Answers for the 2nd sample exam given out for Spring 2002

Part I - 1) In performing an ANOVA, what four assumptions must be satisfied? **The errors must be normally distributed, have mean zero, and equal variances at each treatment level, and must be independent.**

2) Define what is meant by the p-value (or empirical significance level) of a test. **The smallest a-level at which the null hypothesis would be rejected. -or- The probability of observing a test statistic as extreme as the one observed, or more extreme, if the null hypothesis is true.**

3) Consider a one-way ANOVA that has four levels: AI, AII, BI, and BII in that order.
The contrast ½ ½ -½ -½ tests whether the effect of A is significantly different from the effect of B.
The contrast ½ -½ ½ -½ tests whether the effect of I is significantly different from the effect of II.
What does the contrast ½ -½ -½ ½ test? **The interaction between A-B and I-II.**

4) Identify each of the following as either <u>always</u> or <u>not</u>, where always=always controls α_F , and not=does not control α_F.
Bonferroni's formula **<u>always</u>** Fisher's LSD **<u>not</u>** Holm Test **<u>always</u>**

5) Sketch an example of a residual plot that would indicate that taking the square-root or logarithm of the response variable (dependent variable) would be appropriate. <

6) An ANOVA is being performed to examine the effects of herbicide and pesticide on pumpkins. The three most popular brands of herbicide (called A, B, and C) are used. There are no pesticides of particular interest and so two are chosen at random from among 10 leading brands (after choosing them, they are labeled D and E). Twelve plots of land are used, with two for each combination of herbicide and pesticide. This design is **<u>random effect</u> / <u>factorial</u>**.

7) An ANOVA is being performed to examine the effects of herbicide and pesticide on pumpkins. The three most popular brands of herbicide (labeled A, B, and C) and two most popular brands of pesticide (labeled D and E) are used. Twelve plots of land are used, with two for each combination of herbicide and pesticide. This design is **<u>factorial</u>**.

8) An ANOVA is being performed to examine the effects of herbicide and pesticide on pumpkins. The three most popular brands of herbicide (called A, B, and C) are chosen. There are no pesticides of particular interest and so two are chosen at random from among 10 leading brands (after choosing them, they are labeled D and E). Only five plots of land are available, and one each is used to test the combinations A-D, B-D, C-D, A-E, and B-E. This design is **<u>random effect</u> / <u>without replication</u>**.

9) Construct a 95% confidence interval for the difference between the average of the two A treatments and the average of the two B treatments. **5.3333 +/- (2.228)2.5517 (The 2.228 is from the t-table with df=10.)**

10) Write down the contrast you would use to compare the effect of the `A_solid` hormone treatment to the effect of no hormone treatment. **0 0 1 0 -1**

Part II - 1) Below is the data for a two-way ANOVA. There are two factors (factor A has A=2 levels, factor 2 has C=4 levels), and there are replications (n=2). NO random effects.

	Factor C				
Factor A	1	2	3	4	Fact 1 Means
1	2.31 y_{111}	2.20 y_{121}	1.85	2.31	2.20 $\bar{y}_{1\bullet\bullet}$
	2.49 y_{112}	2.36	1.92	2.17	
2	2.22 y_{211}	1.91	2.16	2.25	2.14
	2.09	1.99	2.34	2.16	
Factor 2 Means	2.28	2.11	2.07	2.22	2.17 $\bar{y}_{\bullet\bullet\bullet}$.

b) Write the model equation for the two-way ANOVA with interactions, and identify the parameters you used.

$y_{ijk}=\mu_{baseline}+\alpha_i+\gamma_j+(\alpha\gamma)_{ij}+\varepsilon_{ijk}$ **for $i=1,...2$, $j=1,...4$, and $k=1,...2$**

where the y_{ijk} are the observations, $\mu_{baseline}$ is the baseline

α_1, α_2 are the main effects for the levels of factor A

γ_1, γ_2, ... γ_4 are the main effects for the levels of factor C

$(\alpha\gamma)_{11}$, $(\alpha\gamma)_{12}$,... $(\alpha\gamma)_{21}$,... $(\alpha\gamma)_{24}$ are the interactions for the combinations of A and C

and the ε_{ijk} are the errors

c) Below is the incomplete set of formulas you would use to construct an ANOVA table for analyzing this problem.

Give the formula for SS_{wit} in terms of the x_{ijk} and the various sample means. $SSW = \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2$

Give the formula for SS_{AC} in terms of the other SS. **SSAC= SSB- SSA - SSC**
Give the degrees of freedom for Factor C in terms of A, C, and n. df= **c-1**
Give the F-statistic for testing that there is no interaction in terms of the MS. **F=MSAC/MSW**

2) The average number of flower heads per plant are counted for four species of wild flower in three different habitats. The data is reported as follows, and analyzed as a 2-way ANOVA.

	Species			
	A	B	C	D
Habitat 1	3.6	2.2	1.2	0.0
Habitat 2	2.6	1.5	2.0	0.0
Habitat 3	3.4	1.6	0.5	0.6

a) It is desired to say which of the habitats are significantly different from each other, and which we don't have sufficient evidence to say that they are significantly different from each other. **Holm's test performed on all pairs of habitats**

b) It is desired to see if species C generally has more flowers per plant than species B.
a contrast on the species with coefficients **0 1 -1 0**

c) It is desired to see if there is an interaction between species and habitat **cannot test on this data because there are no replications!**

d) It is desired to determine which species produce significantly more flowers per plants than the others. **Holm's test performed on all pairs of Species**

3) A two-factorial experiment was conducted concerning the affect of different qualities of drinking water on cattle. Four different qualities of water (the treatment) were used, and the experiment was conducted in both the Spring and Fall (season) with sixteen cattle used in each. The response variable was the weight of the cattle. The data and analysis using SAS are given on the attached pages.

a) Check the assumptions for performing this two-way ANOVA. Say how you checked them and whether they were satisfied.
From the normal probability plot we see that there is some slight non-normality at the extremes, but it is not too bad since the procedure is fairly robust. From the residual vs. predicted plot we see that the mean of the errors is approximately zero (but that is always true for a two-way ANOVA with interactions), and that the variances are of the errors seem approximately constant (which is reaffirmed by failing to reject the hypothesis in Levene's test). We are not told if there was random assignment of the cattle, and so can't judge the independence of the errors.

b) Report the single p-value for testing that there is no significant effect due to treatment, season, or interaction. **<.0001 from the ANOVA Table.**

c) For $\alpha_F = 0.05$, construct a display showing which treatment levels are significantly different from each other, and which are not.

Treatment 1	2.1050	A	
		A	
Treatment 3	1.8513	A	B
		A	B
Treatment 2	1.8363	A	B
			B
Treatment 4	1.6700		B

d) What about this data set makes it so that the differences between the treatment levels found in part c hold regardless of the season?

There is no significant interaction.