Stat 541 Homework #6

Kenny Flagg

April 4, 2016

1. *BIBD for wool cleaning agents.*

   (a) *Verify that this study has the appropriate structure for a BIBD.*

   There are $a = 4$ treatments, $b = 16$ blocks, $k = 3$ treatments applied to each block. Each treatment appears in $r = 12$ blocks and each pair of treatments appears in $\lambda = 8$ blocks. Then
   $$\frac{r(k-1)}{a-1} = \frac{(12)(2)}{3} = 8 = \lambda$$
   so this a balanced incomplete block design.

   (b) *State the model and the hypotheses to be tested.*

   The model is
   $$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}; \qquad \epsilon_{ij} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$$
   where $i = 1, 2, 3, 4$ indexes treatments A, B, C, D, and $j = 1, 2, \ldots, 16$ indexes the batches.

   The hypotheses being tested are
   $$H_0\colon\ \tau_1 = \tau_2 = \tau_3 = \tau_4,$$
   $$H_1\colon\ \tau_i \neq \tau_{i'} \text{ for some } i \neq i'.$$

   The ANOVA table produced by SAS apears on the next page. The test statistic is $F = 8.80$ with p-value $= 0.0003$, so at $\alpha = 0.05$ there is enough evidence to conclude that there are differences in mean weight loss among the different cleaning agents.

```
                         Problem 1: BIBD for Wool Cleaning

                                The GLM Procedure

                          Dependent Variable: clean

                                         Sum of
Source                      DF          Squares    Mean Square   F Value   Pr > F

Model                       18      289.4352083     16.0797338     11.48   <.0001

Error                       29       40.6295833      1.4010201

Corrected Total             47      330.0647917


           R-Square     Coeff Var      Root MSE    clean Mean

           0.876904      6.189678      1.183647      19.12292


Source                      DF      Type III SS    Mean Square   F Value   Pr > F

batch                       15      249.2495833     16.6166389     11.86   <.0001
process                      3       36.9970833     12.3323611      8.80   0.0003
```

(c) *Perform Bonferroni's multiple comparison test using both the sample means and the least squares means. For each case, state which pairs of means are considered statistically significantly different.*

Since there is evidence that the treatment means are not all equal, we can perform Bonferroni's multiple comparisons procedure to test the hypotheses

$$H_0\colon \mu_i = \mu_{i'},$$
$$H_1\colon \mu_i \neq \mu_{i'}$$

for each pair, $i \neq i'$.

SAS output appears on the next page. When applied to the sample means, at $\alpha = 0.05$, we conclude that the mean of B is significantly different from the means of A, C, and D. We have little evidence that there are differences among the means of any of A, C, and D.

At $\alpha = 0.05$, the Bonferroni-adjusted p-values for comparisons of the least squares means indicate significant differences between A and C, A and D, B and C, and B and D.

```
                    Problem 1: BIBD for Wool Cleaning

                           The GLM Procedure

                  Bonferroni (Dunn) t Tests for clean

NOTE: This test controls the Type I experimentwise error rate, but it
      generally has a higher Type II error rate than REGWQ.


              Alpha                                  0.05
              Error Degrees of Freedom                 29
              Error Mean Square                   1.40102
              Critical Value of t                 2.83155
              Minimum Significant Difference       1.3683


      Means with the same letter are not significantly different.


        Bon Grouping          Mean      N    process

                   A       20.7000     12    B

                   B       18.7417     12    A
                   B
                   B       18.5500     12    D
                   B
                   B       18.5000     12    C

                    Problem 1: BIBD for Wool Cleaning

                           The GLM Procedure
                          Least Squares Means
              Adjustment for Multiple Comparisons: Bonferroni

                                                 LSMEAN
                  process    clean LSMEAN      Number

                  A            20.0135417           1
                  B            20.0885417           2
                  C            18.2947917           3
                  D            18.0947917           4


                 Least Squares Means for effect process
                 Pr > |t| for H0: LSMean(i)=LSMean(j)

                      Dependent Variable: clean

      i/j             1             2             3             4

        1                      1.0000        0.0134        0.0048
        2        1.0000                      0.0092        0.0032
        3        0.0134        0.0092                      1.0000
        4        0.0048        0.0032        1.0000
```

(d) *What are the sample mean and lsmean values for Agent A, and why are they so different?*

From the SAS output in part (c), the sample mean for Agent A is $\bar{y}_{1\cdot} = 18.7417$ mg and the least squares mean for Agent A is $\widehat{\mu}_1 = 20.0135$ mg. The sample mean is low because the Agent A appears in blocks with low response values. The least squares mean adjusts for the blocks.

(e) *Find estimates of the agent effects assuming the sum of the effects equals 0.*

Under the sum-to-zero constraint, the estimated effects are $\widehat{\tau}_1 = 0.8906$, $\widehat{\tau}_2 = 0.9656$, $\widehat{\tau}_3 = -0.8281$, and $\widehat{\tau}_4 = -1.0281$.

```
                    Problem 1: BIBD for Wool Cleaning

                           The GLM Procedure

                      Dependent Variable: clean

                                           Standard
Parameter                    Estimate         Error    t Value    Pr > |t|

Agent A effect             0.89062500    0.31386180       2.84      0.0082
Agent B effect             0.96562500    0.31386180       3.08      0.0045
Agent C effect            -0.82812500    0.31386180      -2.64      0.0133
Agent D effect            -1.02812500    0.31386180      -3.28      0.0027
```

2. *Factorial experiment for effects of sulphur and nitrogen on red clover growth. Analyze the data and assess the validity of the model assumptions. Include a Bonferroni multiple comparison test comparing the 8 cell means.*

Model and F-Test

The interaction model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad \epsilon_{ijk} \overset{iid}{\sim} \mathrm{N}(0\sigma^2)$$

where $i = 1, 2$ indexes the nitrogen levels, $j = 1, 2, 3, 4$ indexes the level of sulfur, and $k = 1, 2, 3$ indexes the replicates in each combination of nitrogen and sulfur.

We start by testing for interaction effects,

$$H_0\colon\ (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{24},$$
$$H_1\colon\ \text{some } (\alpha\beta)_{ij} \neq (\alpha\beta)_{i'j'}.$$

The ANOVA table shows a test statistic of $F = 349.78$ with p-value < 0.0001. At $\alpha = 0.05$ there is enough evidence to conclude that there are interaction effects and that the effect of nitrogen on clover growth depends upon the sulfur level.

```
                   Factorial Design for Clover Growth

                           The GLM Procedure

                      Dependent Variable: growth

                                          Sum of
Source                      DF         Squares    Mean Square   F Value   Pr > F

Model                        7     14.66146250     2.09449464    584.51   <.0001

Error                       16      0.05733333     0.00358333

Corrected Total             23     14.71879583


           R-Square     Coeff Var      Root MSE    growth Mean

           0.996105      1.058471      0.059861       5.655417


Source                      DF     Type III SS    Mean Square   F Value   Pr > F

nitrogen                     1      7.83183750     7.83183750   2185.63   <.0001
sulfur                       3      3.06947917     1.02315972    285.53   <.0001
nitrogen*sulfur              3      3.76014583     1.25338194    349.78   <.0001
```

Bonferroni's MCP

Since the ANOVA F-test provides evidence of interactions, it is of interest to perform pariwise comparisons of the mean growth for all pairs of nitrogren-sulphur combinations,

$$H_0\colon\ \mu_{ij} = \mu_{i'j'},$$
$$H_1\colon\ \mu_{ij} \neq \mu_{i'j'}.$$

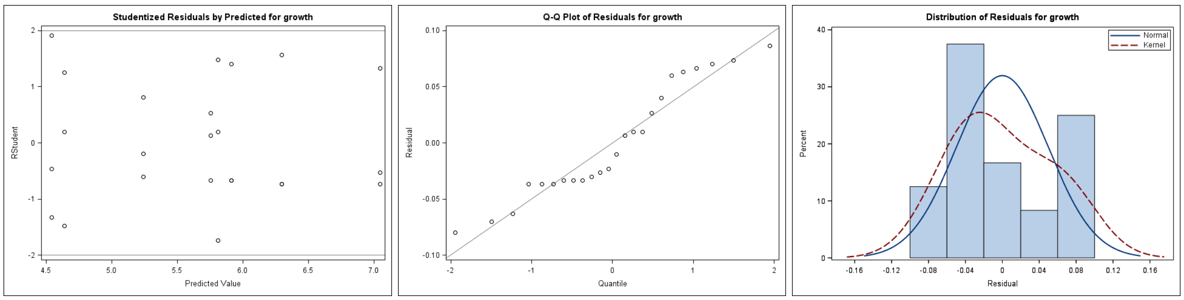
SAS output with Bonferroni-adjusted p-values appears below. At $\alpha = 0.05$, we find significant differences between the means of all pairs except

- 0 lbs/acre nitrogen, 0 lbs/acre sulfur vs 0 lbs/acre nitrogen, 3 lbs/acre sulfur,
- 0 lbs/acre nitrogen, 9 lbs/acre sulfur vs 20 lbs/acre nitrogen, 0 lbs/acre sulfur,
- 0 lbs/acre nitrogen, 9 lbs/acre sulfur vs 20 lbs/acre nitrogen, 6 lbs/acre sulfur, and
- 20 lbs/acre nitrogen, 0 lbs/acre sulfur vs 20 lbs/acre nitrogen, 6 lbs/acre sulfur.

```
                  Factorial Design for Clover Growth

                          The GLM Procedure
                         Least Squares Means
               Adjustment for Multiple Comparisons: Bonferroni

                                          growth      LSMEAN
               nitrogen     sulfur        LSMEAN      Number

               0            0         4.54333333           1
               0            3         4.64000000           2
               0            6         5.24000000           3
               0            9         5.91333333           4
               20           0         5.75333333           5
               20           3         7.04666667           6
               20           6         5.81000000           7
               20           9         6.29666667           8


               Least Squares Means for effect nitrogen*sulfur
                    Pr > |t| for H0: LSMean(i)=LSMean(j)

                      Dependent Variable: growth

        i/j              1             2             3             4

          1                       1.0000        <.0001        <.0001
          2         1.0000                      <.0001        <.0001
          3         <.0001        <.0001                      <.0001
          4         <.0001        <.0001        <.0001
          5         <.0001        <.0001        <.0001        0.1338
          6         <.0001        <.0001        <.0001        <.0001
          7         <.0001        <.0001        <.0001        1.0000
          8         <.0001        <.0001        <.0001        <.0001
```

Model Assumptions and Cramer-Von Mises Test

The plot of studentized residuals against their fitted values shows no patterns indicating non-homogeneous variance. However, the Normal Q-Q plot and the histogram of residuals suggest violations of the Normality assumption. A test of Normailty will be performed on the residuals.



The Cramer-Von Mises Goodness of Fit Test tests the hypotheses

H_0: The residuals follow a Normal CDF,

H_1: The residuals do not follow a Normal CDF.

Output from PROC UNIVARIATE appears below. The Cramer-Von Mises statistic is $W^2 = 0.1319$ with p-value $= 0.0402$. At $\alpha = 0.05$, we reject H_0 and conclude that the distibutions of the residuals is significantly different from Normal. The Normality assumption is violated and the ANOVA model is inappropriate for these data.

```
                    Factorial Design for Clover Growth

                         The UNIVARIATE Procedure
                             Variable:  resid

                           Tests for Normality

Test                  --Statistic---    -----p Value------

Shapiro-Wilk          W      0.923164    Pr < W       0.0686
Kolmogorov-Smirnov    D      0.179873    Pr > D       0.0435
Cramer-von Mises      W-Sq   0.131907    Pr > W-Sq    0.0402
Anderson-Darling      A-Sq   0.770651    Pr > A-Sq    0.0405
```

3. *Each of the following tables represents the cell means from a balanced 2×2 factorial completely randomized design with n replicates per cell. For each table tell which of the following summs of squares would be zero for that table: SS_A, SS_B, S_{AB}.*

For any table,

- $SS_A = 0$ if $\bar{y}_{\cdot 1 \cdot} = \bar{y}_{\cdot 2 \cdot}$,
- $SS_B = 0$ if $\bar{y}_{1 \cdot \cdot} = \bar{y}_{2 \cdot \cdot}$,
- $SS_{AB} = 0$ if $\bar{y}_{11} - \bar{y}_{12} = \bar{y}_{21} - \bar{y}_{22}$ and $\bar{y}_{11 \cdot} - \bar{y}_{21 \cdot} = \bar{y}_{12 \cdot} - \bar{y}_{22 \cdot}$.

Table 1

Factor A

Factor B	1	3
	5	3

- $\bar{y}_{\cdot 1 \cdot} = 3$, $\bar{y}_{\cdot 2 \cdot} = 3 \implies SS_A = 0$
- $\bar{y}_{1 \cdot \cdot} = 2$, $\bar{y}_{2 \cdot \cdot} = 4 \implies SS_B \neq 0$
- $\bar{y}_{11 \cdot} - \bar{y}_{12 \cdot} = -2$, $\bar{y}_{21 \cdot} - \bar{y}_{22 \cdot} = 2$,
  $\bar{y}_{11 \cdot} - \bar{y}_{21 \cdot} = -4$, $\bar{y}_{12 \cdot} - \bar{y}_{22 \cdot} = 0$
  $\implies SS_{AB} \neq 0$

Table 2

Factor A

Factor B	1	3
	5	7

- $\bar{y}_{\cdot 1 \cdot} = 3$, $\bar{y}_{\cdot 2 \cdot} = 5 \implies SS_A \neq 0$
- $\bar{y}_{1 \cdot \cdot} = 2$, $\bar{y}_{2 \cdot \cdot} = 6 \implies SS_B \neq 0$
- $\bar{y}_{11} - \bar{y}_{12} = -2$, $\bar{y}_{21} - \bar{y}_{22} = -2$,
  $\bar{y}_{11} - \bar{y}_{21} = -4$, $\bar{y}_{12} - \bar{y}_{22} = -4$
  $\implies SS_{AB} = 0$

Table 3

Factor A

Factor B	1	3
	5	5

- $\bar{y}_{\cdot 1 \cdot} = 3$, $\bar{y}_{\cdot 2 \cdot} = 4 \implies SS_A \neq 0$
- $\bar{y}_{1 \cdot \cdot} = 2$, $\bar{y}_{2 \cdot \cdot} = 5 \implies SS_B \neq 0$
- $\bar{y}_{11} - \bar{y}_{12} = -2$, $\bar{y}_{21} - \bar{y}_{22} = 0$,
  $\bar{y}_{11} - \bar{y}_{21} = -4$, $\bar{y}_{12} - \bar{y}_{22} = -2$
  $\implies SS_{AB} \neq 0$

Table 4

Factor A

Factor B	5	3
	5	3

- $\bar{y}_{\cdot 1 \cdot} = 5$, $\bar{y}_{\cdot 2 \cdot} = 3 \implies SS_A \neq 0$
- $\bar{y}_{1 \cdot \cdot} = 4$, $\bar{y}_{2 \cdot \cdot} = 4 \implies SS_B = 0$
- $\bar{y}_{11} - \bar{y}_{12} = 2$, $\bar{y}_{21} - \bar{y}_{22} = 2$,
  $\bar{y}_{11} - \bar{y}_{21} = 0$, $\bar{y}_{12} - \bar{y}_{22} = 0$
  $\implies SS_{AB} = 0$

Appendix: SAS Code

Problem 1

```
DM 'LOG; CLEAR; OUT; CLEAR;';
OPTIONS CENTER NONUMBER NODATE LS=75 PS=60 FORMDLIM='=';

ODS LISTING FILE='hw6p1.lst';
ODS GRAPHICS ON / IMAGENAME='bibd' RESET=INDEX;

DATA bibd;
  INPUT batch process $ clean @@;
DATALINES;
 1 A 18.1   1 B 18.3   1 C 17.5
 2 A 19.5   2 B 17.8   2 D 17.2
 3 A 21.1   3 C 17.1   3 D 19.2
 4 B 25.2   4 C 22.5   4 D 20.5
 5 A 17.8   5 B 19.8   5 C 15.2
 6 A 20.8   6 B 18.7   6 D 17.8
 7 A 14.9   7 C 16.3   7 D 15.0
 8 B 23.3   8 C 22.2   8 D 21.8
 9 A 18.7   9 B 20.8   9 C 16.4
10 A 19.9  10 B 17.6  10 D 16.8
11 A 18.9  11 C 17.3  11 D 16.0
12 B 24.0  12 C 22.9  12 D 22.6
13 A 18.4  13 B 20.3  13 C 15.9
14 A 21.5  14 B 19.5  14 D 18.3
15 A 15.3  15 C 16.8  15 D 15.7
16 B 23.1  16 C 21.9  16 D 21.7
;

PROC GLM DATA=bibd PLOTS(UNPACK)=ALL;
  CLASS batch process;
  MODEL clean = batch process / SS3;
  MEANS process / BON;
  LSMEANS process / ADJUST=BON;
  ESTIMATE 'Agent A effect' process  3 -1 -1 -1 / DIVISOR=4;
  ESTIMATE 'Agent B effect' process -1  3 -1 -1 / DIVISOR=4;
  ESTIMATE 'Agent C effect' process -1 -1  3 -1 / DIVISOR=4;
  ESTIMATE 'Agent D effect' process -1 -1 -1  3 / DIVISOR=4;
TITLE 'Problem 1: BIBD for Wool Cleaning';

RUN;
ODS GRAPHICS OFF;
ODS LISTING CLOSE;

QUIT;
```

Problem 2

```
DM 'LOG; CLEAR; OUT; CLEAR;';
OPTIONS CENTER NONUMBER NODATE LS=75 PS=60 FORMDLIM='=';

ODS LISTING FILE='hw6p2glm.lst';
ODS GRAPHICS ON / IMAGENAME='clover' RESET=INDEX;

DATA in;
  DO nitrogen = 0 to 20 by 20;
    DO sulfur = 0 to 9 by 3;
      DO rep = 1 to 3;
        INPUT growth @@;
        OUTPUT;
      END;
    END;
  END;
DATALINES;
4.48  4.52  4.63
4.70  4.65  4.57
5.21  5.23  5.28
5.88  5.98  5.88
5.76  5.72  5.78
7.01  7.11  7.02
5.88  5.82  5.73
6.26  6.26  6.37
;

PROC GLM DATA=in PLOTS(UNPACK)=ALL;
  CLASS nitrogen sulfur;
  MODEL growth = nitrogen|sulfur / SS3;
  LSMEANS nitrogen|sulfur / ADJUST=BON;
  OUTPUT OUT=diag R=resid;
TITLE 'Factorial Design for Clover Growth';

RUN;
ODS LISTING CLOSE;
ODS LISTING FILE='hw6p2univ.lst';

PROC UNIVARIATE NORMAL DATA=diag;
  VAR resid;
ODS SELECT TestsForNormality;

RUN;
ODS GRAPHICS OFF;
ODS LISTING CLOSE;

QUIT;
```