EXPERIMENTAL STATISTICS 7005 NAME ____
April 15, 1999 EXAM 3 All sections : Geaghan

Read Carefully. Give an answer in the form of a number or numeric expression where possible. Show all calculations for possible partial credit. Use a value of 0.05 for α if not specified. NO tables are provided, you should have them all memorized by now.

1) 3 points – An Analysis of Variance (ANOVA) with only two treatment levels can also be done as a two-sample t-test. There is one assumption made in ANOVA that is more easily handled in a two-sample t-test by using alternate calculations when the assumption is shown to be violated. Which assumption is that? (Choose one).

a) Normality of ε_i.

b) Homogeneity of ε_i.

c) The calculations were done correctly.

d) Independence of the ε_i.

2) 3 points – Which of the following is a test of homogeneity of variance?

a) F test of Mean Square Treatments with Means square error

b) Shapiro-Wilk's test

c) Hartley's F_{max} test

d) Chi square test of Independence

3) 3 points – Which assumption for ANOVA is usually tested with PROC UNIVARIATE in SAS. Which assumptions are those? (Choose two).

a) Normality of ε_i.

b) Homogeneity of ε_i.

c) The calculations were done correctly.

d) Independence of the ε_i.

4) 21 points – Answer TRUE (T) or FALSE (F) to the following questions.

a) _____ Effects are RANDOM when all of the effects of interest are included in the experiment.

b) _____ Analyses of Variance (ANOVAs) must be balanced (i.e. have an equal number of observations in each treatment level).

c) _____ Each contrast has a Type I error rate of α on each individual test, just like the LSD.

d) _____ R. A. Fisher derived the LSD post-ANOVA test from the two-sample t-test.

e) _____ Bonferroni's post ANOVA test gives an exact probability of error for any specified number of tests or comparisons.

f) _____ Dunnett's post ANOVA test compares a control treatment to the other treatments with a α probability of error for those tests.

g) _____ The analysis we call "Analysis of Variance" was developed by Bonferroni.

5) 21 points – There are many choices among the multiple range tests for ANOVA, each with its own characteristics. For each characteristic given below circle the one range test that best fits the description.

a) The least powerful is? **LSD Tukey's Scheffé's Bonferroni's**

b) The most powerful is? **LSD Tukey's Scheffé's Bonferroni's**

c) Which one gives a lower bound on the probability of Type I error for a predetermined number of comparisons? **LSD Tukey's Scheffé's Bonferroni's**

d) Which test did R.A. Fisher develop? **LSD Tukey's Scheffé's Bonferroni's**

e) Which is intended for all pairwise tests? **LSD Tukey's Scheffé's Bonferroni's**

f) Which one is most likely to make a Type I error? **LSD Tukey's Scheffé's Bonferroni's**

g) Which one allows all possible pairwise tests and keeps an overall TYPE I error rate equal to α? **LSD Tukey's Scheffé's Bonferroni's**

6) 4 points – A fisheries biologist is experimenting with 5 different trap types to catch Blue Crabs. He builds 6 traps of each type (30 traps in all) and fishes each trap in the same area for two weeks. The dependent variable for the analysis is the total catch of crabs over the two weeks period. In the space below give the appropriate sources (including the Total) and numeric degrees of freedom for this Analysis of Variance.

Source	**numeric d.f.**
________________________	__________
________________________	__________
________________________	__________
________________________	__________
________________________	__________
________________________	__________

7) 6 points – Four treatment combinations were studied in an experiment on the effect of drugs and exercise on blood pressure. The four treatments were (a) NO exercise and NO drug (control group), (b) exercise only, (c), drug only, (d) drug and exercise both. Write contrasts below to test the effects named.

	control	exercise	drug	exercise+drug
a) Effect of exercise	________	________	________	________
b) Effect of drug	________	________	________	________
c) Interaction of exercise and drug	________	________	________	________

For each analysis described below, name the type of analysis, the degrees of freedom for the test (both numerator and denominator for F tests) and any other indicated question. The types of analysis will be CRD with *a priori* treatments (One-way ANOVA) and CRD with factorial treatments (Two-way ANOVA).

8) 7 points – A professor in the marketing department wants to know if supermarket sales can be increased if the store lighting is increased. A large chain or supermarkets is upgrading its facilities and has agreed to accommodate his study with 42 of its stores. He randomly allocates 14 stores to each of 3 treatments. The treatments are (a) normal florescent lighting, (b) florescent lighting increased 50% by adding higher watt bulbs and extra fixtures, (c) increase lighting by 50 percent **and** add sodium lamps to side areas. All changes were made in February, and the dependent variable is total sales by each supermarket in March.

a) What type of analysis would this be?

Circle one: [CRD with *a priori*] [CRD with factorial]

b) What are the degrees of freedom for the test of treatments?

Circle one: [1, 13] [2, 13] [1, 39] [2, 39] [3, 39] [1, 41] [2, 41] [3, 41]

c) What is the experimental unit?

Circle one: [the chain] [a supermarket] [light fixture] [the sales] [increased light]

d) What is the sampling unit?

Circle one: [the chain] [a supermarket] [light fixture] [the sales] [increased light]

10) 7 points – A horticulture student is working on a project to increase the diameter of commercially grown roses. She is working on two varieties of roses and is also testing two different commercial fertilizers. She has 24 plants of each variety (48 total) and randomly allocates 12 to each fertilizer treatment, so there are four combinations of treatments and fertilizer: variety A with fertilizer P, variety A with fertilizer M, Variety B with fertilizer P and variety B with fertilizer M. Three flowers are measured on each plant, and the dependent variable is the diameter of a flower.

a) What type of analysis would this be?

Circle one: [CRD with *a priori*] [CRD with factorial]

b) What are the degrees of freedom for the test of 4 treatment combinations together?

Circle one: [1, 11] [3, 11] [1, 44] [3, 44] [4, 44] [1, 47] [3, 47] [4, 47]

c) What is the experimental unit?

Circle one: [a greenhouse] [a fertilizer] [a plant] [the diameter] [a flower]

c) What is the experimental unit?

Circle one: [a greenhouse] [a fertilizer] [a plant] [the diameter] [a flower]

11) The questions below refer to SAS output given separately with a description of the problem. The program log is given at the end of the exam.

a) 3 points – What kind of design and treatment arrangement (with indication of number of levels) is done by this ANOVA?

a) CRD with a priori treatments (one-way ANOVA), Treatment levels = _________ .

b) CRD with a factorial treatments (two-way ANOVA), Treatment levels = _________ .

a) RBD with a priori treatments (two-way ANOVA), Treatment levels = _________ .

b) RBD with a factorial treatments (three-way ANOVA), Treatment levels = _________ .

b) 3 points – Give the linear model for this design (use appropriate notation and subscripting).

__________ = ____

c) 3 points – Does the assumption of normality appear to be met?

Circle one: Yes No Not enough information to evaluate

If you answer Yes, give the P value for the test ________________________________ .

If you answer No, tell me below what part of the output you checked.

If you answer "Not enough information", tell me what output you need.

d) 3 points – If the investigator was interested only in all pairwise comparisons, and he wanted a TYPE I error rate of α, what post-ANOVA output should he be looking at?

Under those conditions he should use ____ .

e) 3 points – Is the mean of the first two varities in this experiment (1 and 2) different from the mean of the last 5 (3, 4, 5, 6 and 7)?

Circle one: Yes No In either case, the P value = ________________ .

```
1          *********************************************;
2          *** Exam 3 Example                        ***;
3          *********************************************;
4
5          OPTIONS PS=61 LS=78 NOCENTER NODATE PAGENO=1;
6          DATA ONE;  INFILE CARDS MISSOVER;
7             INPUT Treat $ Y_Value;
8                TITLE1 'Exam 3 Example';
9          CARDS;
NOTE: The data set WORK.ONE has 63 observations and 2 variables.
NOTE: The DATA statement used 0.22 seconds.
9                             RUN;
73         ;
74         PROC GLM DATA=ONE; CLASSES Treat;
75           MODEL Y_Value = Treat / SS3;
76           MEANS Treat  / LSD BON TUKEY SCHEFFE;
77         ****** Contrasts ====> Tasks are:   1    2    3   4   5   6   7;
78                   CONTRAST 'A '  Treat     -5   -5    2   2   2   2   2;
79                   CONTRAST 'B '  Treat     -2    1    1   0   0   0   0;
80                   CONTRAST 'C '  Treat      0    0   -2  -2  -2   3   3;
81                   CONTRAST 'D '  Treat      0    0   -4   1   1   1   1;
82                   CONTRAST 'E '  Treat      0    0    0   0  -2   1   1;
83           OUTPUT OUT=NEXT RESIDUALS=E;
84          RUN;
84               QUIT;
NOTE: The data set WORK.NEXT has 63 observations and 3 variables.
NOTE: The PROCEDURE GLM printed pages 1-7.
NOTE: The PROCEDURE GLM used 0.77 seconds.
85         PROC UNIVARIATE DATA=NEXT NORMAL PLOT; VAR E; RUN;
NOTE: The PROCEDURE UNIVARIATE printed pages 8-9.
NOTE: The PROCEDURE UNIVARIATE used 0.0 seconds.
```

An agronomist is evaluating seven different varieties of corn on standard sized plots. In addition to all pairwise tests among the varieties, he has some particular contrasts he is interested in conducting. Questions concerning this output and the program is are on the exam, the computer output is below.

```
Exam 3 Example                                                                     1

General Linear Models Procedure
Class Level Information
Class    Levels    Values
TREAT         7    1 2 3 4 5 6 7

Number of observations in data set = 63


General Linear Models Procedure

Dependent Variable: Y_VALUE
                                    Sum of           Mean
Source                 DF          Squares         Square   F Value     Pr > F
Model                   6      1965.602222     327.600370     94.58     0.0001
Error                  56       193.960000       3.463571
Corrected Total        62      2159.562222

                  R-Square            C.V.       Root MSE         Y_VALUE Mean
                  0.910186        4.619307       1.861067             40.28889

Source                 DF      Type III SS    Mean Square   F Value     Pr > F
TREAT                   6      1965.602222     327.600370     94.58     0.0001



General Linear Models Procedure

T tests (LSD) for variable: Y_VALUE
NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.
Alpha= 0.05  df= 56  MSE= 3.463571
Critical Value of T= 2.00
Least Significant Difference= 1.7575
Means with the same letter are not significantly different.

T Grouping              Mean      N  TREAT
                      A           50.1333      9   5

                      B           44.5000      9   6
                      B
                      B           43.0111      9   2

                      C           39.6000      9   1

                      D           37.1333      9   4

                      E           34.8889      9   3

                      F           32.7556      9   7
```

```
General Linear Models Procedure

Tukey's Studentized Range (HSD) Test for variable: Y_VALUE
NOTE: This test controls the type I experimentwise error rate, but generally
      has a higher type II error rate than REGWQ.
Alpha= 0.05  df= 56  MSE= 3.463571
Critical Value of Studentized Range= 4.325
Minimum Significant Difference= 2.6828
Means with the same letter are not significantly different.

Tukey Grouping              Mean      N  TREAT
                  A           50.1333      9  5

                  B           44.5000      9  6
                  B
                  B           43.0111      9  2

                  C           39.6000      9  1
                  C
          D       C           37.1333      9  4
          D
          D       E           34.8889      9  3
                  E
                  E           32.7556      9  7


Bonferroni (Dunn) T tests for variable: Y_VALUE
NOTE: This test controls the type I experimentwise error rate, but generally
      has a higher type II error rate than REGWQ.
Alpha= 0.05  df= 56  MSE= 3.463571
Critical Value of T= 3.18
Minimum Significant Difference= 2.7923
Means with the same letter are not significantly different.

Bon Grouping              Mean      N  TREAT
                    A           50.1333      9  5

                    B           44.5000      9  6
                    B
                    B           43.0111      9  2

                    C           39.6000      9  1
                    C
            D       C           37.1333      9  4
            D
            D       E           34.8889      9  3
                    E
                    E           32.7556      9  7
```

```
Scheffe's test for variable: Y_VALUE
NOTE: This test controls the type I experimentwise error rate but generally
      has a higher type II error rate than REGWF for all pairwise comparisons
Alpha= 0.05  df= 56  MSE= 3.463571
Critical Value of F= 2.26557
Minimum Significant Difference= 3.2346
Means with the same letter are not significantly different.

Scheffe Grouping              Mean      N  TREAT
                A           50.1333      9  5

                B           44.5000      9  6
                B
                B           43.0111      9  2

                C           39.6000      9  1
                C
        D       C           37.1333      9  4
        D
        D       E           34.8889      9  3
                E
                E           32.7556      9  7


General Linear Models Procedure

Dependent Variable: Y_VALUE
Contrast               DF     Contrast SS    Mean Square   F Value     Pr > F
A                       1      26.0470000     26.0470000      7.52     0.0082
B                       1       2.5350000      2.5350000      0.73     0.3959
C                       1      47.2089259     47.2089259     13.63     0.0005
D                       1     280.5005000    280.5005000     80.99     0.0001
E                       1     794.2668519    794.2668519    229.32     0.0001


Univariate Procedure
Variable=E
                 Moments
 N                63  Sum Wgts         63
 Mean              0  Sum               0
 Std Dev    1.768725  Variance   3.128387
 Skewness   -0.20646  Kurtosis   -0.51571
 USS          193.96  CSS          193.96
 CV                .  Std Mean   0.222838
 T:Mean=0          0  Pr>|T|       1.0000
 Num ^= 0         63  Num > 0          33
 M(Sign)         1.5  Pr>=|M|      0.8013
 Sgn Rank         30  Pr>=|S|      0.8392
 W:Normal   0.963695  Pr<W         0.1444

            Quantiles(Def=5)
 100% Max  4.244444       99%  4.244444
  75% Q3   1.488889       95%  2.411111
  50% Med       0.2       90%       2.1
  25% Q1   -1.33333       10%      -2.5
   0% Min  -3.45556        5%      -3.2
                           1%  -3.45556
 Range          7.7
 Q3-Q1     2.822222
 Mode      -0.41111
```

```
                Extremes
   Lowest    Obs     Highest    Obs
 -3.45556(      57) 2.366667(      41)
 -3.38889(      21) 2.411111(      25)
 -3.33333(      44)      2.6(      47)
     -3.2(       2) 2.611111(      24)
 -2.95556(      58) 4.244444(      63)

  Stem Leaf                     #  Boxplot
     4 2                        1     |
     3                                |
     3                                |
     2 66                       2     |
     2 001244                   6     |
     1 5556899                  7  +-----+
     1 0123                     4  |     |
     0 66799                    5  |     |
     0 22223333                 8  *--+--*
    -0 4433211110              10  |     |
    -0 75                       2  |     |
    -1 321                      3  +-----+
    -1 955                      3     |
    -2 32220                    5     |
    -2 85                       2     |
    -3 4320                     4     |
    -3 5                        1     |
       ----+----+----+----+


                      Normal Probability Plot
   4.25+                                               *++
       |                                            +++
       |                                         +++
       |                                       ++* *
       |                                    +*+**
       |                               *****
       |                             ***+
       |                           ***
   0.25+                        ***
       |                   *****
       |                   *++
       |                ***
       |             ++**
       |          +****
       |        +* *
       |  *   *+**
  -3.75+   +++
        +----+----+----+----+----+----+----+----+----+----+
            -2        -1         0        +1        +2
```