Mixed Model: Randomized Complete Block Design (RCBD)

Penicillin Example


Comparison of four processes for producing penicillin.

We have a random sample of five batches which will each be split into four parts.

- run each process on one part
- randomize the order in which the process are run within each batch

The main goal of the analysis is perform comparison tests of the processes.

SAS statements for data input and Proc GLM:

data set1;
  infile "Y:\penicillin.dat";
  input batch process $ yield;
run;

/* Compute the ANOVA table, formulas for
   expectations of mean squares, process
   means and their standard errors (these
   standard errors for process means are
   not correct... use Proc Mixed instead.)*/

proc glm data=set1;
  class batch process;
  model yield = batch process;
  random batch / test;
  lsmeans process / stderr pdiff;
  output out=set2 r=resid p=yhat;
run;

NOTE: I recommend using residual plots from PROC MIXED for diagnostics rather then PROC GLM.

The GLM Procedure
Class Level Information

Class       Levels Values
batch       5 1 2 3 4 5
process     4 A B C D

Dependent Variable: yield

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>334.0000000</td>
<td>47.7142857</td>
<td>2.53</td>
<td>0.0754</td>
</tr>
<tr>
<td>Error</td>
<td>12</td>
<td>226.0000000</td>
<td>18.8333333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>19</td>
<td>560.0000000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The GLM Procedure
Dependent Variable: yield

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>4</td>
<td>264.0000000</td>
<td>66.0000000</td>
<td>3.50</td>
<td>0.0407</td>
</tr>
<tr>
<td>process</td>
<td>3</td>
<td>70.0000000</td>
<td>23.3333333</td>
<td>1.24</td>
<td>0.3387</td>
</tr>
</tbody>
</table>
The GLM Procedure
Source Type III Expected Mean Square
batch Var(Error) + 4 Var(batch)
process Var(Error) + Q(process)

Because process was a fixed effects factor, the EMS_{process} = \sigma^2 + \frac{\sum_{i=1}^{5} \alpha_i^2}{3}.

The GLM Procedure
Tests of Hypotheses for Mixed Model Analysis of Variance
Dependent Variable: yield

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>4</td>
<td>264.0000000</td>
<td>66.000000</td>
<td>3.50</td>
<td>0.0407</td>
</tr>
<tr>
<td>process</td>
<td>3</td>
<td>70.0000000</td>
<td>23.3333333</td>
<td>1.24</td>
<td>0.3387</td>
</tr>
<tr>
<td>Error: MS(Error)</td>
<td>12</td>
<td>226.000000</td>
<td>18.833333</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Least Squares Means
process yield LSMEAN Standard Error Pr > |t| Number
A 84.0000000 1.9407902 <.0001 1
B 85.0000000 1.9407902 <.0001 2
C 89.0000000 1.9407902 <.0001 3
D 86.0000000 1.9407902 <.0001 4
Least Squares Means for effect process
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: yield

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7219</td>
<td>0.0935</td>
<td>0.4802</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7219</td>
<td></td>
<td>0.1707</td>
<td>0.7219</td>
</tr>
<tr>
<td>3</td>
<td>0.0935</td>
<td>0.1707</td>
<td></td>
<td>0.2958</td>
</tr>
<tr>
<td>4</td>
<td>0.4802</td>
<td>0.7219</td>
<td>0.2958</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
SAS statements for Proc Mixed:

/* Fit the same model using PROC MIXED. Compute REML estimates of variance components. Note that PROC MIXED provides appropriate standard errors for process means. When block effects are random, PROC GLM does not provide correct standard errors for process means */

proc mixed data=set1 plots(only)=Residualpanel(conditional);
class process batch;
model yield = process / ddfm=satterth solution;
random batch;
lsmeans process / pdiff tdiff;
run;

The Mixed Procedure
Model Information

Data Set WORK.SET1
Dependent Variable yield
Covariance Structure Variance Components
Estimation Method REML
Residual Variance Method Profile
Fixed Effects SE Method Model-Based
Degrees of Freedom Method Satterthwaite

Class Level Information

Class    Levels   Values
process   4       A B C D
batch    5       1 2 3 4 5

Dimensions

Covariance Parameters    2
Columns in X             5
Columns in Z             5
Subjects                 1
Max Obs Per Subject      20

Covariance Parameter Estimates

Cov Parm    Estimate
batch       11.7917
Residual    18.8333
### Solution for Fixed Effects

#### Standard

| Effect  | process | Estimate | Error | DF | t Value | Pr > |t| |
|---------|---------|----------|-------|----|---------|-------|---|
| Intercept |         | 86.0000  | 2.4749 | 11.1 | 34.75 | <.0001 |
| process A | -2.0000 | 2.7447  | 12 | -0.73 | 0.4802 |
| process B | -1.0000 | 2.7447  | 12 | -0.36 | 0.7219 |
| process C | 3.0000  | 2.7447  | 12 | 1.09  | 0.2958 |
| process D |         | 0       | .    | .    | .      | .         |

#### Type 3 Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num</th>
<th>Den</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>process</td>
<td>3</td>
<td>12</td>
<td>1.24</td>
<td>0.3387</td>
</tr>
</tbody>
</table>

### The Mixed Procedure

#### Least Squares Means

#### Standard

| Effect | process | Estimate | Error | DF | t Value | Pr > |t| |
|--------|---------|----------|-------|----|---------|-------|---|
| process A | 84.0000 | 2.4749  | 11.1 | 33.94 | <.0001 |
| process B | 85.0000 | 2.4749  | 11.1 | 34.35 | <.0001 |
| process C | 89.0000 | 2.4749  | 11.1 | 35.96 | <.0001 |
| process D | 86.0000 | 2.4749  | 11.1 | 34.75 | <.0001 |

#### Differences of Least Squares Means

#### Standard

| Effect | process | _process | Estimate | Error | DF | t Value | Pr > |t| |
|--------|---------|----------|----------|-------|----|---------|-------|---|
| process A | B  | -1.0000  | 2.7447  | 12 | -0.36 | 0.7219 |
| process A | C  | -5.0000  | 2.7447  | 12 | -1.82 | 0.0935 |
| process A | D  | -2.0000  | 2.7447  | 12 | -0.73 | 0.4802 |
| process B | C  | -4.0000  | 2.7447  | 12 | -1.46 | 0.1707 |
| process B | D  | -1.0000  | 2.7447  | 12 | -0.36 | 0.7219 |
| process C | D  | 3.0000   | 2.7447  | 12 | 1.09  | 0.2958 |
Residual ($\hat{\epsilon}$) vs. conditional mean ($\hat{y}$), conditioning on random effect.
Normal probability plot of residuals for normality in $N(0, \sigma^2)$. 

**Conditional Residuals for yield**

- **Residual Statistics**
  - Observations: 28
  - Minimum: -0.1441
  - Mean: -3E-15
  - Maximum: 5.5707
  - Std Dev: 3.6092

- **Fit Statistics**
  - Objective: 103.83
  - AIC: 107.83
  - AICC: 108.75
  - BIC: 107.05

![Conditional Residuals for yield](image)