Stat 541 Homework #5

Kenny Flagg

March 4, 2016

1. *Problem 3.42 (a), sample size to detect a difference in means of 10 with power 0.90.*

   (a) If $\sigma = 2$, a total of 9 observations (3 batteries of each brand) are needed to detect a difference in mean battery life of 10 hours with a power of at least 0.90. The actual power achieved is 0.998.

   (b) If $\sigma = 3$, a total of 9 observations (3 batteries of each brand) are needed to detect a difference in mean battery life of 10 hours with a power of at least 0.90. The actual power achieved is 0.902.

```
                  Problem 1:
  Sample Sizes Needed to Detect Battery Life
       Difference of 10 with Power 0.90

            The GLMPOWER Procedure

            Fixed Scenario Elements

 Dependent Variable              effect
 Source                           level
 Alpha                             0.05
 Nominal Power                      0.9
 Test Degrees of Freedom              2


                Computed N Total

            Std    Error    Actual        N
  Index     Dev       DF     Power    Total

      1       2        6     0.998        9
      2       3        6     0.902        9
```

2. *Problem 3.42 (a) again, sample size to detect a difference in means of 10 with power 0.95.*

   (a) If $\sigma = 2$, a total of 9 observations (3 batteries of each brand) are needed to detect a difference in mean battery life of 10 hours with a power of at least 0.95. The actual power achieved is 0.998.

   (b) If $\sigma = 2$, a total of 12 observations (4 batteries of each brand) are needed to detect a difference in mean battery life of 10 hours with a power of at least 0.95. The actual power achieved is 0.987.

```
                    Problem 2:
     Sample Sizes Needed to Detect Battery Life
          Difference of 10 with Power 0.95

               The GLMPOWER Procedure

               Fixed Scenario Elements

  Dependent Variable                effect
  Source                             level
  Alpha                               0.05
  Nominal Power                       0.95
  Test Degrees of Freedom                2


                  Computed N Total

               Std    Error    Actual        N
     Index     Dev       DF     Power    Total

         1       2        6     0.998        9
         2       3        9     0.987       12
```

3. *In Problem 3.26, was random assignment used?*

   Random assignment was not used. The treatment is brand, and it would not be possible to randomly assign batteries to brands.

4. *Orthogonal polynomial contrasts for 2mg, 4mg, 8mg, and 10mg.*

The mean is $\bar{x} = 6$, so the initial vectors are:

$$\mathbf{v}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{v_1} = \begin{bmatrix} -4 \\ -2 \\ 2 \\ 4 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 16 \\ 4 \\ 4 \\ 16 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} -64 \\ -8 \\ 8 \\ 64 \end{bmatrix}$$

$$|\mathbf{v}_0| = \sqrt{1+1+1+1} = 2 \qquad \Longrightarrow \qquad \mathbf{u}_0 = \frac{\mathbf{v}_0}{|\mathbf{v}_0|} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Linear Contrast Vector:

$$\mathbf{v}_1 \cdot \mathbf{u}_0 = \frac{1}{2}(-4-2+2+4) = 0$$

$$\mathbf{w}_1 = \mathbf{v}_1 - (\mathbf{v}_1 \cdot \mathbf{u}_0)\mathbf{u}_0 = \mathbf{v}_1 - 0\mathbf{u}_0 = \begin{bmatrix} -4 \\ -2 \\ 2 \\ 4 \end{bmatrix}$$

$$|\mathbf{w}_1| = \sqrt{16+4+4+16} = 2\sqrt{10} \qquad \Longrightarrow \qquad \mathbf{u}_1 = \frac{\mathbf{w}_1}{|\mathbf{w}_1|} = \frac{1}{\sqrt{10}}\begin{bmatrix} -2 \\ -1 \\ 1 \\ 2 \end{bmatrix}$$

Quadratic Contrast Vector:

$$\mathbf{v}_2 \cdot \mathbf{u}_0 = \frac{1}{2}(16+4+4+16) = 20$$

$$\mathbf{v}_2 \cdot \mathbf{u}_1 = \frac{1}{\sqrt{10}}(-32-4+4+32) = 0$$

$$\mathbf{w}_2 = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_0)\mathbf{u}_0 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 = \mathbf{v}_2 - 20\mathbf{u}_0 - 0\mathbf{u}_1 = \begin{bmatrix} 16 \\ 4 \\ 4 \\ 16 \end{bmatrix} - \begin{bmatrix} 10 \\ 10 \\ 10 \\ 10 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ -6 \\ 6 \end{bmatrix}$$

$$|\mathbf{w}_2| = \sqrt{36+36+36+36} = 12 \qquad \Longrightarrow \qquad \mathbf{u}_2 = \frac{\mathbf{w}_2}{|\mathbf{w}_2|} = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

Cubic Contrast Vector:

$$\mathbf{v}_3 \cdot \mathbf{u}_0 = \frac{1}{2}(-64-8+8+64) = 0$$

$$\mathbf{v}_3 \cdot \mathbf{u}_1 = \frac{1}{\sqrt{10}}(128+8+8+128) = \frac{272}{\sqrt{10}}$$

$$\mathbf{v}_3 \cdot \mathbf{u}_2 = \frac{1}{2}(-64 + 8 - 8 + 64) = 0$$

$$\mathbf{w}_3 = \mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{u}_0)\mathbf{u}_0 - (\mathbf{v}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{v}_3 \cdot \mathbf{u}_2)\mathbf{u}_2 = \mathbf{v}_3 - 0\mathbf{u}_0 - \frac{288}{\sqrt{10}}\mathbf{u}_1 - 0\mathbf{u}_2$$

$$= \frac{1}{5}\begin{bmatrix} -320 \\ -40 \\ 40 \\ 320 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} -272 \\ -136 \\ 136 \\ 272 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -48 \\ 96 \\ -96 \\ 48 \end{bmatrix}$$

$$|\mathbf{w}_3| = \frac{48}{5}\sqrt{1+4+4+1} = \frac{48\sqrt{10}}{5} \qquad \Longrightarrow \qquad \mathbf{u}_3 = \frac{\mathbf{w}_3}{|\mathbf{w}_3|} = \frac{1}{48\sqrt{10}}\begin{bmatrix} -1 \\ 2 \\ -2 \\ 1 \end{bmatrix}$$

So the contrasts are:

$$\begin{aligned} \Gamma_L &= -2\mu_1 - \mu_2 + \mu_3 + 2\mu_4 \\ \Gamma_Q &= \mu_1 - \mu_2 - \mu_3 + \mu_4 \\ \Gamma_C &= -1\mu_1 + 2\mu_2 - 2\mu_3 + \mu_4 \end{aligned}$$

5. *Cell count experiment with runs as blocks.*

(a) *Model and parameters.*
The model is

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}; \quad \epsilon_{ij} \stackrel{iid}{\sim} \mathrm{N}(0, \sigma^2)$$

where μ is the mean cell count for the control group in experiment 6; τ_i, $i = 2, 3$, are the effects of the drugs on mean cell count; and β_j, $j = 1, \ldots, 5$, are the effects of the experimental runs on the mean cell count.

(b) *ANOVA hypotheses.*
H_0: $\tau_2 = \tau_3 = 0$; Both drugs have no effect on mean cell count.
H_a: $\tau_i \neq 0$ for some i; At least one drug has an effect on mean cell count.

(c) *ANOVA results.*

There is strong evidence ($F_0 = 6.78$, p-value $= 0.0138$) that at least one of the drugs has an effect on mean cell count.

```
                              Problem 5(c):
                               RCBD ANOVA

                            The GLM Procedure

                         Class Level Information

             Class         Levels    Values

             drug               3    _drug1 _drug2 control

             exprment           6    1 2 3 4 5 6


                 Number of Observations Read          18
                 Number of Observations Used          18


                              Problem 5(c):
                               RCBD ANOVA

                            The GLM Procedure

                      Dependent Variable: cell_cnt

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

 Model                        7    224141.7222    32020.2460     4.83  0.0129

 Error                       10     66293.8889     6629.3889

 Corrected Total             17    290435.6111


           R-Square     Coeff Var      Root MSE    cell_cnt Mean

           0.771743      6.626482      81.42106         1228.722


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

 exprment                     5    134189.6111    26837.9222     4.05  0.0286
 drug                         2     89952.1111    44976.0556     6.78  0.0138


                                             Standard
Parameter                 Estimate              Error    t Value    Pr > |t|
```

```
Intercept             1282.444444 B      54.28070616      23.63      <.0001
exprment  1           -222.000000 B      66.48001649      -3.34      0.0075
exprment  2            -45.000000 B      66.48001649      -0.68      0.5138
exprment  3           -118.666667 B      66.48001649      -1.78      0.1046
exprment  4           -202.000000 B      66.48001649      -3.04      0.0125
exprment  5            -22.000000 B      66.48001649      -0.33      0.7475
exprment  6              0.000000 B        .                .          .
drug      _drug1       147.833333 B      47.00847047       3.14      0.0104
drug      _drug2        -4.166667 B      47.00847047      -0.09      0.9311
drug      control        0.000000 B        .                .          .

NOTE: The X'X matrix has been found to be singular, and a generalized
      inverse was used to solve the normal equations.  Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.
```

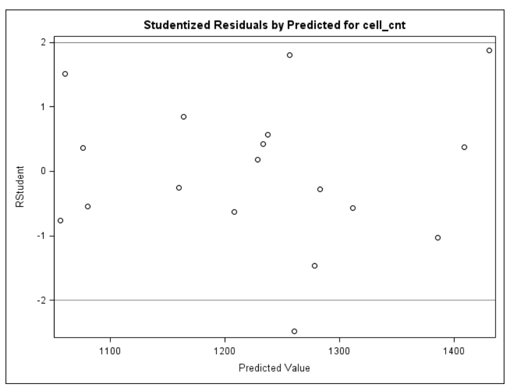
(d) *Did either drug increase cell counts?*

Drug 1 has an estimated effect of $\widehat{\tau}_2 = 147.83$ ($t_0 = 3.14$, right-tailed p-value = 0.0052) so there is strong evidence that drug 1 increases the mean cell count.

Drug 2 has an estimated effect of $\widehat{\tau}_3 = -4.17$ ($t_0 = -0.09$, two-tailed p-value = 0.9311) so there is no evidence that drug 2 has an effect on the mean cell count.

(e) *Is a square root transformation appropriate?*

The square root transformation is not necessary. The studentized residuals by predicted values plot does not show any serious violation of the constant variance assumption.



(f) *ANOVA ignoring blocks.*

There is moderate evidence ($F_0 = 3.37$, p-value = 0.0620) that at least one of the drugs has an effect on mean cell count.

```
                    Problem 5(f):
                ANOVA Ignoring Blocks

                  The GLM Procedure

                Class Level Information
```

```
Class         Levels    Values

drug               3    _drug1 _drug2 control


    Number of Observations Read          18
    Number of Observations Used          18


                 Problem 5(f):
              ANOVA Ignoring Blocks

                The GLM Procedure

         Dependent Variable: cell_cnt

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

Model                        2    89952.1111    44976.0556     3.37  0.0620

Error                       15   200483.5000    13365.5667

Corrected Total             17   290435.6111


          R-Square     Coeff Var      Root MSE    cell_cnt Mean

          0.309714      9.408924      115.6095         1228.722


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

drug                         2   89952.11111   44976.05556     3.37  0.0620


                                            Standard
Parameter                 Estimate             Error    t Value    Pr > |t|

Intercept              1180.833333 B     47.19739870      25.02      <.0001
drug      _drug1        147.833333 B     66.74720136       2.21      0.0427
drug      _drug2         -4.166667 B     66.74720136      -0.06      0.9510
drug      control         0.000000 B       .                .         .

NOTE: The X'X matrix has been found to be singular, and a generalized
      inverse was used to solve the normal equations.  Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.
```

(g) *Did blocking improve the analysis?*

Yes, blocking improved the analysis. In both analyses, the estimated drug effects are the same, but their standard errors differ. Including the blocks accounts for run-to-run variability so the estimates are more precise and we can be more certain in our conclusion that drug 1 is effective.

6. *Power analysis for Problem 3.10.*

Cotton Weight Percent	Tensile Strengths					Mean
15	7	7	15	11	9	9.8
20	12	17	12	18	18	15.4
25	14	19	19	18	18	17.6
30	19	25	22	19	23	21.6
35	7	10	11	15	11	10.8

$$\widehat{\sigma} = \sqrt{MSE} = 2.84$$

(a) A total of 70 observations (14 observations of each cotton weight percentage) are needed to detect a linear trend in mean tensile strength with a power of at least 0.90. The actual power achieved is 0.920.

(b) A total of 10 observations (2 observations of each cotton weight percentage) are needed to detect a quadratic trend in mean tensile strength with a power of at least 0.90. The actual power achieved is 0.904.

```
                    Problem 5:
     Sample Sizes Needed to Linear and Quadratic
     Trends in Tensile Strength with Power 0.90

              The GLMPOWER Procedure

             Fixed Scenario Elements

      Dependent Variable          meanstrength
      Alpha                               0.05
      Error Standard Deviation            2.84
      Nominal Power                        0.9


                  Computed N Total

                                 Test    Error    Actual        N
Index      Type        Source      DF       DF     Power    Total

    1    Effect      pct            4       10     0.976       15
    2    Contrast    Linear         1       65     0.920       70
    3    Contrast    Quadratic      1        5     0.904       10
```

Appendix: SAS Code

Problems 1 and 2

```
options center nodate nonumber ls=75 ps=60 formdlim='';

data prob1;
  do level=1 to 5;
    input effect @@;
    output;
  end;
datalines;
10 0 0
;

ods listing file='hw5p1.lst';
proc glmpower data=prob1;
  class level;
  model effect = level;
  power
    stddev = 2 3
    alpha = 0.05
    ntotal = .
    power = 0.90;
title1 'Problem 1:';
title2 'Sample Sizes Needed to Detect Battery Life';
title3 'Difference of 10 with Power 0.90';
run;

ods listing file='hw5p2.lst';
proc glmpower data=prob1;
  class level;
  model effect = level;
  power
    stddev = 2 3
    alpha = 0.05
    ntotal = .
    power = 0.95;
title1 'Problem 2:';
title2 'Sample Sizes Needed to Detect Battery Life';
title3 'Difference of 10 with Power 0.95';
run;
```

Problem 5

```
options center nodate nonumber ls=75 formdlim=' ';

*****************************************;
*** A RANDOMIZED COMPLETE BLOCK DESIGN ***;
*****************************************;

data in;
  input drug $ exprment cell_cnt @@;
datalines;
control 1 1147  _drug1 1 1169 _drug2 1 1009
control 2 1273  _drug1 2 1323 _drug2 2 1260
```

```
control 3 1216  _drug1 3 1276 _drug2 3 1143
control 4 1046  _drug1 4 1240 _drug2 4 1099
control 5 1138  _drug1 5 1432 _drug2 5 1355
control 6 1265  _drug1 6 1532 _drug2 6 1194
;

ods listing file='hw5p5c.lst';
ods graphics on / imagename='hw5p5c' reset=index;
proc glm data=in plots(unpack)=all;
  class drug exprment;
  model cell_cnt = exprment drug / ss3 solution;
title1 'Problem 5(c):';
title2 'RCBD ANOVA';
run;

ods listing file='hw5p5f.lst';
ods graphics on / imagename='hw5p5f' reset=index;
proc glm data=in;
  class drug;
  model cell_cnt = drug / ss3 solution;
title1 'Problem 5(f):';
title2 'ANOVA Ignoring Blocks';
run;
```

Problem 6

```
options center nodate nonumber ls=75 formdlim='';

data prob6;
  input pct meanstrength @@;
datalines;
15 9.8  20 15.4  25 17.6  30 21.6 35 10.8
;

proc glmpower data=prob6;
  class pct;
  model meanstrength = pct;
  contrast 'Linear   ' pct -2 -1  0  1  2;
  contrast 'Quadratic' pct  2 -1 -2 -1  2;
  power
    stddev = 2.84
    alpha = 0.05
    ntotal = .
    power = 0.90;
title1 'Problem 5:';
title2 'Sample Sizes Needed to Linear and Quadratic';
title3 'Trends in Tensile Strength with Power 0.90';
run;
```