combinations of more than one mean square, and so require *df* approximation. For example, consider the family effect variance component:

   (d) Which, if any, of the three pairwise differences among feeding method means are significant at comparisonwise error rate $\alpha = 0.05$?

   $$MSD = t(.025, 10)\sqrt{\frac{2}{12}MS(AB)} = 2.23\sqrt{5.97} = 5.44$$

   Only differences involving `demand` exceed this MSD.

TABLET PROBLEM

```
proc glm;
   class sample batch;
   model y=sample|batch;
run;
```

The SAS System 1

The GLM Procedure

Class	Levels	Values
Sample	3	1 2 3
Batch	3	1 2 3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	9544.12698	1193.01587	38.01	<.0001
Error	54	1694.85714	31.38624		
Corrected Total	62	11238.98413			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sample	2	19.936508	9.968254	0.32	0.7292
Batch	2	9307.555556	4653.777778	148.27	<.0001
Sample*Batch	4	216.634921	54.158730	1.73	0.1578

TROUT PROBLEM

```
proc glm;
   class feed family;
   model wtgain=feed|family;
   lsmeans feed|family;
run;
```

The SAS System

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	7121.888889	418.934641	18.04	<.0001
Error	18	418.000000	23.222222		
Corrected Total	35	7539.888889			

R-Square	Coeff Var	Root MSE	wtgain Mean
0.944562	7.092477	4.818944	67.94444

Source	DF	Type III SS	Mean Square	F Value	Pr > F
feed	2	1893.555556	946.777778	40.77	<.0001
family	5	4869.888889	973.977778	41.94	<.0001
feed*family	10	358.444444	35.844444	1.54	0.2031

feed	wtgain LSMEAN
demand	78.0000000
hand	64.6666667
pellet	61.1666667

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