Stat 541 Homework #7

Kenny Flagg

April 18, 2016

1. *Problem 13.1, pages 601. Also provide a practical interpretation of the variance component estimates in the context of the study.*

   Since the operators and parts were randomly selected, random effects should be used. The appropriate model is

$$\begin{aligned} y_{ijk} &= \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}; \\ \tau_i &\overset{iid}{\sim} \mathrm{N}\left(0, \sigma_\tau^2\right), \\ \beta_j &\overset{iid}{\sim} \mathrm{N}\left(0, \sigma_\beta^2\right), \\ (\tau\beta)_{ij} &\overset{iid}{\sim} \mathrm{N}\left(0, \sigma_{\tau\beta}^2\right), \\ \epsilon_{ijk} &\overset{iid}{\sim} \mathrm{N}\left(0, \sigma^2\right) \end{aligned}$$

   where τ_i is the operator effect, β_j is the part effect, and $(\tau\beta)_{ij}$ is the operator by part interaction.

   The random effects ANOVA output appears on the next page. We first test for the interaction effect with the hypotheses

$$\begin{aligned} H_0&\colon\ \sigma_{\tau\beta}^2 = 0, \\ H_1&\colon\ \sigma_{\tau\beta}^2 > 0. \end{aligned}$$

   The exact F-statistic is $F = 0.40$ with p-value $= 0.9270$. There is little to no evidence of an interaction effect.

   Now we test the main effects. The hypotheses for the operator effect are

$$\begin{aligned} H_0&\colon\ \sigma_{\tau}^2 = 0, \\ H_1&\colon\ \sigma_{\tau}^2 > 0. \end{aligned}$$

   The exact F-test has test statistic $F = 0.69$ with p-value $= 0.4269$ so there is little evidence that the mean response varies among operators.

   For the part effect, the hypotheses are

$$\begin{aligned} H_0&\colon\ \sigma_{\tau}^2 = 0, \\ H_1&\colon\ \sigma_{\tau}^2 > 0. \end{aligned}$$

The exact F-statistic is $F = 18.28$ with p-value < 0.0001. There is strong evidence that the mean measurements vary among parts.

Problem 13.1 Random Effects ANOVA

The GLM Procedure
Tests of Hypotheses for Random Model Analysis of Variance

Dependent Variable: response

Source	DF	Type III SS	Mean Square	F Value	Pr > F
operator	1	0.416667	0.416667	0.69	0.4269
part	9	99.016667	11.001852	18.28	<.0001
Error	9	5.416667	0.601852		
Error: MS(operator*part)					

Source	DF	Type III SS	Mean Square	F Value	Pr > F
operator*part	9	5.416667	0.601852	0.40	0.9270
Error: MS(Error)	40	60.000000	1.500000		

The variance estmates for the operator and interaction effects are $\widehat{\sigma}_{\tau}^2 = 0$ and $\widehat{\sigma}_{\tau\beta}^2 = 0$. On average, the measurements do not differ from one operator to another. The estimated variance for the part effect is $\widehat{\sigma}_{\beta}^2 = 1.614$, which is small relative to the observed response values and may be acceptably low. However, the estimated residual variance is $\widehat{\sigma}^2 = 1.317$, so the variability among measurements on the same part is almost as large as the variability among parts. This implies that the measurement equipment is not precise enough to measure these parts adequately.

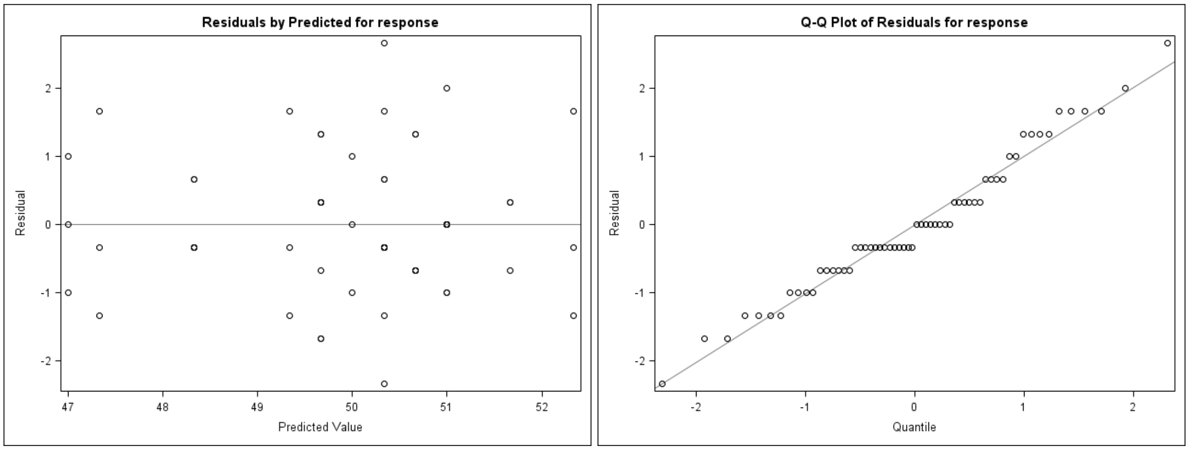
The residual plot shows no concerning patterns and the Normal Q-Q plot shows little deviation from the diagonal line, so we can assume constant variance and Normality of the residuals.

Problem 13.1 Random Effect Variance Components

Variance Components Estimation Procedure

REML Estimates	
Variance Component	Estimate
Var(operator)	0
Var(part)	1.61420
Var(operator*part)	0
Var(Error)	1.31667

The residual plot shows no concerning patterns and the Normal Q-Q plot shows little deviation from the diagonal line, so we can assume constant variance and Normality of the residuals.



2. *Reconsider Problem 13.1. Consider two scenarios:*

   (i) *Suppose that when each measurement was taken, each operator knew the part number being measured and also recorded the data him/herself.*

   (ii) *Suppose that when each measurement was taken, each operator were "blinded" to the part number being measured, and a third party recorded the measurements.*

   How might this impact the data analysis?

   In situation (i), measurements on the same part would not be independent. The observed values could be biased if, for example, an operator remembered previous measurements of the part and adjusted the values to agree with previous readings. In situation (ii), all observations would be made independently and the model used in problem 1 would be appropriate.

3. *Suppose the two operators in Problem 13.1 were the only two available (that is, Operators represents a fixed effect). Which model effects (if any) now have different F-test statistics? And, if an effect's F-test is different, what would be used for the denominator of the F-statistic?*

   In this case, none of the F-tests would change. The denominator of the test for the interaction would still be MS_E and the denominator for the tests of the main effects would still be MS_{AB}.

Problem 13.1 Expected Mean Squares Assuming Operator is a Fixed Effect

The GLM Procedure

Source	Type III Expected Mean Square
operator	Var(Error) + 3 Var(operator*part) + Q(operator)
part	Var(Error) + 3 Var(operator*part) + 6 Var(part)
operator*part	Var(Error) + 3 Var(operator*part)

4. *Consider the information on pages 192–194 for Example 5.1 for the battery design experiment. Suppose the researcher is planning another* 3×3 *factorial experiment. The goal is to determine the sample size required so that the power of each* F*-test (for Material Type, Temperature, and their interaction) is* ≥ 0.90.

   (a) *Based on this experiment, what is the estimate of* σ*?*

   The estimate is $\sqrt{MS_E} = 25.9848$.

   (b) *Use SAS to determine the desired sample size assuming you use the sample mean* $\bar{y}_{ij}$ *as an estimate of* μ_{ij} *for* $i = 1, 2, 3$ *and* $j = 1, 2, 3$*. You can assume* $\alpha = 0.05$.

   Four replicates are needed to achieve a power of at least 0.90 for testing the material effect. Only two replicates are needed for testing the temperature effect with a power of at least 0.90. For testing the interaction, five replicates are needed. Thus we need at least five replicates for a total sample size of 45 in order to obtain a power of 0.90 for all the F-tests.

Sample Size Determination for Example 5.1

The GLMPOWER Procedure

Fixed Scenario Elements	
Dependent Variable	meanest
Alpha	0.05
Error Standard Deviation	25.9848
Nominal Power	0.9

Computed N Total

Index	Source	Test DF	Error DF	Actual Power	N Total
1	material	2	27	0.930	36
2	temperature	2	9	0.985	18
3	material*temperature	4	36	0.904	45

5. *Factors A and B are fixed factors both having 3 fixed levels, and factors C and D are random factors each having 5 randomly-selected levels. Suppose $n = 4$ replicates are taken for each $A*B*C*D$ combination. In the Proc GLM model statement, $A|B|C|D$ will generate the full four-factor factorial model.*

   (a) *Assuming a completely randomized design, generate a table of expected mean squares for this four-factor factorial experiment.*

Expected Mean Squares for Problem 5a

The GLM Procedure

Source	Type III Expected Mean Square
A	Var(Error) + 4 Var(A*B*C*D) + 12 Var(A*C*D) + 20 Var(A*B*D) + 60 Var(A*D) + 20 Var(A*B*C) + 60 Var(A*C) + Q(A,A*B)
B	Var(Error) + 4 Var(A*B*C*D) + 12 Var(B*C*D) + 20 Var(A*B*D) + 60 Var(B*D) + 20 Var(A*B*C) + 60 Var(B*C) + Q(B,A*B)
A*B	Var(Error) + 4 Var(A*B*C*D) + 20 Var(A*B*D) + 20 Var(A*B*C) + Q(A*B)
C	Var(Error) + 4 Var(A*B*C*D) + 12 Var(B*C*D) + 12 Var(A*C*D) + 36 Var(C*D) + 20 Var(A*B*C) + 60 Var(B*C) + 60 Var(A*C) + 180 Var(C)
A*C	Var(Error) + 4 Var(A*B*C*D) + 12 Var(A*C*D) + 20 Var(A*B*C) + 60 Var(A*C)
B*C	Var(Error) + 4 Var(A*B*C*D) + 12 Var(B*C*D) + 20 Var(A*B*C) + 60 Var(B*C)
A*B*C	Var(Error) + 4 Var(A*B*C*D) + 20 Var(A*B*C)
D	Var(Error) + 4 Var(A*B*C*D) + 12 Var(B*C*D) + 12 Var(A*C*D) + 36 Var(C*D) + 20 Var(A*B*D) + 60 Var(B*D) + 60 Var(A*D) + 180 Var(D)
A*D	Var(Error) + 4 Var(A*B*C*D) + 12 Var(A*C*D) + 20 Var(A*B*D) + 60 Var(A*D)
B*D	Var(Error) + 4 Var(A*B*C*D) + 12 Var(B*C*D) + 20 Var(A*B*D) + 60 Var(B*D)
A*B*D	Var(Error) + 4 Var(A*B*C*D) + 20 Var(A*B*D)
C*D	Var(Error) + 4 Var(A*B*C*D) + 12 Var(B*C*D) + 12 Var(A*C*D) + 36 Var(C*D)
A*C*D	Var(Error) + 4 Var(A*B*C*D) + 12 Var(A*C*D)
B*C*D	Var(Error) + 4 Var(A*B*C*D) + 12 Var(B*C*D)
A*B*C*D	Var(Error) + 4 Var(A*B*C*D)

   (b) *For what effects are there exact F-tests? Provide the F-statistic in terms of mean squares for these exact tests.*

   The following are exact F-tests:

Source	F-statistic
A*B*C	$F = \frac{MS_{ABC}}{MS_{ABCD}}$
A*B*D	$F = \frac{MS_{ABD}}{MS_{ABCD}}$
A*C*D	$F = \frac{MS_{ACD}}{MS_{ABCD}}$
B*C*D	$F = \frac{MS_{BCD}}{MS_{ABCD}}$
A*B*C*D	$F = \frac{MS_{ABCD}}{MS_E}$

(c) *For what effects are there approximate F-tests? Provide the F-statistic in terms of mean squares for these approximate tests.*

SAS used the following linear combinations for the approximate F-tests:

Source	F-statistic
A	$F = \frac{MS_A}{MS_{AC}+MS_{AD}-MS_{ACD}}$
B	$F = \frac{MS_B}{MS_{BC}+MS_{BD}-MS_{BCD}}$
A*B	$F = \frac{MS_{AB}}{MS_{ABC}+MS_{ABD}-MS_{ABCD}}$
C	$F = \frac{MS_C}{MS_{AC}+MS_{BC}-MS_{ABC}+MS_{CD}-MS_{ACD}-MS_{BCD}+MS_{ABCD}}$
A*C	$F = \frac{MS_{AC}}{MS_{ABC}+MS_{ACD}-MS_{ABCD}}$
B*C	$F = \frac{MS_{BC}}{MS_{ABC}+MS_{BCD}-MS_{ABCD}}$
D	$F = \frac{MS_D}{MS_{AD}+MS_{BD}-MS_{ABD}+MS_{CD}-MS_{ACD}-MS_{BCD}+MS_{ABCD}}$
A*D	$F = \frac{MS_{AD}}{MS_{ABD}+MS_{ACD}-MS_{ABCD}}$
B*D	$F = \frac{MS_{BD}}{MS_{ABD}+MS_{BCD}-MS_{ABCD}}$
C*D	$F = \frac{MS_{CD}}{MS_{ACD}+MS_{BCD}-MS_{ABCD}}$

(d) *Suppose the four-factor interaction is removed from the model. Answer parts (b) and (c) again.*

Expected Mean Squares for Problem 5d

The GLM Procedure

Source	Type III Expected Mean Square
A	Var(Error) + 12 Var(A*C*D) + 20 Var(A*B*D) + 60 Var(A*D) + 20 Var(A*B*C) + 60 Var(A*C) + Q(A,A*B)
B	Var(Error) + 12 Var(B*C*D) + 20 Var(A*B*D) + 60 Var(B*D) + 20 Var(A*B*C) + 60 Var(B*C) + Q(B,A*B)
A*B	Var(Error) + 20 Var(A*B*D) + 20 Var(A*B*C) + Q(A*B)
C	Var(Error) + 12 Var(B*C*D) + 12 Var(A*C*D) + 36 Var(C*D) + 20 Var(A*B*C) + 60 Var(B*C) + 60 Var(A*C) + 180 Var(C)
A*C	Var(Error) + 12 Var(A*C*D) + 20 Var(A*B*C) + 60 Var(A*C)
B*C	Var(Error) + 12 Var(B*C*D) + 20 Var(A*B*C) + 60 Var(B*C)
A*B*C	Var(Error) + 20 Var(A*B*C)
D	Var(Error) + 12 Var(B*C*D) + 12 Var(A*C*D) + 36 Var(C*D) + 20 Var(A*B*D) + 60 Var(B*D) + 60 Var(A*D) + 180 Var(D)
A*D	Var(Error) + 12 Var(A*C*D) + 20 Var(A*B*D) + 60 Var(A*D)
B*D	Var(Error) + 12 Var(B*C*D) + 20 Var(A*B*D) + 60 Var(B*D)
A*B*D	Var(Error) + 20 Var(A*B*D)
C*D	Var(Error) + 12 Var(B*C*D) + 12 Var(A*C*D) + 36 Var(C*D)
A*C*D	Var(Error) + 12 Var(A*C*D)
B*C*D	Var(Error) + 12 Var(B*C*D)

When the four-factor interaction is removed, these are exact F-tests:

Source	F-statistic
A*B*C	$F = \frac{MS_{ABC}}{MS_E}$
A*B*D	$F = \frac{MS_{ABD}}{MS_E}$
A*C*D	$F = \frac{MS_{ACD}}{MS_E}$
B*C*D	$F = \frac{MS_{BCD}}{MS_E}$

These are the approximate F-tests performed by SAS:

Source	F-statistic
A	$F = \frac{MS_A}{MS_{AC}+MS_{AD}-MS_{ACD}}$
B	$F = \frac{MS_B}{MS_{BC}+MS_{BD}-MS_{BCD}}$
A*B	$F = \frac{MS_{AB}}{MS_{ABC}+MS_{ABD}-MS_E}$
C	$F = \frac{MS_C}{MS_{AC}+MS_{BC}-MS_{ABC}+MS_{CD}-MS_{ACD}-MS_{BCD}+MS_E}$
A*C	$F = \frac{MS_{AC}}{MS_{ABC}+MS_{ACD}-MS_E}$
B*C	$F = \frac{MS_{BC}}{MS_{ABC}+MS_{BCD}-MS_E}$
D	$F = \frac{MS_D}{MS_{AD}+MS_{BD}-MS_{ABD}+MS_{CD}-MS_{ACD}-MS_{BCD}+MS_E}$
A*D	$F = \frac{MS_{AD}}{MS_{ABD}+MS_{ACD}-MS_E}$
B*D	$F = \frac{MS_{BD}}{MS_{ABD}+MS_{BCD}-MS_E}$
C*D	$F = \frac{MS_{CD}}{MS_{ACD}+MS_{BCD}-MS_E}$

(e) *Suppose the four-factor and all three-factor interactions are removed from the model. Answer parts (b) and (c) again.*

Expected Mean Squares for Problem 5e

The GLM Procedure

Source	Type III Expected Mean Square
A	Var(Error) + 60 Var(A*D) + 60 Var(A*C) + Q(A,A*B)
B	Var(Error) + 60 Var(B*D) + 60 Var(B*C) + Q(B,A*B)
A*B	Var(Error) + Q(A*B)
C	Var(Error) + 36 Var(C*D) + 60 Var(B*C) + 60 Var(A*C) + 180 Var(C)
A*C	Var(Error) + 60 Var(A*C)
B*C	Var(Error) + 60 Var(B*C)
D	Var(Error) + 36 Var(C*D) + 60 Var(B*D) + 60 Var(A*D) + 180 Var(D)
A*D	Var(Error) + 60 Var(A*D)
B*D	Var(Error) + 60 Var(B*D)
C*D	Var(Error) + 36 Var(C*D)

These tests are exact:

Source	F-statistic
A*B	$F = \frac{MS_{AB}}{MS_E}$
A*C	$F = \frac{MS_{AC}}{MS_E}$
B*C	$F = \frac{MS_{BC}}{MS_E}$
A*D	$F = \frac{MS_{AD}}{MS_E}$
B*D	$F = \frac{MS_{BD}}{MS_E}$
C*D	$F = \frac{MS_{CD}}{MS_E}$

These are approximate F-tests:

Source	F-statistic
A	$F = \frac{MS_A}{MS_{AC}+MS_{AD}-MS_E}$
B	$F = \frac{MS_B}{MS_{BC}+MS_{BD}-MS_E}$
C	$F = \frac{MS_C}{MS_{AC}+MS_{BC}+MS_{CD}-2MS_E}$
D	$F = \frac{MS_D}{MS_{AD}+MS_{BD}+MS_{CD}-2MS_E}$

6. *See Problem 13.21, page 603.*

   (a) *Answer the second question only: "If the three-factor and $(\tau\beta)_{ij}$ interactions do not exist, can all remaining effects be tested?" That is, for each model effect, do we have an exact F-test or do we have to consider an approximate F-test?*

   The expected mean squares are:

Source	Expected Mean Square
A	$\sigma^2 + b\sigma^2_{\tau\gamma} + bc\sigma^2_{\tau}$
B	$\sigma^2 + a\sigma^2_{\beta\gamma} + ac\sigma^2_{\beta}$
C	$\sigma^2 + a\sigma^2_{\beta\gamma} + b\sigma^2_{\tau\gamma} + ab\sigma^2_{\gamma}$
AC	$\sigma^2 + b\sigma^2_{\tau\gamma}$
BC	$\sigma^2 + a\sigma^2_{\beta\gamma}$

   There are exact F-tests for factor A (denominator is MS_{AC}), factor B (denominator is MS_{BC}), and the AC and BC interactions (denominator is MS_E).

   (b) *For each effect having an associated approximate F-test, provide a linear combination of mean squares appearing in the numerator and in the denominator of the F-statistic.*

   The only approximate F-test is the test for the factor C effect. One possible F-statistic for this test is:

$$F = \frac{MS_C + MS_E}{MS_{AC} + MS_{BC}}$$

Appendix: SAS Code

Problems 1–3

```
dm 'log; clear; out; clear;';
options center nonumber nodate;
ods trace on;

ods pdf file='hw7prob1a.pdf';
ods graphics on / imagename='hw7p1a' reset=index;

data prob1;
  do part = 1 to 10;
    do operator = 1 to 2;
      do rep = 1 to 3;
        input response @@;
        output;
      end;
    end;
  end;
datalines;
50 49 50  50 48 51
52 52 51  51 51 51
53 50 50  54 52 51
49 51 50  48 50 51
48 49 48  48 49 48
52 50 50  52 50 50
51 51 51  51 50 50
52 50 49  53 48 50
50 51 50  51 48 49
47 46 49  46 47 48
;

proc glm data=prob1 plots(unpack)=all;
  class operator part;
  model response = operator|part / SS3;
  random operator|part / test;
title 'Problem 13.1 Random Effects ANOVA';
ods pdf select RandomModelANOVA;

run;
ods graphics off;
ods pdf close;

ods pdf file='hw7prob1b.pdf';

proc varcomp data=prob1 method=REML;
  class operator part;
  model response = operator|part;
title 'Problem 13.1 Random Effect Variance Components';
ods select Estimates;

run;
ods pdf close;

ods pdf file='hw7prob3.pdf';
```

```
proc glm data=prob1 plots=none;
  class operator part;
  model response = operator|part / SS3;
  random part operator*part / test;
title 'Problem 13.1 Expected Mean Squares Assuming Operator is a Fixed Effect';
ods pdf select ExpectedMeanSquares;

run;
ods pdf close;

quit;
```

Problem 4

```
dm 'log; clear; out; clear;';
options center nonumber nodate;
ods trace on;

ods pdf file='hw7prob4.pdf';

data prob4;
  do material = 1 to 3;
    do temperature = 15 to 125 by 55;
      input meanest @@;
      output;
    end;
  end;
datalines;
134.75  57.25  57.50
155.75 119.75  49.50
144.00 145.75  85.50
;

proc glmpower data=prob4;
  class material temperature;
  model meanest = material|temperature;
  power
    stddev =  25.9848
    alpha = 0.05
    ntotal = .
    power = 0.90;
title 'Sample Size Determination for Example 5.1';
run;
ods pdf close;

quit;
```

Problem 5

```
dm 'log; clear; out; clear;';
options center nonumber nodate;
ods trace on;

ods pdf file='hw7prob5.pdf';

data prob5;
  do A = 1 to 3;
    do B = 1 to 3;
      do C = 1 to 5;
        do D = 1 to 5;
          do rep = 1 to 4;
            mu = A + B + C + D;
            y = mu + rannor(68632);
            output;
          end;
        end;
      end;
    end;
  end;

proc glm data=prob5;
  class A B C D;
  model y = A|B|C|D / SS3;
  random C D A*C A*D B*C B*D C*D A*B*C A*B*D A*C*D B*C*D A*B*C*D / test;
title 'Expected Mean Squares for Problem 5a';
run;

proc glm data=prob5;
  class A B C D;
  model y = A|B|C|D @3 / SS3;
  random C D A*C A*D B*C B*D C*D A*B*C A*B*D A*C*D B*C*D / test;
title 'Expected Mean Squares for Problem 5d';
run;

proc glm data=prob5;
  class A B C D;
  model y = A|B|C|D @2 / SS3;
  random C D A*C A*D B*C B*D C*D / test;
title 'Expected Mean Squares for Problem 5e';
run;

ods pdf close;

quit;
```

Problem 6

```
dm 'log; clear; out; clear;';
options center nonumber nodate;
ods trace on;

ods pdf file='hw7prob6.pdf';

data prob6;
  do tau = 1 to 2;
    do beta = 1 to 3;
      do gamma = 1 to 5;
        mu = tau + beta + gamma;
        y = mu + rannor(7635);
        output;
      end;
    end;
  end;

proc glm data=prob6;
  class tau beta gamma;
  model y = tau beta gamma tau*gamma beta*gamma / SS3;
  random tau beta gamma tau*gamma beta*gamma / test;
title 'Expected Mean Squares for Problem 6';
run;

ods pdf close;

quit;
```