**Statistics 516 - Spring 2003 - Practice Exam 2**

**Part I: Answer the two following questions. Eight points each.**

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.

**Part II: Answer 12 of the following 13 questions. Seven Points each.**

1) Consider a one-way ANOVA with three factor levels red, blue, and green. Because the SSB for this ANOVA would have two-degrees of freedom we would need to use two dummy variables if we wanted to perform the ANOVA using dummy variables. Give an example of two dummy variables that would work here, being careful to specify when each would take the value zero or one.

Problems 2 refers to the partial analysis below `faces` that is based on an article that appeared in the Fall 1996 issue of the *Journal of Nonverbal Behavior*. A sample of 36 students was randomly divided into six groups and each group was assigned to view one of six slides showing a person making a facial expression. The six expressions were `Angry`, `Disgusted`, `Fearful`, `Happy`, `Sad`, or `Neutral`. After viewing the faces the students were asked to rate the degree of dominance they inferred from the facial expressions (a scale ranging from -15 to 15).

```
DATA faces;
INPUT expression $ dominance @@;
CARDS;
Angry        2.10         Angry  0.64         Angry        0.47
Angry        0.37         Angry        1.62         Angry  -0.08
Disgusted    0.40         Disgusted    0.73         Disgusted    -0.07
Disgusted    -0.25        Disgusted    0.89         Disgusted    1.93
Fearful      0.82         Fearful      -2.93        Fearful      -0.74
Fearful      0.79         Fearful      -0.77        Fearful      -1.60
Happy        1.71         Happy        -0.04        Happy        1.04
Happy        1.44         Happy        1.37         Happy        0.59
Sad          0.74         Sad          -1.26        Sad          -2.27
Sad          -0.39        Sad          -2.65        Sad          -0.44
Neutral      1.69         Neutral      -0.60        Neutral      -0.55
Neutral      0.27         Neutral      -0.57        Neutral      -2.16
;

PROC GLM ORDER=DATA;
CLASS expression;
MODEL dominance = expression;
ESTIMATE ‘Angry vs. Disgusted’   expression  1    -1    0     0 0    0;
RUN;
```

The GLM Procedure

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 23.08522222 | 4.61704444 | 3.96 | 0.0071 |
| Error | 30 | 34.98700000 | 1.16623333 | | |
| Corrected Total | 35 | 58.07222222 | | | |

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Angry vs. Disgusted | 0.24833333 | 0.62349374 | 0.40 | 0.6932 |

2) Construct a 95% confidence interval for the difference between the true average dominance rating of the angry and disgusted groups.

Problems 3 and 4 refer to the following partial analysis below. Seven different types of material (labeled A-F) were sent out to a sample of 13 laboratories for stress testing (since different laboratories use different testing methods). The `PROC GLM` code used was:

```
PROC GLM;
CLASS lab material;
MODEL stress = lab material lab*material;
RANDOM lab lab*material;
RUN;
```

And the output was:

```
                                         Sum of
Source                       DF         Squares     Mean Square    F Value    Pr > F
Model                        90     322913.2482       3587.9250     177.01    <.0001
Error                       273       5533.5800         20.2695
Corrected Total             363     328446.8282

Source                       DF     Type III SS     Mean Square    F Value    Pr > F
lab                          12      30328.0547       2527.3379     124.69    <.0001
material                      6     268778.0771      44796.3462    2210.03    <.0001
lab*material                 72      23807.1165        330.6544      16.31    <.0001

    Source                   Type III Expected Mean Square
    lab                      Var(Error) + 4 Var(lab*material) + 28 Var(lab)
    material                 Var(Error) + 4 Var(lab*material) + Q(material)
    lab*material             Var(Error) + 4 Var(lab*material)
```

3) Find the value of the F statistic for testing that $\sigma^2_{lab}=0$ against $\sigma^2_{lab} >0$.

4) Find an estimate of $\sigma^2_{lab}$.

Problems 5 through 10 refer to the following data for a two-way ANOVA and SAS output shown below. There are two factors (factor A has a=3 levels, factor C has c=4 levels), and there are replications (n=2). NO random effects.

| | Factor C | | | | |
|---|---|---|---|---|---|
| Factor A | 1 | 2 | 3 | 4 | Factor A Means |
| 1 | 26.4 | 38.3 | 40.5 | 19.8 | 33.1 |
| | 32.6 | 31.7 | 46.9 | 28.7 | |
| 2 | 34.0 | 27.7 | 40.3 | 32.9 | 34.4 |
| | 20.7 | 37.2 | 44.7 | 37.6 | |
| 3 | 43.8 | 54.4 | 49.0 | 43.8 | 47.7 |
| | 49.4 | 50.6 | 44.3 | 46.6 | |
| Factor C Means | 34.5 | 40.0 | 44.3 | 34.9 | 38.4 |

**From PROC INSIGHT**

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| Model | 11 | 1767.3413 | 160.6674 | 6.57 | 0.0015 |
| Error | 12 | 293.2450 | 24.4371 | | |
| C Total | 23 | 2060.5862 | | | |

Type III Tests

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| A | 2 | 1049.9700 | 524.9850 | 21.48 | 0.0001 |
| C | | | 129.4204 | 5.30 | 0.0148 |
| A*C | 6 | 329.1100 | 54.8517 | 2.24 | 0.1099 |

**From PROC MULTTEST**

Continuous Variable Tabulations

| Variable | A | NumObs | Mean | Standard Deviation |
|---|---|---|---|---|
| value | A1 | 2 | 29.5000 | 4.3841 |
| value | A2 | 2 | 27.3500 | 9.4045 |
| value | A3 | 2 | 46.6000 | 3.9598 |

p-Values

| Variable | Contrast | Raw | Stepdown Bonferroni |
|---|---|---|---|
| value | A1 vs. A2 | 0.7595 | 0.7595 |
| value | A1 vs. A3 | 0.0759 | 0.1727 |
| value | A2 vs. A3 | 0.0576 | 0.1727 |

5) Justify that the above design is factorial, balanced, with replications.

6) On the above data set, using the notation from class, identify $y_{111}$, $y_{112}$, $y_{121}$, $y_{342}$, $\bar{y}_{\bullet 1 \bullet}$, and $\bar{y}_{\bullet\bullet\bullet}$.

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

$y_{ijk}$ =

8) The DF and SS for Factor C were deleted. What values should they have?

SS= df=

9) Use the Holm procedure to construct a display showing which levels of Factor A are significantly different from each other at a family-wise $\alpha_T$=0.10.

10) 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 test whether the Factor C has any effect on average on the output values.

b) It is desired to test whether Factor C has the same effect on the output values regardless of the level of Factor A.

c) It is desired to test whether level 1 of Factor C differs from level 4 of Factor C.

d) It is desired to simultaneously test whether there is an effect due to Factor A, Factor C, or an interaction.

Problems 11-13 use the attached partial analysis of the data set `vitaminb`. It is similar to results reported in the July 1995 issue of *Journal of Nutrition*. It concerns the effect of a vitabin B supplement on the weights of the kidneys of Zucker rats. Half of the rats were classified as `obese`, and half were classified as `lean`. The two groups of rats were then randomly assigned to receive either the regular diet or the diet with the vitamin b supplement. At the end of twenty weeks, the weights of the rats' kidneys were measured in grams.

11) Check the assumptions for performing this two-way ANOVA. Say how you checked them and whether they were satisfied.

12) Identify which group was used as the baseline group.

13) Consider the three p-values from the box of Type III Tests. Assuming that all of the assumptions are met, use an $\alpha$=0.05 for each of these three tests and report the conclusion you can draw from each one of them. Phrase your conclusions in terms of what the scientists were looking for in the problems. (e.g. There is/is not a significant effect on the kidney weight due to the choice of diet.)

```
DATA vitaminb;
INPUT diet $ size $   kidney @@;
CARDS;
Regular     Lean  1.62 Regular     Lean  1.47 Regular     Lean  1.80
Regular     Lean  1.37 Regular     Lean  1.71 Regular     Lean  1.71
Regular     Lean  1.81
Bsupp       Lean  1.51 Bsupp       Lean  1.63 Bsupp       Lean  1.65
Bsupp       Lean  1.35 Bsupp       Lean  1.45 Bsupp       Lean  1.66
Bsupp       Lean  1.44
Regular     Obese 2.35 Regular     Obese 2.84 Regular     Obese 2.97
Regular     Obese 2.05 Regular     Obese 2.54 Regular     Obese 2.82
Regular     Obese 2.93
Bsupp       Obese 2.93 Bsupp       Obese  2.63 Bsupp      Obese 2.72
Bsupp       Obese  2.61 Bsupp      Obese 2.99 Bsupp       Obese  2.64
Bsupp       Obese 2.19
;

PROC INSIGHT;
OPEN vitaminb;
FIT kidney = diet size diet*size;
RUN;
```

Analysis of Variance

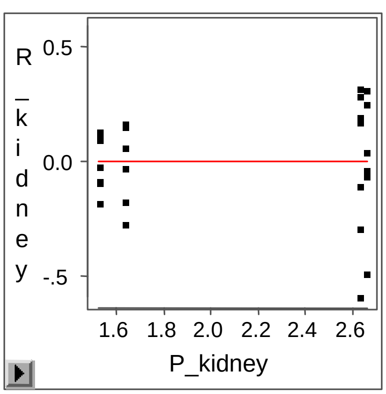
| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 8.1168 | 2.7056 | 47.34 | <.0001 |
| Error | 24 | 1.3715 | 0.0571 | | |
| C Total | 27 | 9.4883 | | | |

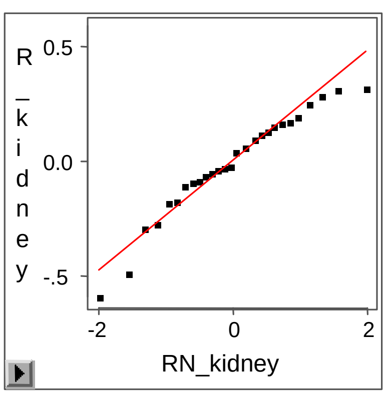
Type III Tests

| Source | DF | Sum of Squares | Mean Square | F Stat | Pr > F |
|---|---|---|---|---|---|
| diet | 1 | 0.0124 | 0.0124 | 0.22 | 0.6451 |
| size | 1 | 8.0679 | 8.0679 | 141.18 | <.0001 |
| diet*size | 1 | 0.0364 | 0.0364 | 0.64 | 0.4324 |

Parameter Estimates

| Variable | diet | size | DF | Estimate | Std Error | t Stat | Pr >\|t\| | Tolerance | Var Inflation |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | | | 1 | 2.6429 | 0.0904 | 29.25 | <.0001 | . | 0 |
| diet | Bsupp | | 1 | 0.0300 | 0.1278 | 0.23 | 0.8164 | 0.5000 | 2.0000 |
| | Regular | | 0 | 0 | . | . | . | . | . |
| size | | Lean | 1 | -1.0014 | 0.1278 | -7.84 | <.0001 | 0.5000 | 2.0000 |
| | | Obese | 0 | 0 | . | . | . | . | . |
| diet*size | Bsupp | Lean | 1 | -0.1443 | 0.1807 | -0.80 | 0.4324 | 0.3333 | 3.0000 |
| | Bsupp | Obese | 0 | 0 | . | . | . | . | . |
| | Regular | Lean | 0 | 0 | . | . | . | . | . |
| | Regular | Obese | 0 | 0 | . | . | . | . | . |



```
DATA vitb2;
SET vitaminb;
KEEP kidney block;
block = trim(diet)||trim(size);


PROC GLM DATA=vitb2 ORDER=DATA;
CLASS block;
MODEL kidney = block;
MEANS block / HOVTEST=bf;
RUN;
```

The GLM Procedure

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

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| block | 3 | 0.1018 | 0.0339 | 1.04 | 0.3945 |
| Error | 24 | 0.7860 | 0.0328 | | |