HANDOUT # 7

CRD WITH FACTORIAL TREATMENT STRUCTURE

1. Model for a CRD with Two Factors
2. Example
3. Three Types of Effects: Simple, Main, and Interaction
4. Estimation of Treatment Means
5. AOV for Experiments with Equal Number of Replications
6. AOV for Experiments with Unequal Number of Replications
7. Tests of Contrasts and Multiple Comparisons
8. Experiments with a Qualitative and Quantitative Factor Levels
9. Experiments with Three Factors
10. Determination of Number of Replications
11. Evaluation of Model Conditions

CRD WITH Two FACTORS

Example of CRD with a 4x3 Factorial Treatment Structure with 3 Reps per treatment

An experiment was conducted to compare the effect of different soil pH (Factor F_1) and calcium additives (Factor F_2) on the increase in trunk diameters of orange trees. Annual applications of elemental sulfur, gypsum, soda ash, and other minerals were applied to provide the soil pH levels of 4, 5, 6, and 7. Three levels of a calcium supplement (100, 200, and 300 pounds per acre) were also applied. All factor-level combinations of the two variables were used in the experiment. At the end of a 2-year study period, three tree diameters were recorded at each factor-level combination. The data along with all relevant means are given in the following table. The treatment means $\bar{y}_{ij.}$ are given in parentheses.

	pH level				
Calcium Level	4	5	6	7	Mean
100	5.2, 5.9, 6.3 (5.8)	7.1, 7.3, 7.5 (7.3)	7.6, 7.2, 7.4 (7.4)	7.2, 7.5, 7.2 (7.3)	6.95
200	7.6, 7.0, 7.6 (7.4)	7.5, 7.3, 7.1 (7.3)	7.6, 7.4, 7.8 (7.6)	7.4, 7.0, 6.9 (7.1)	7.35
300	6.4, 6.7, 6.1 (6.4)	7.3, 7.5, 7.4 (7.4)	7.2, 7.3, 7.1 (7.2)	6.8, 6.6, 6.4 (6.6)	6.9
Mean	6.53	7.33	7.4	7.0	7.07

The cell means model for this type of experiment is

$$y_{ijk} = \mu_{ij} + e_{ijk}, \text{ where}$$

1. $i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, n_{ij} \quad n = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$

   In the above experiment, $a = 3, \ b = 4, \ n_{ij} = 3, \ \Rightarrow \ n = (3)(4)(3) = 36$

2. e_{ijk}'s are iid $N(0, \sigma_e^2)$, that is,

   independent data from $t = ab$ populations having normal distributions with equal variances

3. μ_{ij} is the mean response of the treatment consisting of the

   ith level of Factor F_1 and *jth* level of Factor F_2;

Effect Representation: $\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$

Model is over parameterized, there are $t = ab$ population means but the effects representation has $(a+1)(b+1)$ parameters: $1 \to \mu, \quad a \to \tau_i s, \quad b \to \beta_j s, \quad ab \to (\tau\beta)_{ij} s \Rightarrow 1 + a + b + ab = (a+1)(b+1)$ parameters in the model

1. Constraints on parameters:
2. $\tau_a = 0; \quad \beta_b = 0;$
3. $(\tau\beta)_{ib} = 0$ for $i = 1, \ldots, a;$
4. $(\tau\beta)_{aj} = 0;$ for $j = 1, \ldots, b.$

With the constraints, there are now 1 + (a-1) + (b-1) + [ab-a-b+1] = ab parameters in the model.

Relate Effects model parameters to Cell Means model parameters:

1. $\mu_{ab} = \mu + \tau_a + \beta_b + (\tau\beta)_{ab} = \mu + 0 + 0 + 0 \Rightarrow \mu = \mu_{ab};$
2. For $i \neq a; \quad \mu_{ib} = \mu + \tau_i + \beta_b + (\tau\beta)_{ib} = \mu + \tau_i + 0 + 0 \Rightarrow \tau_i = \mu_{ib} - \mu_{ab}$;
3. For $j \neq b, \quad \mu_{aj} = \mu + \tau_a + \beta_j + (\tau\beta)_{aj} = \mu + 0 + \beta_j + 0 \Rightarrow \beta_j = \mu_{aj} - \mu_{ab}$
4. For $i \neq a, \quad j \neq b, \quad \mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$

   $\Rightarrow (\tau\beta)_{ij} = \mu_{ij} - \mu - \tau_i - \beta_j = \mu_{ij} - \mu_{ab} - (\mu_{ib} - \mu_{ab}) - (\mu_{aj} - \mu_{ab})$

   $= (\mu_{ij} - \mu_{ib}) - (\mu_{aj} - \mu_{ab}) = (\mu_{ij} - \mu_{aj}) - (\mu_{ib} - \mu_{ab})$

There are three types of differences in the treatment means of interest to researchers:

1. **Simple Effects** of a Factor are contrasts between the levels of one Factor at <u>fixed</u> levels of the second Factor.

   Fix factor F_1 at ith level, then construct contrast across levels of factor F_2 or vice versa.

   For example: $\mu_{i1} - \mu_{i2}$ or $\mu_{i2} - \frac{1}{2}(\mu_{i1} + \mu_{i3})$

2. **Main Effects** of the Factors are contrasts between levels of one Factor averaged over the levels of the second Factor.

   Construct contrast in the means of factor F_1: $\bar{\mu}_{1.}, \ \bar{\mu}_{2.}, \ \ldots, \bar{\mu}_{a.}$

   where $\bar{\mu}_{..} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij}; \quad \bar{\mu}_{i.} = \frac{1}{b}\sum_{j=1}^{b}\mu_{ij}; \quad \bar{\mu}_{.j} = \frac{1}{a}\sum_{i=1}^{a}\mu_{ij};.$

   Main Effect of Factor F_1: contrasts in $\bar{\mu}_{i.}$, for example, $\bar{\mu}_{i.} - \bar{\mu}_{h.}$

   Main Effects of Factor F_2: contrasts in $\bar{\mu}_{.j}$, for example, $\bar{\mu}_{.j} - \bar{\mu}_{.k}$

Example: Suppose Factor F_1 has two levels and Factor F_2 has three levels

Main Effects of Factor F_1: $\bar{\mu}_{1.} - \bar{\mu}_{2.}$

Main Effects of Factor F_2:

a. $\frac{2}{3}\bar{\mu}_{.1} - \frac{1}{3}\bar{\mu}_{.2} - \frac{1}{3}\bar{\mu}_{.3}$

b. $-\frac{1}{3}\bar{\mu}_{.1} + \frac{2}{3}\bar{\mu}_{.2} - \frac{1}{3}\bar{\mu}_{.3}$

c. $-\frac{1}{3}\bar{\mu}_{.1} - \frac{1}{3}\bar{\mu}_{.2} + \frac{2}{3}\bar{\mu}_{.3}$

Are the Main Effect Contrasts for Factor F_2 orthogonal?

3. **Interaction Effects** between two Factors measure differences between the Simple Effects of one Factor at different levels of the second Factor.

   Example: Suppose Factor F_1 has two levels and Factor F_2 has three levels. Compare the two levels of Factor F_1 at the three different levels of Factor F_2 or vice versa:

   a. $(\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22}) = d_1 - d_2$

   b. $(\mu_{12} - \mu_{22}) - (\mu_{13} - \mu_{23}) = d_2 - d_3$

   c. $(\mu_{13} - \mu_{23}) - (\mu_{11} - \mu_{21}) = d_3 - d_1$

   d. Let $C_1 = \mu_{11} - 2\mu_{12} + \mu_{13}$ and $C_2 = \mu_{21} - 2\mu_{22} + (\mu_{23}$

   C_1 and C_2 are two simple effects contrasts in the levels of Factor F_2.

   An interaction contrast would be $C_1 - C_2$.

   Note $C_1 - C_2 = (\mu_{11} - \mu_{21}) - 2(\mu_{12} - \mu_{22}) + (\mu_{13} - \mu_{23})$

   $C_1 - C_2$ is a comparison of the simple effects of Factor 1 across the levels of Factor 2.

Definition of Interaction: Two factors F_1 and F_2 are said to **Interact** if the differences in the average response at all pairs of levels F_1 are NOT the same at ALL levels of F_2.

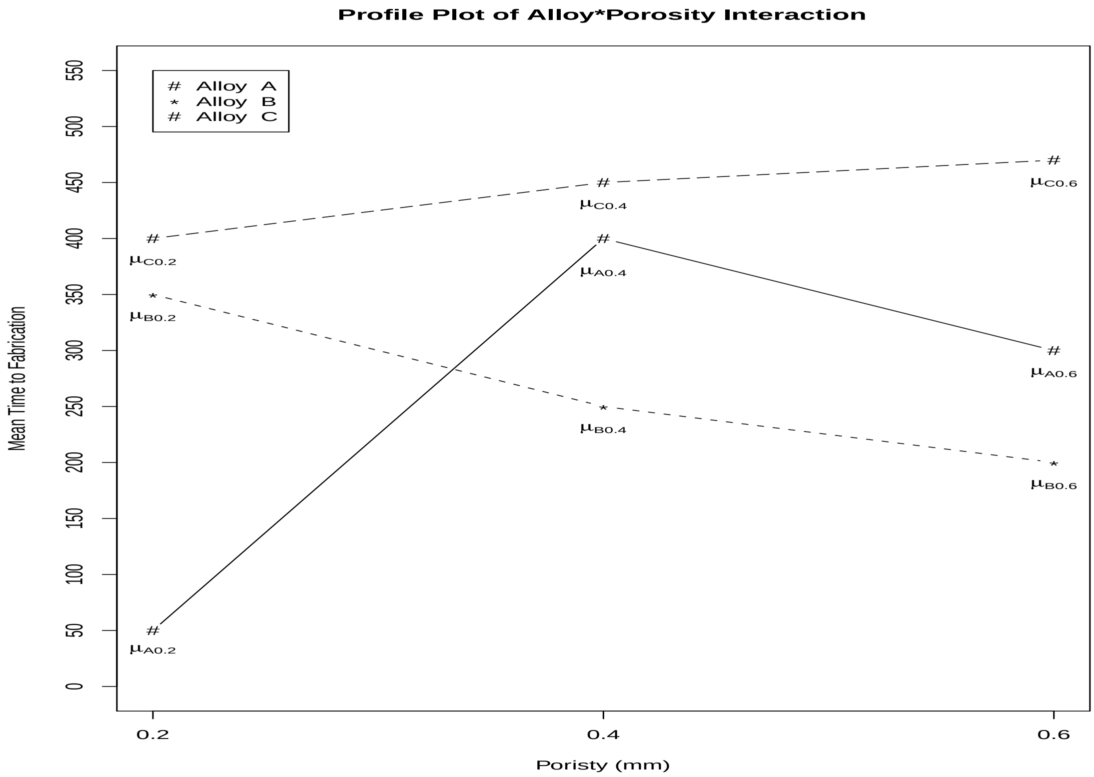
There is an interaction between factors F_1 and F_2 if

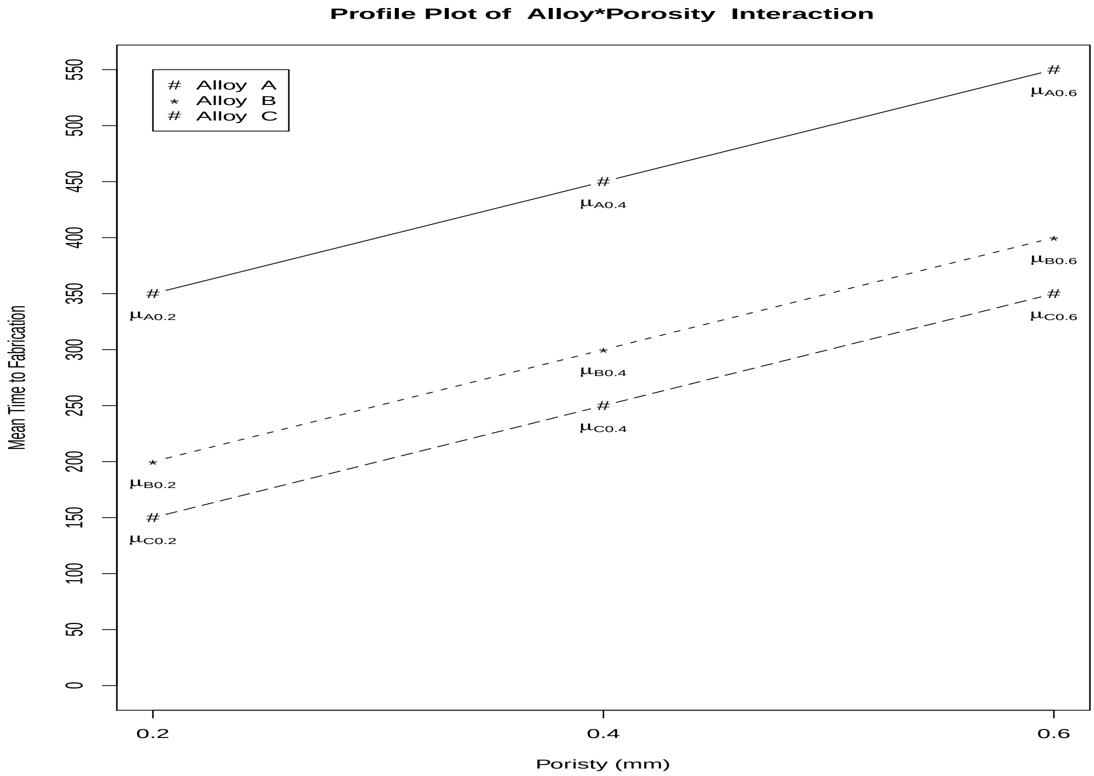
$$(\mu_{ij} - \mu_{kj}) \neq (\mu_{ih} - \mu_{kh}) \quad \text{for at least one choice of} \quad (i, k, j, h)$$

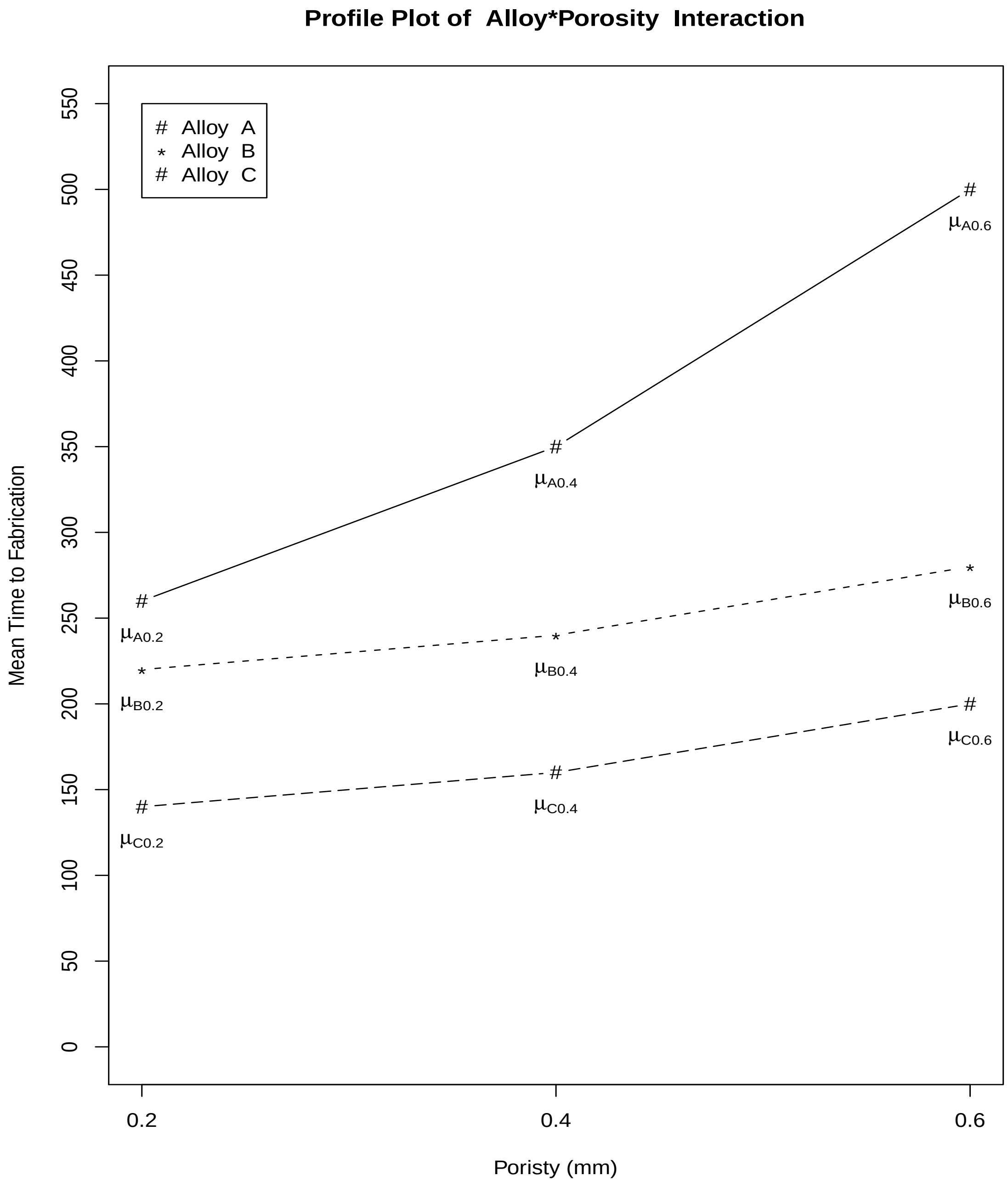
In the plot of the treatment means, the **profile plot**, if $d_1 = d_2 = d_3$, both in magnitude and sign, then Factors F_1 and F_2 do NOT interact.

If at least one of the d_i differs from the remaining d_js then there is an interaction between the Factors.

Profile Plots of the Treatment Means for two Factors F_1 and F_2






Profile Plot of Alloy*Porosity Interaction
Alloy A
* Alloy B
# Alloy C
Mean Time to Fabrication
Poristy (mm)
0
50
100
150
200
250
300
350
400
450
500
550
0.2
0.4
0.6
μA0.2
μA0.4
μA0.6
μB0.2
μB0.4
μB0.6
μC0.2
μC0.4
μC0.6

Matrix Representation of CRD Factorial Treatment Model

Cell Means Model:

Suppose we have two factors: F_1 with a levels and F_2 with b levels with n_{ij} EUs randomly assigned to each of the $t = ab$ treatments yielding a total of $n = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$ responses.

Cell Means Model: $y_{ijk} = \mu_{ij} + e_{ijk}; \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad t = ab;$
$\quad k = 1, \ldots, n_{ij}; \quad n = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$

Model in matrix form: $\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{e}$,

where $\mathbf{Y}$ is the n x 1 data vector, $\mathbf{X}$ is the n x t design matrix, $\boldsymbol{\beta}$ is the t x 1 vector of parameters, and $\mathbf{e}$ is the n x 1 vector of errors.

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ \vdots \\ Y_{11n_{11}} \\ Y_{121} \\ Y_{122} \\ \vdots \\ Y_{12n_{12}} \\ \vdots \\ Y_{ab1} \\ Y_{ab2} \\ \vdots \\ Y_{abn_{ab}} \end{bmatrix}_{\text{n x 1}} \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{\text{n x t}} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{ab} \end{bmatrix}_{\text{t x 1}} \quad \mathbf{e} = \begin{bmatrix} e_{111} \\ e_{112} \\ \vdots \\ e_{11n_{11}} \\ e_{121} \\ e_{122} \\ \vdots \\ e_{12n_{12}} \\ \vdots \\ e_{ab1} \\ e_{ab2} \\ \vdots \\ e_{abn_{ab}} \end{bmatrix}_{\text{n x 1}}$$

The design matrix can also be written as:

$$\mathbf{X} = \begin{bmatrix} 1_{n_1} & 0_{n_1} & 0_{n_1+n_2} & \cdots & 0_{n-n_t} \\ 0_{n-n_1} & 1_{n_2} & 1_{n_3} & \cdots & 1_{n_t} \\ & 0_{n-n_1-n_2} & 0_{n-n_1-n_2-n_3} & & \end{bmatrix}_{\text{n x t}}$$

A third way that design matrix can be expressed:

$$\mathbf{X} = Diag\left[\mathbf{J}_{n_{11}}, \mathbf{J}_{n_{12}}, \cdots, \mathbf{J}_{n_{ab}}\right]_{n \times t}; \quad \text{where} \quad \mathbf{J}_{n_{ij}} \quad \text{is a vector of} \quad n_{ij} \quad 1's$$

The parameters in the matrix model are the treatment means:

$$\beta_1 = \mu_{1,1}; \quad \beta_2 = \mu_{1,2}; \quad \ldots \quad ; \beta_t = \mu_{a,b}$$

Regression Model Formulation

Let the ab treatments be labeled as

$1 = (1, 1),\ 2 = (1, 2),\ \cdots,\ b = (1, b),\ b + 1 = (2, 1),\ \cdots,\ 2b = (2, b),\ \cdots,\ ab = (a, b)$

The design matrix can also be obtained by considering $t = ab$ "explanatory variables" $(\mathbf{X}1, \mathbf{X}2, \ldots, \mathbf{X}t)$, as is in regression models, with $\mathbf{X}i$ given by

$$Xi_j = \begin{cases} 1 & \text{if } jth \text{ EU receives Trt}\#i \\ 0 & \text{otherwise} \end{cases} \quad \text{for} \quad j = 1, \cdots, n; \quad i = 1, \cdots, t$$

Regression model: $y_j = \beta_1 X1_j + \beta_2 X2_j + \cdots + \beta_t Xt_j + e_j$, where $j = 1, 2, \ldots, n$ and $\beta_i = \mu_i$. Note that there is no intercept term.

The design matrix is of full column rank from which we conclude that the rank of $\left(\mathbf{X}'\mathbf{X}\right)$ is t. Thus, $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ exists. The least squares estimators of the model parameters are obtained from the equation:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Y}\right) = \begin{bmatrix} n_{11} & 0 & \cdots & 0 \\ 0 & n_{12} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & n_{ab} \end{bmatrix}^{-1} \begin{bmatrix} y_{11.} \\ y_{12.} \\ \vdots \\ y_{ab.} \end{bmatrix} = \begin{bmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \vdots \\ \bar{y}_{ab.} \end{bmatrix} = \begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{12} \\ \vdots \\ \hat{\mu}_{ab} \end{bmatrix}$$

This yields the following estimators:

Least Squares Estimation of Model Parameters

1. $\hat{\mu}_{ij} = \bar{y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$
2. $\widehat{\bar{\mu}}_{i.} = \frac{1}{b} \sum_{j=1}^{b} \widehat{\mu}_{ij} = \frac{1}{b} \sum_{j=1}^{b} \bar{y}_{ij.} = \frac{1}{b} \sum_{j=1}^{b} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$
3. $\widehat{\bar{\mu}}_{.j} = \frac{1}{a} \sum_{i=1}^{a} \widehat{\mu}_{ij} = \frac{1}{a} \sum_{i=1}^{a} \bar{y}_{ij.} = \frac{1}{a} \sum_{i=1}^{a} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$
4. $\widehat{\bar{\mu}}_{..} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{a} \widehat{\mu}_{ij} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \bar{y}_{ij.} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$

   Note that

a. $\widehat{\mu}_{ij} = \bar{y}_{ij.}$ whether or not the number of reps are equal

b. $\widehat{\bar{\mu}}_{i.} = \bar{Y}_{i..}$ only if $n_{ij} = r$

c. $\widehat{\bar{\mu}}_{.j} = \bar{Y}_{.j.}$ only if $n_{ij} = r$

d. $\widehat{\bar{\mu}}_{..} = \bar{Y}_{...}$ only if $n_{ij} = r$

EXAMPLE: Consider an experiment which has Two Factors F_1 with $a = 3$ levels and F_2 with $b = 2$ levels and $r = 3$ reps for each of the $t = 6$ treatments yielding $n = (2)(3)(3) = 18$ responses, given in the following vector. The cell means model would be given by

Model in matrix form: $\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{e}$,

where $\mathbf{Y}$ is the 18 x 1 data vector, $\mathbf{X}$ is the 18 x 6 design matrix, $\boldsymbol{\beta}$ is the 6 x 1 vector of parameters, and $\mathbf{e}$ is the 18 x 1 vector of errors.

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{221} \\ Y_{222} \\ Y_{223} \\ Y_{311} \\ Y_{312} \\ Y_{313} \\ Y_{321} \\ Y_{322} \\ Y_{323} \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \\ 0 \\ 10 \\ 8 \\ 6 \\ 8 \\ 6 \\ 4 \\ 14 \\ 10 \\ 6 \\ 4 \\ 2 \\ 0 \\ 15 \\ 12 \\ 9 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \\ \mu_{31} \\ \mu_{32} \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_{111} \\ e_{112} \\ e_{113} \\ e_{121} \\ e_{122} \\ e_{123} \\ e_{211} \\ e_{212} \\ e_{213} \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{311} \\ e_{312} \\ e_{313} \\ e_{321} \\ e_{322} \\ e_{323} \end{bmatrix}$$

The Least Squares estimates of the model parameters are obtained from the equation:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Y}\right) = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 12 \\ 24 \\ 18 \\ 30 \\ 6 \\ 36 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 6 \\ 10 \\ 2 \\ 12 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{12} \\ \hat{\mu}_{13} \\ \hat{\mu}_{21} \\ \hat{\mu}_{22} \\ \hat{\mu}_{23} \end{bmatrix}$$

Design Matrix for the Effects Model

<u>Effects Model</u>: $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}, \quad i = 1,2,3; \quad j = 1,2; \quad k = 1,2,$

where τ_i is the fixed effect of ith level of F1, β_j is the fixed effect of jth level of F2, $(\tau\beta)_{ij}$ are interaction effects between F1 and F2

Constraints on parameters: $\tau_3 = 0,\ \beta_2 = 0,\ (\tau\beta)_{12} = (\tau\beta)_{22} = (\tau\beta)_{31} = (\tau\beta)_{32} = 0$

The model parameters are reduced to $\mu,\ \tau_1,\ \tau_2,\ \beta_1,\ (\tau\beta)_{11},\ (\tau\beta)_{21}$

The relationship between these parameters and the treatment means is as follows:

- $\mu_{ij} \ = \ \mu + \tau_i + \beta_j + (\tau\beta)_{ij} \ \Rightarrow$
- $\mu_{11} \ = \ \mu + \tau_1 + \beta_1 + (\tau\beta)_{11} \ = \ \mu + \tau_1 + \beta_1 + (\tau\beta)_{11}$
- $\mu_{12} \ = \ \mu + \tau_1 + \beta_2 + (\tau\beta)_{12} \ = \ \mu + \tau_1$
- $\mu_{21} \ = \ \mu + \tau_2 + \beta_1 + (\tau\beta)_{21} \ = \ \mu + \tau_2 + \beta_1 + (\tau\beta)_{21}$
- $\mu_{22} \ = \ \mu + \tau_2 + \beta_2 + (\tau\beta)_{22} \ = \ \mu + \tau_2$
- $\mu_{31} \ = \ \mu + \tau_3 + \beta_1 + (\tau\beta)_{31} \ = \ \mu + \beta_1$
- $\mu_{32} \ = \ \mu + \tau_3 + \beta_2 + (\tau\beta)_{32} \ = \ \mu$

The effects model in matrix form is given with design matrix, $\mathbf{X}$, and parameter vector, $\boldsymbol{\beta}$

Model in matrix form: $\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{e}$,

where $\mathbf{Y}$ is the 18 x 1 data vector, $\mathbf{X}$ is the 18 x 6 design matrix, $\boldsymbol{\beta}$ is the 6 x 1 vector of parameters, and $\mathbf{e}$ is the 18 x 1 vector of residuals.

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ \\ Y_{221} \\ Y_{222} \\ Y_{223} \\ \\ Y_{311} \\ Y_{312} \\ Y_{313} \\ \\ Y_{321} \\ Y_{322} \\ Y_{323} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ (\tau\beta)_{11} \\ (\tau\beta)_{21} \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_{111} \\ e_{112} \\ e_{113} \\ \\ e_{121} \\ e_{122} \\ e_{123} \\ \\ e_{211} \\ e_{212} \\ e_{213} \\ \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{311} \\ e_{312} \\ e_{313} \\ \\ e_{321} \\ e_{322} \\ e_{323} \end{bmatrix}$$

Development of Test Statistics

In experiments involving a factorial treatment structure, there are several types of hypotheses that the researchers wants to test. We will state them in terms of the parameters in both models:

(a) Cell Means Model: $y_{ijk} = \mu_{ij} + e_{ijk}$

(b) Effects Model: $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$, with the following constraints on the parameters:

$$\tau_a = 0; \quad \beta_b = 0; \quad (\tau\beta)_{ib} = 0 \text{ for } i = 1, \ldots, a; \quad (\tau\beta)_{aj} = 0; \text{ for } j = 1, \ldots, b.$$

Overall Treatment Differences Hypotheses:

(a) Cell Means Model: $H_o : \mu_{ij} = \mu$ for all (i, j) vs $H_1 : \mu_{ij}$ not all equal

(b) Effects Model: $H_o : \tau_i = 0, \beta_j = 0, (\tau\beta)_{ij} = 0$ for all (i, j) vs $H_1 : \tau_i; \beta_j; (\tau\beta)_{ij}$ not all 0

Interactions between the Factors:

(a) Cell Means Model: $H_o : (\mu_{ij} - \mu_{kj}) = (\mu_{ih} - \mu_{kh})$ for all (i,k,j,h) in the set $\{(i, k, j, h) : (i \neq k = 1, \ldots, a; j \neq h = 1, \ldots, b)\}$
vs $H_1 : (\mu_{ij} - \mu_{kj}) \neq (\mu_{ih} - \mu_{kh})$ for at least one set (i,k,j,h)

(b) Effects Model: $H_o : (\tau\beta)_{ij} = 0$ for all (i, j) vs $H_1 : (\tau\beta)_{ij}$ not all 0

Main Effect of Factor F_1: (Given no interaction) $y_{ijk} = \mu + \tau_i + \beta_j + e_{ijk}$

(a) Cell Means Model: $H_o : \bar{\mu}_{1.} = \bar{\mu}_{2.} = \cdots = \bar{\mu}_{a.}$ vs $H_1 : \bar{\mu}_{i.}$ not all equal

(b) Effects Model: $H_o : \tau_i = 0$ for all $(i = 1, \ldots, a)$ vs $H_1 : \tau_i$ not all 0

Main Effect of Factor F_2: (Given no interaction) $y_{ijk} = \mu + \tau_i + \beta_j + e_{ijk}$

(a) Cell Means Model: $H_o : \bar{\mu}_{.1} = \bar{\mu}_{.2} = \cdots = \bar{\mu}_{.b}$ vs $H_1 : \bar{\mu}_{.j}$ not all equal

(b) Effects Model: $H_o : \beta_j = 0$ for all $(j = 1, \ldots, b)$ vs $H_1 : \beta_j$ not all 0

Case 1: Equal Number of Replications: $\mathbf{n_{ij} = r}$

When the experiment has an equal number of replications per treatment ($n_{ij} = r$), the test statistics are obtained by a decomposition of SS_{TOT} into its several components: SS_{TRT} and SSE. Then, SS_{TRT} is decomposed into SS_{F_1}, SSF_2, and SS_{F1*F_2} vnn

$$SS_{TOT} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(y_{ijk} - \bar{y}_{...})^2 \quad \text{with} \quad df = n - 1$$

$$SS_{TRT} = \mathbf{Y}^T\mathbf{A}\mathbf{Y} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{y}_{ij.} - \bar{y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} r(\bar{y}_{ij.} - \bar{y}_{...})^2$$

with $df = Rank(\mathbf{A}) = ab - 1$

$$SS_{F_1} = \mathbf{Y}^T\mathbf{B}\mathbf{Y} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{y}_{i..} - \bar{y}_{...})^2 = \sum_{i=1}^{a} br(\bar{y}_{i..} - \bar{y}_{...})^2$$

with $df = Rank(\mathbf{B}) = a - 1$

$$SS_{F_2} = \mathbf{Y}^T\mathbf{C}\mathbf{Y} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{y}_{.j.} - \bar{y}_{...})^2 = \sum_{j=1}^{b} ar(\bar{y}_{.j.} - \bar{y}_{...})^2$$

with $df = Rank(\mathbf{C}) = b - 1$

$$\begin{aligned} SS_{F_1*F_2} &= \mathbf{Y}^T\mathbf{D}\mathbf{Y} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}((\bar{y}_{ij.} - \bar{y}_{...}) - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}))^2 \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b} r((\bar{y}_{ij.} - \bar{y}_{...}) - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}))^2 \\ &= SS_{Trt} - SS_{F_1} - SS_{F_2} \end{aligned}$$

with $df = Rank(\mathbf{D}) = ab - 1 - (a-1) - (b-1) = (a-1)(b-1)$

$$\begin{aligned} SS_E &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\hat{e}_{ijk}^2 \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(y_{ijk} - \hat{\mu}_{ij})^2 \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(y_{ijk} - \bar{y}_{ij.})^2 \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}((y_{ijk} - \bar{y}_{...}) - (\bar{y}_{ij.} - \bar{y}_{...}))^2 \\ &= SS_{TOT} - SS_{TRT} \quad \text{with} \quad df = \sum_{i=1}^{a}\sum_{j=1}^{b}(r-1) = ab(r-1) = n - ab \end{aligned}$$

Note: $df_{TOT} = n - 1 = abr - 1 = ab - 1 + ab(r-1) = df_{TRT} + df_E$

Note: $df_{TRT} = ab - 1 = (a-1) + (b-1) + (a-1)(b-1) = df_{F_1} + df_{F_2} + df_{F_1*F_2}$

Note: $df_{TOT} = n - 1 = (a-1) + (b-1) + (a-1)(b-1) + (n-ab) = df_{F_1} + df_{F_2} + df_{F_1*F_2} + df_E$

The degrees of freedom df for each of the above sum of squares can be interpreted in terms of the hypotheses for which they are a component of the test statistics. The df are the reduction in the number parameters in the model when the null hypothesis is true. That is,
$df_{EFFECT}=$ (Number of model parameters) - (Number of model parameters under H_o)
$= ab - N_{H_o}$

Overall Treatment Differences Hypotheses:

(a) Cell Means Model: $y_{ij} = \mu_{ij} + e_{ijk}$

To test: $H_o : \mu_{ij} = \mu \quad \Rightarrow \quad df_{TRT} = ab - N_{H_o} = ab - 1$

(b) Effects Model: $y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$

To test: $H_o : \tau_i = 0, \beta_j = 0, (\tau\beta)_{ij} = 0$ for all $(i, j) \Rightarrow \quad df_{TRT} = ab - N_{H_o} = ab - 1$

Interactions between the Factors:

(a) Cell Means Model: $y_{ij} = \mu_{ij} + e_{ijk}$

To test: $H_o : (\mu_{ij} - \mu_{kj}) = (\mu_{ih} - \mu_{kh})$ for all (i,k,j,h) in the set

$\{(i, k, j, h) : (i \neq k = 1, \ldots, a; j \neq h = 1, \ldots, b)\} \quad \Rightarrow$

$df_{A*B} = ab - N_{H_o} = ab - [b + (a - 1)] = (a - 1)(b - 1)$

(b) Effects Model: $y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$

To test: $H_o : (\tau\beta)_{ij} = 0$ for all $(i, j) \quad \Rightarrow$

$df_{A*B} = ab - N_{H_o} = ab - [1 + (a - 1) + (b - 1)] = (a - 1)(b - 1)$

Main Effect of Factor F_1:

(a) Cell Means Model: $y_{ij} = \mu_{ij} + e_{ijk}$

To test: $H_o : \bar{\mu}_{1.} = \bar{\mu}_{2.} = \cdots = \bar{\mu}_{a.} \quad \Rightarrow$

$df_A = ab - N_{H_o} = ab - [a(b-1)+1] = a-1$

(b) Effects Model: $y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$

To test: $H_o : \tau_i = 0$ for all $(i = 1, \ldots, a) \quad \Rightarrow$

$df_A = ab - N_{H_o} = ab - [1 + (b-1) + (a-1)(b-1)] = a-1$

Main Effect of Factor F_2:

(a) Cell Means Model:$y_{ij} = \mu_{ij} + e_{ijk}$

To test: $H_o : \bar{\mu}_{.1} = \bar{\mu}_{.2} = \cdots = \bar{\mu}_{.b} \quad \Rightarrow$

$df_B = ab - N_{H_o} = ab - [b(a-1)+1] = b-1$

(b) Effects Model: $y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$

To test: $H_o : \beta_j = 0$ for all $(j = 1, \ldots, b) \quad \Rightarrow$

$df_{A*B} = ab - N_{H_o} = ab - [1 + (a-1) - (a-1)(b-1)] = b-1$

We summarize the information from the experiment in an AOV Table:

ANOVA TABLE

Source	df	SumSquares	MeanSquares	F-value	p-value*
Model	$ab-1$	SS_{TRT}	$\frac{SS_{TRT}}{ab-1}$	$\frac{MS_{TRT}}{MSE}$	$1-G\left(\frac{MS_{TRT}}{MSE}\right)$
F_1	$a-1$	SS_{F_1}	$\frac{SS_{F_1}}{(a-1)}$	$\frac{MS_{F_1}}{MSE}$	$1-G\left(\frac{MS_{F_1}}{MSE}\right)$
F_2	$b-1$	SS_{F_2}	$\frac{SS_{F_2}}{(b-1)}$	$\frac{MS_{F_2}}{MSE}$	$1-G\left(\frac{MS_{F_2}}{MSE}\right)$
$F_1 * F_2$	$(a-1)(b-1)$	$SS_{F_1*F_2}$	$\frac{SS_{F_1*F_2}}{(a-1)(b-1)}$	$\frac{MS_{F_1*F_2}}{MSE}$	$1-G\left(\frac{MS_{F_1*F_2}}{MSE}\right)$
Error	$ab(r-1)$	SSE	$\frac{SSE}{ab(r-1)}$		
Total	$abr-1$	SS_{TOT}			

Where G is the cdf of a central F-distribution with df=ν_1, ν_2; ν_1=df of MS in numerator and ν_2=df of MS in denominator

EXAMPLE From the Calcium-pH Example,

Because there was an equal number of reps, $r = 3$, the above formulas are applicable with $a = 3$ and $b = 4$:

$$\bar{y}_{...} = 7.067 \quad \bar{y}_{11} = 5.8 \quad \bar{y}_{12} = 7.3 \quad \bar{y}_{13} = 7.4 \quad \bar{y}_{14} = 7.3 \quad \bar{y}_{21} = 7.4 \quad \bar{y}_{22} = 7.3 \quad \bar{y}_{23} = 7.6$$
$$\bar{y}_{24} = 7.1 \bar{y}_{31} = 6.4 \quad \bar{y}_{32} = 7.4 \quad \bar{y}_{33} = 7.2 \quad \bar{y}_{34} = 6.6$$

$$\bar{y}_{1.} = 6.95 \quad \bar{y}_{2.} = 7.35 \quad \bar{y}_{3.} = 6.9 \quad \bar{y}_{.1} = 6.533 \quad \bar{y}_{.2} = 7.333 \quad \bar{y}_{.3} = 7.4 \quad \bar{y}_{.4} = 7.0$$

1. $SS_{TOT} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (y_{ijk} - \bar{y}_{...})^2 = (5.2-7.067)^2 + (5.9-7.067)^2 + \cdots + (6.4-7.067)^2$

   =10.88

   with $df = abr - 1 = (3)(4)(3) - 1 = 35$

2. $SS_{TRT} = \sum_{i=1}^{a}\sum_{j=1}^{b} r(\bar{y}_{ij.} - \bar{y}_{...})^2$

$$= 3[(5.8-7.067)^2 + (7.3-7.067)^2 + (7.4-7.067)^2 + (7.3-7.067)^2 + (7.4-7.067)^2$$
$$+(7.3-7.067)^2 + (7.6-7.067)^2 + (7.1-7.067)^2 + (6.4-7.067)^2$$
$$+(7.4-7.067)^2 + (7.2-7.067)^2 + (6.6-7.067)^2] = 9.200004$$

with $df = ab - 1 = (3)(4) - 1 = 11$

3. $SS_{Ca} = \sum_{i=1}^{a} br(\bar{y}_{i..} - \bar{y}_{...})^2$

$$= (4)(3)[(6.95-7.067)^2 + (7.35-7.067)^2 + (6.9-7.067)^2] = 1.460004$$

with $df = a - 1 = 3 - 1 = 2$

4. $SS_{pH} = \sum_{j=1}^{b} ar(\bar{y}_{.j.} - \bar{y}_{...})^2$

$$= (3)(3)[(6.93-7.067)^2 + (7.33-7.067)^2 + (7.4-7.067)^2 + (7.0-7.067)^2] = 4.24161$$

with $df = b - 1 = 4 - 1 = 3$

5. $SS_{Ca*pH} = \sum_{i=1}^{a}\sum_{j=1}^{b} r((\bar{y}_{ij.} - \bar{y}_{...}) - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}))^2$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b} r((\bar{y}_{ij.} - \bar{y}_{i..}\bar{y}_{.j.} + \bar{y}_{...})^2$$
$$= SS_{TRT} - SS_{Ca} - SS_{pH} = 7.200004 - 1.460004 - 4.24161 = 1.49839$$

with $df = (a-1)(b-1) = (3-1)(4-1) = 6$

6. $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(y_{ijk} - \hat{\mu}_{ij})^2$

$$= (5.2-5.8)^2 + (5.9-5.8)^2 + (6.3-5.8)^2 + (7.1-7.3)^2 + (7.3-7.3)^2$$
$$+(7.5-7.3)^2 + (7.6-7.4)^2 + (7.2-7.4)^2 + (7.4-7.4)^2 + (7.2-7.3)^2$$
$$+(7.5-7.3)^2 + (7.2-7.3)^2 + (7.6-7.4)^2 + (7.0-7.4)^2 + (7.6-7.4)^2$$
$$+(7.5-7.3)^2 + (7.3-7.3)^2 + (7.1-7.3)^2 + (7.6-7.6)^2 + (7.4-7.6)^2$$
$$+(7.8-7.6)^2 + (7.4-7.1)^2 + (7.0-7.1)^2 + (6.9-7.1)^2 + (6.4-6.4)^2$$
$$+(6.7-6.4)^2 + (6.1-6.4)^2 + (7.3-7.4)^2 + (7.5-7.4)^2 + (7.4-7.4)^2$$
$$+(7.2-7.2)^2 + (7.3-7.2)^2 + (7.1-7.2)^2 + (6.8-6.6)^2 + (6.6-6.6)^2 + (6.4-6.6)^2$$
$$= 1.68$$

with $df = ab(r-1) = (3)(4)(3-1) = 24$

ANOVA TABLE

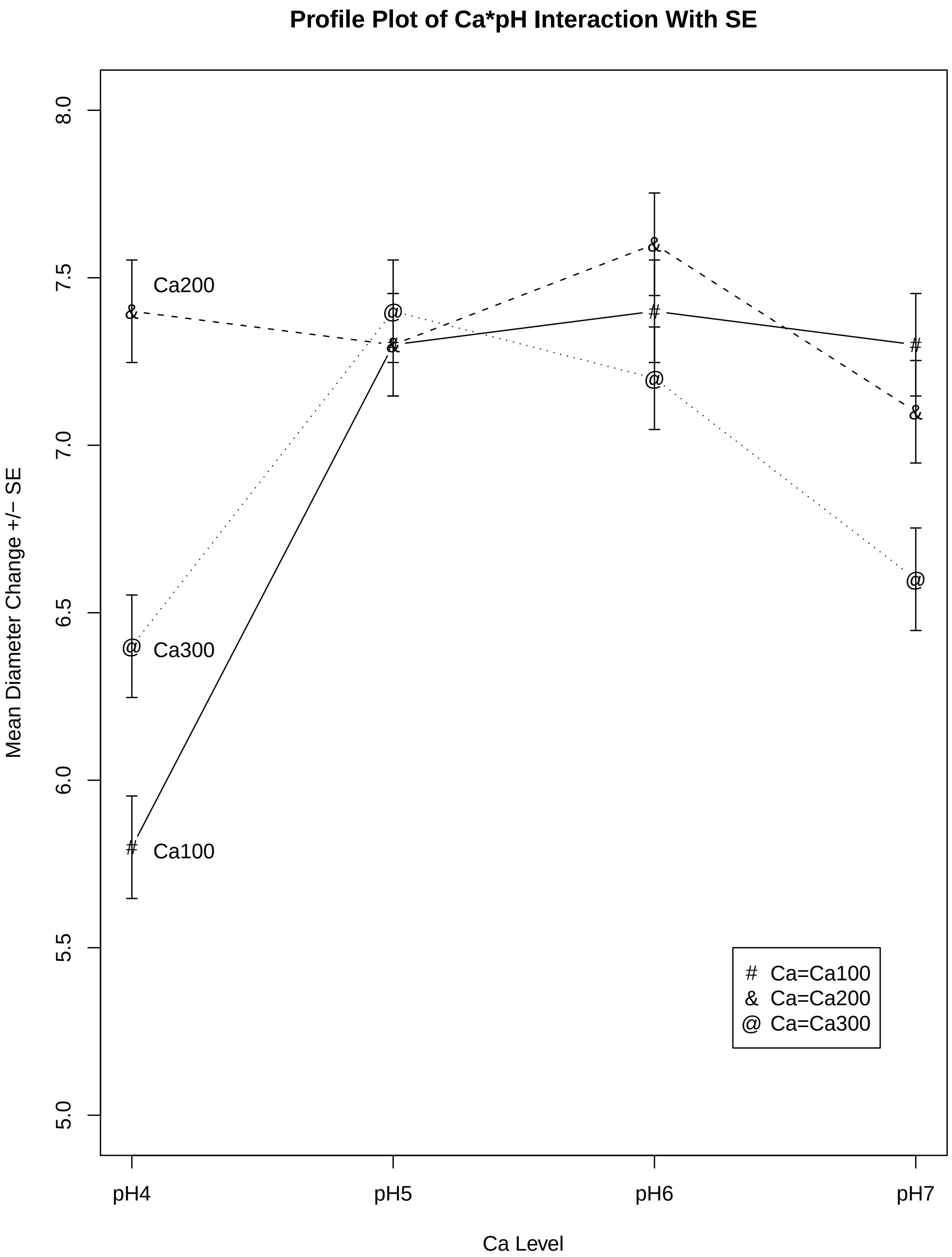
Source	df	SumSquares	MeanSquares	F-value	p-value
Model(TRT)	11	9.200	.8364	11.95	3.14×10^{-7}
Ca	2	1.460	.7300	10.43	0.00055
pH	3	4.242	1.414	20.20	9.41×10^{-7}
$Ca * pH$	6	3.498	.5833	8.33	6.08×10^{-5}
Error	24	1.680	.0700	.	.
Total	35	10.88	.	.	.

The p-value for the overall model (treatment effect) is obtained using the R-function:

1-pf(11.95,11,24) = 0.000000314

The following R Code, Profile_withbars.R, was used to produce a profile plot of sample means:

```
Ca100 =  c(5.8,7.3,7.4,7.3)
Ca200 =  c(7.4,7.3,7.6,7.1)
Ca300 =  c(6.4,7.4,7.2,6.6)
SE  = .153
Ca = cbind(Ca100,Ca200,Ca300)
x = c(1,2,3,4)
LxCa100 = c(1,2,3,4)
UxCa100 = c(1,2,3,4)
LyCa100 = Ca100 - SE
UyCa100 = Ca100 + SE
LxCa200 = c(1,2,3,4)
UxCa200 = c(1,2,3,4)
LyCa200 = Ca200 - SE
UyCa200 = Ca200 + SE
LxCa300 = c(1,2,3,4)
UxCa300 = c(1,2,3,4)
LyCa300 = Ca300 - SE
UyCa300 = Ca300 + SE
postscript(file="u:/meth2/output/Profile_bars.ps",horizontal=FALSE)
par(lab=c(3,15,4))
matplot(x,Ca,type="b",xlab="Ca Level",ylab="Mean Diameter Change +/- SE",
        main="Profile Plot of Ca*pH Interaction With SE",cex=.99,
        ylim=c(5,8),lab=c(3,10,5),col="black",pch=c("#&@"),xaxt="n")
segments(x,LyCa100,x,UyCa100)
segments(x,LyCa200,x,UyCa200)
segments(x,LyCa300,x,UyCa300)
BxL=c(.98,1.98,2.98,3.98)
BxR=c(1.02,2.02,3.02,4.02)
segments(BxL,LyCa100,BxR,LyCa100) segments(BxL,UyCa100,BxR,UyCa100)
segments(BxL,LyCa200,BxR,LyCa200) segments(BxL,UyCa200,BxR,UyCa200)
segments(BxL,LyCa300,BxR,LyCa300) segments(BxL,UyCa300,BxR,UyCa300)
axis(side=1,at=c(1,2,3,4),labels=c("pH4","pH5","pH6","pH7"))
legend(3.3,5.5,pch=c("#&@"),legend=c("Ca=Ca100","Ca=Ca200","Ca=Ca300"))
text(1.2,5.79,"Ca100") text(1.2,6.39,"Ca300") text(1.2,7.48,"Ca200")
graphics.off()
```


Profile Plot of Ca*pH Interaction With SE
Mean Diameter Change +/− SE
8.0
7.5
7.0
6.5
6.0
5.5
5.0
Ca200
Ca300
Ca100
Ca=Ca100
& Ca=Ca200
@ Ca=Ca300
pH4
pH5
pH6
pH7
Ca Level

The types of inferences that are conducted post AOV depend on the conclusions of the test for an interaction between the two factors:

1. If there is NOT significant evidence of an interaction between F_1 and F_2, then inferences about the treatment means can be made concerning the differences in the mean treatment responses across the levels of F_1 averaged over the levels of F_2. This is possible because when there is not an interaction between F_1 and F_2, the differences in the responses across the levels of F_1 are the same at all levels of F_2, hence we can examine the differences in the levels of F_1 averaged across the levels of F_2, and vice versa.

   (a) Multiple comparisons on $\bar{\mu}_{1.}, \bar{\mu}_{2.}, \ldots, \bar{\mu}_{a.}$ and on $\bar{\mu}_{.1}, \bar{\mu}_{.2}, \ldots, \bar{\mu}_{.b}$.

   Use Tukey HSD, Dunnett, Hsu procedures modified for the estimation of the standard errors of the differences in the LSE of the marginal means, $\bar{\mu}_{i.} - \bar{\mu}_{k.}$ or the differences in $\bar{\mu}_{.j} - \bar{\mu}_{.h}$. For comparing the marginal means of $F1$:

   For Tukey, $HSD = q(\alpha_o, k, \nu_2)\hat{\sigma}_e\sqrt{\frac{1}{rb}}$ with $k = a, \nu_2 = ab(r-1), \hat{\sigma}_e = \sqrt{MSE}$

   For Dunnett, $D(t-1, \nu_2) = d(\alpha_o, t-1, \nu_2)\hat{\sigma}_e\sqrt{\frac{2}{rb}}$ with $t = a, \nu_2 = ab(r-1), \hat{\sigma}_e = \sqrt{MSE}$

   In SAS, use **LSMEANS F1/ADJUST=TUKEY** to obtain pairwise comparisons of F1

   (b) Contrasts in $\bar{\mu}_{i.}$ or contrasts in $\bar{\mu}_{.j}$ can be constructed to answer particular research questions posed by the researcher.

2. If there IS significant evidence of an interaction between F_1 and F_2, then inferences concerning the differences in the mean treatment responses for F_1 must be conducted separately for each level of F_2. Because when there is an interaction, the differences in the treatment mean responses across the levels of F_1 may differ depending on the level of F_2.

   (a) Multiple comparisons on $\mu_{1j}, \mu_{2j}, \ldots, \mu_{aj}$ separately for each value $j = 1, \ldots, b$ or

   Multiple comparisons on $\mu_{i1}, \mu_{i2}, \ldots, \mu_{ib}$ separately for each value $i = 1, \ldots, a$

   Use Tukey HSD, Dunnett, Hsu procedures modified for the estimation of the standard errors of the differences in the LSE of the marginal means, $\mu_{ij} - \mu_{kj}$ for $j = 1, \ldots, b$ or the differences in $\mu_{ij} - \mu_{ih}$ for $i = 1, \ldots, a$.

   For Tukey, $HSD = q(\alpha_o/b, k, \nu_2)\hat{\sigma}_e\sqrt{\frac{1}{r}}$ with $k = a, \nu_2 = ab(r-1), \hat{\sigma}_e = \sqrt{MSE}$

   For Dunnett, $D(t-1, \nu_2) = d(\alpha_o/b, t-1, \nu_2)\hat{\sigma}_e\sqrt{\frac{2}{r}}$ with $t = a, \nu_2 = ab(r-1), \hat{\sigma}_e = \sqrt{MSE}$

   In SAS, use **LSMEANS F1*F2/ADJUST=TUKEY** to obtain pairwise comparisons

   (b) Contrasts in $\mu_{1j}, \mu_{2j}, \ldots, \mu_{aj}$, can be constructed to answer particular research questions posed by the researcher.

   (c) Multiple Comparisons or Contrasts on $\bar{\mu}_{i.}$s or $\bar{\mu}_{.j}$s should be avoided because they can lead to misleading conclusions.

Contrasts to Decompose SS in AOV

Suppose we have 2 factors A with 3 levels and B with 4 levels. We then have 12 treatments:

	Treatments											
Contrast	A1B1	A1B2	A1B3	A1B4	A2B1	A2B2	A2B3	A2B4	A3B1	A3B2	A3B3	A3B4
C1	-1	-1	-1	-1	0	0	0	0	1	1	1	1
C2	1	1	1	1	-2	-2	-2	-2	1	1	1	1
C3	-3	-1	1	3	-3	-1	1	3	-3	-1	1	3
C4	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1
C5	-1	3	-3	1	-1	3	-3	1	-1	3	-3	1
C6	3	1	-1	-3	0	0	0	0	-3	-1	1	3
C7	-1	1	1	-1	0	0	0	0	1	-1	-1	1
C8	1	-3	3	-1	0	0	0	0	-1	3	-3	1
C9	-3	-1	1	3	6	2	-2	-6	-3	-1	1	3
C10	1	-1	-1	1	-2	2	2	-2	1	-1	-1	1
C11	-1	3	-3	1	2	-6	6	-2	-1	3	-3	1

Note:

1. Two contrasts for Main A are $C_1 = \bar{\mu}_{1.} - \bar{\mu}_{3.}$ and $C_2 = \bar{\mu}_{1.} - 2\bar{\mu}_{2.} + \bar{\mu}_{3.}$
2. Three contrasts for Main B are $C_3 = -3\bar{\mu}_{.1} - \bar{\mu}_{.2} + \bar{\mu}_{.3} + 3\bar{\mu}_{.4}$,

   $C_4 = -\bar{\mu}_{.1} - \bar{\mu}_{.2} - \bar{\mu}_{.3} + \bar{\mu}_{.4}$, and $C_5 = -\bar{\mu}_{.1} + 3\bar{\mu}_{.2} - 3\bar{\mu}_{.3} + \bar{\mu}_{.4}$
3. These contrasts are not unique. There are many different choices for the contrasts.
4. The 6 contrasts for the A*B interaction are obtained by multiplying the corresponding coefficients from the Main of A by Main of B.
5. The resulting 11 contrasts are mutually orthogonal.
6. The 6 interaction contrasts are contrasts in the 4 levels of Factor B then contrasted over the 3 levels of Factor A. For example,

$$C6 = -1*(-3\mu_{11}-1\mu_{12}+1\mu_{13}+3\mu_{14})+0*(-3\mu_{21}-1\mu_{22}+1\mu_{23}+3\mu_{24})+1*(-3\mu_{31}-1\mu_{32}+1\mu_{33}+3\mu_{34})$$

$$C9 = 1*(-3\mu_{11}-1\mu_{12}+1\mu_{13}+3\mu_{14})-2*(-3\mu_{21}-1\mu_{22}+1\mu_{23}+3\mu_{24})+1*(-3\mu_{31}-1\mu_{32}+1\mu_{33}+3\mu_{34})$$

7. If the 11 treatments have the same number of replications, then we will obtain the following decomposition of the sum of squares for Treatment:

- $SS_{Treatment} = SS_A + SS_B + SS_{A*B}$ with the further decompostion
- $SS_A = SS_{C1} + SS_{C2}$;
- $SS_B = SS_{C3} + SS_{C4} + SS_{C5}$;
- $SS_{A*B} = SS_{C6} + SS_{C7} + SS_{C8} + SS_{C9} + SS_{C10} + SS_{C11}$

Recall, the sum of squares for a contrast is given by $SS_C = \frac{\hat{C}^2}{\sum_{i=1}^{t} k_i^2/r}$

Table XI Orthogonal polynomials From Kuehl, "Design of Experiments"

	$t=3$		$t=4$			$t=5$				$t=6$					$t=7$					
X_j	P_1	P_2	P_1	P_2	P_3	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4	P_5	P_1	P_2	P_3	P_4	P_5	P_6
1	−1	1	−3	1	−1	−2	2	−1	1	−5	5	−5	1	−1	−3	5	−1	3	−1	1
2	0	−2	−1	−1	3	−1	−1	2	−4	−3	−1	7	−3	5	−2	0	1	−7	4	−6
3	1	1	1	−1	−3	0	−2	0	6	−1	−4	4	2	−10	−1	−3	1	1	−5	15
4			3	1	1	1	−1	−2	−4	1	−4	−4	2	10	0	−4	0	6	0	−20
5						2	2	1	1	3	−1	−7	−3	−5	1	−3	−1	1	5	15
6										5	5	5	1	1	2	0	−1	−7	−4	−6
7															3	5	1	3	1	1
$\sum_{j=1}^{t} \{P_i(X_j)\}^2$	2	6	20	4	20	10	14	10	70	70	84	180	28	252	28	84	6	154	84	924
λ	1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$	2	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{21}{10}$	1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{7}{20}$	$\frac{77}{60}$

	$t=8$						$t=9$						$t=10$					
X_j	P_1	P_2	P_3	P_4	P_5	P_6	P_1	P_2	P_3	P_4	P_5	P_6	P_1	P_2	P_3	P_4	P_5	P_6
1	−7	7	−7	7	−7	1	−4	28	−14	14	−4	4	−9	6	−42	18	−6	3
2	−5	1	5	−13	23	−5	−3	7	7	−21	11	−17	−7	2	14	−22	14	−11
3	−3	−3	7	−3	−17	9	−2	−8	13	−11	−4	22	−5	−1	35	−17	−1	10
4	−1	−5	3	9	−15	−5	−1	−17	9	9	−9	1	−3	−3	31	3	−11	6
5	1	−5	−3	9	15	−5	0	−20	0	18	0	−20	−1	−4	12	18	−6	−8
6	3	−3	−7	−3	17	9	1	−17	−9	9	9	1	1	−4	−12	18	6	−8
7	5	1	−5	−13	−23	−5	2	−8	−13	−11	4	22	3	−3	−31	3	11	6
8	7	7	7	7	7	1	3	7	−7	−21	−11	−17	5	−1	−35	−17	1	10
9							4	28	14	14	4	4	7	2	−14	−22	−14	−11
10													9	6	42	18	6	3
$\sum_{j=1}^{t} \{P_i(X_j)\}^2$	168	168	264	616	2184	264	60	2772	990	2002	468	1980	330	132	8580	2860	780	660
λ	2	1	$\frac{2}{3}$	$\frac{7}{12}$	$\frac{7}{10}$	$\frac{11}{60}$	1	3	$\frac{5}{6}$	$\frac{7}{12}$	$\frac{3}{20}$	$\frac{11}{60}$	2	$\frac{1}{2}$	$\frac{5}{3}$	$\frac{5}{12}$	$\frac{1}{10}$	$\frac{11}{240}$

Adapted from *Biometrika Tables for Statisticians*, Vol. 1, 1966, edited by E. S. Pearson and H. O. Hartley, with permission of the Biometrica Trustees.

623

SAS code needed to analyze the Ca-pH Tree Growth experiment:

```
ods html; ods graphics on;
* factorialCapH.sas;

option ls=80 ps=50 nocenter nodate;
OPTIONS FORMCHAR="|----|+|---+=|-/\<>*";
title 'Effect of Ca and pH on Tree Growth';
data RAW;
input TRT $ pH $ Ca $ y @@;
cards;
T01 4 100 5.2  T05 4 200 7.6  T09  4 300 6.4
T01 4 100 5.9  T05 4 200 7.0  T09  4 300 6.7
T01 4 100 6.3  T05 4 200 7.6  T09  4 300 6.1
T02 5 100 7.1  T06 5 200 7.5  T10 5 300 7.3
T02 5 100 7.3  T06 5 200 7.3  T10 5 300 7.5
T02 5 100 7.5  T06 5 200 7.1  T10 5 300 7.4
T03 6 100 7.6  T07 6 200 7.6  T11 6 300 7.2
T03 6 100 7.2  T07 6 200 7.4  T11 6 300 7.3
T03 6 100 7.4  T07 6 200 7.8  T11 6 300 7.1
T04 7 100 7.2  T08 7 200 7.4  T12 7 300 6.8
T04 7 100 7.5  T08 7 200 7.0  T12 7 300 6.6
T04 7 100 7.2  T08 7 200 6.9  T12 7 300 6.4
run;
TITLE "ANALYSIS OF FACTORIAL TREATMENT STRUCTURE";
proc glm ;
class Ca pH;
model Y = pH Ca pH*Ca/SS3;
contrast 'Linear Trend in Ca'  Ca -1  0 1;
contrast 'Quad Trend in Ca'    Ca  1 -2 1;
contrast 'Linear Trend in pH'  pH -3 -1  1 3;
contrast 'Quad Trend in pH'    pH  1 -1 -1 1;
contrast 'Cubic Trend in pH'   pH -1  3 -3 1;
contrast  'INTER1'             Ca*pH   3  1 -1 -3  0  0  0  0 -3 -1  1  3;
contrast  'INTER2'             Ca*pH  -1  1  1 -1  0  0  0  0  1 -1 -1  1;
contrast  'INTER3'             Ca*pH   1 -3  3 -1  0  0  0  0 -1  3 -3  1;
contrast  'INTER4'             Ca*pH  -3 -1  1  3  6  2 -2 -6 -3 -1  1  3;
contrast  'INTER5'             Ca*pH   1 -1 -1  1 -2  2  2 -2  1 -1 -1  1;
contrast  'INTER6'             Ca*pH  -1  3 -3  1  2 -6  6 -2 -1  3 -3  1;
contrast  'MAIN OF Ca'         Ca  -1  0 1,
                               Ca   1 -2 1;
contrast  'Main of pH'         pH -3 -1  1 3,
                               pH  1 -1 -1 1,
                               pH -1  3 -3 1;
contrast  'INTERACTION'        Ca*pH   3  1 -1 -3  0  0  0  0 -3 -1  1  3,
                               Ca*pH  -3 -1  1  3  6  2 -2 -6 -3 -1  1  3,
                               Ca*pH  -1  1  1 -1  0  0  0  0  1 -1 -1  1,
                               Ca*pH   1 -1 -1  1 -2  2  2 -2  1 -1 -1  1,
                               Ca*pH   1 -3  3 -1  0  0  0  0 -1  3 -3  1,
                               Ca*pH  -1  3 -3  1  2 -6  6 -2 -1  3 -3  1;

lsmeans Ca pH Ca*pH/stderr pdiff adjust=tukey;
RUN;
```

```
TITLE "ANALYSIS OF Cell Means Model";
PROC GLM DATA=RAW;
CLASS TRT;
MODEL Y=TRT/SS3;
contrast 'Linear Trend in Ca'  TRT  -1 -1 -1 -1  0  0  0  0  1  1  1  1;
contrast 'Quad Trend in Ca'    TRT   1  1  1  1 -2 -2 -2 -2  1  1  1  1;
contrast 'Linear Trend in pH'  TRT  -3 -1  1  3 -3 -1  1  3 -3 -1  1  3;
contrast 'Quad Trend in pH'    TRT   1 -1 -1  1  1 -1 -1  1  1 -1 -1  1;
contrast 'Cubic Trend in pH'   TRT  -1  3 -3  1 -1  3 -3  1 -1  3 -3  1;
contrast  'INTER1'             TRT   3  1 -1 -3  0  0  0  0 -3 -1  1  3;
contrast  'INTER2'             TRT  -1  1  1 -1  0  0  0  0  1 -1 -1  1;
contrast  'INTER3'             TRT   1 -3  3 -1  0  0  0  0 -1  3 -3  1;
contrast  'INTER4'             TRT  -3 -1  1  3  6  2 -2 -6 -3 -1  1  3;
contrast  'INTER5'             TRT   1 -1 -1  1 -2  2  2 -2  1 -1 -1  1;
contrast  'INTER6'             TRT  -1  3 -3  1  2 -6  6 -2 -1  3 -3  1;
output out=ASSUMP R=resid P=means;
means TRT/hovtest=bf;
lsmeans TRT/stderr pdiff adjust = Tukey;
PROC PLOT;
PLOT RESID*TRT/VREF=0;
PROC UNIVARIATE DEF=5 PLOT NORMAL;
VAR RESID;
RUN;
proc glimmix data=raw;
class trt;
model y=trt;
lsmeans TRT /  plot = meanplot cl;
run;
proc glimmix data=raw;
class Ca pH;
model Y = pH Ca pH*Ca;
lsmeans Ca*pH /  plot = meanplot cl;
run;
ods graphics off; ods html close;
```

```
ANALYSIS OF EFFECTS MODEL

The GLM Procedure

      Class Level Information

Class         Levels    Values
Ca                 3    100 200 300
pH                 4    4 5 6 7

Number of Observations Read          36
Number of Observations Used          36




Dependent Variable: y

                                         Sum of
Source                      DF          Squares    Mean Square   F Value   Pr > F
Model                       11       9.20000000     0.83636364     11.95   <.0001
Error                       24       1.68000000     0.07000000
Corrected Total             35      10.88000000

Source                      DF      Type III SS    Mean Square   F Value   Pr > F
pH                           3       4.24000000     1.41333333     20.19   <.0001
Ca                           2       1.46000000     0.73000000     10.43   0.0006
Ca*pH                        6       3.50000000     0.58333333      8.33   <.0001

Contrast                    DF      Contrast SS    Mean Square   F Value   Pr > F
Linear Trend in Ca           1       0.01500000     0.01500000      0.21   0.6476
Quad Trend in Ca             1       1.44500000     1.44500000     20.64   0.0001
Linear Trend in pH           1       0.96800000     0.96800000     13.83   0.0011
Quad Trend in pH             1       3.24000000     3.24000000     46.29   <.0001
Cubic Trend in pH            1       0.03200000     0.03200000      0.46   0.5054
INTER1                       1       1.32300000     1.32300000     18.90   0.0002
INTER2                       1       0.00000000     0.00000000      0.00   1.0000
INTER3                       1       0.01200000     0.01200000      0.17   0.6825
INTER4                       1       0.96100000     0.96100000     13.73   0.0011
INTER5                       1       0.72000000     0.72000000     10.29   0.0038
INTER6                       1       0.48400000     0.48400000      6.91   0.0147
MAIN OF Ca                   2       1.46000000     0.73000000     10.43   0.0006
Main of pH                   3       4.24000000     1.41333333     20.19   <.0001
INTERACTION                  6       3.50000000     0.58333333      8.33   <.0001
```

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Ca	y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
100	6.95000000	0.07637626	<.0001	1
200	7.35000000	0.07637626	<.0001	2
300	6.90000000	0.07637626	<.0001	3

Least Squares Means for effect Ca
Adjustment for Multiple Comparisons: Tukey
Pr > \|t\| for H0: LSMean(i)=LSMean(j)

i/j	1	2	3
1		0.0031	0.8891
2	0.0031		0.0010
3	0.8891	0.0010	

pH	y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
4	6.53333333	0.08819171	<.0001	1
5	7.33333333	0.08819171	<.0001	2
6	7.40000000	0.08819171	<.0001	3
7	7.00000000	0.08819171	<.0001	4

Least Squares Means for effect pH
Adjustment for Multiple Comparisons: Tukey
Pr > \|t\| for H0: LSMean(i)=LSMean(j)

i/j	1	2	3	4
1		<.0001	<.0001	0.0052
2	<.0001		0.9498	0.0600
3	<.0001	0.9498		0.0185
4	0.0052	0.0600	0.0185	

Ca	pH	y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
100	4	5.80000000	0.15275252	<.0001	1
100	5	7.30000000	0.15275252	<.0001	2
100	6	7.40000000	0.15275252	<.0001	3
100	7	7.30000000	0.15275252	<.0001	4
200	4	7.40000000	0.15275252	<.0001	5
200	5	7.30000000	0.15275252	<.0001	6
200	6	7.60000000	0.15275252	<.0001	7
200	7	7.10000000	0.15275252	<.0001	8
300	4	6.40000000	0.15275252	<.0001	9
300	5	7.40000000	0.15275252	<.0001	10
300	6	7.20000000	0.15275252	<.0001	11
300	7	6.60000000	0.15275252	<.0001	12

Least Squares Means for effect Ca*pH - Tukey Adjusted for Multiple Testing
$Pr > |t|$ for H0: LSMean(i)=LSMean(j)

i/j	1	2	3	4	5	6	7	8	9	10	11	12
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.2499	<.0001	<.0001	0.0404
2	<.0001		1.0000	1.0000	1.0000	1.0000	0.9549	0.9980	0.0142	1.0000	1.0000	0.1067
3	<.0001	1.0000		1.0000	1.0000	1.0000	0.9980	0.9549	0.0048	1.0000	0.9980	0.0404
4	<.0001	1.0000	1.0000		1.0000	1.0000	0.9549	0.9980	0.0142	1.0000	1.0000	0.1067
5	<.0001	1.0000	1.0000	1.0000		1.0000	0.9980	0.9549	0.0048	1.0000	0.9980	0.0404
6	<.0001	1.0000	1.0000	1.0000	1.0000		0.9549	0.9980	0.0142	1.0000	1.0000	0.1067
7	<.0001	0.9549	0.9980	0.9549	0.9980	0.9549		0.4932	0.0005	0.9980	0.7755	0.0048
8	0.0002	0.9980	0.9549	0.9980	0.9549	0.9980	0.4932		0.1067	0.9549	1.0000	0.4932
9	0.2499	0.0142	0.0048	0.0142	0.0048	0.0142	0.0005		0.1067	0.0048	0.0404	0.9980
10	<.0001	1.0000	1.0000	1.0000	1.0000	1.0000	0.9980	0.9549		0.0048	0.9980	0.0404
11	<.0001	1.0000	0.9980	1.0000	0.9980	1.0000	0.7755	1.0000	0.0404		0.9980	0.2499
12	0.0404	0.1067	0.0404	0.1067	0.0404	0.1067	0.0048	0.4932	0.9980	0.0404	0.2499	

Least Squares Means for effect Ca*pH - Unadjusted for Multiple Comparison
$Pr > |t|$ for H0: LSMean(i)=LSMean(j)

Dependent Variable: y

i/j	1	2	3	4	5	6	7	8	9	10	11	12
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0105	<.0001	<.0001	0.0011
2	<.0001		0.6476	1.0000	0.6476	1.0000	0.1777	0.3638	0.0003	0.6476	0.6476	0.0035
3	<.0001	0.6476		0.6476	1.0000	0.6476	0.3638	0.1777	0.0001	1.0000	0.3638	0.0011
4	<.0001	1.0000	0.6476		0.6476	1.0000	0.1777	0.3638	0.0003	0.6476	0.6476	0.0035
5	<.0001	0.6476	1.0000	0.6476		0.6476	0.3638	0.1777	0.0001	1.0000	0.3638	0.0011
6	<.0001	1.0000	0.6476	1.0000	0.6476		0.1777	0.3638	0.0003	0.6476	0.6476	0.0035
7	<.0001	0.1777	0.3638	0.1777	0.3638	0.1777		0.0295	<.0001	0.3638	0.0764	0.0001
8	<.0001	0.3638	0.1777	0.3638	0.1777	0.3638	0.0295		0.0035	0.1777	0.6476	0.0295
9	0.0105	0.0003	0.0001	0.0003	0.0001	0.0003	<.0001	0.0035		0.0001	0.0011	0.3638
10	<.0001	0.6476	1.0000	0.6476	1.0000	0.6476	0.3638	0.1777	0.0001		0.3638	0.0011
11	<.0001	0.6476	0.3638	0.6476	0.3638	0.6476	0.0764	0.6476	0.0011	0.3638		0.0105
12	0.0011	0.0035	0.0011	0.0035	0.0011	0.0035	0.0001	0.0295	0.3638	0.0011	0.0105	

ANALYSIS OF CELL MEANS MODEL

The GLM Procedure

Class Level Information

Class	Levels	Values
TRT	12	T01 T02 T03 T04 T05 T06 T07 T08 T09 T10 T11 T12

Number of Observations Read	36
Number of Observations Used	36

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	9.20000000	0.83636364	11.95	<.0001
Error	24	1.68000000	0.07000000		
Corrected Total	35	10.88000000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	11	9.20000000	0.83636364	11.95	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Linear Trend in Ca	1	0.01500000	0.01500000	0.21	0.6476
Quad Trend in Ca	1	1.44500000	1.44500000	20.64	0.0001
Linear Trend in pH	1	0.96800000	0.96800000	13.83	0.0011
Quad Trend in pH	1	3.24000000	3.24000000	46.29	<.0001
Cubic Trend in pH	1	0.03200000	0.03200000	0.46	0.5054
INTER1	1	1.32300000	1.32300000	18.90	0.0002
INTER2	1	0.00000000	0.00000000	0.00	1.0000
INTER3	1	0.01200000	0.01200000	0.17	0.6825
INTER4	1	0.96100000	0.96100000	13.73	0.0011
INTER5	1	0.72000000	0.72000000	10.29	0.0038
INTER6	1	0.48400000	0.48400000	6.91	0.0147

Brown and Forsythe's Test for Homogeneity of y Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
TRT	11	0.2097	0.0191	0.54	0.8530
Error	24	0.8400	0.0350		

Tests for Normality

```
Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.963024    Pr < W      0.2660
Kolmogorov-Smirnov    D       0.1249    Pr > D     >0.1500
Cramer-von Mises      W-Sq  0.096918    Pr > W-Sq   0.1221
Anderson-Darling      A-Sq  0.597162    Pr > A-Sq   0.1145
```

```
             Plot of resid*TRT.  Legend: A = 1 obs, B = 2 obs, etc.

      resid |
            |
        0.5 +  A
            |
            |
        0.4 +
            |
            |
        0.3 +                                      A    A
            |
            |
        0.2 +       A    A    A    B    A    A                        A
            |
            |
        0.1 +  A                                             A    A
            |
            |
        0.0 +-------A----A-------------A----A---------A----A----A----A--
            |
            |
       -0.1 +                 B                   A         A    A
            |
            |
       -0.2 +       A    A              A    A    A                   A
            |
            |
       -0.3 +                                          A
            |
            |
       -0.4 +                      A
            |
            |
       -0.5 +
            |
            |
       -0.6 +  A
            |
            ---+----+----+----+----+----+----+----+----+----+----+----+--
              T01  T02  T03  T04  T05  T06  T07  T08  T09  T10  T11  T12

                                          TRT
```

```
   Stem Leaf                     #  Boxplot
      5 0                        1     |
      4                                |
      3 00                       2     |
      2 00000000                 8  +-----+
      1 000                      3  |     |
      0 00000000                 8  *--+--*
     -0                             |     |
     -1 00000                    5  +-----+
     -2 000000                   6     |
     -3 0                        1     |
     -4 0                        1     |
     -5                                |
     -6 0                        1     |
        ----+----+----+----+
    Multiply Stem.Leaf by 10**-1
```

```
Variable:  resid

                      Normal Probability Plot
    0.55+                                            * +++
        |                                          ++++
        |                                     ++*++
        |                            *********
        |                          *******
        |                     ******+
   -0.05+                 *****+
        |          * ******+
        |        * ++++
        |     ++*++
        |   ++++
        |+++ *
   -0.65+
         +----+----+----+----+----+----+----+----+----+----+
             -2        -1         0        +1        +2
```

TRT Least Squares Means

TRT	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
T01	5.8000	0.1528	24	37.97	<.0001	0.05	5.4847	6.1153
T02	7.3000	0.1528	24	47.79	<.0001	0.05	6.9847	7.6153
T03	7.4000	0.1528	24	48.44	<.0001	0.05	7.0847	7.7153
T04	7.3000	0.1528	24	47.79	<.0001	0.05	6.9847	7.6153
T05	7.4000	0.1528	24	48.44	<.0001	0.05	7.0847	7.7153
T06	7.3000	0.1528	24	47.79	<.0001	0.05	6.9847	7.6153
T07	7.6000	0.1528	24	49.75	<.0001	0.05	7.2847	7.9153
T08	7.1000	0.1528	24	46.48	<.0001	0.05	6.7847	7.4153
T09	6.4000	0.1528	24	41.90	<.0001	0.05	6.0847	6.7153
T10	7.4000	0.1528	24	48.44	<.0001	0.05	7.0847	7.7153
T11	7.2000	0.1528	24	47.14	<.0001	0.05	6.8847	7.5153
T12	6.6000	0.1528	24	43.21	<.0001	0.05	6.2847	6.9153

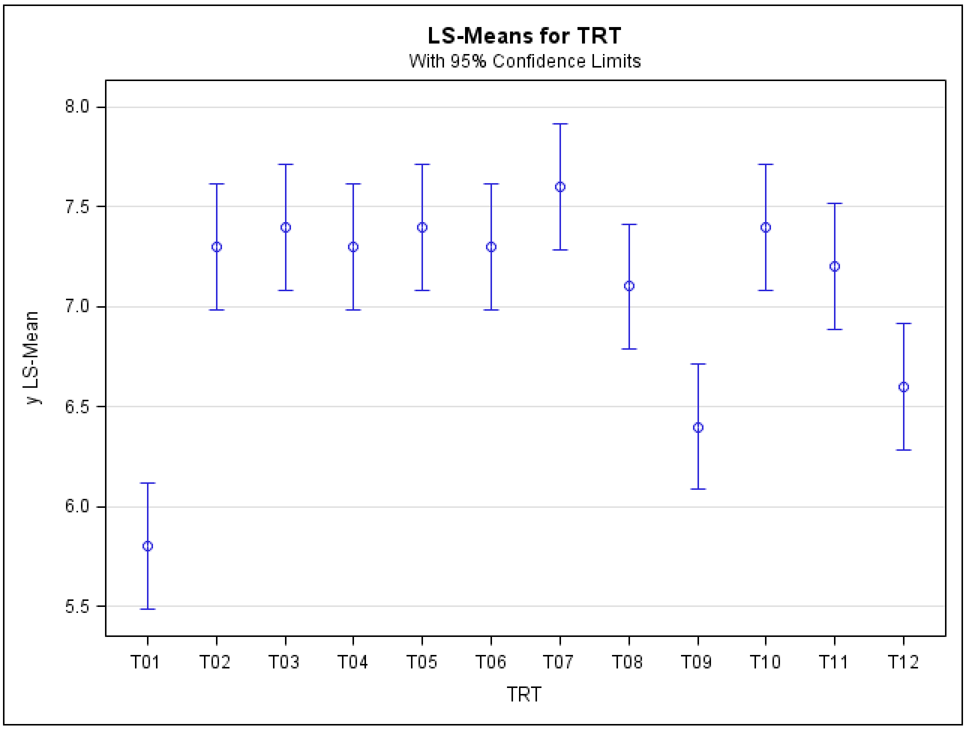
Least Squares Means for effect TRT
Adjustment for Multiple Comparisons: Tukey
Pr > |t| for H0: LSMean(i)=LSMean(j)

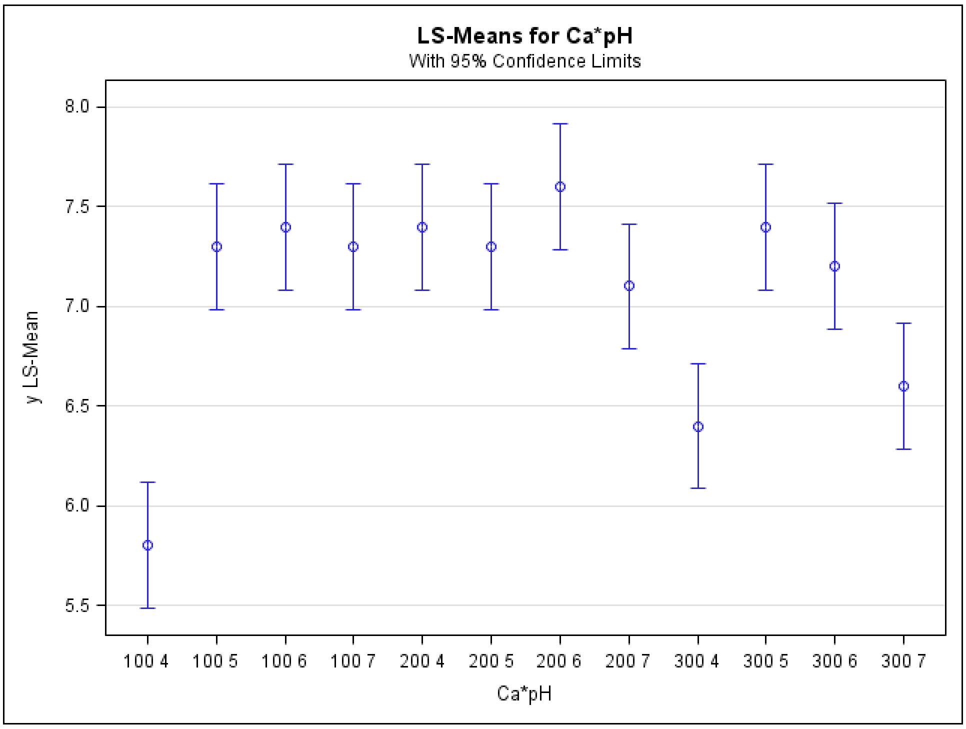
Dependent Variable: y

i/j	1	2	3	4	5	6
1		<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		1.0000	1.0000	1.0000	1.0000
3	<.0001	1.0000		1.0000	1.0000	1.0000
4	<.0001	1.0000	1.0000		1.0000	1.0000
5	<.0001	1.0000	1.0000	1.0000		1.0000
6	<.0001	1.0000	1.0000	1.0000	1.0000	
7	<.0001	0.9549	0.9980	0.9549	0.9980	0.9549
8	0.0002	0.9980	0.9549	0.9980	0.9549	0.9980
9	0.2499	0.0142	0.0048	0.0142	0.0048	0.0142
10	<.0001	1.0000	1.0000	1.0000	1.0000	1.0000
11	<.0001	1.0000	0.9980	1.0000	0.9980	1.0000
12	0.0404	0.1067	0.0404	0.1067	0.0404	0.1067

i/j	7	8	9	10	11	12
1	<.0001	0.0002	0.2499	<.0001	<.0001	0.0404
2	0.9549	0.9980	0.0142	1.0000	1.0000	0.1067
3	0.9980	0.9549	0.0048	1.0000	0.9980	0.0404
4	0.9549	0.9980	0.0142	1.0000	1.0000	0.1067
5	0.9980	0.9549	0.0048	1.0000	0.9980	0.0404
6	0.9549	0.9980	0.0142	1.0000	1.0000	0.1067
7		0.4932	0.0005	0.9980	0.7755	0.0048
8	0.4932		0.1067	0.9549	1.0000	0.4932
9	0.0005	0.1067		0.0048	0.0404	0.9980
10	0.9980	0.9549	0.0048		0.9980	0.0404
11	0.7755	1.0000	0.0404	0.9980		0.2499
12	0.0048	0.4932	0.9980	0.0404	0.2499	

The following plot is produced using Proc glimmix in the SAS code:





Sample Size Calculations - Equal Number of Reps

Factorial Model: $y_{ijk} = \mu_{ij} + e_{ijk}$ for $i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, r$

1. Determine r for specified regions in the alternative hypothesis for the interaction: $F_1 * F_2$

   $H_o : (\mu_{ij} - \mu_{kj}) = (\mu_{ih} - \mu_{kh})$ for all (i, j, k, h) versus

   $H_a : (\mu_{ij} - \mu_{kj}) \neq (\mu_{ih} - \mu_{kh})$ for at least one set (i, j, k, h)

   Reject H_o if $F = \frac{MS_{F_1*F_2}}{MSE} \geq F_{\alpha,(a-1)(b-1),ab(r-1)}$

   Power Function: $\gamma(\lambda) = P[F = \frac{MS_{F_1*F_2}}{MSE} \geq F_{\alpha,(a-1)(b-1),df_E}] = 1 - G(F_{\alpha,(a-1)(b-1),df_E})$

   where $G(\cdot)$ is the cdf of a noncentral F-distribution with $df = (a-1)(b-1), ab(r-1)$ and non-centrality parameter

   $$\lambda = \frac{r}{\sigma_e^2} \sum_{i=1}^{a} \sum_{j=1}^{b} (\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..})^2$$

   For specified values of λ_o, α and γ_o determine the minimum value of r such that $\gamma(\lambda) \geq \gamma_o$ whenever $\lambda \geq \lambda_o$.

   Use the R function for the non-central F-distribution:

   $pf(q(1-\alpha, (a-1)(b-1), ab(r-1)), (a-1)(b-1), ab(r-1), \lambda)$

   or Table IX in Kuehl with $\Phi = \sqrt{\lambda/t}$ and $t = ab$ to determine the value of r.

   The above approach is not very realistic because specifying appropriate values for λ_o are not very intuitive to the researcher or engineer. The following approach follows the methodology used in single factor experiments.

2. Determine r by specifying differences in simple effects of either factor.

   a. Let $D_j = \mu_{ij} - \mu_{kj}$ be the difference in the simple effects of F_1 at the jth level of factor F_2 that the researcher deems important to detect.
   b. From our previous results, we know that the minimum value for λ_j is

$$\lambda_j = \frac{r \sum_{i=1}^{a} (\mu_{ij} - \bar{\mu}_{.j})^2}{\sigma_e^2} = \frac{r D_j^2}{2\sigma_e^2}$$

   c. Determine minimum value of r such that $\gamma(\lambda_j) \geq \gamma_o$ whenever $\lambda_j \geq \lambda_o = \frac{rD_j^2}{2\sigma_e^2}$. The result is obtained by using non-central F-distribution:

      $pf(q(1-\alpha, a-1, t(r-1)), a-1, t(r-1), \lambda)$ with $t = ab$

      or Table IX in Kuehl as was done in Handout #3. A similar calculation holds for simple effects for F_2.

3. Determine $n = ar$ by specifying differences in the - Main Effect of Individual Factors.

   a. Let $D = \mu_{.j} - \mu_{.h}$ be the difference in any two levels or factor F_2 that the researcher deems important to detect.
   b. From our previous results, we know that the minimum value for λ is

$$\lambda = \frac{n \sum_{j=1}^{b} (\mu_{.j} - \bar{\mu}_{..})^2}{\sigma_e^2} = \frac{nD^2}{2\sigma_e^2}$$

   c. Determine minimum value of n such that $\gamma(\lambda) \geq \gamma_o$ whenever $\lambda \geq \lambda_o = \frac{nD^2}{2\sigma_e^2}$. The result is obtained by using non-central F-distribution:

      $pf(q(1-\alpha, b-1, ab(r-1)), b-1, ab(r-1), \lambda)$

      or Table IX in Kuehl as was done in Handout #3.
   d. The number of reps is then $r = n/a$

Examples of Determining Number of Reps in Factorial Experiment

Example 1 Factorial Experiment - Treatment Effect

Suppose we have two factors F1 with 3 levels and F2 with 4 levels. We want to determine the number of reps, r, such that an $\alpha = .05$ F-test will have power of $\gamma \geq .8$ of detecting a difference in the treatment means whenever, $|\mu_{ij} - \mu_{hk}| \geq D$, where $D = 2.5$. The researcher provides an estimate of σ, $\hat{\sigma} = 1.4$

Solution In the SAS program, repsize_Factorial.SAS, we input

$$a = 3; \quad b = 4; \quad t = (4)(3) = 12; \quad \alpha = .05; \quad D = 2.5; \quad S = 1.4$$

and then select the value of r such that the power exceeds $\gamma = .80$

The output from the program is given here:

Obs	t	r	c	u1	u2	L	phi	power
8	12	9	1.88980	11	96	14.3495	1.09352	0.66355
9	12	10	1.87839	11	108	15.9439	1.15267	0.72690
10	12	11	1.86929	11	120	17.5383	1.20893	0.78146
11	12	12	1.86187	11	132	19.1327	1.26269	0.82744
12	12	13	1.85570	11	144	20.7270	1.31425	0.86544
13	12	14	1.85048	11	156	22.3214	1.36386	0.89630
14	12	15	1.84602	11	168	23.9158	1.41173	0.92096

From the above table, the required number of reps $r = 12$ reps per treatment.

Example 2 Factorial Experiment - Main Effect of Individual Factors

Suppose we have two factors F1 with 3 levels and F2 with 4 levels. We want to determine the number of reps, r, such that an $\alpha = .05$ F-test will have power of $\gamma \geq .8$ of detecting a difference in the treatment means of two levels of factor F2 whenever, $|\mu_{.j} - \mu_{.k}| \geq D$, where $D = 2.5$. The researcher provides an estimate of σ, $\hat{\sigma} = 1.4$

Solution In the second part of the SAS program, repsize_Factorial.SAS, we input

$$a = 3; \quad b = 4; \quad \alpha = .05; \quad D = 2.5; \quad S = 1.4$$

and then select the value of r such that the power exceeds $\gamma = .80$

The output from the program is given here:

Obs	b	r	c	u1	u2	L	phi	power
4	4	5	2.79806	3	48	7.9719	1.41173	0.61300
5	4	6	2.75808	3	60	9.5663	1.54647	0.70927
6	4	7	2.73181	3	72	11.1607	1.67038	0.78630
7	4	8	2.71323	3	84	12.7551	1.78571	0.84593
8	4	9	2.69939	3	96	14.3495	1.89404	0.89083
9	4	10	2.68869	3	108	15.9439	1.99649	0.92383

From the above table, the required value for n is $n = 8$ which yields $r = 8/3$ or 3 reps per treatment.

Two Factor Experiment with Unequal Number of Reps per Treatment

When the n_{ij}s are not equal, the estimation of the treatment means and tests of hypotheses becomes somewhat more complicated.

LSE of $\widehat{\mu}_{i.}$, $\widehat{\mu}_{.j}$, $\widehat{\mu}_{ij}$

1. $\hat{\mu}_{ij} = \bar{y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$
2. $\widehat{\mu}_{i.} = \frac{1}{b} \sum_{j=1}^{b} \hat{\mu}_{ij} = \frac{1}{b} \sum_{j=1}^{b} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \neq \bar{y}_{i..}$
3. $\widehat{\mu}_{.j} = \frac{1}{a} \sum_{i=1}^{a} \hat{\mu}_{ij} = \frac{1}{a} \sum_{i=1}^{a} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \neq \bar{y}_{.j.}$
4. $\widehat{\mu}_{..} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \hat{\mu}_{ij} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \neq \bar{y}_{...}$

Standard Errors of LSE

Let $\hat{\sigma}_e^2 = \sqrt{MSE}$

1. $\widehat{SE}(\widehat{\mu}_{ij}) = \frac{\hat{\sigma}_e}{\sqrt{n_{ij}}}$

   $Var(\bar{y}_{ij.}) = \frac{1}{n_{ij}^2} \sum_{k=1}^{n_{ij}} Var(y_{ijk}) = \frac{1}{n_{ij}^2} n_{ij} \sigma_e^2 = \frac{\sigma_e^2}{n_{ij}}$

2. $\widehat{SE}(\widehat{\mu}_{i.}) = \hat{\sigma}_e \sqrt{\frac{1}{b^2} \sum_{j=1}^{b} \frac{1}{n_{ij}}}$

   $Var(\widehat{\mu}_{i.}) = \frac{1}{b^2} \sum_{j=1}^{b} \frac{1}{n_{ij}^2} \sum_{k=1}^{n_{ij}} Var(y_{ijk}) = \frac{1}{b^2} \sum_{j=1}^{b} \frac{1}{n_{ij}^2} \sum_{k=1}^{n_{ij}} \sigma_e^2 = \frac{\sigma_e^2}{b^2} \sum_{j=1}^{b} \frac{1}{n_{ij}^2} n_{ij}$

3. $\widehat{SE}(\widehat{\mu}_{.j}) = \hat{\sigma}_e \sqrt{\frac{1}{a^2} \sum_{i=1}^{a} \frac{1}{n_{ij}}}$

   $Var(\widehat{\mu}_{.j}) = \frac{1}{a^2} \sum_{i=1}^{a} \frac{1}{n_{ij}^2} \sum_{k=1}^{n_{ij}} Var(y_{ijk}) = \frac{1}{a^2} \sum_{i=1}^{a} \frac{1}{n_{ij}^2} \sum_{k=1}^{n_{ij}} \sigma_e^2 = \frac{\sigma_e^2}{a^2} \sum_{i=1}^{a} \frac{1}{n_{ij}^2} n_{ij}$

4. $\widehat{SE}(\widehat{\mu}_{..}) = \hat{\sigma}_e \sqrt{\frac{1}{(ab)^2} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}}}$

   $Var(\widehat{\mu}_{..}) = \frac{1}{(ab)^2} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}^2} \sum_{k=1}^{n_{ij}} Var(y_{ijk}) = \frac{1}{(ab)^2} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}^2} \sum_{k=1}^{n_{ij}} \sigma_e^2$

Confidence Intervals

Using the above estimators of the standard errors of the LSE of the various types of population means, the following $100(1-\alpha)\%$ C.I.s are obtained:

1. $100(1-\alpha)\%$ C.I. on μ_{ij} :

$$\widehat{\mu}_{ij} \pm \left(t_{\alpha/2,df_E}\right) \widehat{SE}(\widehat{\mu}_{ij}) \quad \Rightarrow \quad \widehat{\mu}_{ij} \pm \left(t_{\alpha/2,df_E}\right) \frac{\hat{\sigma}_e}{\sqrt{n_{ij}}}$$

2. $100(1-\alpha)\%$ C.I. on $\bar{\mu}_{i.}$:

$$\widehat{\bar{\mu}}_{i.} \pm \left(t_{\alpha/2,df_E}\right) \widehat{SE}(\widehat{\bar{\mu}}_{i.}) \quad \Rightarrow \quad \widehat{\bar{\mu}}_{i.} \pm \left(t_{\alpha/2,df_E}\right) \hat{\sigma}_e \sqrt{\tfrac{1}{b^2} \textstyle\sum_{j=1}^{b} \tfrac{1}{n_{ij}}}$$

3. $100(1-\alpha)\%$ C.I. on $\bar{\mu}_{.j}$:

$$\widehat{\bar{\mu}}_{.j} \pm \left(t_{\alpha/2,df_E}\right) \widehat{SE}(\widehat{\bar{\mu}}_{i.}) \quad \Rightarrow \quad \widehat{\bar{\mu}}_{.j} \pm \left(t_{\alpha/2,df_E}\right) \hat{\sigma}_e \sqrt{\tfrac{1}{a^2} \textstyle\sum_{i=1}^{a} \tfrac{1}{n_{ij}}}$$

4. $100(1-\alpha)\%$ C.I. on $\mu_{ij} - \mu_{kh}$:

$$\widehat{\mu}_{ij} - \widehat{\mu}_{kh} \pm \left(t_{\alpha/2,df_E}\right) \hat{\sigma}_e \sqrt{\tfrac{1}{n_{ij}} + \tfrac{1}{n_{kh}}}$$

5. $100(1-\alpha)\%$ C.I. on $\bar{\mu}_{i.} - \bar{\mu}_{k.}$:

$$\widehat{\bar{\mu}}_{i.} - \widehat{\bar{\mu}}_{k.} \pm \left(t_{\alpha/2,df_E}\right) \hat{\sigma}_e \sqrt{\tfrac{1}{b^2} \textstyle\sum_{j=1}^{b} \left(\tfrac{1}{n_{ij}} + \tfrac{1}{n_{kj}}\right)}$$

6. $100(1-\alpha)\%$ C.I. on $\bar{\mu}_{.j} - \bar{\mu}_{.h}$:

$$\widehat{\bar{\mu}}_{.j} - \widehat{\bar{\mu}}_{.h} \pm \left(t_{\alpha/2,df_E}\right) \hat{\sigma}_e \sqrt{\tfrac{1}{a^2} \textstyle\sum_{i=1}^{a} \left(\tfrac{1}{n_{ij}} + \tfrac{1}{n_{ih}}\right)}$$

Tests of Hypotheses (See *Linear Models for Unbalanced Data* by S. Searle)

The test of no interaction when the sample sizes are unequal becomes somewhat more complex than in the equal sample size case.

Test of an Interaction between F_1 and F_2:

$$H_o : (\mu_{ij} - \mu_{kj}) = (\mu_{ih} - \mu_{kh}) \quad \text{for all sets} \quad \{(i,k,j,h) : (i \neq k = 1, \dots, a; j \neq h = 1, \dots, b)\}$$

$$H_1 : (\mu_{ij} - \mu_{kj}) \neq (\mu_{ih} - \mu_{kh}) \quad \text{for at least one set of} \quad (i,k,j,h)$$

The test statistic used when $n_{ij} = r$ is no longer valid. The most direct method for obtaining the test statistic is to consider fitting two models:

Model 1: A model with No Constraints on μ_{ij}: $y_{ijk} = \mu_{ij} + e_{ijk}$

Model 2: A model with No Interaction: $y_{ijk} = \mu_{ij} + e_{ijk}$ with the constraints μ_{ij}s satisfy H_o

Fit both models to the data, and obtain

SSE_{NC} the sum of squares error from Model 1, the model with no constraints on μ_{ij}s, with

df = (number of data values) - (number of parameters in model) $= n - ab$.

SSE_{NI} the sum of squares error from Model 2, the model with the constraints that the μ_{ij}s have no interaction between F_1 and F_2, with

df $= n - (b + (a-1)) = n - a - b + 1$.

The test statistic is

$$F_{INT} = \frac{(SSE_{NI} - SSE_{NC})/[(n-a-b+1) - (n-ab)]}{SSE_{NC}/(n-ab)}$$

which under H_o has an F-distribution with $df = (a-1)(b-1), (n-ab)$ after noting that $(n-a-b+1) - (n-ab) = (a-1)(b-1)$.

The sum of squares error under no restrictions is obtained by finding the least squares estimator of μ_{ij} which we have shown is just $\widehat{\mu}_{ij} = \bar{y}_{ij.}$ and computing

$$SSE_{NC} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk} - \widehat{\mu}_{ij})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij.})^2$$

The sum of squares error under H_o, SSE_{NI} is more complex, it requires obtaining the least squares estimator of μ_{ij} under the constraint of

$(\mu_{ij} - \mu_{kj}) = (\mu_{ih} - \mu_{kh}) \quad$ for all sets $\{(i,k,j,h) : (i \neq k = 1, \dots, a; j \neq h = 1, \dots, b)\}$.

Let $\theta_{i,j,k,h} = \mu_{ij} - \mu_{kj} - \mu_{ih} + \mu_{kh}$. H_o and H_1 expressed in terms of $\theta_{i,j,k,h}$ become

$$H_o : \theta_{i,j,k,h} = 0 \;\text{ for all }\; (i,j,k,h) \qquad H_1 : \theta_{i,j,k,h} \neq 0 \;\text{ for at least one set }\; (i,j,k,h)$$

Express Model 2, no interaction, as $\mu_{ij} = \alpha_i + \gamma_j$ for all $i = 1, \ldots, a$ and $j = 1, \ldots, b$. Then, for all (i,j,k,h),

$$\theta_{i,j,k,h} = \mu_{ij} - \mu_{kj} - \mu_{ih} + \mu_{kh} = (\alpha_i + \gamma_j) - (\alpha_k + \gamma_j) - (\alpha_i + \gamma_h) + (\alpha_k + \gamma_h) = 0$$

The model $\mu_{ij} = \alpha_i + \gamma_j$ is not unique, because we could always reparametrize the model by letting

$\alpha_i^{'} = \alpha_i + \delta$ and $\gamma_j^{'} = \gamma_j - \delta$. Then we would have

$$\mu_{ij} = \alpha_i + \gamma_j = (\alpha_i + \delta) + (\gamma_j - \delta) = \alpha_i^{'} + \gamma_j^{'}$$

In particular, we could take $\delta = \gamma_b$, which yields $\gamma_b^{'} = 0$. Thus, only $a + b - 1$ parameters are needed to model μ_{ij} under H_o.

To obtain SSE_{NI}, we need to determine the LSE estimators of α_i and γ_j. That is find the functions of the data $\hat{\alpha}_i$ and $\hat{\gamma}_j$ which minimize

$$SS = \sum_{ijk} (y_{ijk} - \mu_{ij})^2 = \sum_{i=1}^{a} \sum_{j=1}^{b-1} \sum_{k=1}^{n_{ij}} (y_{ijk} - \alpha_i - \gamma_j)^2 + \sum_{i=1}^{a} \sum_{k=1}^{n_{ib}} (y_{ijk} - \alpha_i)^2$$

The solutions to these equations require:

Solving for $\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{b-1}$ in the $b-1$ equations, that is, for $j = 1, 2, \ldots, b-1$:

$$c_{jj}\hat{\gamma}_j + \sum_{h \neq j}^{b-1} c_{jh}\hat{\gamma}_h = r_j$$

where $c_{jj} = n_{.j} - \sum_{i=1}^{a} \frac{n_{ij}^2}{n_{i.}}$, $c_{jh} = -\sum_{i=1}^{a} \frac{n_{ij}n_{ih}}{n_{i.}}$, and $r_j = y_{.j.} - \sum_{i=1}^{a} n_{ij}\bar{y}_{i..}$.

Using the $\hat{\gamma}_j$s obtained, compute for $i = 1, \ldots, a$;

$$\hat{\alpha}_i = \bar{y}_{i..} - \frac{1}{n_{i.}} \sum_{j=1}^{b-1} n_{ij}\hat{\gamma}_i$$

Finally, we obtain

$$SSE_{NI} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} (y_{ijk} - \hat{\alpha}_i - \hat{\gamma}_j)^2$$

In those situations, **where there is not significant evidence of an interaction,** tests for the marginal means can be conducted. A test of differences in the treatment means across the levels of F_1 is given by

Test $H_o : \bar{\mu}_{1.} = \bar{\mu}_{2.} = \cdots = \bar{\mu}_{a.}$ versus H_1 : not all $\bar{\mu}_{i.}$ are equal

Test statistic is given by

$$F = \frac{SS_{F_{1w}}/(a-1)}{MSE}$$

which under H_o has an F-distribution with $df = a-1, n-ab$.

$$SS_{F_{1w}} = \sum_{i=1}^{a} w_i \left(\tilde{y}_{i..} - \frac{\sum_{i=1}^{a} w_i \tilde{y}_{i..}}{\sum_{i=1}^{a} w_i} \right)^2$$

where $\tilde{y}_{i..} = \frac{1}{b}\sum_{j=1}^{b} \bar{y}_{ij.}$ and $\frac{1}{w_i} = \frac{1}{b^2}\sum_{j=1}^{b} \frac{1}{n_{ij}}$

Similarly, a test of differences in the treatment means across the levels of F_2 is given by

Test $H_o : \bar{\mu}_{.1} = \bar{\mu}_{.2} = \cdots = \bar{\mu}_{.b}$ versus H_1 : not all $\bar{\mu}_{.j}$ are equal

Test statistic is given by

$$F = \frac{SS_{F_{2w}}/(b-1)}{MSE}$$

which under H_o has an F-distribution with $df = b-1, n-ab$.

$$SS_{F_{2w}} = \sum_{j=1}^{b} v_j \left(\tilde{y}_{.j.} - \frac{\sum_{j=1}^{b} v_j \tilde{y}_{.j.}}{\sum_{j=1}^{b} v_j} \right)^2$$

where $\tilde{y}_{.j.} = \frac{1}{a}\sum_{i=1}^{a} \bar{y}_{ij.}$ and $\frac{1}{v_j} = \frac{1}{a^2}\sum_{i=1}^{a} \frac{1}{n_{ij}}$ and

$$MSE = \frac{1}{n-ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2$$

Most computer programs calculate the Sum of Squares in AOV using one of two (SAS has four) methods:

Sequential Sum of Squares (Type I Sum of Squares in SAS) which computes full and reduced models sequentially in the order that the terms are given in the model statement. For example, in a two factor experiment with an interaction term, the model statement would be

y = $F_1 \quad F_2 \quad F_1 * F_2$

The Type I Sum of Squares would be

1. SS_{F_1} which is computed by fitting two models,
   (a) the Reduced model without any terms but the intercept yielding $SSE_1 = SS_{TOT}$
   (b) the Full model with the intercept and the term F_1 yielding SSE_2
   (c) $SS_{F_1} = SSE_1 - SSE_2$

2. SS_{F_2} which is computed by fitting two models,
   (a) the Reduced model with the intercept and the term F_1 yielding SSE_2
   (b) the Full model with the intercept, the term F_1, and F_2 yielding SSE_3
   (c) $SS_{F_2} = SSE_2 - SSE_3$

3. $SS_{F_1*F_2}$ which is computed by fitting two models,
   (a) the Reduced model with the intercept, the term F_1, and F_2 yielding SSE_3
   (b) the Full model with with the intercept, the term F_1, F_2, and $F_1 * F_2$ yielding SSE_4
   (c) $SS_{F_1*F_2} = SSE_3 - SSE_4$

The hypotheses that these sum of squares are tested are given below:

Source of Variation	Type I Hypotheses
F_1	$\frac{1}{n_{1.}} \sum_{j=1}^{b} n_{1j}\mu_{1j} = \frac{1}{n_{2.}} \sum_{j=1}^{b} n_{2j}\mu_{2j} = \cdots = \frac{1}{n_{a.}} \sum_{j=1}^{b} n_{aj}\mu_{aj}$
F_2	$\sum_{i=1}^{a} \left(n_{ij} - \frac{n_{ij}^2}{n_{i.}}\right)\mu_{ij} = \sum_{h \neq j}^{b} \sum_{i=1}^{a} \frac{n_{ij}n_{ih}}{n_{i.}}\mu_{ih}$ for $j = 1, 2, \ldots, b$
$F_1 * F_2$	$(\mu_{ij} - \mu_{kj}) - (\mu_{ih} - \mu_{kh}) = 0$ for all (i, j, k, h)

Partial Sum of Squares (Type III Sum of Squares in SAS) which computes full and reduced models with and without each term given in the model statement included in the model. For example, in a two factor experiment with an interaction term, the model statement would be

y = $F_1 \quad F_2 \quad F_1 * F_2$

The Type III Sum of Squares would be

1. SS_{F_1} which is computed by fitting two models,
   (a) the Full model with an intercept and all terms F_1, F_2, and $F_1 * F_2$ yielding SSE_1
   (b) the Reduced model with all terms but F_1 yielding SSE_2
   (c) $SS_{F_1} = SSE_2 - SSE_1$

2. SS_{F_2} which is computed by fitting two models,

   the Full model with an intercept and all terms F_1, F_2, and $F_1 * F_2$ yielding SSE_1
   the Reduced model with all terms but F_2 yielding SSE_3
   $SS_{F_1} = SSE_3 - SSE_1$

3. $SS_{F_1 * F_2}$ which is computed by fitting two models,

   the Full model with an intercept and all terms F_1, F_2, and $F_1 * F_2$ yielding SSE_1
   the Reduced model with all terms but $F_1 * F_2$ yielding SSE_4
   $SS_{F_1 * F_2} = SSE_4 - SSE_1$

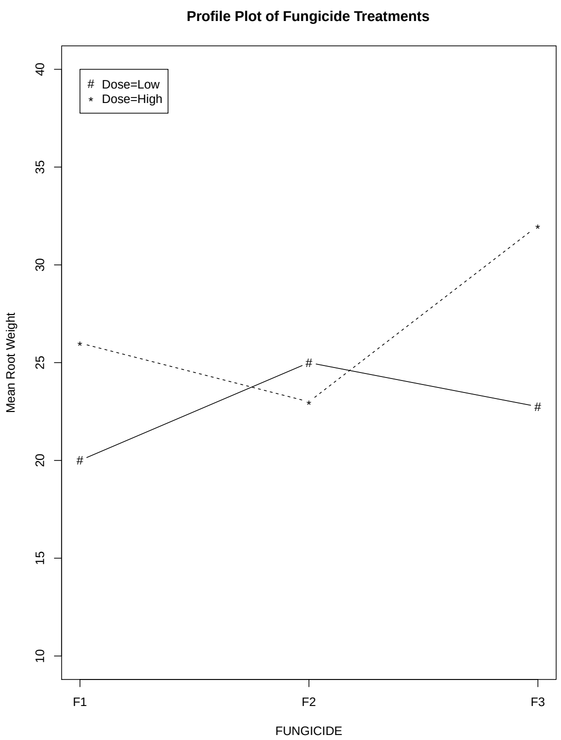
The hypotheses that these sum of squares are tested are given below:

Source of Variation	Type III Hypotheses
F_1	$\bar{\mu}_{1.} = \bar{\mu}_{2.} = \cdots = \bar{\mu}_{a.}$
F_2	$\bar{\mu}_{.1} = \bar{\mu}_{.2} = \cdots = \bar{\mu}_{.b}$
$F_1 * F_2$	$(\mu_{ij} - \mu_{kj}) - (\mu_{ih} - \mu_{kh}) = 0 \quad \text{for all} \quad (i, j, k, h)$

In most situations, Type III Sum of Squares are preferred because the hypotheses being tested do not involve the sample sizes in the particular experiment which would limit the type of inferences made by the researcher.

EXAMPLE A horticulturist is interested in studying the effectiveness of fungicide treatments applied to plots on which roses are grown. The three types of fungicides (F_2) were randomly applied at two dose levels (F_1) to 4 plots before inoculation. This is a CRD with a factorial (2x3) treatment structure and $r = 4$ replications per treatment. Rose plants of the same health, size, and age were inoculated, planted, and after twenty weeks were dug up and the root weights determined. However, a number of plants died during the twenty weeks. This resulted in an unbalanced design with the number of reps per treatment varying from $n_{ij} = 2$ to $n_{ij} = 4$.

F_1	F_2		
	1	2	3
1	19 20 21	24 26	22 25 25 19
2	25 27	21 24 24	31 32 33 32



The following SAS code yields the tests described above for an unbalanced experiment.

```
*twofac,uneqrep.sas;
* This is the SAS code for the Example dealing with unequal number of reps;
* The data is from a CR 2x3 factorial experiment with an unequal number of
  replications:  n11=3, n12=2, n13=4, n21=2, n22=3, n23=4;
options pagesize=60 linesize=75 nocenter nodate;
title 'Unbalanced factorial experiment with no missing treatments';
data raw;
input A B Y @@;
IF A=1 and B=1 THEN TRT=1;
IF A=1 and B=2 THEN TRT=2;
IF A=1 and B=3 THEN TRT=3;
IF A=2 and B=1 THEN TRT=4;
IF A=2 and B=2 THEN TRT=5;
IF A=2 and B=3 THEN TRT=6;

cards;
1  1 19 1 1 20 1 1 21
1  2 24 1 2 26
1  3 22 1 3 25 1 3 25 1 3 19
2  1 25 2 1 27
2  2 21 2 2 24 2 2 24
2  3 31 2 3 32 2 3 33  2 3 32
;
run;
proc glm;
class A B;
model Y=A B A*B/ss1 ss2 ss3 ss4;
means A B A*B;
lsmeans A B A*B/stderr pdiff adjust=Tukey;
RUN;
proc glm;
class TRT;
model y=TRT/ss1 ss2 ss3 ss4;

contrast  'Main A' TRT 1  1  1 -1 -1 -1;
contrast  'Main B' TRT 1 -1  0  1 -1  0,
                   TRT 1  1 -2  1  1 -2;
contrast 'A*B'     TRT 1 -1  0 -1  1  0,
                   TRT 1  1 -2 -1 -1  2;
lsmeans TRT/stderr pdiff adjust=Tukey;
run;
```

Output from SAS for Unbalanced factorial experiment with NO missing treatments-Effects Model

Class	Levels	Values
A	2	1 2
B	3	1 2 3

Number of Observations Read 18
Number of Observations Used 18

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	305.2500000	61.0500000	18.91	<.0001
Error	12	38.7500000	3.2291667		
Corrected Total	17	344.0000000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	128.0000000	128.0000000	39.64	<.0001
B	2	81.5090909	40.7545455	12.62	0.0011
A*B	2	95.7409091	47.8704545	14.82	0.0006

Source	DF	Type II SS	Mean Square	F Value	Pr > F
A	1	123.3840909	123.3840909	38.21	<.0001
B	2	81.5090909	40.7545455	12.62	0.0011
A*B	2	95.7409091	47.8704545	14.82	0.0006

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	81.02884615	81.02884615	25.09	0.0003
B	2	67.92272727	33.96136364	10.52	0.0023
A*B	2	95.74090909	47.87045455	14.82	0.0006

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
A	1	81.02884615	81.02884615	25.09	0.0003
B	2	67.92272727	33.96136364	10.52	0.0023
A*B	2	95.74090909	47.87045455	14.82	0.0006

Level of A	N	Y Mean	Y Std Dev
1	9	22.3333333	2.73861279
2	9	27.6666667	4.41588043

Level of B	N	Y Mean	Y Std Dev
1	5	22.4000000	3.43511281
2	5	23.8000000	1.78885438
3	8	27.3750000	5.31675250

Level of A	Level of B	N	Y Mean	Y Std Dev
1	1	3	20.0000000	1.00000000
1	2	2	25.0000000	1.41421356
1	3	4	22.7500000	2.87228132
2	1	2	26.0000000	1.41421356
2	2	3	23.0000000	1.73205081
2	3	4	32.0000000	0.81649658

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

A	Y LSMEAN	Standard Error	H0:LSMEAN=0 Pr > |t|	H0:LSMean1= LSMean2 Pr > |t|
1	22.5833333	0.6234549	<.0001	0.0003
2	27.0000000	0.6234549	<.0001	

B	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	23.0000000	0.8202092	<.0001	1
2	24.0000000	0.8202092	<.0001	2
3	27.3750000	0.6353313	<.0001	3

Least Squares Means for effect B
Adjustment for Multiple Comparisons: Tukey-Kramer
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Y

i/j	1	2	3
1		0.6732	0.0032
2	0.6732		0.0176
3	0.0032	0.0176	

A	B	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	1	20.0000000	1.0374916	<.0001	1
1	2	25.0000000	1.2706626	<.0001	2
1	3	22.7500000	0.8984941	<.0001	3
2	1	26.0000000	1.2706626	<.0001	4
2	2	23.0000000	1.0374916	<.0001	5
2	3	32.0000000	0.8984941	<.0001	6

Least Squares Means for effect A*B
Adjustment for Multiple Comparisons: Tukey-Kramer
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Y

i/j	1	2	3	4	5	6
1		0.0837	0.3937	0.0302	0.3739	<.0001
2	0.0837		0.7014	0.9922	0.8198	0.0074
3	0.3937	0.7014		0.3535	1.0000	0.0001
4	0.0302	0.9922	0.3535		0.4843	0.0216
5	0.3739	0.8198	1.0000	0.4843		0.0003
6	<.0001	0.0074	0.0001	0.0216	0.0003	

To compare the treatment means, we can use the p-values directly from the above matrix because the multiple comparison have already been adjusted by Tukey-Kramer.

Case 1: Compare means across levels of B at Fixed levels of A:

At A=1, G = {B1, B2, B3}

At A=2, G1 = {B1, B2}, G2 = {B3}

Case 2: Compare all 6 Treatments:

G1 = {(A1,B1), (A1,B2), (A1,B3), (A2,B2)}

G2 = {(A1,B2), (A1,B3), (A2,B1), (A2,B2)}

G3 = {(A2,B3)}

Output from SAS for Unbalanced factorial experiment with NO missing treatments-Cell Means Model

The GLM Procedure

Class Level Information

Class	Levels	Values
TRT	6	1 2 3 4 5 6

Number of Observations Read 18
Number of Observations Used 18

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	305.2500000	61.0500000	18.91	<.0001
Error	12	38.7500000	3.2291667		
Corrected Total	17	344.0000000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	5	305.2500000	61.0500000	18.91	<.0001

Source	DF	Type II SS	Mean Square	F Value	Pr > F
TRT	5	305.2500000	61.0500000	18.91	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	5	305.2500000	61.0500000	18.91	<.0001

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
TRT	5	305.2500000	61.0500000	18.91	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Main A	1	81.02884615	81.02884615	25.09	0.0003
Main B	2	67.92272727	33.96136364	10.52	0.0023
A*B	2	95.74090909	47.87045455	14.82	0.0006

Least Squares Means

TRT	Y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	20.0000000	1.0374916	<.0001	1
2	25.0000000	1.2706626	<.0001	2
3	22.7500000	0.8984941	<.0001	3
4	26.0000000	1.2706626	<.0001	4
5	23.0000000	1.0374916	<.0001	5
6	32.0000000	0.8984941	<.0001	6

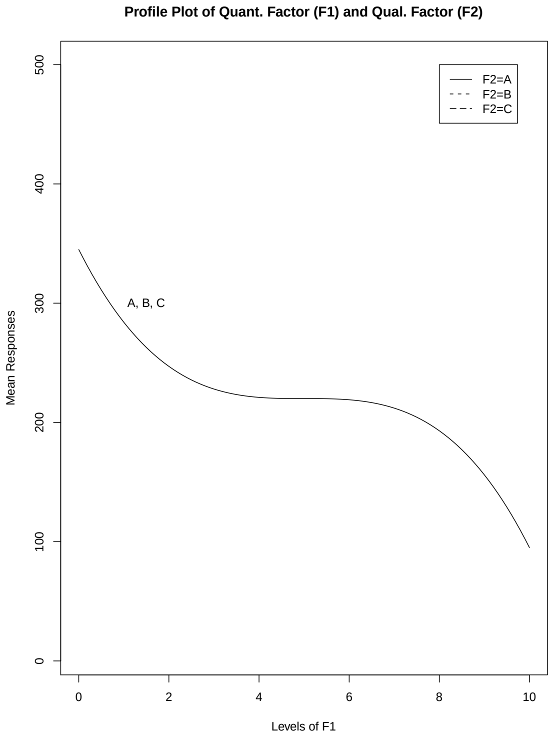
Least Squares Means for effect TRT
Adjustment for Multiple Comparisons: Tukey-Kramer
Pr > |t| for H0: LSMean(i)=LSMean(j)

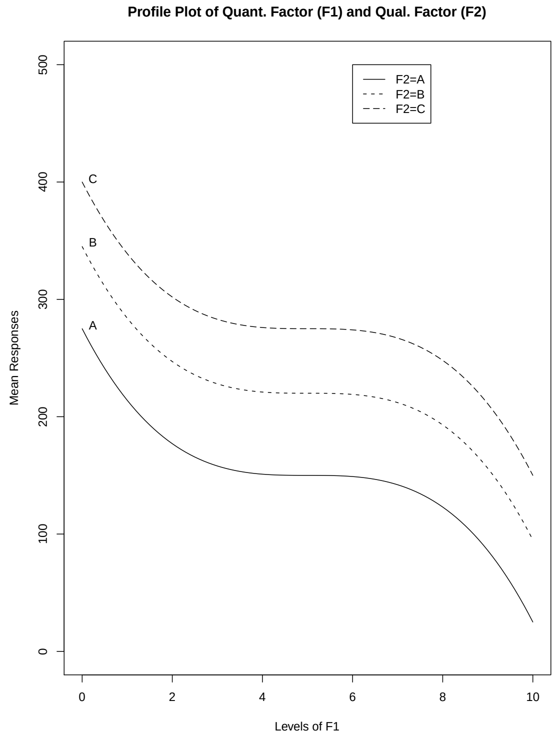
Dependent Variable: Y

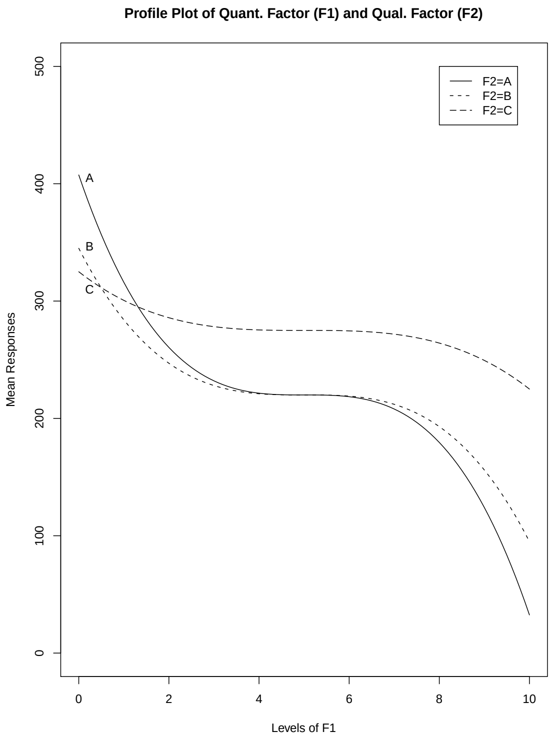
i/j	1	2	3	4	5	6
1		0.0837	0.3937	0.0302	0.3739	<.0001
2	0.0837		0.7014	0.9922	0.8198	0.0074
3	0.3937	0.7014		0.3535	1.0000	0.0001
4	0.0302	0.9922	0.3535		0.4843	0.0216
5	0.3739	0.8198	1.0000	0.4843		0.0003
6	<.0001	0.0074	0.0001	0.0216	0.0003	

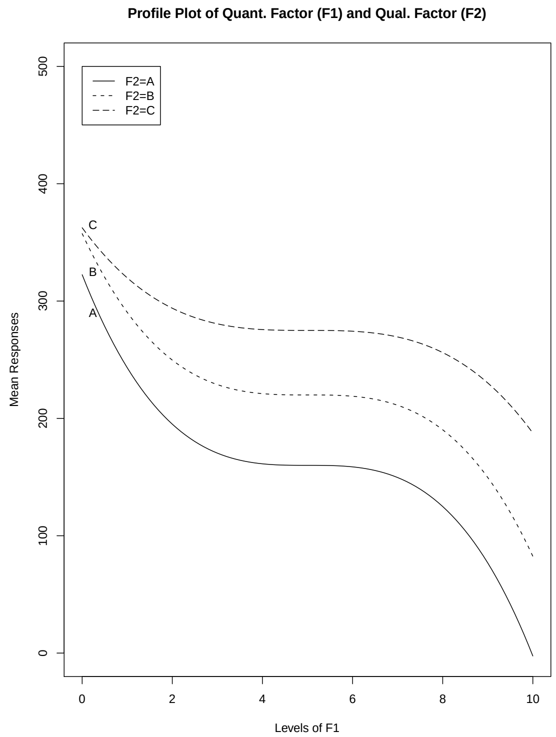
Factorial Experiment With Quantitative and Qualitative Factors

In this experiment factor F_1 has quantitative levels and F_2 has qualitative levels. The goal is to determine the polynomial model relating the levels of F_1 to the mean values of the response variables. Ascertaining whether there is interaction between factors F_1 and F_2 needs to be the first objective in the analysis. This will determine if a single polynomial model is needed (no interaction) or if separate models will be needed for each level of factor F_2 (interaction exists). The model relating the response across the levels of the quantitative factor will be different across the levels of the qualitative factor when the two factors interact. In that case, the response curves for the quantitative factor must be estimated separately for each level of the qualitative factor. In the absence of interaction between the two factors, the response trend for the quantitative factor will be similar at all levels of the qualitative factor except for a possible shift of the curves. If the interaction is nonsignificant **and** the main effect of the qualitative factor is nonsignificant, a single response curve will suffice for describing of the process with respect to the quantitative factor. If the interaction is nonsignificant **but** the main effect of the qualitative factor is significant, then the response curves will be parallel but shifted away from one another.


Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)
F2=A
F2=B
F2=C
A, B, C
Mean Responses
Levels of F1
0
100
200
300
400
500
0
2
4
6
8
10


Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)
F2=A
F2=B
F2=C
C
B
A
Mean Responses
Levels of F1
0
100
200
300
400
500
0
2
4
6
8
10


Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)
F2=A
F2=B
F2=C
A
B
C
Mean Responses
Levels of F1
0
100
200
300
400
500
0
2
4
6
8
10


Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)
F2=A
F2=B
F2=C
C
B
A
Mean Responses
Levels of F1
0
100
200
300
400
500
0
2
4
6
8
10

Example: Heavy Metals in Sewage Sludge

EXAMPLE from Textbook Sludge is a dried product remaining from processed sewage; it contains nutrients beneficial to plant growth. It can be used for fertilizer on agricultural crops provided it does not contain toxic levels of certain elements such as heavy metals (such as zinc, not rock groups). Typically, the levels of metals in sludge are assayed by growing plants in media containing different doses of the sludge.

Research Hypothesis: A soil scientist hypothesized the concentration of certain heavy metals in sludge would differ among the metropolitan areas from which the sludge was obtained. The variation could result from any number of reasons, including different industrial bases surrounding the areas and the efficiency of the sewage treatment facilities. If this were true, then recommendations for applications on crops would have to be preceded by knowledge about the source of the sludge material. An assay was planned to determine whether there was significant variation in heavy metal concentrations among diverse metropolitan areas.

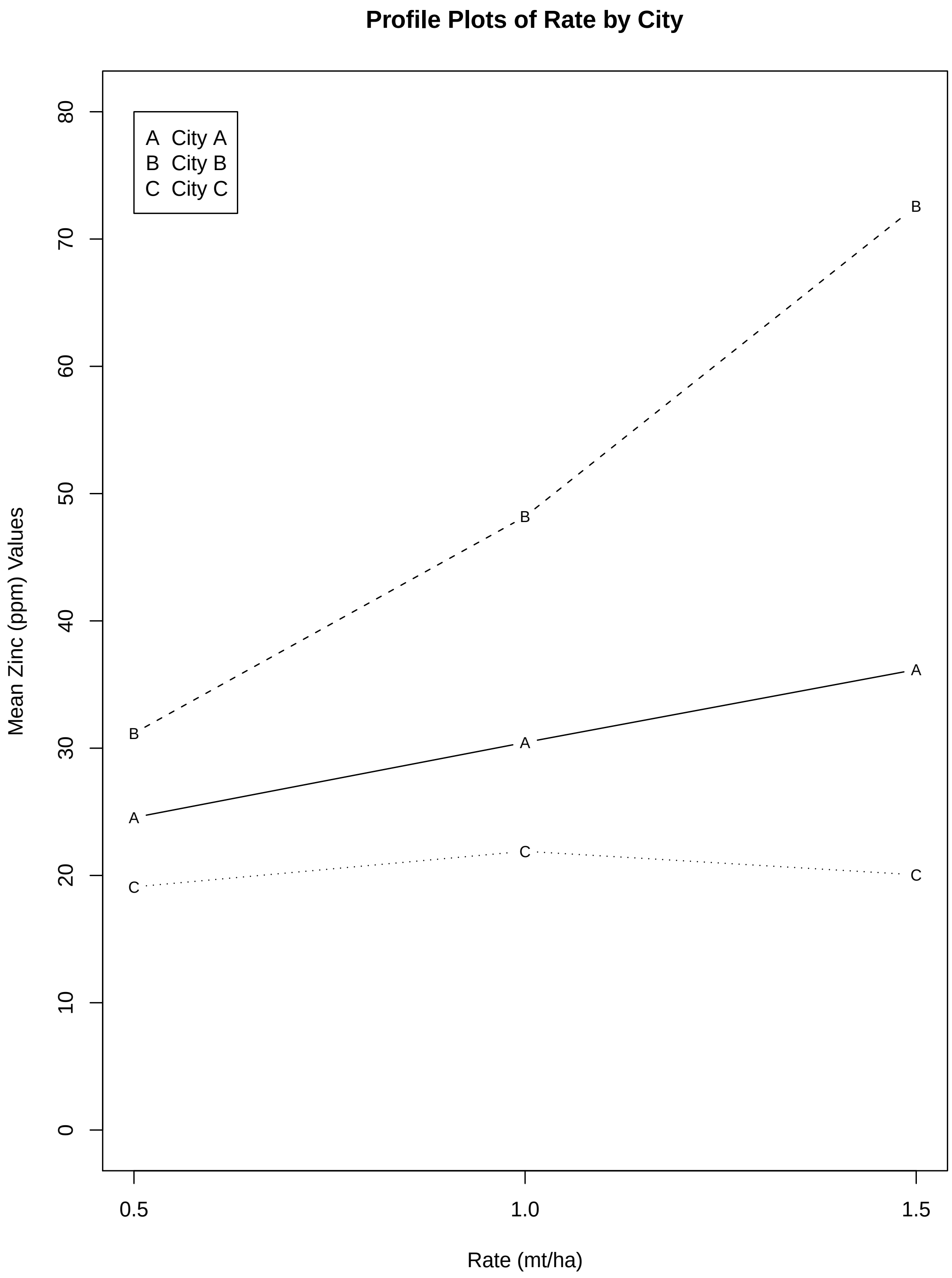
Treatment Design: The investigator obtained sewage sludge from treatment plants located in three different metropolitan areas. Barley plants were grown in a sand medium to which sludge was added as fertilizer. The sludge was added to the sand at three different rates: 0.5, 1.0, and 1.5 metric tons/acre. The factorial arrangement for the treatment design consisted of one qualititative factor, "City" with three fixed levels and one quantitative factor, "Rate" with three fixed levels.

Experiment Design: Each of the nine treatment combinations was randomly assigned to four replicate containers in a completely randomized design. The containers were arranged completely at random in a growth chamber. At a certain stage of growth the zinc content in parts per million was determined for the barley plants grown in each of the containers. The data are shown in Table 1, the manual calculations for linear and quadratic sums of squares partitions are shown in Table 2, and the analysis of variance is provided in Table 3.

Table 1. Zinc content (ppm) of barley plants grown in media containing sludge at three rates from three metropolitan areas.

City and Rate (MT/hectare)

A			B			C		
0.5	1.0	1.5	0.5	1.0	1.5	0.5	1.0	1.5
26.4	25.2	26.0	30.1	47.7	73.8	19.4	23.2	18.9
23.5	39.2	44.6	31.0	39.1	71.1	19.3	21.3	19.8
25.4	25.5	35.5	30.8	55.3	68.4	18.7	23.2	19.6
22.9	31.9	38.6	32.8	50.7	77.1	19.0	19.9	21.9
24.55	32.45	36.175	31.175	48.2	72.6	19.1	21.9	20.05


Profile Plots of Rate by City
A City A
B City B
C City C
Mean Zinc (ppm) Values
Rate (mt/ha)
0
10
20
30
40
50
60
70
80
0.5
1.0
1.5

The model $y_{ijk} = \mu_{ij} + e_{ijk}$ with $i, j = 1, 2, 3$ and $k = 1, 2, 3, 4$ was fit to the data. The researcher wanted to examine trends in $\mu_{i1}, \quad \mu_{i2}, \quad \mu_{i3}, \quad$ for each City i:

The analysis of variance table is given here.

SOURCE VARIATION	D. F.	SUM SQS	MS	F Ratio	P-value
Rate	2	1945.45	972.72	50.71	$< .0001$
Rate_Lin	1	1944	1944	101.35	$< .0001$
Rate_Qud	1	1.45	1.45	0.08	.786
City	2	5720.67	2860.34	149.13	$< .0001$
Rate x City	4	1809.40	452.35	23.58	$< .0001$
Rate_Lin x City	2	1760.15	880.07	45.88	$< .0001$
Rate_Qud x City	2	49.25	24.63	1.28	.293
Error	27	517.86	19.18	.	.

The analysis of variance in the above table indicates a significant interaction between Rate and City and significant main effects for both factors (p-value<.0001).

The 2 degrees of freedom for the Rate sum of squares partition into 1 degree of freedom each for linear and quadratric rate effects.

The 4 degrees of freedom for the interaction between Rate and City partition into 2 degrees of freedom for Linear and 2 degrees of freedom for quadratric.

The interaction between Rate linear trend and City is significant (p-value<.0001) but the interaction between Rate Quadratric and City is not significant (p-value=.293).

The signficant interactions between City and the linear trend on rate of sludge application suggests that interpretations should be based on separately fitted lines for each of the three cities.

The estimated linear regression lines are plotted on the previous page along with the treatment means.

The plot illustrates the Rate (linear) by City interaction. The response to rate is linear for each city. However, the significance of the interaction between City and the linear partition for Rate shows up in the plot as a different linear response of zinc to Rate of application for each of the three cities.

SAS code and output:

```
* Factorial_sludgex.sas;
ods html;  ods graphics on;
option ls=80 ps=50 nocenter nodate;
title 'ZINC IN SLUDGE EXAMPLE-KUEHL EX 6.3';
data raw;
array Y Y1-Y4; INPUT F1 $ F2 $ TRT $ Y1-Y4;   do over Y;
X=Y;
output; end;
      drop  Y1-Y4;
      label F1 = 'CITY' F2 = 'RATE' X = 'ZINC-CONC' TRT = 'CITY-RATE';
cards;
A 0.5 A05  26.4 23.5 25.4 22.9
A 1.0 A10 25.2 39.2 25.5 31.9
A 1.5 A15 26.0 44.6 35.5 38.6
B 0.5 B05  30.1 31.0 30.8 32.8
B 1.0 B10 47.7 39.1 55.3 50.7
B 1.5 B15 73.8 71.1 68.4 77.1
C 0.5 C05  19.4 19.3 18.7 19.0
C 1.0 C10 23.2 21.3 23.2 19.9
C 1.5 C15 18.9 19.8 19.6 21.9
Run;

*Analysis using Effects Model;
proc glm;
class F1 F2;
model X = F1 F2 F1*F2/ss4;
means F1 F2 F1*F2;
output out=ASSUMP r=RESID p=MEANS;
PROC PLOT; PLOT RESID*MEANS=F1/vref=0;
PROC PLOT; PLOT MEANS*F2=F1;
proc univariate def=5 plot normal; var RESID;

*Analysis using Cell Means Model;
PROC GLM;
CLASS TRT;
MODEL X = TRT/SS3;
CONTRAST  'LIN-A' TRT    -1  0  1  0  0  0  0  0  0;
CONTRAST  'LIN-B'  TRT    0  0  0 -1  0  1  0  0  0;
CONTRAST  'LIN-C'  TRT    0  0  0  0  0  0 -1  0  1;
CONTRAST  'QUAD-A' TRT    1 -2  1  0  0  0  0  0  0;
CONTRAST  'QUAD-B'  TRT   0  0  0  1 -2  1  0  0  0;
CONTRAST  'QUAD-C'  TRT   0  0  0  0  0  0  1 -2  1;

*The following commands yield a simultaneous test of
the two trend contrasts for each city;
CONTRAST  'City-A'  TRT   -1  0  1  0  0  0  0  0  0,
                    TRT    1 -2  1  0  0  0  0  0  0;
CONTRAST  'City-B'  TRT    0  0  0 -1  0  1  0  0  0,
                    TRT    0  0  0  1 -2  1  0  0  0;
CONTRAST  'City-C'  TRT    0  0  0  0  0  0 -1  0  1,
                    TRT    0  0  0  0  0  0  1 -2  1;
LSMEANS TRT/pdiff stderr adjust=tukey;
MEANS TRT/TUKEY hovtest=bf;
RUN;
```

ZINC IN SLUDGE EXAMPLE-KUEHL EX 6.3
The GLM Procedure

Class Level Information

Class	Levels	Values
F1	3	A B C
F2	3	0.5 1.0 1.5

Number of Observations Read 36
Number of Observations Used 36

Dependent Variable: X ZINC-CONC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	9475.515000	1184.439375	61.75	<.0001
Error	27	517.865000	19.180185		
Corrected Total	35	9993.380000			

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
F1	2	5720.671667	2860.335833	149.13	<.0001
F2	2	1945.445000	972.722500	50.71	<.0001
F1*F2	4	1809.398333	452.349583	23.58	<.0001

Level of F1	N	X Mean	X Std Dev
A	12	30.3916667	7.3227489
B	12	50.6583333	18.2248311
C	12	20.3500000	1.6312293

Level of F2	N	X Mean	X Std Dev
0.5	12	24.9416667	5.2636763
1.0	12	33.5166667	12.4972603
1.5	12	42.9416667	23.4039025

Level of F1	Level of F2	N	X Mean	X Std Dev
A	0.5	4	24.5500000	1.62992842
A	1.0	4	30.4500000	6.60126251
A	1.5	4	36.1750000	7.76418057
B	0.5	4	31.1750000	1.15000000
B	1.0	4	48.2000000	6.82446579
B	1.5	4	72.6000000	3.72290209
C	0.5	4	19.1000000	0.31622777
C	1.0	4	21.9000000	1.60623784
C	1.5	4	20.0500000	1.29228480

```
                  Plot of RESID*MEANS.  Symbol is value of F1.

RESID |
      |
   10 +
      |
      |                     A     A
      |
      |                                       B
      |
      |
    5 +
      |                                                                B
      |
      |                                       B
      |           C    A          A
      |             C       AB                                         B
      |                A
    0 +----------CC----------B--------------------------------------------------
      |          CC C  A     B    A           B
      |           C    A     B                                         B
      |             C
      |
      |
      |                                                                B
   -5 +                     A
      |
      |
      |
      |
      |
      |                                       B
  -10 +                           A
      |
      --+---------+---------+---------+---------+---------+---------+---------+-
       10        20        30        40        50        60        70        80

                                         MEANS

NOTE: 4 obs hidden.
```

Brown and Forsythe's Test for Homogeneity of X Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
TRT	8	138.4	17.2963	2.70	0.0255
Error	27	173.2	6.4157		

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.914787	Pr < W	0.0089
Kolmogorov-Smirnov	D	0.167312	Pr > D	0.0117
Cramer-von Mises	W-Sq	0.247338	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.351298	Pr > A-Sq	<0.0050

```
   Stem Leaf                     #  Boxplot
      8 48                       2     0
      6 1                        1     0
      4 5                        1     |
      2 45                       2     |
      0 2382334688              10  +--+--+
     -0 65210765444221          14  *-----*
     -2 0                        1     |
     -4 202                      3     0
     -6
     -8 1                        1     *
    -10 2                        1     *
        ----+----+----+----+
```

Variable: RESID

```
                     Normal Probability Plot
        9+                                        *   *****
         |                                      ******
         |                                  +++++
         |                             ++++**
         |                       +********
      -1+              ************
         |             +++++
         |        *+++*
         |   +++++
         |+++++   *
     -11+    *
          +----+----+----+----+----+----+----+----+----+----+
              -2        -1         0        +1        +2
```

ZINC IN SLUDGE EXAMPLE-KUEHL EX 6.3

Class Level Information

Class	Levels	Values
TRT	9	A05 A10 A15 B05 B10 B15 C05 C10 C15

Number of Observations Read 36
Number of Observations Used 36

Dependent Variable: X ZINC-CONC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	9475.515000	1184.439375	61.75	<.0001
Error	27	517.865000	19.180185		
Corrected Total	35	9993.380000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	8	9475.515000	1184.439375	61.75	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LIN-A	1	270.281250	270.281250	14.09	0.0008
LIN-B	1	3432.061250	3432.061250	178.94	<.0001
LIN-C	1	1.805000	1.805000	0.09	0.7614
QUAD-A	1	0.020417	0.020417	0.00	0.9742
QUAD-B	1	36.260417	36.260417	1.89	0.1804
QUAD-C	1	14.415000	14.415000	0.75	0.3936
City-A	2	270.301667	135.150833	7.05	0.0034
City-B	2	3468.321667	1734.160833	90.41	<.0001
City-C	2	16.220000	8.110000	0.42	0.6595

Least Squares Means

CITY-RATE	Estimate	Standard Error	DF	Alpha	Lower	Upper	LSMEAN Number
A05	24.5500	2.1898	27	0.05	20.0570	29.0430	1
A10	30.4500	2.1898	27	0.05	25.9570	34.9430	2
A15	36.1750	2.1898	27	0.05	31.6820	40.6680	3
B05	31.1750	2.1898	27	0.05	26.6820	35.6680	4
B10	48.2000	2.1898	27	0.05	43.7070	52.6930	5
B15	72.6000	2.1898	27	0.05	68.1070	77.0930	6
C05	19.1000	2.1898	27	0.05	14.6070	23.5930	7
C10	21.9000	2.1898	27	0.05	17.4070	26.3930	8
C15	20.0500	2.1898	27	0.05	15.5570	24.5430	9

Least Squares Means for effect TRT
Pr > |t| for H0: LSMean(i)=LSMean(j)
Adjustment for Multiple Comparisons: Tukey

i/j	1	2	3	4	5	6	7	8	9
1		0.6162	0.0203	0.4696	<.0001	<.0001	0.7061	0.9936	0.8666
2	0.6162		0.6516	1.0000	0.0001	<.0001	0.0251	0.1739	0.0507
3	0.0203	0.6516		0.7889	0.0149	<.0001	0.0002	0.0024	0.0005
4	0.4696	1.0000	0.7889		0.0002	<.0001	0.0143	0.1101	0.0297
5	<.0001	0.0001	0.0149	0.0002		<.0001	<.0001	<.0001	<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001
7	0.7061	0.0251	0.0002	0.0143	<.0001	<.0001		0.9908	1.0000
8	0.9936	0.1739	0.0024	0.1101	<.0001	<.0001	0.9908		0.9995
9	0.8666	0.0507	0.0005	0.0297	<.0001	<.0001	1.0000	0.9995	

Tukey's Studentized Range (HSD) Test for X
Means with the same letter are not significantly different.

Tukey Grouping			Mean	N	TRT	B
	A		72.600	4	B15	
	B		48.200	4	B10	
	C		36.175	4	A15	
D	C		31.175	4	B05	
D	C	E	30.450	4	A10	
D	F	E	24.550	4	A05	
D	F	E	21.900	4	C10	
	F	E	20.050	4	C15	
	F		19.100	4	C05	

Because the City*Rate interaction was significant,
the Rates should be grouped separately for each city:

City A - G1 = {R5, R10}, G2 = {R10, R15}

City B - G1 = {R5}, G2 = {R10}, G3 = {R15}

City C - G1 = {R5, R10, R15}

In the above analysis, the BFL test and residual plot both indicated there may be unequal variances across the nine treatments. Also, the residual plots and S-W test indicated non-normality. The Box-Cox procedure yielded the transformation $X = Y^{-.83} \approx 1/Y$. A running of the SAS program with the data transformed yields the following:

ZINC IN SLUDGE EXAMPLE-KUEHL EX 6.3 using X = 1/Y Transformation:
Dependent Variable: X   1/ZINC-CONC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.00547830	0.00068479	45.51	<.0001
Error	27	0.00040631	0.00001505		
Corrected Total	35	0.00588462			

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
F1	2	0.00441535	0.00220768	146.70	<.0001
F2	2	0.00077426	0.00038713	25.73	<.0001
F1*F2	4	0.00028869	0.00007217	4.80	0.0047

```
                       Plot of RESID*MEANS.  Symbol is value of F1.

 RESID |
       |
 0.010 +                          A
       |
       |
       |
       |
       |
       |                                  A
 0.005 +                                  A
       |                B                                 C
       |
       |                                           A            C
       |
       |                               B           A      C        C
       |     B                         B                        C
 0.000 +-----B----------B--------------B---------------------------C------------
       |     B                    A                                C
       |                B              B           A
       |
       |                B         A       A        A      C
       |
       |                                                        C
-0.005 +
       |
       |                          A
       |
       |
       |                                  A
       |
       -+-------------+-------------+-------------+-------------+-------------+-
      0.01          0.02          0.03          0.04          0.05          0.06
```

Brown and Forsythe's Test for Homogeneity of X Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
TRT	8	0.000096	0.000012	2.35	0.0465
Error	27	0.000138	5.101E-6		

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.964985	Pr < W	0.3047
Kolmogorov-Smirnov	D	0.149674	Pr > D	0.0405
Cramer-von Mises	W-Sq	0.097865	Pr > W-Sq	0.1184
Anderson-Darling	A-Sq	0.556962	Pr > A-Sq	0.1439

```
   Stem Leaf                     #  Boxplot
      8 7                        1     0
      6
      4 4537                     4     |
      2 89                       2     |
      0 13345801117             11  +--+--+
     -0 654886631                9  |     |
     -2 008776                   6  +-----+
     -4 4                        1     |
     -6 3                        1     |
     -8 4                        1     0
        ----+----+----+----+
    Multiply Stem.Leaf by 10**-3
```

Variable: RESID

```
                      Normal Probability Plot
   0.009+                                            *  ++
        |                                         ++++++
        |                                  *******
   0.003+                             +++++*
        |                        *********
        |                   *******
  -0.003+             * *****+
        |         ++++++
        |  +++++++
  -0.009+++  *
         +----+----+----+----+----+----+----+----+----+----+
             -2        -1         0        +1        +2
```

Class	Levels	Values
TRT	9	A05 A10 A15 B05 B10 B15 C05 C10 C15

Number of Observations Read 36

Dependent Variable: X 1/ZINC-CONC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.00547830	0.00068479	45.51	<.0001
Error	27	0.00040631	0.00001505		
Corrected Total	35	0.00588462			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	8	0.00547830	0.00068479	45.51	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
LIN-A	1	0.00029417	0.00029417	19.55	0.0001
LIN-B	1	0.00067035	0.00067035	44.55	<.0001
LIN-C	1	0.00001097	0.00001097	0.73	0.4006
QUAD-A	1	0.00000199	0.00000199	0.13	0.7187
QUAD-B	1	0.00000931	0.00000931	0.62	0.4384
QUAD-C	1	0.00007616	0.00007616	5.06	0.0328
City-A	2	0.00029616	0.00014808	9.84	0.0006
City-B	2	0.00067966	0.00033983	22.58	<.0001
City-C	2	0.00008713	0.00004356	2.89	0.0726

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

TRT	X LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
A05	0.04086755	0.00193963	<.0001	1
A10	0.03393910	0.00193963	<.0001	2
A15	0.02873970	0.00193963	<.0001	3
B05	0.03210900	0.00193963	<.0001	4
B10	0.02108671	0.00193963	<.0001	5
B15	0.01380122	0.00193963	<.0001	6
C05	0.05236684	0.00193963	<.0001	7
C10	0.04585163	0.00193963	<.0001	8
C15	0.05002440	0.00193963	<.0001	9

Least Squares Means for effect TRT
Pr > |t| for H0: LSMean(i)=LSMean(j)
Adjustment for Multiple Comparisons: Tukey

i/j	1	2	3	4	5	6	7	8	9
1		0.2640	0.0039	0.0728	<.0001	<.0001	0.0070	0.6713	0.0530
2	0.2640		0.6223	0.9988	0.0020	<.0001	<.0001	0.0048	<.0001
3	0.0039	0.6223		0.9428	0.1647	0.0003	<.0001	<.0001	<.0001
4	0.0728	0.9988	0.9428		0.0107	<.0001	<.0001	0.0009	<.0001
5	<.0001	0.0020	0.1647	0.0107		0.2107	<.0001	<.0001	<.0001
6	<.0001	<.0001	0.0003	<.0001	0.2107		<.0001	<.0001	<.0001
7	0.0070	<.0001	<.0001	<.0001	<.0001	<.0001		0.3363	0.9937
8	0.6713	0.0048	<.0001	0.0009	<.0001	<.0001	0.3363		0.8360
9	0.0530	<.0001	<.0001	<.0001	<.0001	<.0001	0.9937	0.8360	

Because the City*Rate interaction was significant,
the Rates should be grouped separately for each city:

City A - G1 = {R5, R10}, G2 = {R10, R15}

City B - G1 = {R5}, G2 = {R10, R15}

City C - G1 = {R5, R10, R15}

The results are slightly different from the results using the untransformed data.

EXAMPLE OF CONTRASTS IN A CRD 3x3 FACTORIAL EXPERIMENT

Model: $Y_{ijk} = \mu_{ij} + e_{ijk}$; for $i = 1, \cdots, 3$; $\jmath = 1, \cdots, 3$; $k = 1, \cdots, r_{ij}$

$SS_{TOT} = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{r_{ij}} [Y_{ijk} - \bar{Y}_{...}]^2$

$SS_E = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{r_{ij}} [Y_{ijk} - \hat{\mu}_{ij}]^2$

$SS_{MODEL} = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{r_{ij}} [\hat{\mu}_{ij} - \bar{Y}_{...}]^2$

We want to decompose SS_{MODEL} into terms which represent differences in the $t = ab$ treatments: "F1 Main Effect", "F2 Main Effect", and "F1xF2 Interaction."

Construct Contrasts which represent comparisons between treatment means which are main effects and two-way interactions. We will want to simultaneously test the several contrasts that make up the two degrees of freedom for the main effects and the four degrees of freedom for the interaction effects.

Suppose we have $t = ab$ treatment means: $\mu_{11}, \ldots, \mu_{1b}, \mu_{21}, \ldots, \mu_{2b}, \ldots, \mu_{a1}, \ldots, \mu_{ab}$ and construct k contrasts in the t means: $\mathbf{C}_1, \ldots, \mathbf{C}_k$ with

$$\mathbf{C}_i = (c_{i1}, c_{i2}, \ldots, c_{it}), \quad \sum_{j=1}^{t} c_{ij} = 0 \quad \text{for all} \quad i = 1, \ldots, k$$

Form a $k \times t$ contrast matrix: $\mathbf{C}' = (\mathbf{C}_1', \mathbf{C}_2', \ldots, \mathbf{C}_k')$

Let $\hat{\mu}_{ij} = \bar{y}_{ij.}$ and $\hat{\boldsymbol{\mu}}' = (\bar{y}_{11.}, \ldots, \bar{y}_{1b.}, \ldots, y_{ab.})$

where $E[\hat{\mu}_{ij}] = \mu_{ij}$ and $Var(\hat{\mu}_{ij}) = \frac{\sigma_e^2}{n_{ij}}$

Let $\mathbf{D} = Diag(\frac{1}{n_{11}}, \ldots, \frac{1}{n_{ab}})$ then $\hat{\boldsymbol{\mu}}$ has a $N_t(\mu, \sigma_e^2 \mathbf{D})$ distribution.

Then, $\mathbf{C}\hat{\mu}$ has a $N_k(\mathbf{C}\boldsymbol{\mu}, \sigma_e^2 \mathbf{C}DC')$ distribution. This then yields

$\frac{1}{\sigma_e^2} SS_{\mathbf{C}} = \frac{1}{\sigma_e^2} (C\hat{\boldsymbol{\mu}})' [CDC']^{-1} (C\hat{\boldsymbol{\mu}})$ has a noncentral Chi-square distribution with non-centrality parameter $(C\boldsymbol{\mu})'[CDC']^{-1}(C\boldsymbol{\mu})$ and df = Row Rank $\mathbf{C}$ and is independent of MSE.

A test statistic for simultaneously testing $H_o : \mathbf{C}_1\boldsymbol{\mu} = \mathbf{0}, \mathbf{C}_2\boldsymbol{\mu} = \mathbf{0}, \cdots, \mathbf{C}_k\boldsymbol{\mu} = \mathbf{0}$ versus H_1 : at least one $\mathbf{C}_i\boldsymbol{\mu} \neq \mathbf{0}$ is given by

$F = \frac{SS_{\mathbf{C}}/k}{MSE}$ which has under $H_o : \mathbf{C}\boldsymbol{\mu} = 0$ an F-distribution with $df = k, df_E$.

For our example, t=(3)(3)=9 The following table contains t-1 = 9-1=8 mutually orthogonal contrasts which represent the t-1=8 df for decomposing SS_{MODEL} into components for Main Effects and Interaction provided $r_{ij} > 0$ for $i = 1, \cdots, a;\ \jmath = 1, \cdots, b$.

Coefficients for Mutually Orthogonal Contrasts in 9 Treatment Means

		TREATMENT MEANS								
EFFECT	CONTRAST	μ_{11}	μ_{12}	μ_{13}	μ_{21}	μ_{22}	μ_{23}	μ_{31}	μ_{32}	μ_{33}
Main F_1	C_1	1	1	1	-1	-1	-1	0	0	0
	C_2	1	1	1	1	1	1	-2	-2	-2
Main F_2	C_3	1	-1	0	1	-1	0	1	-1	0
	C_4	1	1	-2	1	1	-2	1	1	-2
Interaction	C_5	1	-1	0	-1	1	0	0	0	0
	C_6	1	1	-2	-1	-1	2	0	0	0
	C_7	1	-1	0	1	-1	0	-2	2	0
	C_8	1	1	-2	1	1	-2	-2	-2	4

I. Interaction Contrasts:

$\mu^{\mathbf{t}} = (\mu_{11},\ \ \mu_{12},\ \ \mu_{13},\ \ \mu_{21},\ \ \mu_{22},\ \ \mu_{23},\ \ \mu_{31},\ \ \mu_{32},\ \ \mu_{33})$

$\hat{\mu}^{\mathbf{t}} = (.039258,\ \ .032794,\ \ .027904,\ \ .031109,\ \ .020644,\ \ .013613,\ \ .049761,\ \ .043834,\ \ .047635)$

$\mathbf{C_5} = (1, -1, 0, -1, 1, 0, 0, 0, 0) \qquad \mathbf{C_6} = (1, 1, -2, -1, -1, 2, 0, 0, 0)$

$\mathbf{C_7} = (1, -1, 0, 1, -1, 0, -2, 2, 0) \qquad \mathbf{C_8} = (1, 1, -2, 1, 1, -2, -2, -2, 4)$

$$\mathbf{C_{F_1*F_2}} = \begin{pmatrix} \mathbf{C_5} \\ \mathbf{C_6} \\ \mathbf{C_7} \\ \mathbf{C_8} \end{pmatrix} \qquad \mathbf{C_{F_1*F_2}}\hat{\mu} = \begin{pmatrix} 11.125 \\ 48.475 \\ -17.325 \\ -84.975 \end{pmatrix} \qquad \mathbf{D} = Diag\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

$$\mathbf{C_{F_1*F_2}DC^t_{F_1*F_2}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix} \qquad \Rightarrow \quad \left(C_{F_1*F_2}DC^t_{F_1*F_2}\right)^{-1} = Diag\left(1, \tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{9}\right)$$

$\mathbf{SS_{F_1*F_2}} = (\mathbf{C_{F_1*F_2}}\hat{\mu})^{\mathbf{t}} \left(\mathbf{C_{F_1*F_2}DC^t_{F_1*F_2}}\right)^{\mathbf{-1}} (\mathbf{C_{F_1*F_2}}\hat{\mu}) = \mathbf{.00028869}$ with $df_{F_1*F_2}$=rank($\mathbf{C_{F_1*F_2}}$) = 4

$F = \frac{MS_{F_1*F_2}}{MSE} = \frac{.00028869/4}{.00040631/27} = 4.80 \quad \Rightarrow p - value = Pr[F_{4,27} \geq 4.80] < 0.0047$

This is the sum of squares and F-test given on the SAS output for Interaction.

II. Main Effect of City (F_1):

$\mu^{\mathbf{t}} = (\mu_{11},\ \ \mu_{12},\ \ \mu_{13},\ \ \mu_{21},\ \ \mu_{22},\ \ \mu_{23},\ \ \mu_{31},\ \ \mu_{32},\ \ \mu_{33})$

$\hat{\mu}^{\mathbf{t}} = (.039258,\ \ .032794,\ \ .027904,\ \ .031109,\ \ .020644,\ \ .013613,\ \ .049761,\ \ .043834,\ \ .047635)$

$\mathbf{C_1} = (1, 1, 1, -1, -1, -1, 0, 0, 0)$

$\mathbf{C_2} = (1, 1, 1, 1, 1, 1, -2, -2, -2)$

$$\mathbf{C_{F_1}} = \begin{pmatrix} \mathbf{C_1} \\ \mathbf{C_2} \end{pmatrix} \qquad \mathbf{C_{F_1}}\hat{\mu} = \begin{pmatrix} -60.8 \\ 121.05 \end{pmatrix} \qquad \mathbf{D} = Diag\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

$\mathbf{C_{F_1}DC^t_{F_1}} = \begin{pmatrix} \frac{6}{4} & 0 \\ 0 & \frac{18}{4} \end{pmatrix}$

$SS_{F_1} = (\mathbf{C_{F_1}}\hat{\mu})^{\mathbf{t}} \left(\mathbf{C_{F_1}DC^t_{F_1}}\right)^{\mathbf{-1}} (\mathbf{C_{F_1}}\hat{\mu}) = \mathbf{.00441535}$ with $df_{F_1} = rank(\mathbf{C_{F_1}}) = 2$

$F = \frac{MS_{F_1}}{MSE} = \frac{.00441535/2}{.00040631/27} = 146.70 \quad \Rightarrow p - value = Pr[F_{2,27} \geq 146.70] < 0.0001$

This is the sum of squares and F-test given on the SAS output for Main Effect due to City.

III. Main Effect of Rate (F_2):

$\mu^{\mathbf{t}} = (\mu_{11},\ \ \mu_{12},\ \ \mu_{13},\ \ \mu_{21},\ \ \mu_{22},\ \ \mu_{23},\ \ \mu_{31},\ \ \mu_{32},\ \ \mu_{33})$

$\hat{\mu}^{\mathbf{t}} = (24.55,\ \ 30.45,\ \ 36.175,\ \ 31.175,\ \ 48.2,\ \ 72.6,\ \ 19.1,\ \ 21.9,\ \ 20.05)$

$\mathbf{C_3} = (1, -1, 0, 1, -1, 0, 1, -1, 0)$

$\mathbf{C_4} = (1, 1, -2, 1, 1, -2, 1, 1, -2)$

$$\mathbf{C_{F_2}} = \begin{pmatrix} \mathbf{C_3} \\ \mathbf{C_4} \end{pmatrix} \qquad \mathbf{C_{F_2}}\hat{\mu} = \begin{pmatrix} -25.725 \\ -82.275 \end{pmatrix} \qquad \mathbf{D} = Diag\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

$\mathbf{C_{F_2}DC^t_{F_2}} = \begin{pmatrix} \frac{6}{4} & 0 \\ 0 & \frac{18}{4} \end{pmatrix}$

$SS_{F_2} = (\mathbf{C_{F_2}}\hat{\mu})^{\mathbf{t}} \left(\mathbf{C_{F_2}DC^t_{F_2}}\right)^{\mathbf{-1}} (\mathbf{C_{F_2}}\hat{\mu}) = \mathbf{.00077426}$ with $df_{F_2} = rank(\mathbf{C_{F_2}}) = 2$

$F = \frac{MS_{F_2}}{MSE} = \frac{.00077426/2}{.00040631/27} = 25.73 \quad \Rightarrow p - value = Pr[F_{2,27} \geq 25.73] < 0.0001$

This is the sum of squares and F-test given on the SAS output for Main Effect due to Rate.

The following SAS code is in the Dostat folder SAScode: **Factorial,SimultaneousContrasts.sas;** and output is given on the next pages.

```
*  Factorial,SimultaneousContrasts.sas;
option ls=80 ps=50 nocenter nodate;
title 'ZINC IN SLUDGE EXAMPLE-KUEHL EX 6.3';

data new;
array Y Y1-Y4; INPUT F1 $ F2 $ TRT $ Y1-Y4;   do over Y;
X=1/Y;output; end;
drop  Y1-Y4;label F1 = 'CITY' F2 = 'RATE' X = 'ZINC-CONC'
            TRT = 'CITY-RATE';
cards;
A 0.5 A05 26.4 23.5 25.4 22.9
A 1.0 A10 25.2 39.2 25.5 31.9
A 1.5 A15 26.0 44.6 35.5 38.6
B 0.5 B05 30.1 31.0 30.8 32.8
B 1.0 B10 47.7 39.1 55.3 50.7
B 1.5 B15 73.8 71.1 68.4 77.1
C 0.5 C05 19.4 19.3 18.7 19.0
C 1.0 C10 23.2 21.3 23.2 19.9
C 1.5 C15 18.9 19.8 19.6 21.9
Run;
PROC GLM;
CLASS TRT;
MODEL X = TRT/SS3;
CONTRAST  'F1-1' TRT     1  1  1 -1 -1 -1  0  0  0;
CONTRAST  'F1-2'  TRT    1  1  1  1  1  1 -2 -2 -2;
CONTRAST  'F2-1'  TRT    1 -1  0  1 -1  0  1 -1  0;
CONTRAST  'F2-2'  TRT    1  1 -2  1  1 -2  1  1 -2;
CONTRAST  'I1'     TRT   1 -1  0 -1  1  0  0  0  0;
CONTRAST  'I2'     TRT   1  1 -2 -1 -1  2  0  0  0;
CONTRAST  'I3'     TRT   1 -1  0  1 -1  0 -2  2  0;
CONTRAST  'I4'     TRT   1  1 -2  1  1 -2 -2 -2  4;

*The following commands yield a simultaneous test of
the two trend contrasts for each city;

CONTRAST  'F1' TRT     1  1  1 -1 -1 -1  0  0  0,
               TRT     1  1  1  1  1  1 -2 -2 -2;
CONTRAST  'F2' TRT     1 -1  0  1 -1  0  1 -1  0,
               TRT     1  1 -2  1  1 -2  1  1 -2;
CONTRAST  'I'  TRT     1 -1  0 -1  1  0  0  0  0,
               TRT     1  1 -2 -1 -1  2  0  0  0,
               TRT     1 -1  0  1 -1  0 -2  2  0,
               TRT     1  1 -2  1  1 -2 -2 -2  4;
RUN;
```

```
TRT                 9    A05 A10 A15 B05 B10 B15 C05 C10 C15
```

Dependent Variable: X   1/ZINC-CONC

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	0.00547830	0.00068479	45.51	<.0001
Error	27	0.00040631	0.00001505		
Corrected Total	35	0.00588462			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	8	0.00547830	0.00068479	45.51	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
F1-1	1	0.00089057	0.00089057	59.18	<.0001
F1-2	1	0.00352478	0.00352478	234.23	<.0001
F2-1	1	0.00039906	0.00039906	26.52	<.0001
F2-2	1	0.00037520	0.00037520	24.93	<.0001
I1	1	0.00001676	0.00001676	1.11	0.3006
I2	1	0.00002278	0.00002278	1.51	0.2292
I3	1	0.00000807	0.00000807	0.54	0.4703
I4	1	0.00024109	0.00024109	16.02	0.0004
F1	2	0.00441535	0.00220768	146.70	<.0001
F2	2	0.00077426	0.00038713	25.73	<.0001
I	4	0.00028869	0.00007217	4.80	0.0047

CRD with an axbxc factorial treatment structure and r reps

Suppose we have t treatments which are constructed by combining the levels of three factors: Factor F_1 having a levels, Factor F_2 having b levels, and Factor F_3 having c levels. Hence $t = abc$. To each of the t treatments, r EU's are randomly assigned, yielding $N = rabc$ observations. Further suppose that the levels of the three factors are the only levels of interest and hence we have a fixed treatments experiment. A model for this situation is given by

$$Y_{ijkl} = \mu_{ijk} + e_{ijkl} \quad \text{where,}$$

μ_{ijk} is the mean response from the (i, j, k) treatment and e_{ijkl}'s are iid $N(0, \sigma_e^2)$ representing the variation in the observations at each of the t treatments.

Goal: Assess the size of the difference in the μ_{ijk}'s, if any. Also, assign these differences to various types of interactions and main effects in the three factors.

1. Evaluate 3-Way Interaction in the Factors:

   H_o : No 3-Way Interaction in F_1, F_2, F_3

   $H_o : [(\mu_{ijk} - \mu_{i'jk}) - (\mu_{ij'k} - \mu_{i'j'k})] = [(\mu_{ijk'} - \mu_{i'jk'}) - (\mu_{ij'k'} - \mu_{i'j'k'})]$,
   for all choices of (i, i', j, j', k, k')

   H_1 : Equality Does Not Hold for SOME choices of (i, i', j, j', k, k')

   Note: $(\mu_{ijk} - \mu_{i'jk}) - (\mu_{ij'k} - \mu_{i'j'k})$ is the size of the 2-Way interaction between levels (i, i', j, j') of factors F_1 and F_2 at the kth level of F_3. The 3-Way interaction is evaluating whether the 2-Way interactions between factors F_1 and F_2 are consistent across all levels of F_3.

2. Three sets of 2-Way Interactions:

   (a) $F_1 * F_2$: H_o : No 2-Way Interaction in F_1, F_2
   $H_o : (\bar{\mu}_{ij.} - \bar{\mu}_{i'j.}) = (\bar{\mu}_{ij'.} - \bar{\mu}_{i'j'.})$, for all choices of (i, i', j, j')
   H_1 : Equality Does Not Hold for SOME choices of (i, i', j, j')

   (b) $F_1 * F_3$: H_o : No 2-Way Interaction in F_1, F_3
   $H_o : (\bar{\mu}_{i.k} - \bar{\mu}_{i'.k}) = (\bar{\mu}_{i.k'} - \bar{\mu}_{i'.k'})$, for all choices of (i, i', k, k')
   H_1 : Equality Does Not Hold for SOME choices of (i, i', k, k')

   (c) $F_2 * F_3$: H_o : No 2-Way Interaction in F_2, F_3
   $H_o : (\bar{\mu}_{.jk} - \bar{\mu}_{.j'k}) = (\bar{\mu}_{.jk'} - \bar{\mu}_{.j'k'})$, for all choices of (j, j', k, k')
   H_1 : Equality Does Not Hold for SOME choices of (j, j', k, k')

3. Three Sets of Main Effects

   (a) $F_1 : H_o : \bar{\mu}_{1..} = \bar{\mu}_{2..} = \cdots = \bar{\mu}_{a..}$
   H_1 : Not all $\bar{\mu}_{i..}$ are equal

   (b) $F_2 : H_o : \bar{\mu}_{.1.} = \bar{\mu}_{.2.} = \cdots = \bar{\mu}_{.b.}$
   H_1 : Not all $\bar{\mu}_{.j.}$ are equal

   (c) $F_3 : H_o : \bar{\mu}_{..1} = \bar{\mu}_{..2} = \cdots = \bar{\mu}_{..c}$
   H_1 : Not all $\bar{\mu}_{..k}$ are equal

In order to construct test statistics for the above hypotheses, we first partition SS_{TOT} into SS_{Model} and SSE. Next, we partition SS_{Model} into components representing Main Effects, Two-Way Interactions, and Three-Way Interactions.

The following formulas are valid only if we have equal replication of the abc treatments, i.e., $n_{ijk} = r$ for all (i, j, k).

$SS_{TOT} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{r} (y_{ijkl} - \bar{y}_{....})^2$

$= r \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (\bar{y}_{ijk.} - \bar{y}_{....})^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{r} (y_{ijkl} - \bar{y}_{ijk.})^2$

$= SS_{Model} + SSE$

$SS_{Model} = SS_{F_1} + SS_{F_2} + SS_{F_3} + SS_{F_1*F_2} + SS_{F_1*F_3} + SS_{F_2*F_3} + SS_{F_1*F_2*F_3}$, where

$SS_{F_1} = rbc \sum_{i=1}^{a} (\bar{y}_{i...} - \bar{y}_{....})^2$, with $df = a - 1$

$SS_{F_2} = rac \sum_{j=1}^{b} (\bar{y}_{.j..} - \bar{y}_{....})^2$, with $df = b - 1$

$SS_{F_3} = rab \sum_{k=1}^{c} (\bar{y}_{..k.} - \bar{y}_{....})^2$, with $df = c - 1$

$SS_{F_1*F_2} = rc \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij..} - \bar{y}_{....})^2 - SS_{F_1} - SS_{F_2}$, with $df = (a-1)(b-1)$

$SS_{F_1*F_3} = rb \sum_{i=1}^{a} \sum_{k=1}^{c} (\bar{y}_{i.k.} - \bar{y}_{....})^2 - SS_{F_1} - SS_{F_3}$, with $df = (a-1)(c-1)$

$SS_{F_2*F_3} = ra \sum_{j=1}^{b} \sum_{k=1}^{c} (\bar{y}_{.jk.} - \bar{y}_{....})^2 - SS_{F_2} - SS_{F_3}$, with $df = (b-1)(c-1)$

$SS_{F_1*F_2*F_3} = r \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (\bar{y}_{ijk.} - \bar{y}_{....})^2 - [SS_{F_1*F_2} + SS_{F_1*F_3} + SS_{F_2*F_3}] - [SS_{F_1} + SS_{F_2} + SS_{F_3}]$, with $df = (a-1)(b-1)(c-1)$

Note: $df_{Model} = (t-1) = (abc-1) =$

$= (a-1)+(b-1)+(c-1)+(a-1)(b-1)+(a-1)(c-1)+(b-1)(c-1)+(a-1)(b-1)(c-1)$

We summarize all these numbers in an AOV table:

	ANOVA Table for Balanced 3-Factor CRD				
Source	D.F.	SS	MS=SS/DF	E[MS]	Test Statistic
Model	abc-1	SS_{Model}	MS_{Model}	$\sigma_e^2 + r\theta_{Model}$	$\frac{MS_{Model}}{MSE}$
F_1	a-1	SS_{F_1}	MS_{F_1}	$\sigma_e^2 + rbc\theta_{F_1}$	$\frac{MS_{F_1}}{MSE}$
F_2	b-1	SS_{F_2}	MS_{F_2}	$\sigma_e^2 + rac\theta_{F_2}$	$\frac{MS_{F_2}}{MSE}$
F_3	c-1	SS_{F_3}	MS_{F_3}	$\sigma_e^2 + rab\theta_{F_3}$	$\frac{MS_{F_3}}{MSE}$
$F_1 * F_2$	(a-1)(b-1)	$SS_{F_1*F_2}$	$MS_{F_1*F_2}$	$\sigma_e^2 + rc\theta_{F_1*F_2}$	$\frac{MS_{F_1*F_2}}{MSE}$
$F_1 * F_3$	(a-1)(c-1)	$SS_{F_1*F_3}$	$MS_{F_1*F_3}$	$\sigma_e^2 + rb\theta_{F_1*F_3}$	$\frac{MS_{F_1*F_3}}{MSE}$
$F_2 * F_3$	(b-1)(c-1)	$SS_{F_2*F_3}$	$MS_{F_2*F_3}$	$\sigma_e^2 + ra\theta_{F_2*F_3}$	$\frac{MS_{F_2*F_3}}{MSE}$
$F_1 * F_2 * F_3$	(a-1)(b-1)(c-1)	$SS_{F_1*F_2*F_3}$	$MS_{F_1*F_2*F_3}$	$\sigma_e^2 + r\theta_{F_1*F_2*F_3}$	$\frac{MS_{F_1*F_2*F_3}}{MSE}$
Error	abc(r-1)	SSE	MSE	σ_e^2	.
Total	abcr-1	SS_{TOT}	.	.	.

The parameters in the Expected Mean Squares are given here:

$\theta_{F_1} = \frac{1}{a-1}\sum_{i=1}^{a}(\bar{\mu}_{i..} - \bar{\mu}_{...})^2$ $\quad\theta_{F_2} = \frac{1}{b-1}\sum_{j=1}^{b}(\bar{\mu}_{.j.} - \bar{\mu}_{...})^2$ $\quad\theta_{F_3} = \frac{1}{c-1}\sum_{k=1}^{c}(\bar{\mu}_{..k} - \bar{\mu}_{...})^2$

$\theta_{F_1*F_2} = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{\mu}_{ij.} - \bar{\mu}_{i..} - \bar{\mu}_{.j.} + \bar{\mu}_{...})^2$

$\theta_{F_1*F_3} = \frac{1}{(a-1)(c-1)}\sum_{i=1}^{a}\sum_{k=1}^{c}(\bar{\mu}_{i.k} - \bar{\mu}_{i..} - \bar{\mu}_{..k} + \bar{\mu}_{...})^2$

$\theta_{F_2*F_3} = \frac{1}{(b-1)(c-1)}\sum_{j=1}^{b}\sum_{k=1}^{c}(\bar{\mu}_{.jk} - \bar{\mu}_{.j.} - \bar{\mu}_{..k} + \bar{\mu}_{...})^2$

$\theta_{F_1*F_2*F_3} = \frac{1}{(a-1)(b-1)(c-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\mu_{ijk} - \bar{\mu}_{ij.} - \bar{\mu}_{i.k} - \bar{\mu}_{.jk} + \bar{\mu}_{i..} + \bar{\mu}_{.j.} + \bar{\mu}_{..k} - \bar{\mu}_{...})^2$

The cell means model:

$$y_{ijkl} = \mu_{ijk} + e_{ijkl}$$

can be written as an effects model with

$$\mu_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk}$$

There are $t = abc$ treatment means but there are $(a+1)(b+1)(c+1)$ parameters in the effects representation of the treatment means. Therefore, in order to obtain LSE of these parameters we need to place restrictions on the parameters. One of these types of restrictions: **Highest subscripted value is set equal to 0:**

1. $\tau_a = 0 \qquad \beta_b = 0 \qquad \gamma_c = 0$
2. $(\tau\beta)_{aj} = 0$ for all $j = 1, \cdots, b \qquad (\tau\beta)_{ib} = 0$ for all $i = 1, \cdots, a$
3. $(\tau\gamma)_{ak} = 0$ for all $k = 1, \cdots, c \qquad (\tau\gamma)_{ic} = 0$ for all $i = 1, \cdots, a$
4. $(\beta\gamma)_{bk} = 0$ for all $k = 1, \cdots, c \qquad (\beta\gamma)_{jc} = 0$ for all $j = 1, \cdots, b$
5. $(\tau\beta\gamma)_{ajk} = 0$ for all $j = 1, \cdots, b; k = 1, \cdots, c$
6. $(\tau\beta\gamma)_{ibk} = 0$ for all $i = 1, \cdots, a; k = 1, \cdots, c$
7. $(\tau\beta\gamma)_{ijc} = 0$ for all $i = 1, \cdots, a; j = 1, \cdots, b$

We thus have reduced the number of parameters to

1. μ
2. (a-1) τ_i's $\qquad$ (b-1) β_j's $\qquad$ (c-1) γ_k's
3. (a-1)(b-1) $(\tau\beta)_{ij}$'s $\qquad$ (a-1)(c-1) $(\tau\gamma)_{ik}$'s $\qquad$ (b-1)(c-1) $(\beta\gamma)_{jk}$'s
4. (a-1)(b-1)(c-1) $(\tau\beta\gamma)_{ijk}$'s

Under the restrictions, we have a total of

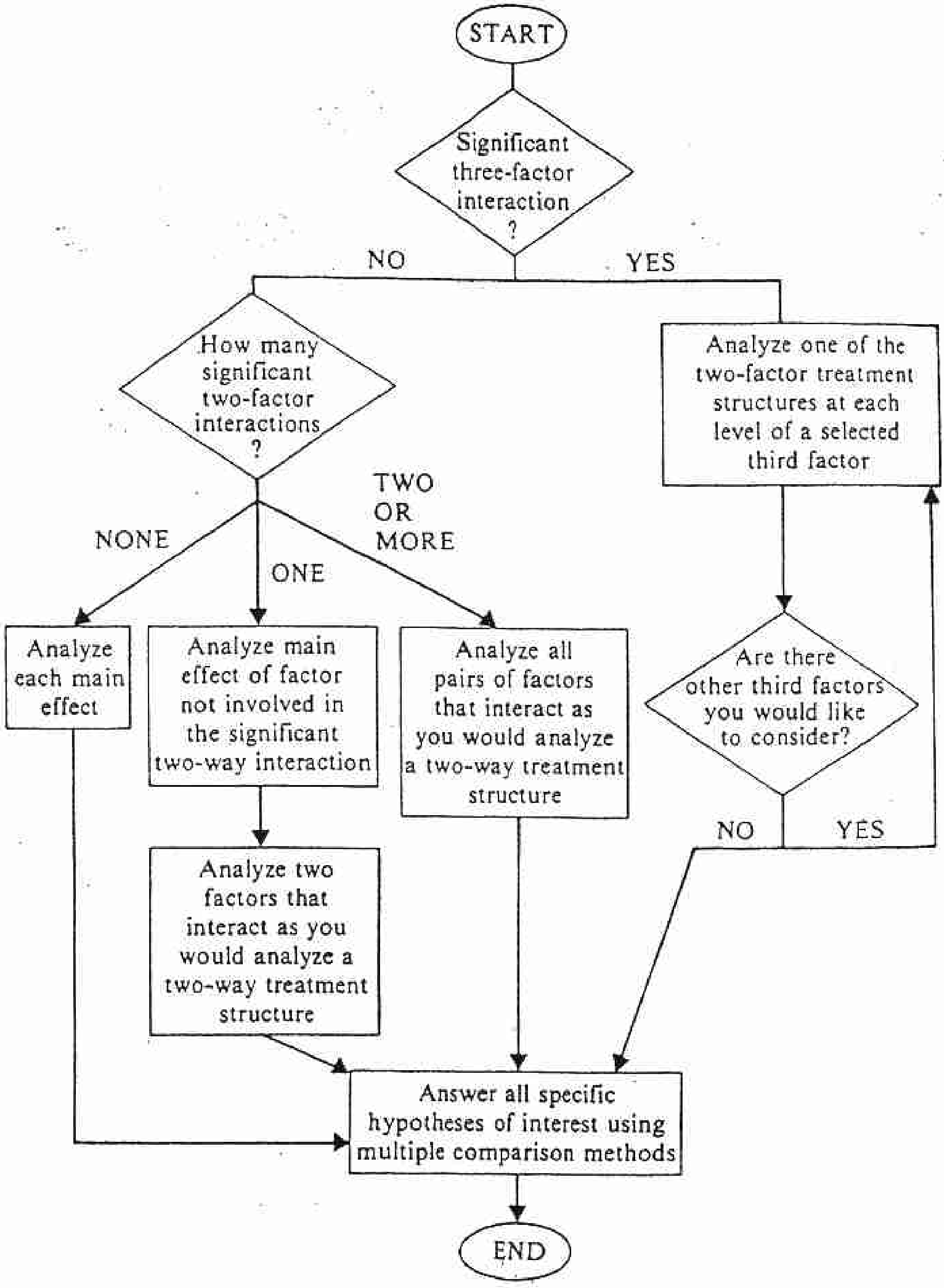
1+(a-1)+(b-1)+(c-1)+(a-1)(b-1)+(a-1)(c-1)+(b-1)(c-1)+(a-1)(b-1)(c-1)=abc

parameters, the same as the total number of treatment means.

The parameters in the effects model represent main effects, two-way interactions, and three-way interactions. However, in the unequal replication case, it is somewhat difficult to describe exactly what they represent.

The flow diagram on the next page will illustrate the manner in which factorial experiments should analyzed.

Flow Diagram from *Analysis of Messy Data*, Milliken & Johnson



Strategy for analyzing three-factor experiments.

Contrast for Main Effects and Interactions

In the following discussion, I will demonstrate contrasts which would represent main effects and interactions for a three factor experiment with

- Factor S with two levels S1, S2
- Factor F with two levels F1, F2
- Factor P with three levels P1, P2, P3

Next, consider the cell means model with 12 treatments:

$$S1F1P1, \quad S1F1P2, \quad S1F1P3, \quad S1F2P1, \quad S1F2P2, \quad S1F2P3,$$

$$S2F1P1, \quad S2F1P2, \quad S2F1P3, \quad S2F2P1, \quad S2F2P2, \quad S2F2P3$$

with μ_{ijk} being defined as the population average response of the treatment consisting of the ith level of S, jth level of F, and the kth level of P.

1. Main effect for factor P would be a contrast in $\bar{\mu}_{..1},\ \bar{\mu}_{..2},\ \bar{\mu}_{..3}$. For example,

- $C_1 = \bar{\mu}_{..1} - 2\bar{\mu}_{..2} + \bar{\mu}_{..3} =$

  $\frac{1}{4}(\mu_{111} + \mu_{121} + \mu_{211} + \mu_{221}) - 2(\frac{1}{4}(\mu_{112} + \mu_{122} + \mu_{212} + \mu_{222}) + \frac{1}{4}(\mu_{113} + \mu_{123} + \mu_{213} + \mu_{223})$

- $C_2 = \bar{\mu}_{..1} - \bar{\mu}_{..3} =$

  $\frac{1}{4}(\mu_{111} + \mu_{121} + \mu_{211} + \mu_{221}) - \frac{1}{4}(\mu_{113} + \mu_{123} + \mu_{213} + \mu_{223})$

2. A 2-way interaction between S and P would consists of writing a contrast in $\bar{\mu}_{i.k}$ across the levels of P at the first level of S and changing the signs of this contrast in P for the second level of S. For example,

- $C_3 = \bar{\mu}_{1.1} - 2\bar{\mu}_{1.2} + \bar{\mu}_{1.3} - \bar{\mu}_{2.1} + 2\bar{\mu}_{2.2} - \bar{\mu}_{2.3} =$

  $\frac{1}{2}(\mu_{111}+\mu_{121})-2(\frac{1}{2}(\mu_{112}+\mu_{122}))+\frac{1}{2}(\mu_{113}+\mu_{123})-\frac{1}{2}(\mu_{211}+\mu_{221})+2(\frac{1}{2}(\mu_{212}+\mu_{222}))-\frac{1}{2}(\mu_{213}+\mu_{223})$

  Alternatively, the contrast could be over the levels of S and then contrast this contrast over the levels of P. For example,

- $C_4 = (\bar{\mu}_{1.1} - \bar{\mu}_{2.1}) - 2(\bar{\mu}_{1.2} - \bar{\mu}_{2.2}) + (\bar{\mu}_{1.3} - \bar{\mu}_{2.3}) =$

  $[\frac{1}{2}(\mu_{111}+\mu_{121})-\frac{1}{2}(\mu_{211}+\mu_{221})]-2[\frac{1}{2}(\mu_{112}+\mu_{122})-(\frac{1}{2}(\mu_{212}+\mu_{222}))]+[\frac{1}{2}(\mu_{113}+\mu_{123})-\frac{1}{2}(\mu_{213}+\mu_{223})]$

3. A 3-way interaction would involve contrasting a 2-way interaction contrast for S and P across the levels of F. For example,

- $C_5 = (\mu_{111} - 2\mu_{112} + \mu_{113} - \mu_{211} + 2\mu_{212} - \mu_{213}) - (\mu_{121} - 2\mu_{122} + \mu_{123} - \mu_{221} + 2\mu_{222} - \mu_{223})$

We can represent these contrast in terms of the cell means, μ_{ijk}, as follows:

Contrast	μ_{111}	μ_{112}	μ_{113}	μ_{121}	μ_{122}	μ_{123}	μ_{211}	μ_{212}	μ_{213}	μ_{221}	μ_{222}	μ_{223}
C_1= Main P	1	-2	1	1	-2	1	1	-2	1	1	-2	1
C_2=Main P	1	0	-1	1	0	-1	1	0	-1	1	0	-1
C_3= S*P	1	-2	1	1	-2	1	-1	2	-1	-1	2	-1
C_4= S*P	1	-2	1	1	-2	1	-1	2	-1	-1	2	-1
C_5= S*F*P	1	-2	1	-1	2	-1	-1	2	-1	1	-2	1

Multiple Comparison in Three Factor Experiments

Case I. If the three-way interaction $F_1 * F_2 * F_3$ is significant then we need to examine the two-way interactions $F_1 * F_2$, $F_1 * F_3$, and $F_2 * F_3$ at each level of the factor not involved in the interaction. For example,

1. Examine using contrasts in the levels of (F_1, F_3) at each level of F_2:

$$C_{F_2=1} = (\mu_{111} - 2\mu_{112} + \mu_{113} - \mu_{211} + 2\mu_{212} - \mu_{213})$$
$$C_{F_2=2} = (\mu_{121} - 2\mu_{122} + \mu_{123} - \mu_{221} + 2\mu_{222} - \mu_{223})$$

2. Examine $F_1 * F_2$ at each level of F_3

    a. If the two-way interaction is not signification then examine each of the two main effects F_1 and F_2.

    b. If the two-way interaction is significant then examine the main effect of F_1 at each level of F_2 or vice versa.

3. Examine $F_1 * F_3$ at each level of F_2 (if of interest).

4. Examine $F_2 * F_3$ at each level of F_1 (if of interest).

Case II. If the three-way interaction $F_1 * F_2 * F_3$ is not significant then we next examine the three two-way interactions $F_1 * F_2$, $F_1 * F_3$, and $F_2 * F_3$.

a. If a two-way interaction, e.g., $F_1 * F_2$, is not signification then examine each of the two main effects F_1 and F_2.

b. If a two-way interaction, e.g., $F_1 * F_2$, is significant then examine the main effect of F_1 at each level of F_2 or vice versa.

CR 2x2x3 Factorial Experiment with r = 3 equal Reps

Example: Shrimp Culture in Aquaria (Kuehl 2000)

The California brown shrimp spawn at sea and the hatched eggs undergo larval transformation while being transported toward the shore. By the time they transform to postlarval stage they enter estuaries, where they grow rapidly into subadults and migrate back offshore as the approach sexual maturity. The shrimp encounter wide temperature and salinity variations in their life cycle as a result of their migrations. Thus, a knowledge of how temperatue and salinity affect their growth and survival is of great importance to understanding their life history and ecology. There was at the time of this experiment great interest in commercial culture of shrimp. From the standpoint of mariculture another important factor was stocking density in the culture tanks that affects intraspecific competition.

Research Hypothesis: The investigators wanted to know how water temperature, water salinity, and density of shrimp populations influenced the growth rate of shrimp raised in aquaria and whether the factors acted independently on the shrimp population.

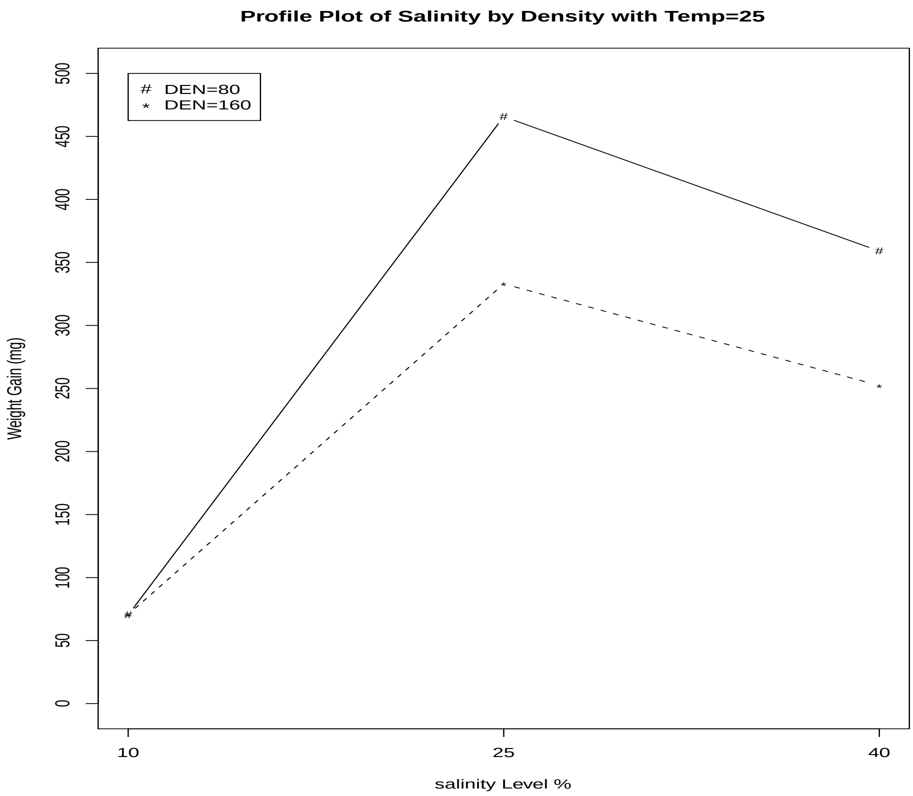
Treatment Design: A factorial arrangement was used with three factors: Temperature ($25°, 35°$); Salinity (10%,25%,40%); and Density of the shrimp in the aquarium (80 shrimp/40 liters, 160 shrimp/40 liters). The levels selected were those considered most likely to exhibit an effect if the factor was influential on shrimp growth.

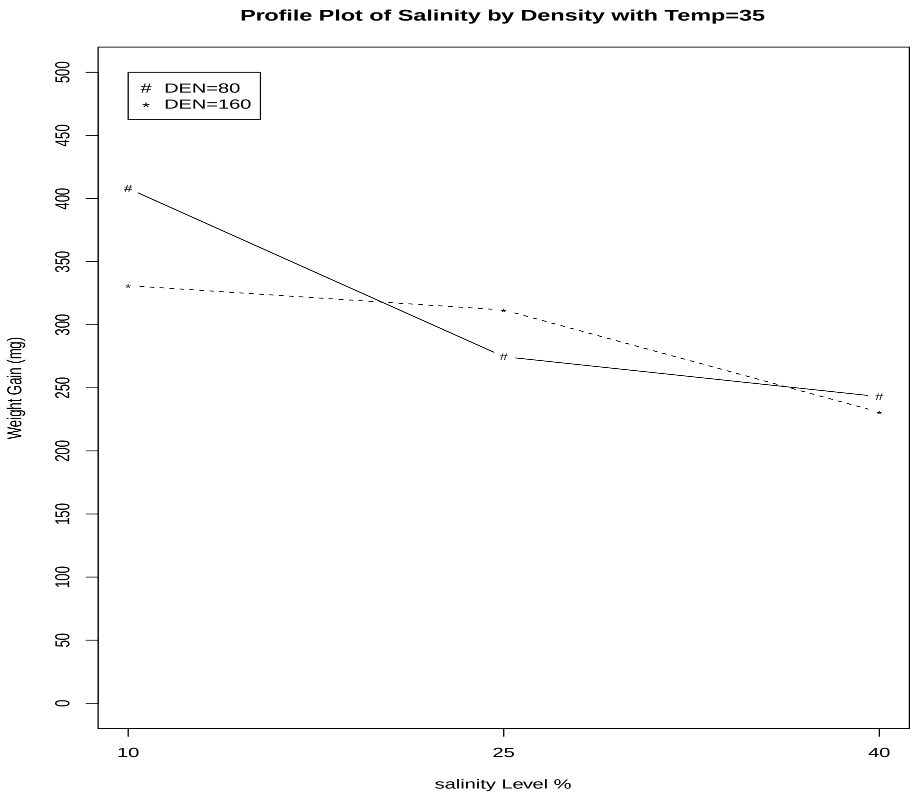
Experiment Design: The experiment design consisted of three replicate aquaria for each of the 12 treatment combinations of the 2x2x3 factorial. Each of the 12 treatment combinations was randomly assigned to three aquaria for a completely randomized design. The 36 aquaria were stocked with post-larval shrimp at the beginning of the experiment. The weight gain of the shrimp in four weeks for each of the 36 aquaria is shown in Table 1 on a per-shrimp basis.

Table 1. Four-week Weight Gain of Shrimp Cultured in Aquaria at Different Levels of Temperature (T), Water Salinity (S) and Density of Shrimp Populations (D)

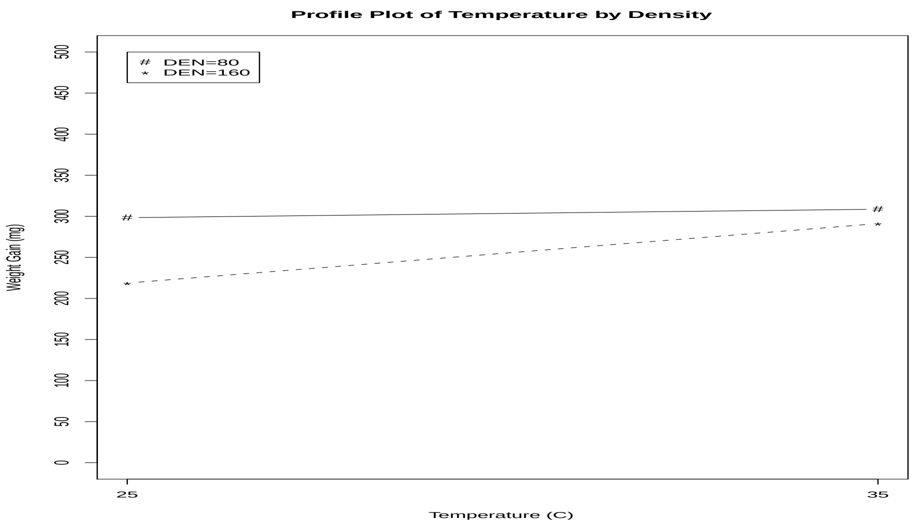
T	D	S	Weight Gain (mg)	$\bar{y}_{ijk.}$	T	D	$\bar{y}_{ij..}$	T	S	$\bar{y}_{i.k.}$	D	S	$\bar{y}_{.jk.}$
25°	80	10%	86,52,73	70.33	25	80	298.3	25	10	70.5	80	10	239.2
		25%	544,371,482	465.67	25	160	218.7	25	25	399.3	80	25	370.2
		40%	390,290,397	359	35	80	308.6	25	40	305.7	80	40	301.0
	160	10%	53,73,86	70.67	35	160	291.1	35	10	369.5	160	10	200.8
		25%	393,398,208	333				35	25	293.2	160	25	322.3
		40%	249,265,243	252.33				35	40	236.8	160	40	241.5
35°	80	10%	439,436,349	408									
		25%	249,245,330	274.67									
		40%	247,277,205	243									
	160	10%	324,305,364	331									
		25%	352,267,316	311.67									
		40%	188,223,281	230.67									

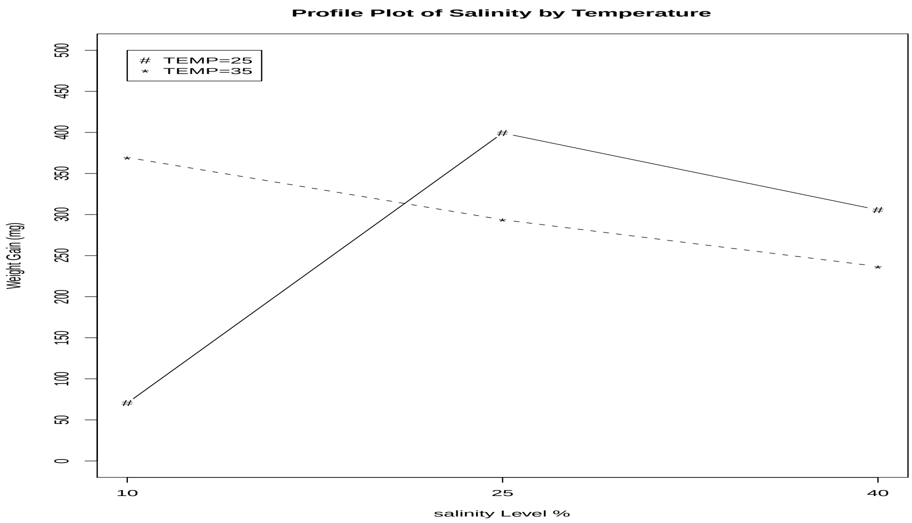
Profile Plot Depicting Three- Way Interaction

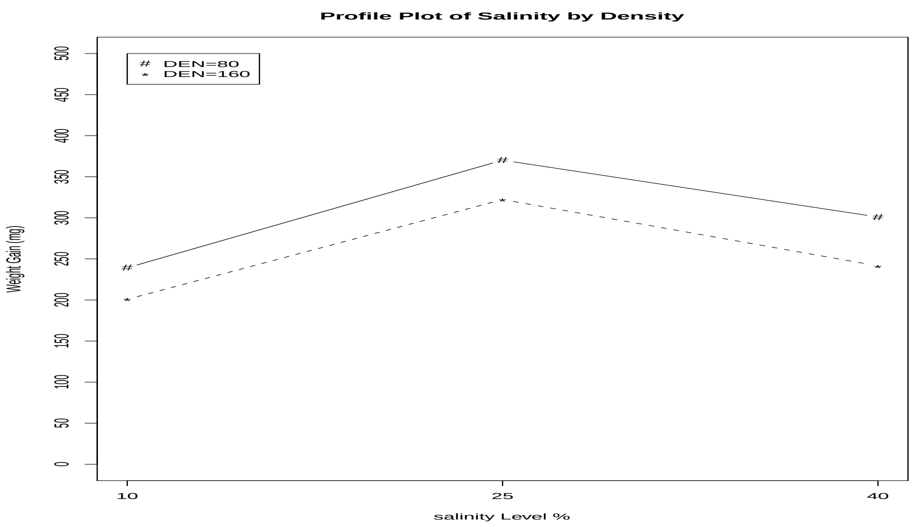




Profile Plots Depicting Two- Way Interactions







The SAS code and output for EXAMPLE 6.5: SHRIMP CULTURE IN AQUARIA experiment are given next.

```
* shrimp_3factors.sas;
ods html; ods graphics on;
option ls=75 ps=55 nocenter nodate;
OPTIONS FORMCHAR="|----|+|---+=|-/\<>*";
title 'Shrimp Cultures-3 Factor Experiment';

data shrimp;
array Y Y1-Y3; INPUT TRT $ T D  S  Y1-Y3;   do over Y;
WG=Y;
output; end; drop  Y1-Y3;
 label T = 'TEMPERATURE' D = 'DENSITY' S = 'SALINITY' WG = 'WEIGHT GAIN';
cards;
TRT01 25  80 10  86  52  73
TRT02 25  80 25 544 371 482
TRT03 25  80 40 390 290 397
TRT04 25 160 10  53  73  86
TRT05 25 160 25 393 398 208
TRT06 25 160 40 249 265 243
TRT07 35  80 10 439 436 349
TRT08 35  80 25 249 245 330
TRT09 35  80 40 247 277 205
TRT10 35 160 10 324 305 364
TRT11 35 160 25 352 267 316
TRT12 35 160 40 188 223 281
run;

proc glm;
class T D S;
model WG = T|D|S/SS3;
means S*D S*T T*D;
lsmeans  S*T*D/stderr pdiff adjust=tukey;
lsmeans  S*T*D/stderr pdiff;
contrast 'Effect Den' D 1 -1;
contrast 'SLinT25-SLinT35'   S*T 1  0 -1 -1  0  1;
contrast 'SQuadT25-SQuadT35' S*T 1 -2  1 -1  2 -1;
output out=ASSUMP r=RESID p=MEANS;
proc plot; plot resid*means;
proc univariate def=5 plot normal; var RESID;
run;

proc glm;
class trt;
model wg=trt;
means trt/hovtest=bf;

proc glimmix data=shrimp;
class T D S;
model WG = T|D|S;
lsmeans T*D*S /  plot = meanplot;
run;
ods graphics off; ods html close;
```

SAS Output from shrimp_3factors.SAS:

Class	Levels	Values
T	2	25 35
D	2	80 160
S	3	10 25 40

Number of observations in data set = 36

Dependent Variable: WG   WEIGHT GAIN

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	467636.333	42512.394	14.64	0.0001
Error	24	69690.667	2903.778		
Corrected Total	35	537327.000			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
T	1	15376.000	15376.000	5.30	0.0304
D	1	21218.778	21218.778	7.31	0.0124
T*D	1	8711.111	8711.111	3.00	0.0961
S	2	96762.500	48381.250	16.66	0.0001
T*S	2	300855.167	150427.583	51.80	0.0001
D*S	2	674.389	337.194	0.12	0.8909
T*D*S	2	24038.389	12019.194	4.14	0.0285

Dependent Variable: WG   WEIGHT GAIN

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
Effect Den	48.555556	2.70	0.0124	17.9622375
SLin25-SLin35	-367.833333	-8.36	0.0001	43.9983165
SQuad25-SQuad35	-442.500000	-5.81	0.0001	76.2073196

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Effect Den	1	21218.778	21218.778	7.31	0.0124
SLin25-SLin35	1	202952.042	202952.042	69.89	0.0001
SQuad25-SQuad35	1	97903.125	97903.125	33.72	0.0001

Least Squares Means

T	D	S	WG LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
25	80	10	70.333333	31.111508	0.0331	1
25	80	25	465.666667	31.111508	<.0001	2
25	80	40	359.000000	31.111508	<.0001	3
25	160	10	70.666667	31.111508	0.0324	4
25	160	25	333.000000	31.111508	<.0001	5
25	160	40	252.333333	31.111508	<.0001	6
35	80	10	408.000000	31.111508	<.0001	7
35	80	25	274.666667	31.111508	<.0001	8
35	80	40	243.000000	31.111508	<.0001	9
35	160	10	331.000000	31.111508	<.0001	10
35	160	25	311.666667	31.111508	<.0001	11
35	160	40	230.666667	31.111508	<.0001	12

Least Squares Means for effect T*D*S Adjustment for Multiple Comparison: Tukey

i/j	1	2	3	4	5	6	7	8	9	10	11	12
1		<.0001	<.0001	1.0000	0.0002	0.0152	<.0001	0.0046	0.0247	0.0002	0.0006	0.0460
2	<.0001		0.4277	<.0001	0.1643	0.0028	0.9695	0.0095	0.0017	0.1509	0.0626	0.0009
3	<.0001	0.4277		<.0001	1.0000	0.4277	0.9909	0.7388	0.3139	0.9999	0.9931	0.1964
4	1.0000	<.0001	<.0001		0.0002	0.0155	<.0001	0.0047	0.0252	0.0002	0.0006	0.0468
5	0.0002	0.1643	1.0000	0.0002		0.7854	0.8494	0.9670	0.6614	1.0000	1.0000	0.4863
6	0.0152	0.0028	0.4277	0.0155	0.7854		0.0578	1.0000	1.0000	0.8092	0.9629	1.0000
7	<.0001	0.9695	0.9909	<.0001	0.8494	0.0578		0.1597	0.0365	0.8280	0.5711	0.0194
8	0.0046	0.0095	0.7388	0.0047	0.9670	1.0000	0.1597		0.9998	0.9741	0.9992	0.9962
9	0.0247	0.0017	0.3139	0.0252	0.6614	1.0000	0.0365	0.9998		0.6893	0.9070	1.0000
10	0.0002	0.1509	0.9999	0.0002	1.0000	0.8092	0.8280	0.9741	0.6893		1.0000	0.5142
11	0.0006	0.0626	0.9931	0.0006	1.0000	0.9629	0.5711	0.9992	0.9070	1.0000		0.7813
12	0.0460	0.0009	0.1964	0.0468	0.4863	1.0000	0.0194	0.9962	1.0000	0.5142	0.7813	

Least Squares Means for effect T*D*S UnAdjusted for Multiple Comparisons

i/j	1	2	3	4	5	6	7	8	9	10	11	12
1		<.0001	<.0001	0.9940	<.0001	0.0004	<.0001	0.0001	0.0006	<.0001	<.0001	0.0013
2	<.0001		0.0232	<.0001	0.0060	<.0001	0.2024	0.0002	<.0001	0.0054	0.0018	<.0001
3	<.0001	0.0232		<.0001	0.5601	0.0232	0.2764	0.0673	0.0145	0.5305	0.2927	0.0076
4	0.9940	<.0001	<.0001		<.0001	0.0004	<.0001	0.0001	0.0006	<.0001	<.0001	0.0013
5	<.0001	0.0060	0.5601	<.0001		0.0792	0.1012	0.1974	0.0519	0.9641	0.6322	0.0288
6	0.0004	<.0001	0.0232	0.0004	0.0792		0.0017	0.6164	0.8338	0.0864	0.1901	0.6269
7	<.0001	0.2024	0.2764	<.0001	0.1012	0.0017		0.0058	0.0010	0.0929	0.0385	0.0005
8	0.0001	0.0002	0.0673	0.0001	0.1974	0.6164	0.0058		0.4786	0.2127	0.4087	0.3273
9	0.0006	<.0001	0.0145	0.0006	0.0519	0.8338	0.0010	0.4786		0.0569	0.1317	0.7816
10	<.0001	0.0054	0.5305	<.0001	0.9641	0.0864	0.0929	0.2127	0.0569		0.6643	0.0318
11	<.0001	0.0018	0.2927	<.0001	0.6322	0.1901	0.0385	0.4087	0.1317	0.6643		0.0780
12	0.0013	<.0001	0.0076	0.0013	0.0288	0.6269	0.0005	0.3273	0.7816	0.0318	0.0780	

Because the T*D*S interaction was significant,
The grouping of the levels of S will be done separately at each of the four combinations (T, G):

T = 25, D = 80
G1 = {10}, G2 = {25, 40}

T = 35, D = 80
G1 = {10, 25}, G2 = {25, 40}

T = 25, D = 160
G1 = {10}, G2 = {25, 40}

T = 35, D = 160
G1 = {10, 25, 40}

```
         Plot of RESID*MEANS.  Legend: A = 1 obs, B = 2 obs, etc.

RESID |
      |
  100 +
      |
      |
      |                                                              A
   75 +
      |
      |                                            B
      |                                     A
   50 +                               A
      |
      |                                          A     A
      |                                 A          A   A     A
   25 +                                                      A
      |          A                                                   A
      |          A                       A
      |                                 A        A
    0 +----------B-------------------------------------------------------------
      |                               A  B         A
      |
      |          B
  -25 +                                     A      A
      |                                     A
      |                                 A
      |                               A          A
  -50 +
      |                                                      A
      |
      |                                                A
  -75 +
      |
      |
      |                                                              A
 -100 +
      |
      |
      |
 -125 +                                            A
      |
      --+------------+------------+------------+------------+------------+-
        0           100          200          300          400          500
                                      MEANS
```

```
Variable=RESID

   Stem Leaf                     #  Boxplot
      6 058                      3     |
      4 005                      3     |
      2 811348                   6  +-----+
      0 23443566                 8  *--+--*
     -0 889873                   6  |     |
     -2 8066                     4  +-----+
     -4 953                      3     |
     -6 9                        1     |
     -8 5                        1     |
    -10
    -12 5                        1     0
        ----+----+----+----+
    Multiply Stem.Leaf by 10**+1


                       Normal Probability Plot
      70+                                       +**+   *
        |                                  +****
        |                              *+***
        |                        ******
        |                    ****+
     -30+                ****+
        |            *+***
        |        ++*+
        |   +++++*
        |+++
    -130+    *
         +----+----+----+----+----+----+----+----+----+----+
             -2        -1         0        +1        +2

                   Tests for Normality

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

Shapiro-Wilk          W     0.969147    Pr < W      0.4027
Kolmogorov-Smirnov    D     0.083827    Pr > D     >0.1500
Cramer-von Mises      W-Sq  0.041607    Pr > W-Sq  >0.2500
Anderson-Darling      A-Sq  0.296413    Pr > A-Sq  >0.2500


  Brown and Forsythe's Test for Homogeneity of WG Variance
       ANOVA of Absolute Deviations from Group Medians

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

TRT           11      9805.6       891.4       0.44    0.9233
Error         24     48977.3      2040.7
```

An examination of the residuals reveals that the conditions of equal variance and normality do not appear to be violated. There is no natural spatial or time ordering in the data so we have no valid method to assess correlation in the data. However, we would need to discuss such issues with the experimenter.

From the profile plots, there appears to be a 3-Way interaction since the pattern in the 2-Way interaction between Salinity and Density at TEMP=25 is quite different from the pattern at TEMP=35. An examination of the AOV table confirms this conclusion. The test for a significant 3-Way interaction has p-value=0.0285. Any further inferences about the effects of Salinity, Density, or Temperature will be made conditionally. For example, if we wanted to determine the best combination of Density and Salinity, we would have to do the analysis twice, once at TEMP=25 and again at TEMP=35.

If we use the LSMEANS output, we can compare levels of SALINITY and DENSITY combinations using a Bonferroni $\alpha = \frac{.05}{12} = .0042$ which yields:

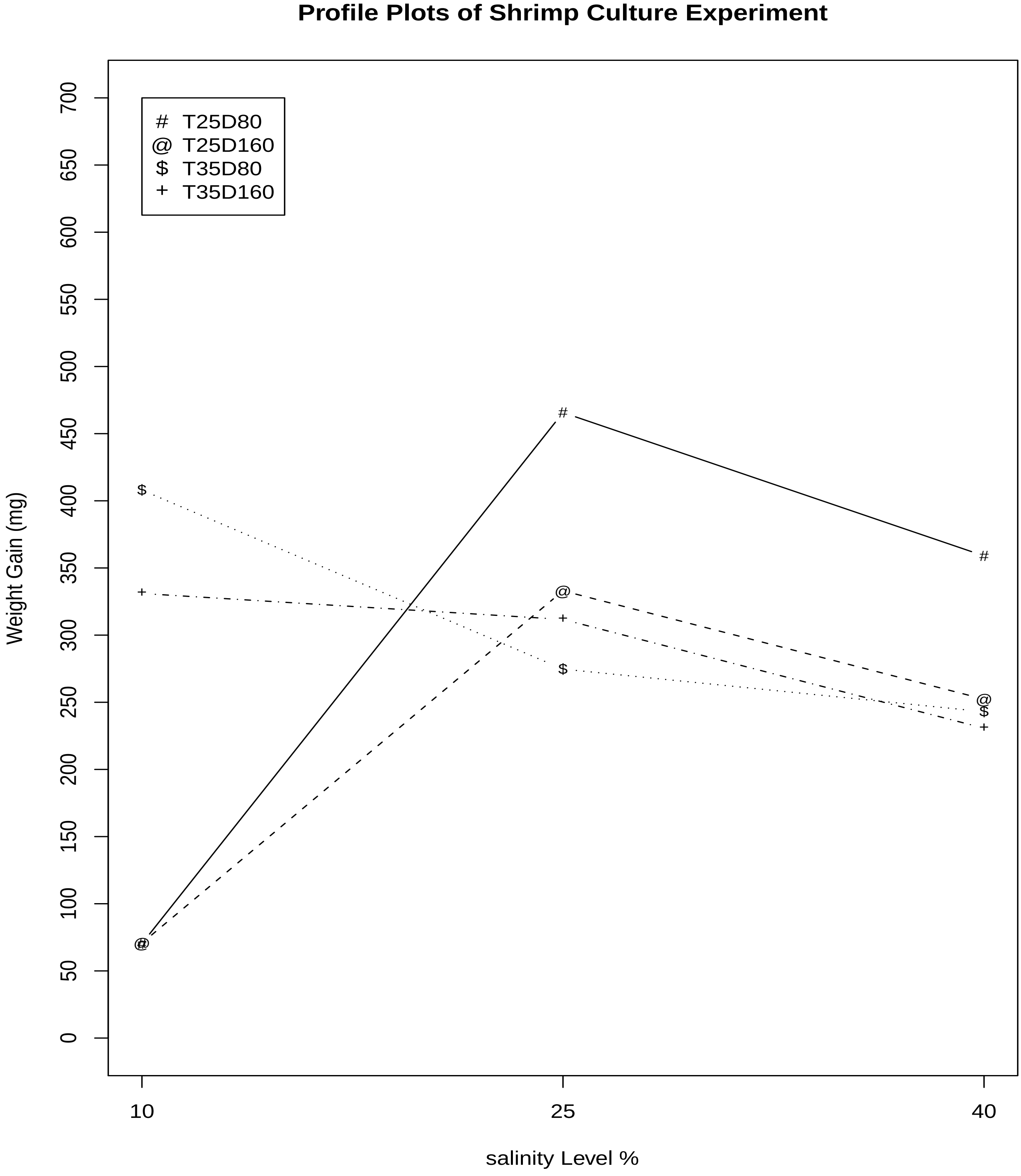
At TEMP=25, (S,D)=(25%,80) and (40%,80) have the highest mean weight gain and they are not significantly different.

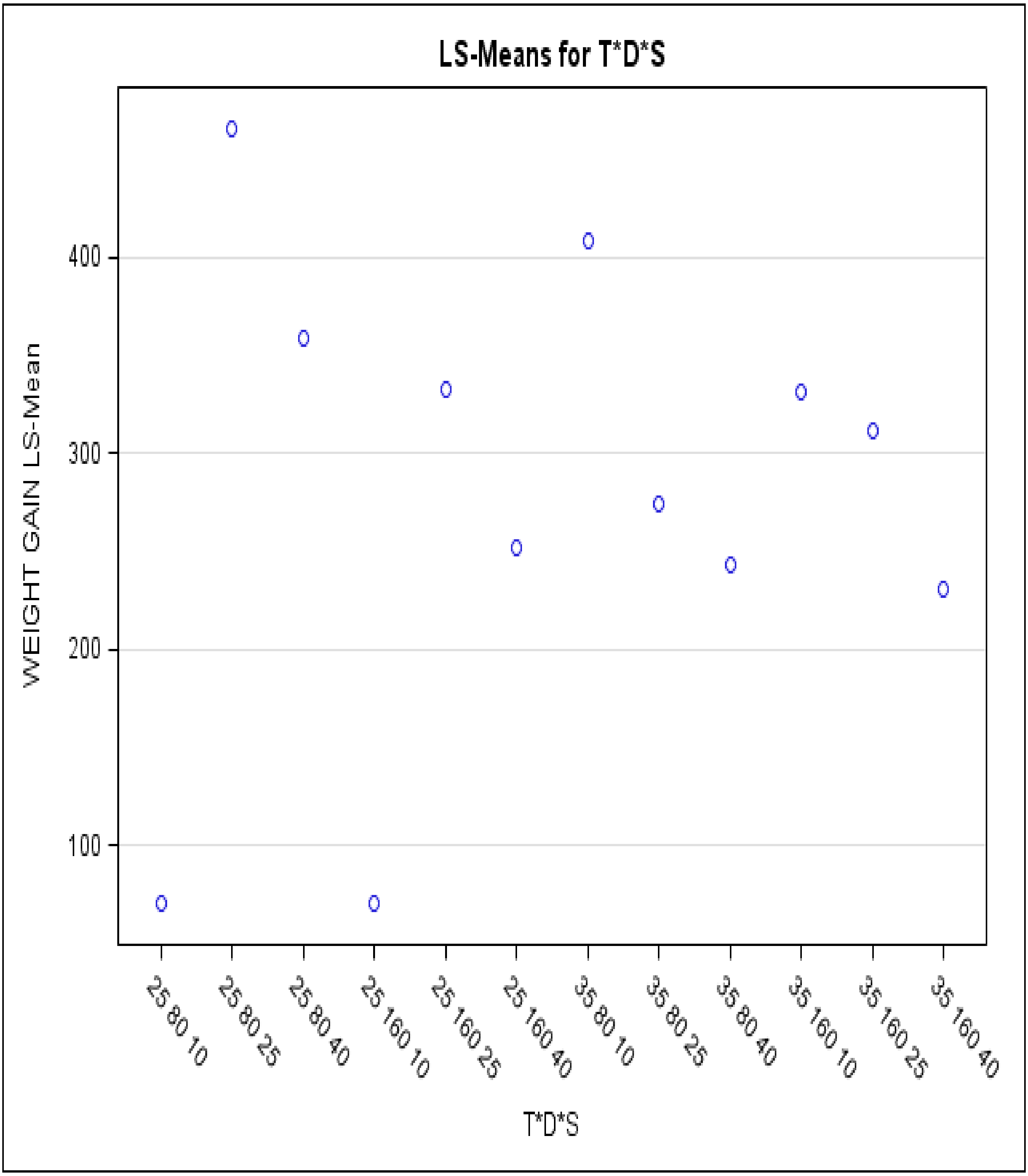
At TEMP=35, (S,D)=(10%,80), (10%,160) and (25%,160) have the highest mean weight gain and they are not significantly different.

Thus, we observe that conclusions about levels of Salinity and Density depend on the Temperature.

Further note, that when we have a significant 3-Way interaction, the interpretation of the tests for significant 2-Way interactions and Main Effects may not be meaningful. From the AOV table the test for a significant 2-Way interaction between DENSITY and SALINITY is 0.8909 which would indicate that differences in Mean Weight Gain at the two levels of DENSITY are consistent across the 3 levels of SALINITY. However, we know this is not true since at TEMP=25, there is essentially no difference in the means at the two DENSITY levels at a SALINITY level of 10% but there are large differences at SALINITY levels of 25% and 40%. Furthermore, at TEMP=35, this pattern is completely reversed with a large difference in the means of the two density levels at a SALINITY level of 10% but very small differences at SALINITY levels of 25% and 40%. This demonstrates how conclusions about Main Effects and 2-Way interactions are inconsistent when there is a significant 3-Way interaction.

If we want to test for trends in mean weight gains across the levels of Salinity, we would have to construct the the Linear and Quadratic contrasts across the levels of Salinity separately at the four combinations of Density and Temperature.


Profile Plots of Shrimp Culture Experiment
T25D80
@ T25D160
$ T35D80
+ T35D160
Weight Gain (mg)
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
10
25
40
salinity Level %


LS-Means for T*D*S
WEIGHT GAIN LS-Mean
400
300
200
100
25 80 10
25 80 25
25 80 40
25 160 10
25 160 25
25 160 40
35 80 10
35 80 25
35 80 40
35 160 10
35 160 25
35 160 40
T*D*S

The SAS code **shrimpcontrasts.sas** is used to test for the significance of the trends in mean weight gain across the levels of salinity are given next:

```
* shrimpcontrasts.sas;
option ls=75 ps=55 nocenter nodate;
title 'Shrimp Cultures-3 Factor Exp';
data shrimp;
array Y Y1-Y3; INPUT T D  S TRT $ Y1-Y3;   do over Y;
WG=Y;
output; end;
      drop  Y1-Y3;
      label T = 'TEMPERATURE' D = 'DENSITY' S = 'SALINITY'
            WG = 'WEIGHT GAIN';
cards;
25  80 10 T1D1S1   86  52  73
25  80 25 T1D1S2  544 371 482
25  80 40 T1D1S3  390 290 397
25 160 10 T1D2S1   53  73  86
25 160 25 T1D2S2  393 398 208
25 160 40 T1D2S3  249 265 243
35  80 10 T2D1S1  439 436 349
35  80 25 T2D1S2  249 245 330
35  80 40 T2D1S3  247 277 205
35 160 10 T2D2S1  324 305 364
35 160 25 T2D2S2  352 267 316
35 160 40 T2D2S3  188 223 281
run;
PROC GLM; CLASS TRT;
MODEL WG = TRT;
ESTIMATE 'SLIN2580'  TRT  -1  0 1  0  0 0  0  0 0  0  0 0;
ESTIMATE 'SLIN25160' TRT   0  0 0 -1  0 1  0  0 0  0  0 0;
ESTIMATE 'SLIN3580'  TRT   0  0 0  0  0 0 -1  0 1  0  0 0;
ESTIMATE 'SLIN35160' TRT   0  0 0  0  0 0  0  0 0 -1  0 1;
ESTIMATE 'SQUAD2580'  TRT  1 -2 1  0  0 0  0  0 0  0  0 0;
ESTIMATE 'SQUAD25160' TRT  0  0 0  1 -2 1  0  0 0  0  0 0;
ESTIMATE 'SQUAD3580'  TRT  0  0 0  0  0 0  1 -2 1  0  0 0;
ESTIMATE 'SQUAD35160' TRT  0  0 0  0  0 0  0  0 0  1 -2 1;
run;
```

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

LINEAR SALINITY-T=25,D=80       288.666667   43.9983165       6.56      <.0001
LINEAR SALINITY-T=25,D=160      181.666667   43.9983165       4.13      0.0004
LINEAR SALINITY-T=35,D=80      -165.000000   43.9983165      -3.75      0.0010
LINEAR SALINITY-T=35,D=160     -100.333333   43.9983165      -2.28      0.0318
QUAD SALINITY-T=25,D=80        -502.000000   76.2073196      -6.59      <.0001
QUAD SALINITY-T=25,D=160       -343.000000   76.2073196      -4.50      0.0001
QUAD SALINITY-T=35,D=80         101.666667   76.2073196       1.33      0.1947
QUAD SALINITY-T=35,D=160        -61.666667   76.2073196      -0.81      0.4264
```

Finding the Best Treatment

Suppose the researcher was interested in determining which combination of (T,D,S) produced shrimp having the largest average weight gain. If there was no interactions we could apply Hsu's procedure individual to each of the three factors and find the best level of T, best level of D, and best level of S based on the marginal means: $\bar{y}_{i...}, \bar{y}_{.j..}$, and $\bar{y}_{..k.}$. However, in the above experiment due to significant evidence of a three way interaction between T, D, and S, it will be necessary to consider the 12 treatments which have the combined levels of the three factors.

Hsu's Procedure for finding group of treatment producing largest average weight gain:

1. $t = abc = (2)(2)(3) = 12$ treatments $\Rightarrow k = t - 1 = 11$
2. $r = 3$ Reps per treatment $\Rightarrow \nu = df_{MSE} = abc(r-1) = 24$
3. From Table VI with $\alpha = .05$ 1-sided, we have $d_{\alpha,k,\nu} = 2.64$
4. $M = d_{\alpha,k,\nu}\sqrt{\frac{2MSE}{r}} = 2.64\sqrt{\frac{2(2903.778)}{3}} = 116.16$
5. $D_{ijk} = \bar{y}_{ijk} - max\{\bar{y}_{lmn} : l \neq i, m \neq j, n \neq k\}$
6. Declare Treatment (T_i, D_j, S_k) in Group of Best Treatments if $D_{ijk} + M > 0$

T	D	S	WG LSMEAN	D_{ijk}	$D_{ijk} + M$	In Best Group?
25	80	10	70.33	-395.34	-279.18	No
25	80	25	465.67	57.67	116.168	Yes
25	80	40	359.00	-106.67	-279.18	No
25	160	10	70.67	-395.00	-278.84	No
25	160	25	333.00	-132.67	-16.51	No
25	160	40	252.33	-213.33	-97.18	No
35	80	10	408.00	-57.67	58.49	Yes
35	80	25	274.67	-191.00	-74.84	No
35	80	40	243.00	-222.67	-106.51	No
35	160	10	331.00	-134.67	-18.51	No
35	160	25	311.67	-154.00	-37.84	No
35	160	40	230.67	-235.00	-118.84	No

With 95% confidence, the treatments yielding the "Best Average Weight Gains" are

$(Temp = 25°,\ Density = 80,\ Salinity = 25\%)$ and

$(Temp = 35°,\ Density = 80,\ Salinity = 10\%)$