## Analysis of a Split-Plot Experiment

Soy protein isolates (SPI) are widely used in the food industry. Their applications are based upon their functional properties related to solubility, emulsification, and viscosity control. SPI are usually stored in dry powder form to enhance shelf life and make them easier to distribute. The drying method of choice in industrial plants is spray-drying while the usual method used in bench-scale isolate production for research purposes is freeze-drying. A study was conducted at ISU to determine how the functional properties of SPI are affected by the method used to dry them and to compare the functional properties of dried SPI to fresh (undried) or frozen-thawed SPI. Another factor that may effect the functional properties of SPI is the temperature used in the extraction process to create SPI. Thus a two-factor experiment was conducted. The two factors and their levels are listed below.

**A=Temperature (temp)** : 25, 40, 60, 80 degrees Fahrenheit

**B=Method (meth)** : 1=fresh, 2=frozen and then thawed, 3=freeze dried, 4=spray dried

For each temperature, three SPI were created independently. Each SPI was split into four parts. The four methods were assigned to the four parts of each SPI in a completely randomized manner. Many response variables were measured for each part of each SPI. We will consider an analysis of $y$ = emulsion capacity (grams of oil emulsified by 1 gram of product).

This is an example of a split-plot experiment. The whole-plot experimental units are the SPI. The whole-plot factor is temperature. The split-plot experimental units are the parts of each SPI to which the levels of the split-plot factor (method) were assigned. We may consider the following general model.

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + (wp)_{ik} && \text{(whole-plot portion)} \\ &\quad +\beta_j + (\alpha\beta)_{ij} + (sp)_{ijk} && \text{(split-plot portion)} \\ &\quad (i = 1, \ldots, a \quad j = 1, \ldots, b \quad k = 1, \ldots, r) \end{aligned}$$

where $a$ is the number of levels of factor $A$, $b$ is the number of levels of factor $B$, $r$ is the number of whole-plot experimental units per level of the whole-plot factor, $(wp)_{ik} \sim N(0, \sigma^2_{wp})$ are the random effects associated with the whole-plot experimental units, $(sp)_{ijk} \sim N(0, \sigma^2_{sp})$ are the random effects associated with the split-plot experimental units, and all random effects are assumed to be independent.

1. What are $a$, $b$, and $r$ in this experiment?

We may partition the degrees of freedom as follows:

**Whole Plot Partitioning**

| SOURCE | DF | DF |
|---|---|---|
| $A$ | $a-1$ | 3 |
| W.P. Error | $(r-1)a$ | 8 |
| C. Total(wp) | $ra-1$ | 11 |

**Split Plot Partitioning**

| SOURCE | DF | DF |
|---|---|---|
| Whole Plot | $ra-1$ | 11 |
| $B$ | $b-1$ | 3 |
| $AB$ | $(a-1)(b-1)$ | 9 |
| S.P. Error | $(r-1)a(b-1)$ | 24 |
| C. Total(sp) | $rab-1$ | 47 |

2. What are good names for *W.P. Error* and *S.P. Error* in this case?

SAS code for the analysis using *proc glm* is provided below. Output can be found on the back of this sheet.

```
proc glm;
  class temp meth spi;
  model y=temp spi(temp) meth temp*meth;
  random spi(temp);
run;
```

The GLM Procedure

Class Level Information

| Class | Levels | Values |
|---|---|---|
| temp | 4 | 25 40 60 80 |
| meth | 4 | 1 2 3 4 |
| spi | 12 | 1 2 3 4 5 6 7 8 9 10 11 12 |

Number of observations 48

Dependent Variable: EC EC

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 23 | 143713.0844 | 6248.3950 | 13.61 | <.0001 |
| Error | 24 | 11016.0043 | 459.0002 | | |
| Corrected Total | 47 | 154729.0887 | | | |

| R-Square | Coeff Var | Root MSE | EC Mean |
|---|---|---|---|
| 0.928805 | 4.035427 | 21.42429 | 530.9051 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| temp | 3 | 129975.1720 | 43325.0573 | 94.39 | <.0001 |
| spi(temp) | 8 | 3750.2369 | 468.7796 | 1.02 | 0.4470 |
| meth | 3 | 8613.2860 | 2871.0953 | 6.26 | 0.0027 |
| temp*meth | 9 | 1374.3896 | 152.7100 | 0.33 | 0.9551 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| temp | 3 | 129975.1720 | 43325.0573 | 94.39 | <.0001 |
| spi(temp) | 8 | 3750.2369 | 468.7796 | 1.02 | 0.4470 |
| meth | 3 | 8613.2860 | 2871.0953 | 6.26 | 0.0027 |
| temp*meth | 9 | 1374.3896 | 152.7100 | 0.33 | 0.9551 |

| Source | Type III Expected Mean Square |
|---|---|
| temp | Var(Error) + 4 Var(spi(temp)) + Q(temp,temp*meth) |
| spi(temp) | Var(Error) + 4 Var(spi(temp)) |
| meth | Var(Error) + Q(meth,temp*meth) |
| temp*meth | Var(Error) + Q(temp*meth) |

3. Was there a significant interaction between temperature and method? Conduct one test to answer this question. Provide a test statistic, its degrees of freedom, a $p$-value, and a brief conclusion.

4. Were there significant method main effects? Conduct one test to answer this question. Provide a test statistic, its degrees of freedom, a $p$-value, and a brief conclusion.

5. Were there significant temperature main effects? Conduct one test to answer this question. Provide a test statistic, its degrees of freedom, a $p$-value, and a brief conclusion.

6. For split-plot designs like the one considered here, it can be shown that the variance of a mean corresponding to a level of factor $A$ is

$$\text{Var}(\bar{y}_{i\cdot\cdot}) = \frac{\sigma^2_{wp}}{r} + \frac{\sigma^2_{sp}}{rb}$$

The mean corresponding to the temperature 25 degrees Fahrenheit was 571.74. Determine a 95% confidence interval for the 25-degree mean.

7. For split-plot designs like the one considered here, it can be shown that the variance of a mean corresponding to a level of factor $B$ is

$$\text{Var}(\bar{y}_{\cdot j\cdot}) = \frac{\sigma^2_{wp} + \sigma^2_{sp}}{ra}$$

Provide the standard error of the mean corresponding to the spray-dry method.

8. The approximate degrees of freedom associated with a linear combination of mean squares is given by Satterthwaite's method as

$$\text{d.f. of } \sum_{i=1}^{k} c_i \text{MS}_i \approx \frac{\left(\sum_{i=1}^{k} c_i \text{MS}_i\right)^2}{\sum_{i=1}^{k} (c_i \text{MS}_i)^2 / \text{df}_i} \text{ where df}_i \text{ is the d.f. for MS}_i.$$

The mean for the spray-dry method was 538.97. Find an approximate 95% confidence interval for the mean of the spray-dry method.

9. The variance of a **contrast** of factor B means is

$$\text{Var}\left(\sum_{j=1}^{b} c_j \bar{y}_{\cdot j\cdot}\right) = \frac{\sigma^2_{sp}}{ra} \sum_{j=1}^{b} c_j^2.$$

The spray-dry and freeze-dry means were 538.97 and 521.41, respectively. Was there a significant difference the spray-dry and freeze-dry methods? Provide a test statistic, its degrees of freedom, a $p$-value, and a brief conclusion.

```
proc mixed method=type3 data=one;
  class temp meth spi;
  model ec=temp meth temp*meth / ddfm=satterthwaite;
  random spi(temp);
  lsmeans temp meth;
  estimate 'temp 25 - temp 80' temp 1 0 0 -1;
  estimate 'spray dry - freeze dry' meth 0 0 -1 1;
run;
```

Type 3 Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | Expected Mean Square |
|---|---|---|---|---|
| temp | 3 | 129975 | 43325 | Var(Residual) + 4 Var(spi(temp)) + Q(temp,temp*meth) |
| meth | 3 | 8613.285966 | 2871.095322 | Var(Residual) + Q(meth,temp*meth) |
| temp*meth | 9 | 1374.389645 | 152.709961 | Var(Residual) + Q(temp*meth) |
| spi(temp) | 8 | 3750.236867 | 468.779608 | Var(Residual) + 4 Var(spi(temp)) |
| Residual | 24 | 11016 | 459.000179 | Var(Residual) |

| Source | Error Term | Error DF | F Value | Pr > F |
|---|---|---|---|---|
| temp | MS(spi(temp)) | 8 | 92.42 | <.0001 |
| meth | MS(Residual) | 24 | 6.26 | 0.0027 |
| temp*meth | MS(Residual) | 24 | 0.33 | 0.9551 |
| spi(temp) | MS(Residual) | 24 | 1.02 | 0.4470 |
| Residual | . | . | . | . |

Covariance Parameter Estimates

| Cov Parm | Estimate |
|---|---|
| spi(temp) | 2.4449 |
| Residual | 459.00 |

Estimates

| Label | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| temp 25 - temp 80 | 104.99 | 8.8391 | 8 | 11.88 | <.0001 |
| spray dry - freeze dry | 17.5611 | 8.7464 | 24 | 2.01 | 0.0560 |

Least Squares Means

| Effect | temp | meth | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|---|
| temp | 25 | | 571.74 | 6.2502 | 8 | 91.48 | <.0001 |
| temp | 40 | | 591.43 | 6.2502 | 8 | 94.63 | <.0001 |
| temp | 60 | | 493.70 | 6.2502 | 8 | 78.99 | <.0001 |
| temp | 80 | | 466.75 | 6.2502 | 8 | 74.68 | <.0001 |
| meth | | 1 | 514.87 | 6.2011 | 32 | 83.03 | <.0001 |
| meth | | 2 | 548.38 | 6.2011 | 32 | 88.43 | <.0001 |
| meth | | 3 | 521.41 | 6.2011 | 32 | 84.08 | <.0001 |
| meth | | 4 | 538.97 | 6.2011 | 32 | 86.91 | <.0001 |