Questions from the 2nd practice exam given for Spring 2002

Part I:

1) In performing an ANOVA, what four assumptions must be satisfied?

2) Define what is meant by the p-value (or empirical significance level) of a test.

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?

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 _______ Fisher's LSD_______ Holm Test_______

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) (Circle all that apply) 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>unbalanced</u> / <u>random effect</u> / <u>without replication</u> / <u>factorial</u>**.

7) (Circle all that apply) 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>unbalanced</u> / <u>random effect</u> / <u>without replication</u> / <u>factorial</u>**.

8) (Circle all that apply) 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>unbalanced</u> / <u>random effect</u> / <u>without replication</u> / <u>factorial</u>**.

Problems 9 and 10 refer to the attached data set `growth` that concerns the effect of growth hormones for plants. There are two types of hormones (labeled `A` and `B`) provided to the plant as either `liquids` or `solids`. There is also a control sample that received no hormones, labeled simply `control`. The response is the `weight` in grams of the randomly assigned plants at the end of the treatment period.

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.

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.

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	2.20	1.85	2.31	2.20
	2.49	2.36	1.92	2.17	
2	2.22	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

a) On the above data set, using the notation from class, identify x_{111}, x_{112}, x_{211}, x_{121}, $\bar{y}_{1\bullet}$, and $\bar{y}_{\bullet\bullet}$.

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

$y_{ijk} =$

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

Give the formula for TSS in terms of the y_{ijk} and the various sample means. TSS=______________________________.

Give the formula for SSAC in terms of the other SS. SSAC=____________________.

Give the degrees of freedom for Factor C in terms of a, c, and n. df=______________________.

Give the F-statistic for testing that there is no interaction in terms of the MS. F=______________________.

Source	SS	df	MS
Between	$SSB = \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{k=1}^{n}(\bar{y}_{ij\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$	*ac-1*	$MSB = SSB/(ac-1)$
► Factor A	$SSA = \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{k=1}^{n}(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$	*a-1*	$MSA = SSA/(a-1)$
► Factor C	$SSC = \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{k=1}^{n}(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$		$MSC = SSC/(c-1)$
► AC Interaction	$SSAC =$	*(a-1)(c-1)*	$MSAC = SSAC/((a-1)(c-1))$
Within	$SSW = \sum_{i=1}^{a}\sum_{j=1}^{c}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij\bullet})^2$	*ac(n-1)*	$MSW = SSW/(ac(n-1))$
Total	$TSS =$	*acn-1*	

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

For each of the cases below, determine which test is appropriate:

- the overall p-value from the ANOVA table
- one of the type III tests (say which factor or interaction)
- a contrast (say which factor, and what the coefficients would be)
- Holm's test performed on all pairs of factor levels (say which factor)
- cannot be tested for this data-set

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.

b) It is desired to see if species C generally has more flowers per plant than species B.

c) It is desired to see if there is an interaction between species and habitat

d) It is desired to determine which species produce significantly more flowers per plants than the others.

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.

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

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

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?

```
DATA growth;
INPUT treatment $ weight @@;
CARDS;
A_liquid 21 B_liquid 15  A_solid 22 B_solid 10 Control 6
A_liquid 13 B_liquid 18  A_solid 19 B_solid 14 Control 11
A_liquid 20 B_liquid  9  A_solid 24 B_solid 21 Control 15
;

PROC GLM ORDER=DATA;
CLASS treatment;
MODEL weight=treatment;
ESTIMATE ‘A vs. B’  treatment  0.5 -0.5 0.5 -0.5 0;
RUN;
```

The GLM Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	208.4000000	52.1000000	2.67	0.0949
Error	10	195.3333333	19.5333333		
Corrected Total	14	403.7333333			

R-Square	Coeff Var	Root MSE	weight Mean
0.516182	27.85496	4.419653	15.86667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treatment	4	208.4000000	52.1000000	2.67	0.0949

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treatment	4	208.4000000	52.1000000	2.67	0.0949

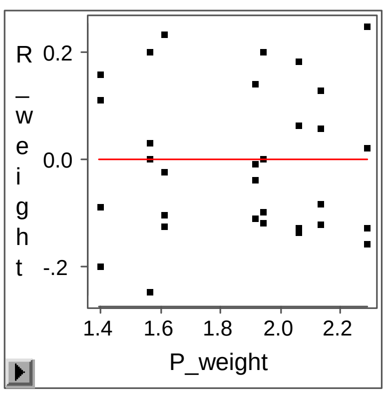
Parameter	Estimate	Standard Error	t Value	Pr > |t|
A vs. B	5.33333333	2.55168789	2.09	0.0631

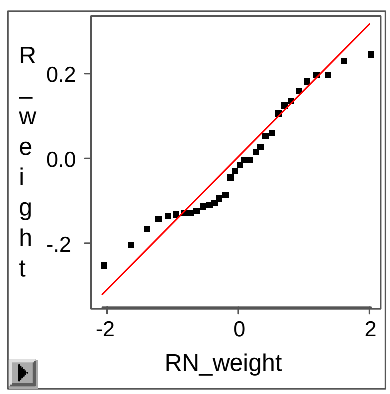
```
DATA beef;
INPUT treat $ season $  weight @@;
CARDS;
1      spring      1.81         2      spring      1.85
1      spring      1.88         2      spring      1.59
1      spring      2.06         2      spring      1.51
1      spring      1.91         2      spring      1.49
1      winter      2.14         2      winter      2.13
1      winter      2.32         2      winter      1.93
1      winter      2.17         2      winter      2.25
1      winter      2.55         2      winter      1.94
3      spring      1.77         4      spring      1.51
3      spring      1.60         4      spring      1.56
3      spring      1.57         4      spring      1.31
3      spring      1.32         4      spring      1.20
3      winter      2.06         4      winter      1.83
3      winter      2.20         4      winter      2.15
3      winter      2.27         4      winter      1.85
3      winter      2.02         4      winter      1.95
;

PROC INSIGHT;
OPEN beef;
FIT weight = treat season treat*season;
RUN;

DATA beef2;
SET beef;
KEEP block weight;
block = trim(treat)||trim(season);


PROC GLM DATA=beef2 ORDER=DATA;
CLASS block;
MODEL weight = block;
MEANS block / HOVTEST=bf;
RUN;
```





The GLM Procedure

Brown and Forsythe's Test for Homogeneity of weight Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
block	7	0.0155	0.00221	0.27	0.9604
Error	24	0.1981	0.00826		

```
PROC GLM DATA=beef ORDER=DATA;
CLASS treat season;
MODEL weight = treat season treat*season;
;

PROC MULTTEST DATA=beef ORDER=DATA HOLM;
CLASS treat;
CONTRAST ‘1 vs. 2’ 1 –1 0 0;
CONTRAST ‘1 vs. 3’ 1 0 –1 0;
CONTRAST ‘1 vs. 4’ 1 0 0 –1;
CONTRAST ‘2 vs. 3’ 0 1 –1 0;
CONTRAST ‘2 vs. 4’ 0 1 0 –1;
CONTRAST ‘3 vs. 4’ 0 0 1 –1;
TEST mean(weight);
RUN;
```

The GLM Procedure

Dependent Variable: weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	2.73193750	0.39027679	15.93	<.0001
Error	24	0.58785000	0.02449375		
Corrected Total	31	3.31978750			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
treat	3	0.77311250	0.25770417	10.52	0.0001
season	1	1.91101250	1.91101250	78.02	<.0001
treat*season	3	0.04781250	0.01593750	0.65	0.5903

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	3	0.77311250	0.25770417	10.52	0.0001
season	1	1.91101250	1.91101250	78.02	<.0001
treat*season	3	0.04781250	0.01593750	0.65	0.5903

Continuous Variable Tabulations

Variable	treat	NumObs	Mean	Standard Deviation
weight	1	8	2.1050	0.2471
weight	2	8	1.8363	0.2839
weight	3	8	1.8513	0.3381
weight	4	8	1.6700	0.3284

The Multtest Procedure
p-Values

Variable	Contrast	Raw	Stepdown Bonferroni
weight	1 vs. 2	0.0856	0.4278
weight	1 vs. 3	0.1035	0.4278
weight	1 vs. 4	0.0075	0.0447
weight	2 vs. 3	0.9215	0.9215
weight	2 vs. 4	0.2796	0.7183
weight	3 vs. 4	0.2394	0.7183