

START Friday 3/11/12 (Week 8, Sec 22)

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

- **Supplemental Reading:** Design & ANOVA Book - Ch. 6 & 7
-

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.

Calcium Level	pH level				Mean
	4	5	6	7	
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

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

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

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 $t = ab = 12$ populations having normal distributions with equal variances

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

i th level of Factor F_1 and j th level of Factor F_2 ;

Thus, we have 12 populations, each having a normal distribution and the same variance but possibly different means.

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: $1 \rightarrow \mu$, $a \rightarrow \tau_i s$, $b \rightarrow \beta_j s$, $ab \rightarrow (\tau\beta)_{ij} s \Rightarrow 1 + a + b + ab = (a+1)(b+1)$ parameters in the model

1. Constraints on parameters:

2. $\tau_a = 0; \beta_b = 0;$

3. $(\tau\beta)_{ib} = 0$ for $i = 1, \dots, a;$

4. $(\tau\beta)_{aj} = 0$; for $j = 1, \dots, b.$

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 $i \neq a; \mu_{ib} = \mu + \tau_i + \beta_b + (\tau\beta)_{ib} = \mu + \tau_i + 0 + 0 \Rightarrow \tau_i = \mu_{ib} - \mu_{ab};$

3. For $j \neq b, \mu_{aj} = \mu + \tau_a + \beta_j + (\tau\beta)_{aj} = \mu + 0 + \beta_j + 0 \Rightarrow \beta_j = \mu_{aj} - \mu_{ab}$

4. For $i \neq a, j \neq b, \mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$

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

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

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 fixed levels of the second Factor.

Fix factor F_1 at i th 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 : $\bar{\mu}_{1..}, \bar{\mu}_{2..}, \dots, \bar{\mu}_{a..}$

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

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 : $\bar{\mu}_{1\cdot} - \bar{\mu}_{2\cdot}$

Main Effects of Factor F_2 :

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

Are the Main Effect Contrasts for Factor F_2 orthogonal?

NO

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. $(\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22}) = d_1 - d_2$ $\cancel{||} = 0 \Rightarrow \text{NO Interaction}$
- b. $(\mu_{12} - \mu_{22}) - (\mu_{13} - \mu_{23}) = d_2 - d_3$ $\cancel{||}$
- c. $(\mu_{13} - \mu_{23}) - (\mu_{11} - \mu_{21}) = d_3 - d_1$ $\cancel{||}$
- 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 .

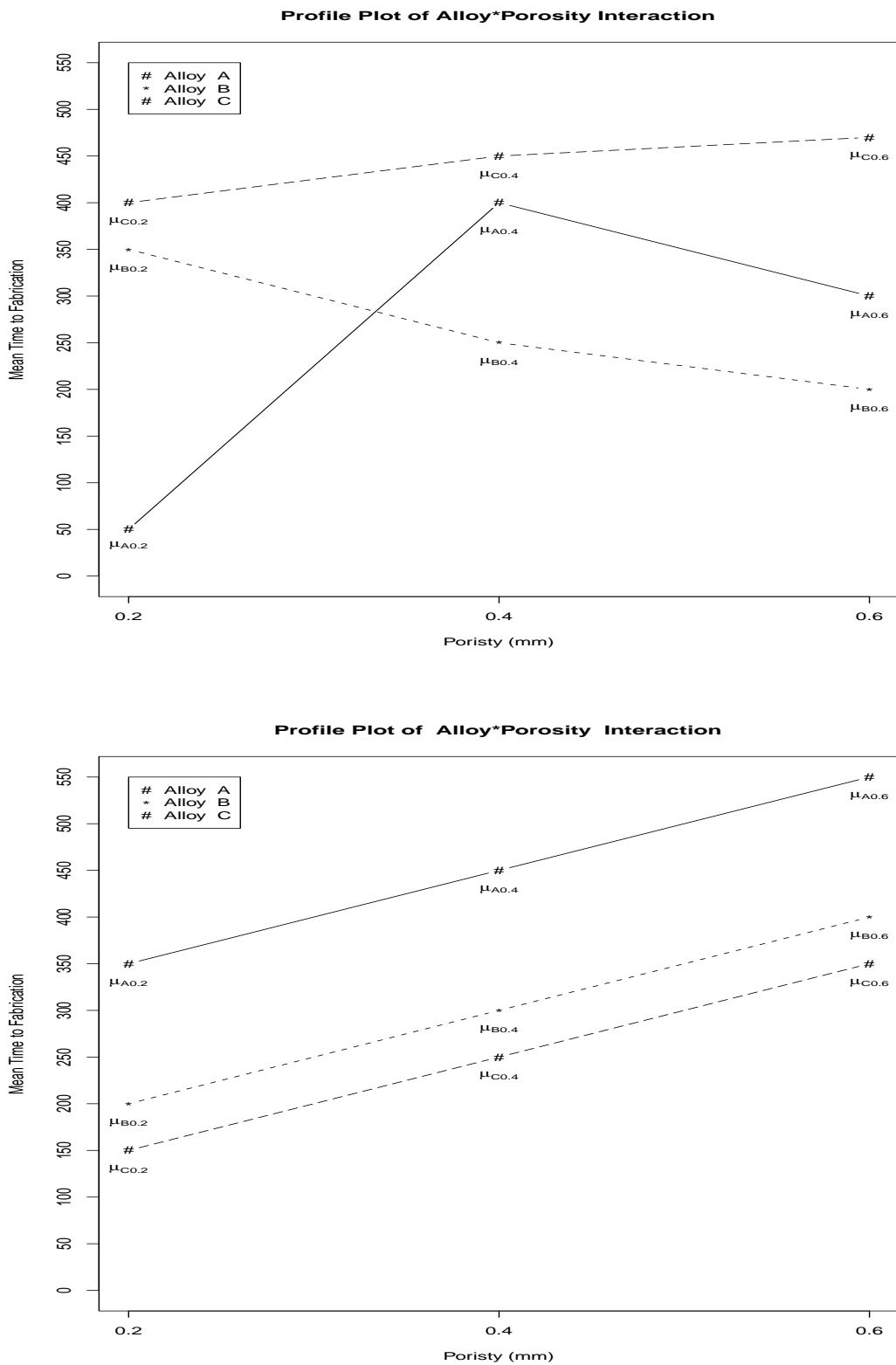
There is an interaction between factors F_1 and F_2 if

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

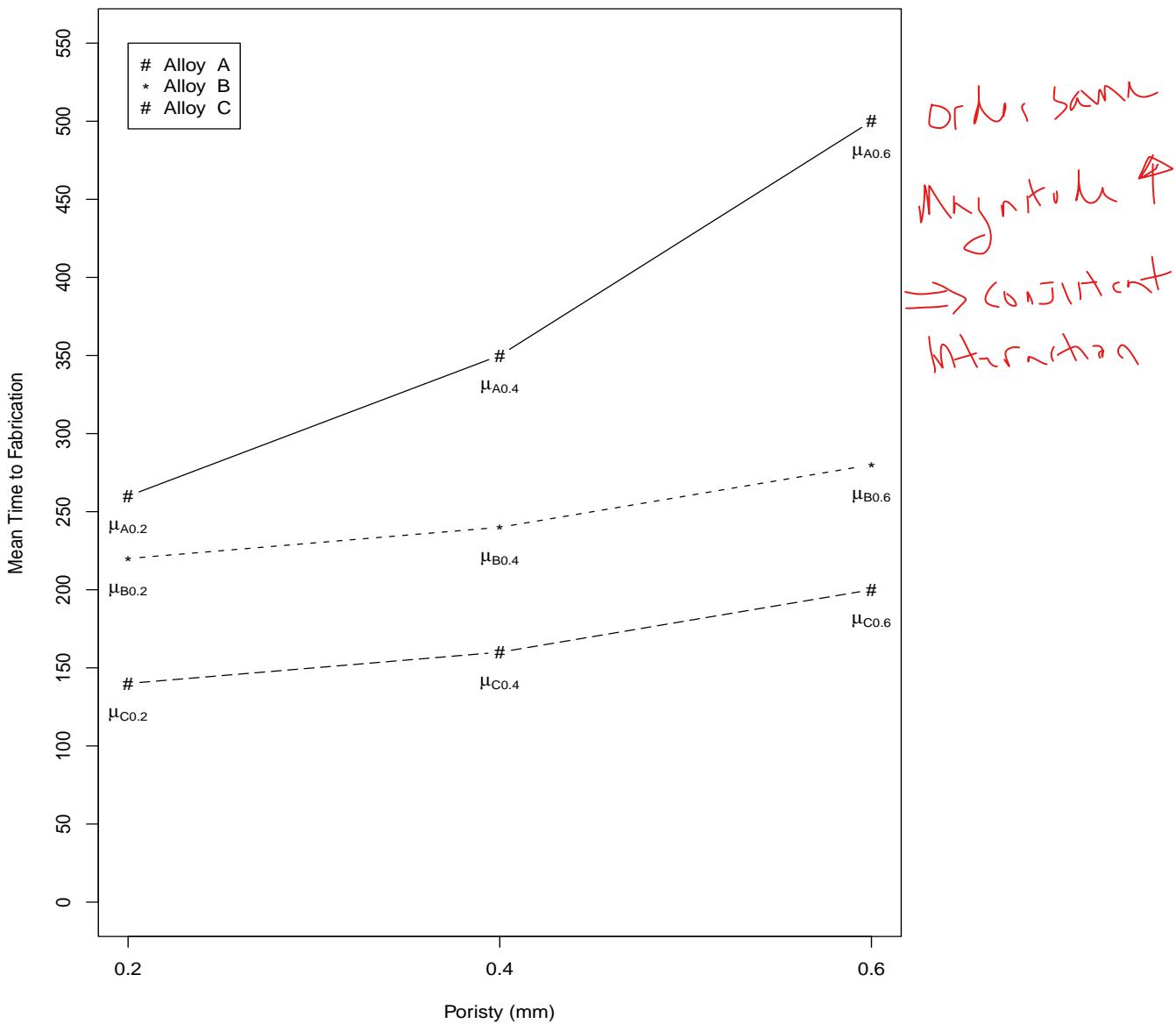
In the plot of the treatment means, the **profile plot**, if $d_1 = d_2 = d_3$, both in magnitude and sign, then Factors F_1 and F_2 do NOT interact.

If at least one of the d_i differs from the remaining d_j s 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



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: $y_{ijk} = \mu_{ij} + e_{ijk}; \quad i = 1, \dots, a; \quad j = 1, \dots, b; \quad t = ab;$

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

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

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

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

The design matrix can also be written as:

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

A third way that design matrix can be expressed:

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

The parameters in the matrix model are the treatment means:

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

Regression Model Formulation

Let the ab treatments be labeled as

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

The design matrix can also be obtained by considering $t = ab$ “explanatory variables” $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t)$, as is in regression models, with \mathbf{X}_i given by

$$X_{ij} = \begin{cases} 1 & \text{if } j\text{th EU receives Trt\#}i \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = 1, \dots, n; i = 1, \dots, t$$

Regression model: $y_j = \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_t X_{tj} + e_j$, where $j = 1, 2, \dots, n$ 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 $(\mathbf{X}' \mathbf{X})$ is t . Thus, $(\mathbf{X}' \mathbf{X})^{-1}$ exists. The least squares estimators of the model parameters are obtained from the equation:

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

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

Note that

- a. $\hat{\mu}_{ij} = \bar{y}_{ij.}$ whether or not the number of reps are equal
- b. $\hat{\mu}_{i.} = \bar{Y}_{i..}$ only if $n_{ij} = r$
- c. $\hat{\mu}_{.j} = \bar{Y}_{.j..}$ only if $n_{ij} = r$
- d. $\hat{\mu}_{..} = \bar{Y}_{...}$ 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: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$,

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

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 12 \\ 24 \\ 18 \\ 30 \\ 6 \\ 36 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 6 \\ 10 \\ 2 \\ 12 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_{11} \\ \hat{\mu}_{12} \\ \hat{\mu}_{21} \\ \hat{\mu}_{22} \\ \hat{\mu}_{31} \\ \hat{\mu}_{32} \end{bmatrix}$$

Design Matrix for the Effects Model

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

where τ_i is the fixed effect of i th level of F1, β_j is the fixed effect of j th 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 μ , τ_1 , τ_2 , β_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: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$,

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

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

Overall Treatment Differences Hypotheses:

- (a) Cell Means Model: $H_0 : \mu_{ij} = \mu$ for all (i, j) vs $H_1 : \mu_{ij}$ not all equal
- (b) Effects Model: $H_0 : \tau_i = 0, \beta_j = 0, (\tau\beta)_{ij} = 0$ 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: $H_0 : (\mu_{ij} - \mu_{kj}) = (\mu_{ih} - \mu_{kh})$ for all (i, k, j, h) in the set $\{(i, k, j, h) : (i \neq k = 1, \dots, a; j \neq h = 1, \dots, b)\}$
vs $H_1 : (\mu_{ij} - \mu_{kj}) \neq (\mu_{ih} - \mu_{kh})$ for at least one set (i, k, j, h)
- (b) Effects Model: $H_0 : (\tau\beta)_{ij} = 0$ 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: $H_0 : \bar{\mu}_{1.} = \bar{\mu}_{2.} = \dots = \bar{\mu}_{a.}$ vs $H_1 : \bar{\mu}_{i.}$ not all equal
- (b) Effects Model: $H_0 : \tau_i = 0$ for all $(i = 1, \dots, a)$ 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: $H_0 : \bar{\mu}_{.1} = \bar{\mu}_{.2} = \dots = \bar{\mu}_{.b}$ vs $H_1 : \bar{\mu}_{.j}$ not all equal
- (b) Effects Model: $H_0 : \beta_j = 0$ for all $(j = 1, \dots, b)$ vs $H_1 : \beta_j$ not all 0

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

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} , SS_{F_2} , and $SS_{F_1*F_2}$

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

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

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

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

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

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

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

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

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

$$= SS_{Trt} - SS_{F_1} - SS_{F_2}$$

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

$$SS_E = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r \hat{e}_{ijk}^2$$

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

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

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

$$= SS_{TOT} - SS_{TRT} \quad \text{with } df = \sum_{i=1}^a \sum_{j=1}^b (r - 1) = ab(r - 1) = n - ab$$

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

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

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

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,

$$df_{EFFECT} = (\text{Number of model parameters}) - (\text{Number of model parameters under } H_o)$$

$$= ab - N_{H_o}$$

Overall Treatment Differences Hypotheses:

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

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

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

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

Interactions between the Factors:

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

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

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

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

Under H_o , for $j = 1, \dots, b$, $\mu_{2j} - \mu_{1j} = C_1$; $\mu_{3j} - \mu_{2j} = C_2; \dots; \mu_{aj} - \mu_{a-1,j} = C_{a-1}$

Given $\mu_{11}, \mu_{12}, \dots, \mu_{1b}$ and C_1, C_2, \dots, C_{a-1}

$$\mu_{ij} = \mu_{1j} + (\mu_{2j} - \mu_{1j}) + (\mu_{3j} - \mu_{2j}) + \dots + (\mu_{ij} - \mu_{i-1,j}) = \mu_{1j} + C_1 + C_2 + \dots + C_{i-1}$$

Under H_o , there are only $b + a - 1$ parameters needed to identify μ_{ij}

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

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

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

Main Effect of Factor F_1 :

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

To test: $H_o : \bar{\mu}_{1\cdot} = \bar{\mu}_{2\cdot} = \dots = \bar{\mu}_{a\cdot} = C \Rightarrow$

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

Given C and $\mu_{i1}, \mu_{i2}, \dots, \mu_{ib-1}$ for each $i = 1, 2, \dots, a$

$$\bar{\mu}_{i\cdot} = \frac{1}{b} \sum_{k=1}^b \mu_{ik} = C \Rightarrow \mu_{ij} + \sum_{k \neq j}^b \mu_{ik} = bC \Rightarrow$$

$$\mu_{ij} = bC - \sum_{k \neq j}^b \mu_{ik}$$

Thus under H_o , only need $1 + a(b-1)$ parameters to describe μ_{ij}

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

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

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

Main Effect of Factor F_2 :

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

To test: $H_o : \bar{\mu}_{1\cdot} = \bar{\mu}_{2\cdot} = \dots = \bar{\mu}_{b\cdot} \Rightarrow$

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

Same argument as for the degrees of freedom for Main Effect of Factor F_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 $(j = 1, \dots, b) \Rightarrow$

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

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

STOP Friday 3/11/22 (week 8, lecture 22)

START Monday 3/21/22 (Week 9, Lecture 22)

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 $\text{df}=\nu_1, \nu_2$; $\nu_1=\text{df of MS in numerator}$ and $\nu_2=\text{df of MS in denominator}$

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 $b = 4$:

$$\bar{y}_{...} = 7.067 \quad \bar{y}_{11} = 5.8 \quad \bar{y}_{12} = 7.3 \quad \bar{y}_{13} = 7.4 \quad \bar{y}_{14} = 7.3 \quad \bar{y}_{21} = 7.4 \quad \bar{y}_{22} = 7.3 \quad \bar{y}_{23} = 7.6$$

$$\bar{y}_{24} = 7.1 \quad \bar{y}_{31} = 6.4 \quad \bar{y}_{32} = 7.4 \quad \bar{y}_{33} = 7.2 \quad \bar{y}_{34} = 6.6$$

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

$$1. \quad SS_{TOT} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (y_{ijk} - \bar{y}_{...})^2 = (5.2 - 7.067)^2 + (5.9 - 7.067)^2 + \dots + (6.4 - 7.067)^2 \\ = 10.88$$

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

average of treatment (ij) over the r replicates.

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

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

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

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

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

$$= SS_{TRT} - SS_{Ca} - SS_{pH} = 9.200004 - 1.460004 - 4.24161 = 3.49839$$

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}
<i>Ca</i>	2	1.460	.7300	10.43	0.00055
<i>pH</i>	3	4.242	1.414	20.20	9.41×10^{-7}
<i>Ca * pH</i>	6	3.498	.5833	8.33	6.08×10^{-5}
Error	24	1.680	.0700	.	.
Total	35	10.88	.	.	.

$\text{df}_1 = 2$
 $\text{df}_2 = 24$
 $\text{df}_3 = 3$
 $\text{df}_{\text{error}} = 24$
 SEC level
 are Spec'd.

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

$$1 - \text{pf}(11.95, 11, 24) = 0.000000314$$

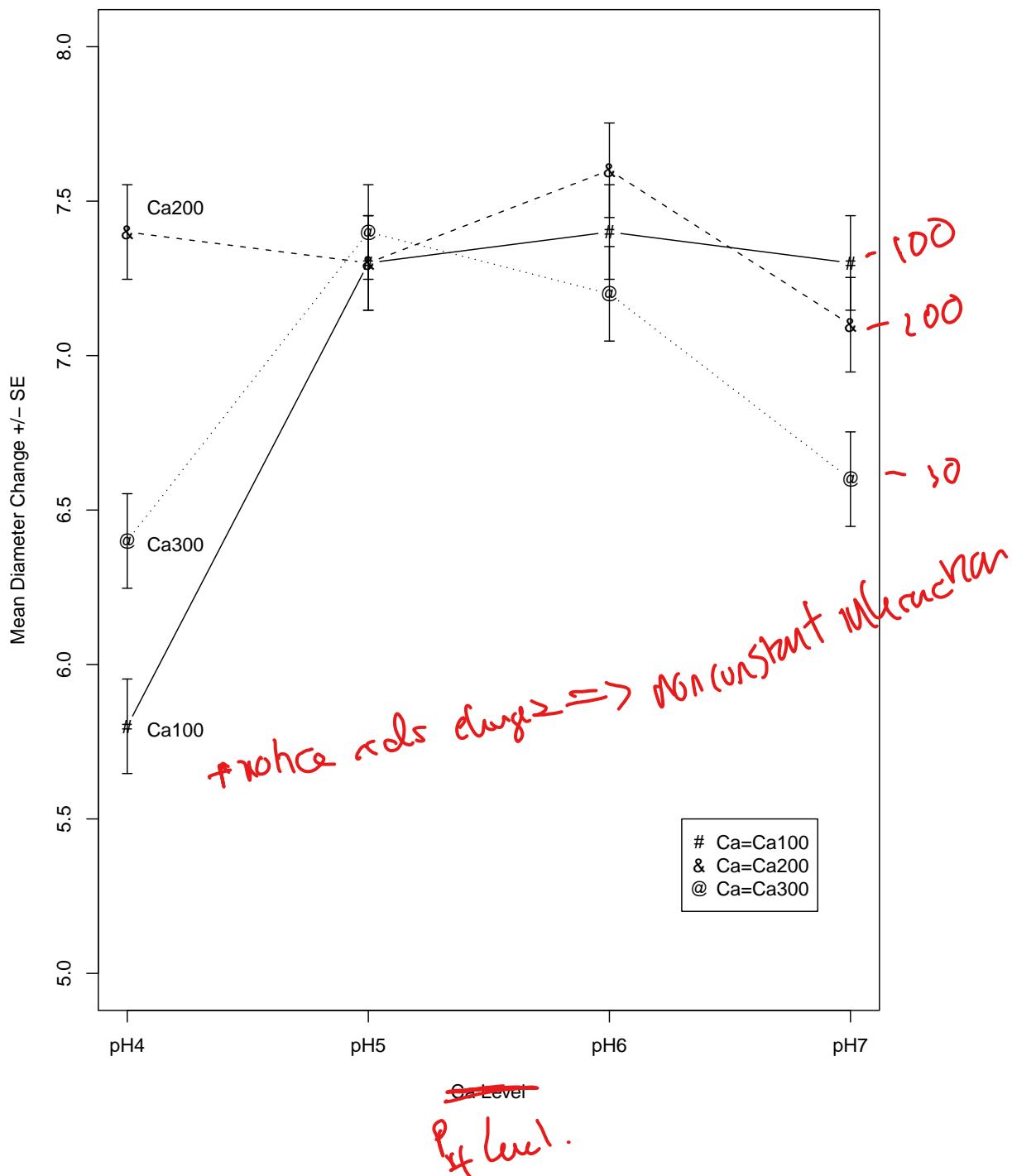
The following R Code, **Profile_withbars.R**, was used to produce a profile plot of sample means:

```

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



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

Case 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 $\bar{\mu}_{1..}, \bar{\mu}_{2..}, \dots, \bar{\mu}_{a..}$ and on $\bar{\mu}_{.1}, \bar{\mu}_{.2}, \dots, \bar{\mu}_{.b..}$

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 F_1 :

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

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

In SAS, use **LSMEANS F1/ADJUST=TUKEY** to obtain pairwise comparisons of F_1

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

Case 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 $\mu_{1j}, \mu_{2j}, \dots, \mu_{aj}$ separately for each value $j = 1, \dots, b$ or

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

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

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

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

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

- (b) Contrasts in $\mu_{1j}, \mu_{2j}, \dots, \mu_{aj}$, 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. We can then form 11 mutually orthogonal contrasts in the 12 treatment means:

Effect	Contrast	Treatments									
		A1B1	A1B2	A1B3	A1B4	A2B1	A2B2	A2B3	A2B4	A3B1	A3B2
Main A	C1	-1	-1	-1	-1	0	0	0	0	1	1
Main A	C2	1	1	1	1	-2	-2	-2	-2	1	1
Main B	C3	-3	-1	1	3	-3	-1	1	3	-3	-1
Main B	C4	1	-1	-1	1	1	-1	-1	1	1	-1
Main B	C5	-1	3	-3	1	-1	3	-3	1	-1	3
A*B	C6=C1*C3	3	1	-1	-3	0	0	0	0	-3	-1
A*B	C7=C1*C4	-1	1	1	-1	0	0	0	0	1	-1
A*B	C8=C1*C5	1	-3	3	-1	0	0	0	0	-1	3
A*B	C9=C2*C3	-3	-1	1	3	6	2	-2	-6	-3	-1
A*B	C10=C2*C4	1	-1	-1	1	-2	2	2	-2	1	-1
A*B	C11=C2*C5	-1	3	-3	1	2	-6	6	-2	-1	3

Note:

- Two contrasts for Main A are $C_1 = \bar{\mu}_3 - \bar{\mu}_1$, and $C_2 = \bar{\mu}_1 - 2\bar{\mu}_2 + \bar{\mu}_3$.
- Three contrasts for Main B are $C_3 = -3\bar{\mu}_{.1} - \bar{\mu}_{.2} + \bar{\mu}_{.3} + 3\bar{\mu}_{.4}$, and $C_5 = -\bar{\mu}_{.1} + 3\bar{\mu}_{.2} - 3\bar{\mu}_{.3} + \bar{\mu}_{.4}$
- These contrasts are not unique. There are many different choices for the contrasts. ★
- The 6 contrasts for the A*B interaction are obtained by multiplying the corresponding coefficients from the Main of A by Main of B.
- The resulting 11 contrasts are mutually orthogonal.
- The 6 interaction contrasts are contrasts in the 4 levels of Factor B then contrasted over the 3 levels of Factor A. For example, $c_1 = (-1, 0, 1)$ $\Rightarrow c_6 = c_1 + c_3$

$$C_6 = -1*(-3\mu_{11} - 1\mu_{12} + 1\mu_{13} + 3\mu_{14}) + 0*(-3\mu_{21} - 1\mu_{22} + 1\mu_{23} + 3\mu_{24}) + 1*(-3\mu_{31} - 1\mu_{32} + 1\mu_{33} + 3\mu_{34})$$

$$C_9 = 1*(-3\mu_{11} - 1\mu_{12} + 1\mu_{13} + 3\mu_{14}) - 2*(-3\mu_{21} - 1\mu_{22} + 1\mu_{23} + 3\mu_{24}) + 1*(-3\mu_{31} - 1\mu_{32} + 1\mu_{33} + 3\mu_{34})$$

- 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 $SS_C = \frac{\hat{C}^2}{\sum_{i=1}^t k_i^2/r}$

See Tables for Exams-Homework in Review Materials for Exams folder in eCampus.

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

X_j	$t=3$		$t=4$			$t=5$				$t=6$					$t=7$						
	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	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	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{7}{20}$	$\frac{77}{60}$	
<hr/>																					
X_j		$t=8$					$t=9$					$t=10$					<hr/>				
P_1		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_7	P_8	P_9	P_10	P_11	P_12	P_13	
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			
$\sum_{j=1}^{10} \{P_i(X_j)\}^2$		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. I, 1966, edited by E. S. Pearson and H. O. Hartley, with permission of the Biometrika Trustees.

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"; X
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 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 1;
contrast 'Quad Trend in Ca' TRT 1 1 1 1 -2 -2 -2 -2 1 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; — run this so we can see this
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;

```

— look at homogeneity of variances

— look at residuals to see if normality is satisfied.

— run this so we can see this

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

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
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	<.0001
Linear Trend in pH	1	0.96800000	0.96800000	13.83	<.0001
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	<.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	<.0001
INTER5	1	0.72000000	0.72000000	10.29	<.00038
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

Model is significant (Scree 2% 19)

(Interaction significant, so we don't look at these)

- II (contrast)
 Mult var
 $\alpha_{pc} = \frac{\alpha_{\alpha}}{II}$
 $\alpha_{\alpha} = 0.05$
 $\alpha_{pc} = 0.0015$

$\alpha = \text{significant}$

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

- Don't look at this
b/c our interaction is significant
(see case 2 pg 1A)

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	

- Don't look at this b/c interaction is significant
(see case 2 pg 1A)

p	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	

- only look at if we want to normalize or minimize our response we can look at this. But there is no comparison here.

Ca	H	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

max diameter achieved w/ Ca=200, pH=6

- * but we are more interested in comparing them b/c we want to be able to talk about it. for each level of Ca what are the differences in the responses for each level of pH.
~ mean

$$\alpha_{PC} = \frac{0.05}{18} \text{ p.t. levels}$$

$$3(\text{levels})^2 = 18$$

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	0.0002	0.2499	<.0001	<.0001	0.0404		
2	<.0001		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	0.9980	0.9549	0.0048	1.0000	0.9980	0.0404	
4	<.0001	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	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	

look at unadjusted p-vls

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

$$\alpha_{PC} = \frac{0.05}{18} = 0.002778$$

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	

$$\mu_1 = \{1, 2\}$$

$$\mu_2 = \{3, 4, 5, 6, 7\}$$

$$\mu_3 = \{8, 9, 10, 11, 12\}$$

$$\alpha = 0.05$$

$$\alpha = 0.05$$

$$i_1 = \{4, 5, 6, 7\}$$

$$i_2 = \{1, 2, 3\}$$

$$i_3 = \{8, 9, 10\}$$

$$i_4 = \{6, 7, 11\}$$

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

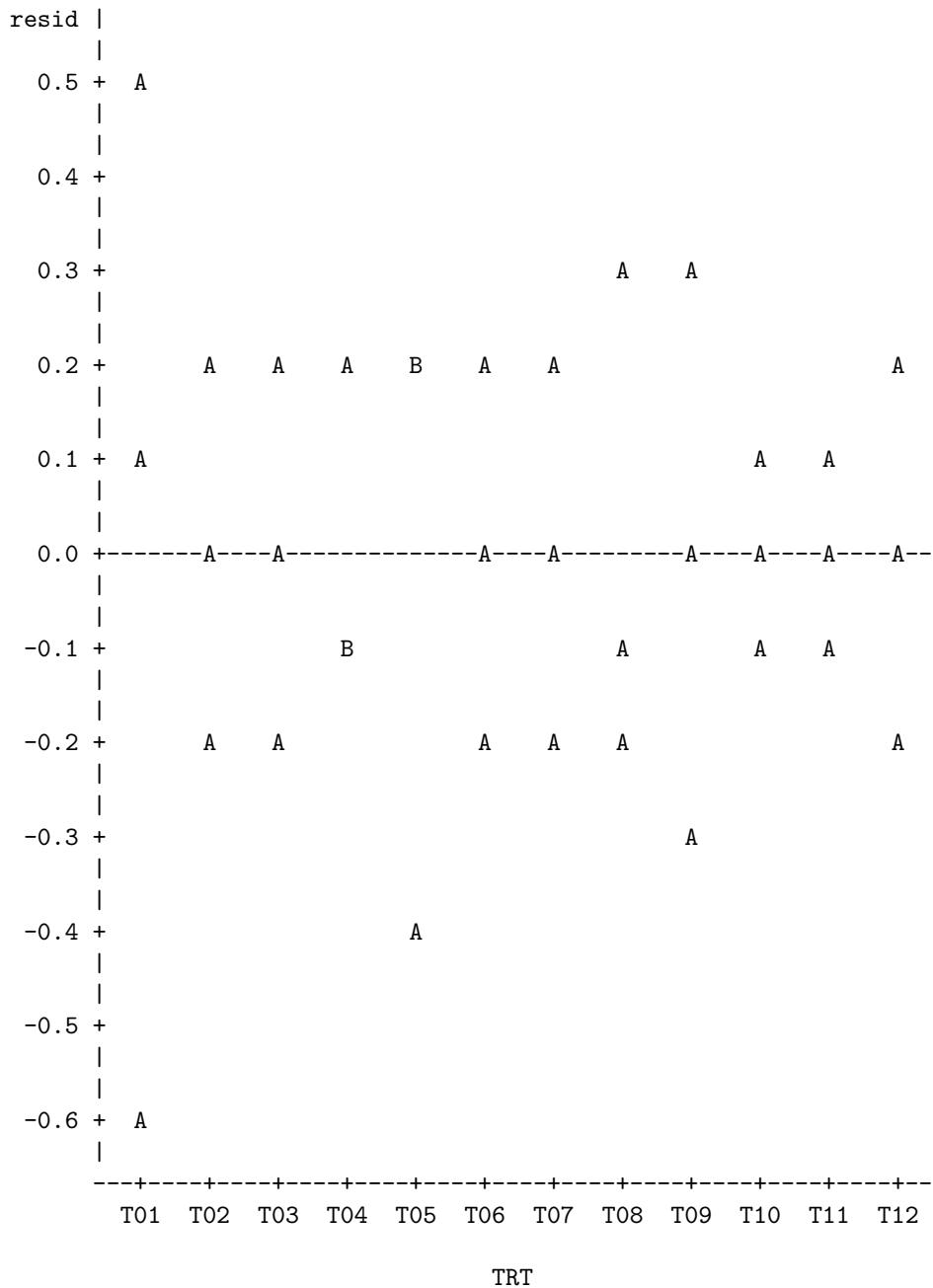
Brown and Forsythe's Test for Homogeneity of y Variance ANOVA of Absolute Deviations from Group Medians

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

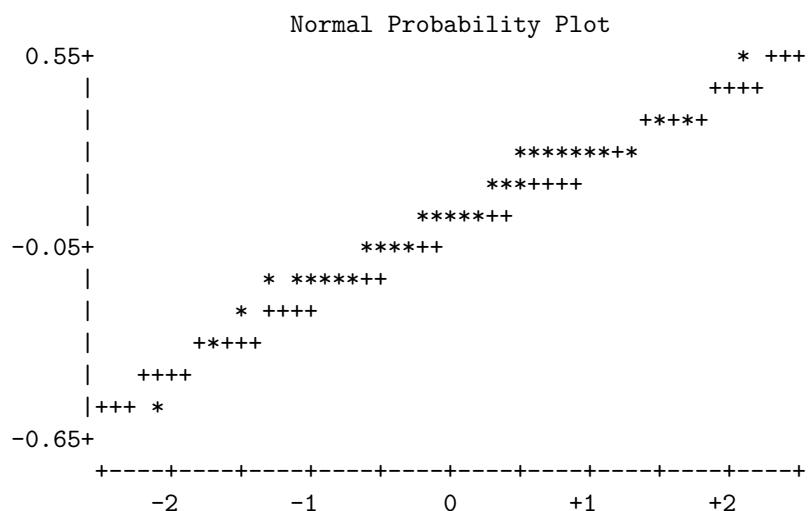
Plot of resid*TRT. Legend: A = 1 obs, B = 2 obs, etc.



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



TRT Least Squares Means

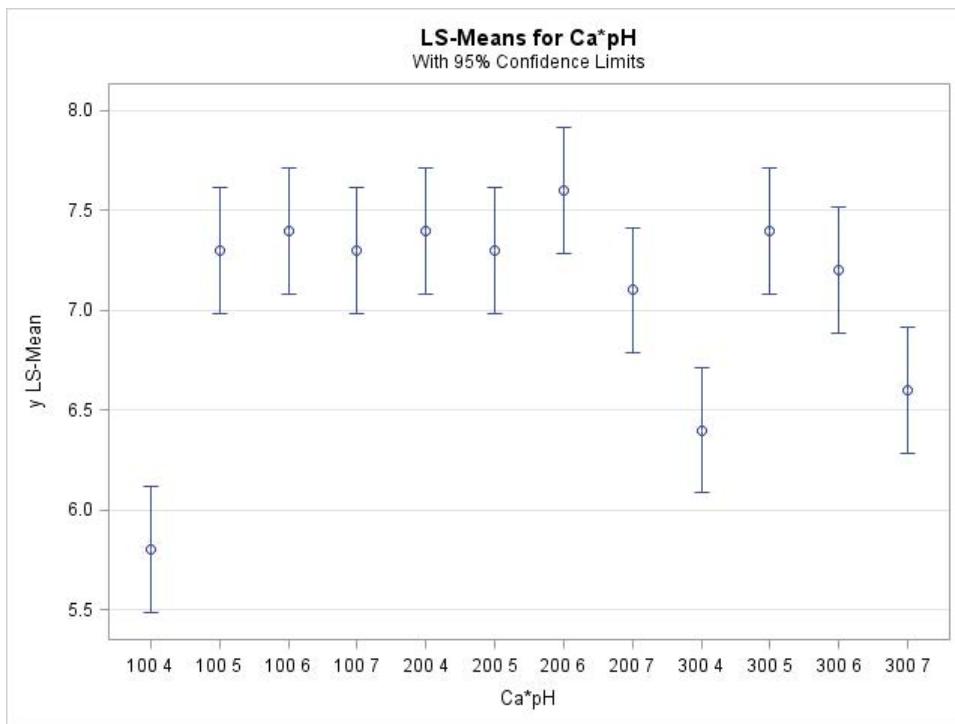
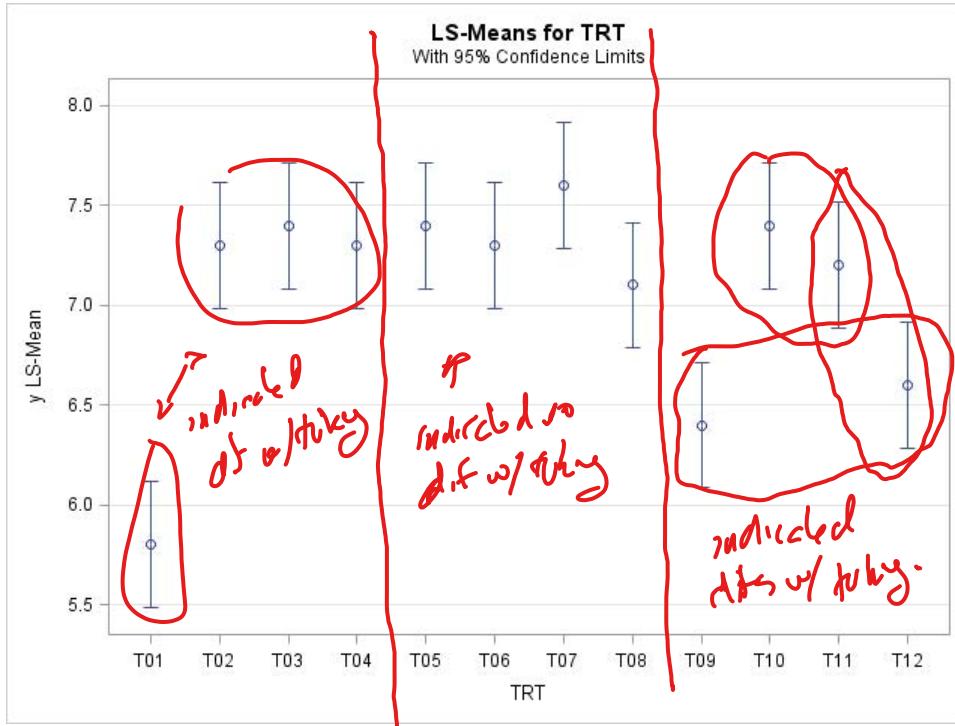
TRT	Estimate	Standard		DF	t Value	Pr > t	Alpha	Lower	Upper
		Error							
T01	5.8000	0.1528		24	37.97	<.0001	0.05	5.4847	6.1153
T02	7.3000	0.1528		24	47.79	<.0001	0.05	6.9847	7.6153
T03	7.4000	0.1528		24	48.44	<.0001	0.05	7.0847	7.7153
T04	7.3000	0.1528		24	47.79	<.0001	0.05	6.9847	7.6153
T05	7.4000	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:



The following R code - **factorialCapH.R** - will provide a similar analysis as the SAS output:

```

y = c(5.2, 5.9, 6.3, 7.1, 7.3, 7.5, 7.6, 7.2, 7.4, 7.2, 7.5, 7.2,
     7.6, 7.0, 7.6, 7.5, 7.3, 7.1, 7.6, 7.4, 7.8, 7.4, 7.0, 6.9,
     6.4, 6.7, 6.1, 7.3, 7.5, 7.4, 7.2, 7.3, 7.1, 6.8, 6.6, 6.4)
Calcium =c(rep("100",12), rep("200",12), rep("300",12))
Ca = factor(Calcium)
ph = c(rep("4",3),rep("5",3),rep("6",3),rep("7",3),
       rep("4",3),rep("5",3),rep("6",3),rep("7",3),
       rep("4",3),rep("5",3),rep("6",3),rep("7",3))
pH = factor(ph)
CapH = c(rep("100-4",3), rep("100-5",3), rep("100-6",3), rep("100-7",3),
         rep("200-4",3), rep("200-5",3), rep("200-6",3), rep("200-7",3),
         rep("300-4",3), rep("300-5",3), rep("300-6",3), rep("300-7",3))
CapH = factor(CapH)
rawE = cbind(y,Ca,pH)
rawE = factor(rawE)
data_Effects = data.frame(rawE)
rawC = cbind(y,CapH)
rawC = factor(rawC)
data_CellMs = data.frame(rawC)

#Effects Model
modelE = lm(y~pH*Ca,data=data_Effects)
summary(aov(modelE))

#Cell Means Model
modelC = lm(y~CapH,data=data_CellMs)
summary(aov(modelC))

#Contrast matrix for linear, quad, and cubic trends

trendmatrix = matrix(c(-1, -1, -1, -1, 0, 0, 0, 0, 1, 1, 1, 1,
                       1, 1, 1, 1, -2, -2, -2, -2, 1, 1, 1, 1,
                       -3, -1, 1, 3, -3, -1, 1, 3, -3,-1, 1, 3,
                       1, -1, -1, 1, 1, -1, -1, 1, 1,-1,-1, 1,
                       -1, 3, -3, 1, -1, 3, -3, 1, -1, 3,-3, 1),12,5)

trendmatrix = t(trendmatrix)

rownames(trendmatrix) = c("Linear Trend Ca", "Quad Trend Ca", "Linear Trend pH", "Quad Trend pH", "Cubic Trend pH")

fit.contrast(modelC, "CapH", trendmatrix)

#Pairwise Comparisons of Ca and pH means:

lsmeans(modelC,"CapH")

#Comparisons with Unadjusted p-values
pairwise.t.test(y,CapH,p.adjust="none")

#Pairwise comparisons yielding Tukey Adjusted p-values
TukeyHSD(aov(y~CapH,data=data_CellMs))

Output from R code:

```

```

> summary(aov(modelE))
      Df Sum Sq Mean Sq F value    Pr(>F)
pH       3   4.24   1.413   20.19 9.4e-07 ***
Ca       2   1.46   0.730   10.43 0.00055 ***
pH:Ca    6   3.50   0.583    8.33 6.1e-05 ***
Residuals 24   1.68   0.070

> summary(aov(modelC))
      Df Sum Sq Mean Sq F value    Pr(>F)
CapH     11   9.20   0.836   11.9 3.1e-07 ***
Residuals 24   1.68   0.070

> fit.contrast(modelC, "CapH", trendmatrix)
            Estimate Std. Error t value Pr(>|t|)
CapHLinear Trend Ca     -0.2      0.432 -0.463 6.48e-01
CapHQuad Trend Ca      -3.4      0.748 -4.543 1.33e-04
CapHLinear Trend pH      4.4      1.183  3.719 1.07e-03
CapHQuad Trend pH      -3.6      0.529 -6.803 4.90e-07
CapHCubic Trend pH      0.8      1.183  0.676 5.05e-01

> pairwise.t.test(y,CapH,p.adjust="none")
Pairwise comparisons using t tests with pooled SD

data: y and CapH

  100-4 100-5 100-6 100-7 200-4 200-5 200-6 200-7 300-4 300-5 300-6
100-5 4e-07 - - - - - - - - - - - -
100-6 1e-07 0.648 - - - - - - - - - - - -
100-7 4e-07 1.000 0.648 - - - - - - - - - -
200-4 1e-07 0.648 1.000 0.648 - - - - - - - -
200-5 4e-07 1.000 0.648 1.000 0.648 - - - - - - - -
200-6 2e-08 0.178 0.364 0.178 0.364 0.178 - - - - - - - -
200-7 3e-06 0.364 0.178 0.364 0.178 0.364 0.030 - - - - - -
300-4 0.010 3e-04 1e-04 3e-04 1e-04 3e-04 1e-05 0.003 - - - -
300-5 1e-07 0.648 1.000 0.648 1.000 0.648 0.364 0.178 1e-04 - - -
300-6 1e-06 0.648 0.364 0.648 0.364 0.648 0.076 0.648 0.001 0.364 - -
300-7 0.001 0.003 0.001 0.003 0.001 0.003 1e-04 0.030 0.364 0.001 0.010

P value adjustment method: none

> TukeyHSD(aov(y~CapH,data=data_CellMs))
  Tukey multiple comparisons of means
  95% family-wise confidence level

Fit: aov(formula = y ~ CapH, data = data_CellMs)

$CapH

      diff      lwr      upr p adj
100-5-100-4 1.500e+00 0.7211  2.2789 0.0000
100-6-100-4 1.600e+00 0.8211  2.3789 0.0000
100-7-100-4 1.500e+00 0.7211  2.2789 0.0000
200-4-100-4 1.600e+00 0.8211  2.3789 0.0000
200-5-100-4 1.500e+00 0.7211  2.2789 0.0000
200-6-100-4 1.800e+00 1.0211  2.5789 0.0000
200-7-100-4 1.300e+00 0.5211  2.0789 0.0002
300-4-100-4 6.000e-01 -0.1789  1.3789 0.2499
300-5-100-4 1.600e+00 0.8211  2.3789 0.0000
300-6-100-4 1.400e+00 0.6211  2.1789 0.0001
300-7-100-4 8.000e-01 0.0211  1.5789 0.0404

```

100-6-100-5	1.000e-01	-0.6789	0.8789	1.0000
100-7-100-5	-8.882e-16	-0.7789	0.7789	1.0000
200-4-100-5	1.000e-01	-0.6789	0.8789	1.0000
200-5-100-5	-8.882e-16	-0.7789	0.7789	1.0000
200-6-100-5	3.000e-01	-0.4789	1.0789	0.9549
200-7-100-5	-2.000e-01	-0.9789	0.5789	0.9980
300-4-100-5	-9.000e-01	-1.6789	-0.1211	0.0142
300-5-100-5	1.000e-01	-0.6789	0.8789	1.0000
300-6-100-5	-1.000e-01	-0.8789	0.6789	1.0000
300-7-100-5	-7.000e-01	-1.4789	0.0789	0.1067
100-7-100-6	-1.000e-01	-0.8789	0.6789	1.0000
200-4-100-6	-8.882e-16	-0.7789	0.7789	1.0000
200-5-100-6	1.000e-01	-0.8789	0.6789	1.0000
200-6-100-6	2.000e-01	-0.5789	0.9789	0.9980
200-7-100-6	-3.000e-01	-1.0789	0.4789	0.9549
300-4-100-6	-1.000e+00	-1.7789	-0.2211	0.0048
300-5-100-6	0.000e+00	-0.7789	0.7789	1.0000
300-6-100-6	-2.000e-01	-0.9789	0.5789	0.9980
300-7-100-6	-8.000e-01	-1.5789	-0.0211	0.0404
200-4-100-7	1.000e-01	-0.6789	0.8789	1.0000
200-5-100-7	0.000e+00	-0.7789	0.7789	1.0000
200-6-100-7	3.000e-01	-0.4789	1.0789	0.9549
200-7-100-7	-2.000e-01	-0.9789	0.5789	0.9980
300-4-100-7	-9.000e-01	-1.6789	-0.1211	0.0142
300-5-100-7	1.000e-01	-0.6789	0.8789	1.0000
300-6-100-7	-1.000e-01	-0.8789	0.6789	1.0000
300-7-100-7	-7.000e-01	-1.4789	0.0789	0.1067
200-5-200-4	-1.000e-01	-0.8789	0.6789	1.0000
200-6-200-4	2.000e-01	-0.5789	0.9789	0.9980
200-7-200-4	-3.000e-01	-1.0789	0.4789	0.9549
300-4-200-4	-1.000e+00	-1.7789	-0.2211	0.0048
300-5-200-4	8.882e-16	-0.7789	0.7789	1.0000
300-6-200-4	-2.000e-01	-0.9789	0.5789	0.9980
300-7-200-4	-8.000e-01	-1.5789	-0.0211	0.0404
200-6-200-5	3.000e-01	-0.4789	1.0789	0.9549
200-7-200-5	-2.000e-01	-0.9789	0.5789	0.9980
300-4-200-5	-9.000e-01	-1.6789	-0.1211	0.0142
300-5-200-5	1.000e-01	-0.6789	0.8789	1.0000
300-6-200-5	-1.000e-01	-0.8789	0.6789	1.0000
300-7-200-5	-7.000e-01	-1.4789	0.0789	0.1067
200-7-200-6	-5.000e-01	-1.2789	0.2789	0.4932
300-4-200-6	-1.200e+00	-1.9789	-0.4211	0.0005
300-5-200-6	-2.000e-01	-0.9789	0.5789	0.9980
300-6-200-6	-4.000e-01	-1.1789	0.3789	0.7755
300-7-200-6	-1.000e+00	-1.7789	-0.2211	0.0048
300-4-200-7	-7.000e-01	-1.4789	0.0789	0.1067
300-5-200-7	3.000e-01	-0.4789	1.0789	0.9549
300-6-200-7	1.000e-01	-0.6789	0.8789	1.0000
300-7-200-7	-5.000e-01	-1.2789	0.2789	0.4932
300-5-300-4	1.000e+00	0.2211	1.7789	0.0048
300-6-300-4	8.000e-01	0.0211	1.5789	0.0404
300-7-300-4	2.000e-01	-0.5789	0.9789	0.9980
300-6-300-5	-2.000e-01	-0.9789	0.5789	0.9980
300-7-300-5	-8.000e-01	-1.5789	-0.0211	0.0404
300-7-300-6	-6.000e-01	-1.3789	0.1789	0.2499

SCRT Wednesday 3/23/22 (week 9, lecture 23)

Sample Size Calculations - Equal Number of Reps

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

We're talking about 2 factor design.

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

⇒ no interaction

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

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

Interaction

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

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

not ideal approach.

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

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 Tables for Homework-Exams 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. Need to specify the values of the following:

$$\mu_{ij} - \bar{\mu}_{i.}, \quad \bar{\mu}_{.j}, \quad \bar{\mu}_{..}, \quad \sigma_e^2$$

The following approach follows the methodology used in single factor experiments.

most preferred.

2. Determine r by specifying differences between any two treatment means: $\mu_{ij} - \mu_{hk}$

- a. Let D be the difference between any two treatment means: $\mu_{ij} - \mu_{hk}$ that the researcher deems important to detect.
- b. From our previous results, we know that the minimum value for λ is

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

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

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

or Table IX in Tables for Homework-Exams as was done in Handout #3.

3. Determine r by specifying differences in the - Main Effect of Individual Factors.

↳ assumes no interaction effects.

- 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

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

- c. Determine minimum value of r such that $\gamma(\lambda) \geq \gamma_o$ whenever $\lambda \geq \lambda_o = \frac{ar D^2}{2\sigma_e^2}$. 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 Tables for Homework-Exams as was done in Handout #3.

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 $\sigma, \hat{\sigma} = 1.4$

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

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

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
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

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

Alternatively, compute $\lambda = \frac{rD^2}{2\sigma_e^2} = \frac{r(2.5)^2}{2(1.4)^2} = 1.5944r$, $\nu_1 = ab - 1 = 11$, $\nu_2 = ab(r - 1) = 12(r - 1)$

With $r=12$, $\lambda = 19.133$, $\nu_1 = 11$, $\nu_2 = 132$,

$$\text{power} = 1 - pf(qf(.95, 11, 132), 11, 132, 19.133) = .82744$$

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 $\gamma \geq .9$ 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 $\sigma, \hat{\sigma} = 1.4$

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

$$a = 3; \quad b = 4; \quad \alpha = .05; \quad D = 2.5; \quad S = \hat{\sigma}_e = 1.4$$

and then select the value of r such that the power exceeds $\gamma = .90$

The output from the program yields a $r = 4$, that is 4 reps per treatment.

t	r	c	u1	u2	L	phi	power
12	2	3.49029	3	12	9.5663	1.54647	0.5847
12	3	3.00879	3	24	14.3495	1.89404	0.8459
12	4	3.00879	3	36	14.3495	1.89404	0.9496

With $r=4$, $\lambda = \frac{3rD^2}{2\sigma_e^2} = 19.133$, $\nu_1 = 3$, $\nu_2 = 36$,

$$\text{power} = 1 - pf(qf(.95, 3, 36), 3, 36, 19.133) = .9496$$

Two Factor Experiment with Unequal Number of Reps per Treatment

Let n_{ij} be the number of EUs assigned to Treatment (i,j) and suppose the n_{ij} are not all equal, the estimation of the treatment means and tests of hypotheses becomes somewhat more complicated.

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

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

If we have unequal rep. levels

Standard Errors of LSE

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

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

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

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

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

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

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

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

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

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: $\hat{\theta} \pm (t_{\alpha/2, df})(\widehat{SE}(\hat{\theta}))$,

where df are the degrees of freedom associated with $\widehat{SE}(\hat{\theta})$

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

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

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

$$\hat{\mu}_i \pm (t_{\alpha/2, df_E}) \widehat{SE}(\hat{\mu}_i) \Rightarrow \hat{\mu}_i \pm (t_{\alpha/2, df_E}) \hat{\sigma}_e \sqrt{\frac{1}{b^2} \sum_{j=1}^b \frac{1}{n_{ij}}}$$

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

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

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

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

5. $100(1 - \alpha)\%$ C.I. on $\bar{\mu}_i - \bar{\mu}_k$:

$$\hat{\mu}_i - \hat{\mu}_k \pm (t_{\alpha/2, df_E}) \hat{\sigma}_e \sqrt{\frac{1}{b^2} \sum_{j=1}^b \left(\frac{1}{n_{ij}} + \frac{1}{n_{kj}} \right)}$$

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

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

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 :

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

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

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

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

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

$$\text{df} = n - (b + (a - 1)) = n - a - b + 1.$$

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 $\hat{\mu}_{ij} = \bar{y}_{ij}$. and computing

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

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

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

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

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

Express Model 2, no interaction, as $\mu_{ij} = \alpha_i + \gamma_j$ for all $i = 1, \dots, a$ and $j = 1, \dots, b$. 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

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

The solutions to these equations require:

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

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

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 $i = 1, \dots, a$;

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

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 $H_o : \bar{\mu}_1 = \bar{\mu}_2 = \dots = \bar{\mu}_a$. versus $H_1 : \text{not all } \bar{\mu}_i \text{ are equal}$

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

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

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

When there is
interaction, we don't
look at main
effects

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

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

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 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	$\frac{1}{n_{1.}} \sum_{j=1}^b n_{1j} \mu_{1j} = \frac{1}{n_{2.}} \sum_{j=1}^b n_{2j} \mu_{2j} = \dots = \frac{1}{n_{a.}} \sum_{j=1}^b n_{aj} \mu_{aj}$
F_2	$\sum_{i=1}^a \left(n_{ij} - \frac{n_{ij}^2}{n_{i.}} \right) \mu_{ij} = \sum_{h \neq j}^b \sum_{i=1}^a \frac{n_{ij} n_{ih}}{n_{i.}} \mu_{ih} \quad \text{for } j = 1, 2, \dots, b$
$F_1 * F_2$	$(\mu_{ij} - \mu_{kj}) - (\mu_{ih} - \mu_{kh}) = 0 \quad \text{for all } (i, j, k, h)$

If no $\alpha_{ij} = 0$ then Type III = Type IV

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

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

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

$$SS_{F_2} = SSE_3 - SSE_1$$

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:

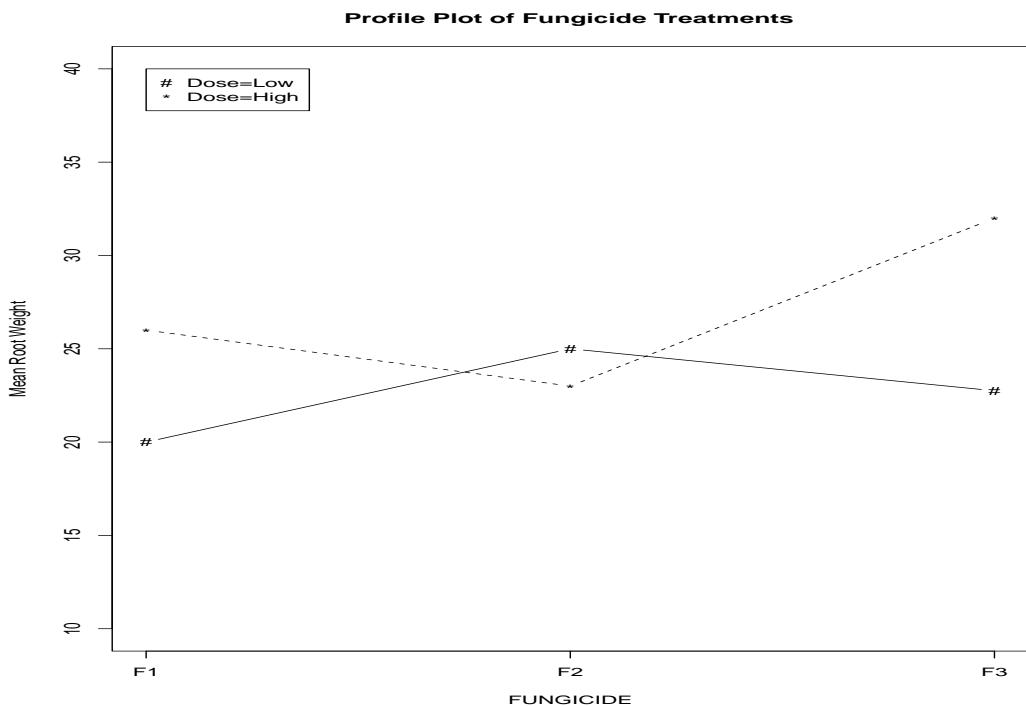
X	Source of Variation	Type III Hypotheses
	F_1	$\bar{\mu}_{1.} = \bar{\mu}_{2.} = \cdots = \bar{\mu}_{a.}$
	F_2	$\bar{\mu}_{.1} = \bar{\mu}_{.2} = \cdots = \bar{\mu}_{.b}$
	$F_1 * F_2$	$(\mu_{ij} - \mu_{kj}) - (\mu_{ih} - \mu_{kh}) = 0 \quad \text{for all } (i, j, k, h)$

X 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) were randomly applied at two dose levels (D) 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 $n_{ij} = 4$.

D	F		
	1	2	3
1	19	24	22
	20	26	25
	21		25
2			19
	25	21	31
	27	24	32
	24	33	
			32

$y = \text{root weight}$
 $\Sigma y = \text{sum of plant}$
 $F \rightarrow 3 \text{ fixed levels}$
 $D - 2 \text{ fixed levels}$
 $3 \times 2 = 6 \text{ treatments}$



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 D F Y @@;
IF D=1 and F=1 THEN TRT=1;
IF D=1 and F=2 THEN TRT=2;
IF D=1 and F=3 THEN TRT=3;
IF D=2 and F=1 THEN TRT=4;
IF D=2 and F=2 THEN TRT=5;
IF D=2 and F=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 D F;
model Y=D F D*F/ssi ss2 ss3 ss4;
means D F D*F;
lsmeans D F D*F/stderr pdiff adjust=Tukey;
lsmeans D F D*F/stderr pdiff ;
RUN;
proc glm;
class TRT;
model y=TRT/ssi ss2 ss3 ss4;

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

(Cell means mode!)

(Effects model)

so we can look at all differences

adjusted so we can look at sum of treatment effects

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

Class	Levels	Values
D	2	1 2
F	3	1 2 3

$\Rightarrow \text{Type III} = \text{Type IV}$

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
D	1	128.0000000	128.0000000	39.64	<.0001
F	2	81.5090909	40.7545455	12.62	0.0011
D*F	2	95.7409091	47.8704545	14.82	0.0006

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

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

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

① look at interaction
since significant
we don't look at

②

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

individ
means & SD's
 y_i

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

\bar{y}_i

Level of		Y		
D	F	N	Mean	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

\bar{y}_{ij}

Least Squares Means

Adjustment for Multiple Comparisons: Tukey

D	Y LSMEAN	Standard Error	H0: LSMEAN=0	
			Pr > t	
1	22.5833333	0.6234549	<.0001	
2	27.0000000	0.6234549	<.0001	

H0: LSMean1 =
LSMean2
Pr > |t|
0.0003

F	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	

not gonna talk abt
floc bc we
have an interaction.

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	

D	F	Y LSMEAN	Standard		Pr > t	LSMEAN Number
			Error	Pr > t		
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

$$\alpha_{pc} = \frac{0.05}{6} = 0.0083$$

Least Squares Means for effect A*B
UnAdjusted p-values
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Y						
i/j	1	2	3	4	5	6
1		0.0101	0.0682	0.0033	0.0635	<.0001
2	0.0101		0.1738	0.5881	0.2462	0.0007
3	0.0682	0.1738		0.0587	0.8585	<.0001
4	0.0033	0.5881	0.0587		0.0924	0.0023
5	0.0635	0.2462	0.8585	0.0924		<.0001
6	<.0001	0.0007	<.0001	0.0023	<.0001	

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	

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

Compare the p-values from the UnAdjusted p-value matrix to $0.05/6 = .0083$ because we only want to compare $(2)(3) = 6$ pairs of means not all 15 possible pairs of means.

At $D=1$, $G_1 = \{F_1, F_2\}$, $G_2 = \{F_3\}$ $\Rightarrow G = \{F_1, F_2, F_3\}$

At $D=2$, $G_1 = \{F_1, F_2\}$; $G_2 = \{F_3\}$

Case 2: Compare all 6 Treatments: Compare the p-values from the Adjusted p-value matrix to 0.05 because the multiple comparison have already been adjusted by Tukey-Kramer.

$G_1 = \{D1F1, D1F2, D1F3, D2F2\}$;
 $G_2 = \{D1F2, D1F3, D2F1, D2F2\}$; $G_3 = \{D2F3\}$

$\Rightarrow (2,1), (2,3)$

when $D=2$ there is a d.f. for bfr treatment of F_2 \Rightarrow

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 D	1	81.02884615	81.02884615	25.09	0.0003
Main F	2	67.92272727	33.96136364	10.52	0.0023
D*F	2	95.74090909	47.87045455	14.82	0.0006

Least Squares Means

TRT	Y LSMEAN	Standard		LSMEAN Number
		Error	Pr > t	
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
 $\text{Pr} > |t| \text{ for } H_0: \text{LSMean}(i) = \text{LSMean}(j)$

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	

R-Code - factorial-unequaln.R -

```
options(contrasts = c("contr.sum", "contr.poly"))

rw = c(19,20,21,24,26,22,25,25,19,25,27,21,24,24,31,32,33,32)
dose = c(rep("D1",9), rep("D2",9))
dosefac = factor(dose)
fung = c(rep("F1",3), rep("F2",2), rep("F3",4),
         rep("F1",2), rep("F2",3), rep("F3",4))
fungfac = factor(fung)
datafac = cbind(rw,dosefac,fungfac)

modelfac = lm(rw~dosefac + fungfac + dosefac:fungfac)

#outputs Type I Sum of Squares
anova(modelfac)

#outputs Type III Sum of Squares
drop1(modelfac,.~.,test="F")

#Cell Means model

trt = c(rep("D1F1",3),rep("D1F2",2),rep("D1F3",4),
        rep("D2F1",2),rep("D2F2",3),rep("D2F3",4))

trtfac = factor(trt)
datatr = cbind(rw,trtfac)

modeltrt = lm(rw~trtfac)
anova(modeltrt)

#Pairwise Comparisons of Treatment means:

lsmeans(modeltrt,"trtfac")
#Comparisons with Unadjusted p-values
pairwise.t.test(rw,trtfac,p.adjust="none")
#Pairwise comparisons yielding Tukey Adjusted p-values

TukeyHSD(aov(rw~trtfac))

Output from R:
```

```

> modelfac = lm(rw~dosefac + fungfac + dosefac:fungfac)
>
> #outputs Type I Sum of Squares
> anova(modelfac)
Analysis of Variance Table

Response: rw
          Df Sum Sq Mean Sq F value Pr(>F)
dosefac      1 128.0 128.0   39.6  4e-05 ***
fungfac      2   81.5   40.8   12.6  0.00112 **
dosefac:fungfac 2   95.7   47.9   14.8  0.00057 ***
Residuals    12   38.8    3.2
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

>
> #outputs Type III Sum of Squares
>
> drop1(modelfac,.~.,test="F")
Single term deletions

```

Model:

```

rw ~ dosefac + fungfac + dosefac:fungfac
          Df Sum of Sq   RSS   AIC F value Pr(>F)
<none>            38.8 25.8
dosefac      1     81.0 119.8 44.1    25.1 0.00030 ***
fungfac      2     67.9 106.7 40.0    10.5 0.00230 **
dosefac:fungfac 2     95.7 134.5 44.2    14.8 0.00057 ***

```

```

#Cell Means model
> modeltrt = lm(rw~trtfac)
> anova(modeltrt)
Analysis of Variance Table

Response: rw
  Df Sum Sq Mean Sq F value Pr(>F)
trtfac     5 305.2   61.0    18.9 2.6e-05 ***
Residuals 12  38.8    3.2
> pairwise.t.test(rw,trtfac,p.adjust="none")

  Pairwise comparisons using t tests with pooled SD

data: rw and trtfac

      D1F1  D1F2  D1F3  D2F1  D2F2
D1F2 0.010 -     -     -     -
D1F3 0.068 0.174 -     -     -
D2F1 0.003 0.588 0.059 -     -
D2F2 0.063 0.246 0.859 0.092 -
D2F3 1e-06 7e-04 1e-05 0.002 3e-05

P value adjustment method: none
> #Pairwise comparisons yielding Tukey Adjusted p-values
>
> TukeyHSD(aov(rw~trtfac))
  Tukey multiple comparisons of means
  95% family-wise confidence level

Fit: aov(formula = rw ~ trtfac)

$trtfac
      diff     lwr      upr   p adj
D1F2-D1F1 5.00 -0.5100 10.510 0.0837
D1F3-D1F1 2.75 -1.8600  7.360 0.3937
D2F1-D1F1 6.00  0.4900 11.510 0.0302
D2F2-D1F1 3.00 -1.9283  7.928 0.3739
D2F3-D1F1 12.00  7.3900 16.610 0.0000
D1F3-D1F2 -2.25 -7.4773  2.977 0.7014
D2F1-D1F2  1.00 -5.0359  7.036 0.9922
D2F2-D1F2 -2.00 -7.5100  3.510 0.8198
D2F3-D1F2  7.00  1.7727 12.227 0.0074
D2F1-D1F3  3.25 -1.9773  8.477 0.3535
D2F2-D1F3  0.25 -4.3600  4.860 1.0000
D2F3-D1F3  9.25  4.9819 13.518 0.0001
D2F2-D2F1 -3.00 -8.5100  2.510 0.4843
D2F3-D2F1  6.00  0.7727 11.227 0.0216
D2F3-D2F2  9.00  4.3900 13.610 0.0003

```

~~start~~ Friday 3/25/22 (Week 9, Lecture 24)

EXAMPLE OF CONTRASTS IN A CRD 3x3 FACTORIAL EXPERIMENT

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

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

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

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

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: $\mu_{11}, \dots, \mu_{1b}, \mu_{21}, \dots, \mu_{2b}, \dots, \mu_{a1}, \dots, \mu_{ab}$ and construct k contrasts in the t means: $\mathbf{C}_1, \dots, \mathbf{C}_k$ with

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

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

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

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

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

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

$\frac{1}{\sigma_e^2} SS_{\mathbf{C}} = \frac{1}{\sigma_e^2} (\mathbf{C}\hat{\boldsymbol{\mu}})' [\mathbf{C}\mathbf{D}\mathbf{C}']^{-1} (\mathbf{C}\hat{\boldsymbol{\mu}})$ has a noncentral Chi-square distribution

with non-centrality parameter $(\mathbf{C}\boldsymbol{\mu})' [\mathbf{C}\mathbf{D}\mathbf{C}']^{-1} (\mathbf{C}\boldsymbol{\mu})$ and $df = \text{Row Rank } \mathbf{C}$ and is independent of MSE .

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

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

Example: Heavy Metals in Sewage Sludge

EXAMPLE from Kuehl's Book 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.

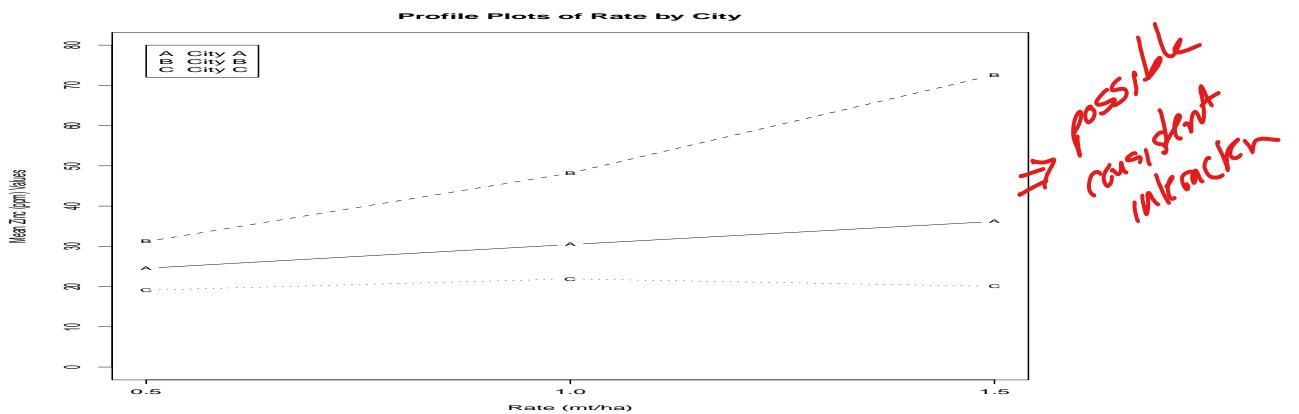
Research Hypothesis: A soil scientist hypothesized the concentration of certain heavy metals in sludge would differ among the metropolitan areas from which the sludge was obtained. The variation could result from any number of reasons, including different industrial bases surrounding the areas and the 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 qualitative 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.

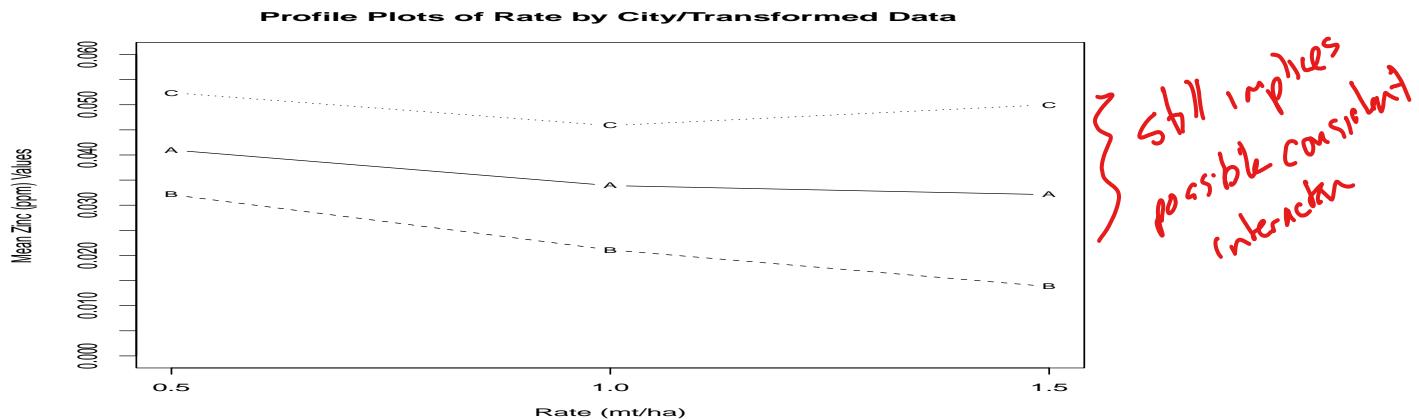
**Table 1. Zinc content (ppm) of barley plants
grown in media containing sludge at three rates
from three metropolitan areas.**

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	



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 $\mu_{i1}, \mu_{i2}, \mu_{i3}$, for each City i :

In a preliminary analysis of the data, the residual plots and SW test indicated that the residuals were from a non-normal distribution. The Box-Cox procedure yielded the transformation $X = Y^{-0.83} \approx X = 1/Y$. The analysis of variance table is given here.



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

Dependent Variable: X 1/ZINC-CONC

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

Model
Significant

Interaction
Significant

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

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 $\mathbf{H}_{ij} > 0$ for $i = 1, \dots, a$; $j = 1, \dots, b$.

$$\mathbf{H}_{ij} > 0$$

Coefficients for Mutually Orthogonal Contrasts in 9 Treatment Means

EFFECT	CONTRAST	TREATMENT MEANS								
		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:

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

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

$$\mathbf{C}_5 = (1, -1, 0, -1, 1, 0, 0, 0, 0) \quad \mathbf{C}_6 = (1, 1, -2, -1, -1, 2, 0, 0, 0)$$

$$\mathbf{C}_7 = (1, -1, 0, 1, -1, 0, -2, 2, 0) \quad \mathbf{C}_8 = (1, 1, -2, 1, 1, -2, -2, -2, 4)$$

$$\mathbf{C}_{F_1 * F_2} = \begin{pmatrix} \mathbf{C}_5 \\ \mathbf{C}_6 \\ \mathbf{C}_7 \\ \mathbf{C}_8 \end{pmatrix} \quad \mathbf{C}_{F_1 * F_2} \hat{\mu} = \begin{pmatrix} 11.125 \\ 48.475 \\ -17.325 \\ -84.975 \end{pmatrix} \quad \mathbf{D} = Diag \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right)$$

$$\mathbf{C}_{F_1 * F_2} \mathbf{D} \mathbf{C}_{F_1 * F_2}^t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix} \quad \Rightarrow \quad (C_{F_1 * F_2} D C_{F_1 * F_2}^t)^{-1} = Diag \left(1, \frac{1}{3}, \frac{1}{3}, \frac{1}{9} \right)$$

$$SS_{F_1 * F_2} = (\mathbf{C}_{F_1 * F_2} \hat{\mu})^t (\mathbf{C}_{F_1 * F_2} \mathbf{D} \mathbf{C}_{F_1 * F_2}^t)^{-1} (\mathbf{C}_{F_1 * F_2} \hat{\mu}) = .00028869 \text{ with } df_{F_1 * F_2} = \text{rank}(\mathbf{C}_{F_1 * F_2}) = 4$$

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

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

II. Main Effect of City (F_1):

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

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

$$\mathbf{C}_1 = (1, 1, 1, -1, -1, -1, 0, 0, 0)$$

$$\mathbf{C}_2 = (1, 1, 1, 1, 1, 1, -2, -2, -2)$$

$$\mathbf{C}_{\mathbf{F}_1} = \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{pmatrix} \quad \mathbf{C}_{\mathbf{F}_1}\hat{\mu} = \begin{pmatrix} -60.8 \\ 121.05 \end{pmatrix} \quad \mathbf{D} = Diag\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

$$\mathbf{C}_{\mathbf{F}_1} \mathbf{D} \mathbf{C}_{\mathbf{F}_1}^t = \begin{pmatrix} \frac{6}{4} & 0 \\ 0 & \frac{18}{4} \end{pmatrix}$$

$$SS_{F_1} = (\mathbf{C}_{\mathbf{F}_1}\hat{\mu})^t (\mathbf{C}_{\mathbf{F}_1} \mathbf{D} \mathbf{C}_{\mathbf{F}_1}^t)^{-1} (\mathbf{C}_{\mathbf{F}_1}\hat{\mu}) = .00441535 \text{ with } df_{F_1} = rank(\mathbf{C}_{\mathbf{F}_1}) = 2$$

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

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):

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

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

$$\mathbf{C}_3 = (1, -1, 0, 1, -1, 0, 1, -1, 0)$$

$$\mathbf{C}_4 = (1, 1, -2, 1, 1, -2, 1, 1, -2)$$

$$\mathbf{C}_{\mathbf{F}_2} = \begin{pmatrix} \mathbf{C}_3 \\ \mathbf{C}_4 \end{pmatrix} \quad \mathbf{C}_{\mathbf{F}_2}\hat{\mu} = \begin{pmatrix} -25.725 \\ -82.275 \end{pmatrix} \quad \mathbf{D} = Diag\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

$$\mathbf{C}_{\mathbf{F}_2} \mathbf{D} \mathbf{C}_{\mathbf{F}_2}^t = \begin{pmatrix} \frac{6}{4} & 0 \\ 0 & \frac{18}{4} \end{pmatrix}$$

$$SS_{F_2} = (\mathbf{C}_{\mathbf{F}_2}\hat{\mu})^t (\mathbf{C}_{\mathbf{F}_2} \mathbf{D} \mathbf{C}_{\mathbf{F}_2}^t)^{-1} (\mathbf{C}_{\mathbf{F}_2}\hat{\mu}) = .00077426 \text{ with } df_{F_2} = rank(\mathbf{C}_{\mathbf{F}_2}) = 2$$

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

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 ~~Factorial~~ folder SAScode: **Factorial,SimultaneousContrasts.sas**; and output is given on the next pages.

```

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 yield simultaneous tests of trends 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;
OUTPUT FROM SAS:

```

TRT 9 A05 A10 A15 B05 B10 B15 C05 C10 C15

Dependent Variable: X 1/ZINC-CONC

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

$$H_o : \text{No 3-Way Interaction in } F_1, F_2, F_3$$

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

$$H_1 : \text{Equality Does Not Hold for SOME choices of } (i, i', j, j', k, k')$$

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 k th 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:

$$(a) F_1 * F_2: H_o : \text{No 2-Way Interaction in } F_1, F_2$$

$$H_o : (\bar{\mu}_{ij.} - \bar{\mu}_{i'j.}) = (\bar{\mu}_{ij'.} - \bar{\mu}_{i'j'.}), \quad \text{for all choices of } (i, i', j, j')$$

$$H_1 : \text{Equality Does Not Hold for SOME choices of } (i, i', j, j')$$

$$(b) F_1 * F_3: H_o : \text{No 2-Way Interaction in } F_1, F_3$$

$$H_o : (\bar{\mu}_{i.k} - \bar{\mu}_{i'.k}) = (\bar{\mu}_{i.k'} - \bar{\mu}_{i'.k'}), \quad \text{for all choices of } (i, i', k, k')$$

$$H_1 : \text{Equality Does Not Hold for SOME choices of } (i, i', k, k')$$

$$(c) F_2 * F_3: H_o : \text{No 2-Way Interaction in } F_2, F_3$$

$$H_o : (\bar{\mu}_{jk} - \bar{\mu}_{j'k}) = (\bar{\mu}_{jk'} - \bar{\mu}_{j'k'}), \quad \text{for all choices of } (j, j', k, k')$$

$$H_1 : \text{Equality Does Not Hold for SOME choices of } (j, j', k, k')$$

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) .

$$\begin{aligned} SS_{TOT} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^r (y_{ijkl} - \bar{y}_{....})^2 \\ &= r \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\bar{y}_{ijk.} - \bar{y}_{....})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^r (y_{ijkl} - \bar{y}_{ijk.})^2 \\ &= SS_{Model} + SSE \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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:

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

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

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

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

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

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

We thus have reduced the number of parameters to

- 1. μ
- 2. $(a-1) \tau_i$'s $(b-1) \beta_j$'s $(c-1) \gamma_k$'s
- 3. $(a-1)(b-1) (\tau\beta)_{ij}$'s $(a-1)(c-1) (\tau\gamma)_{ik}$'s $(b-1)(c-1) (\beta\gamma)_{jk}$'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.

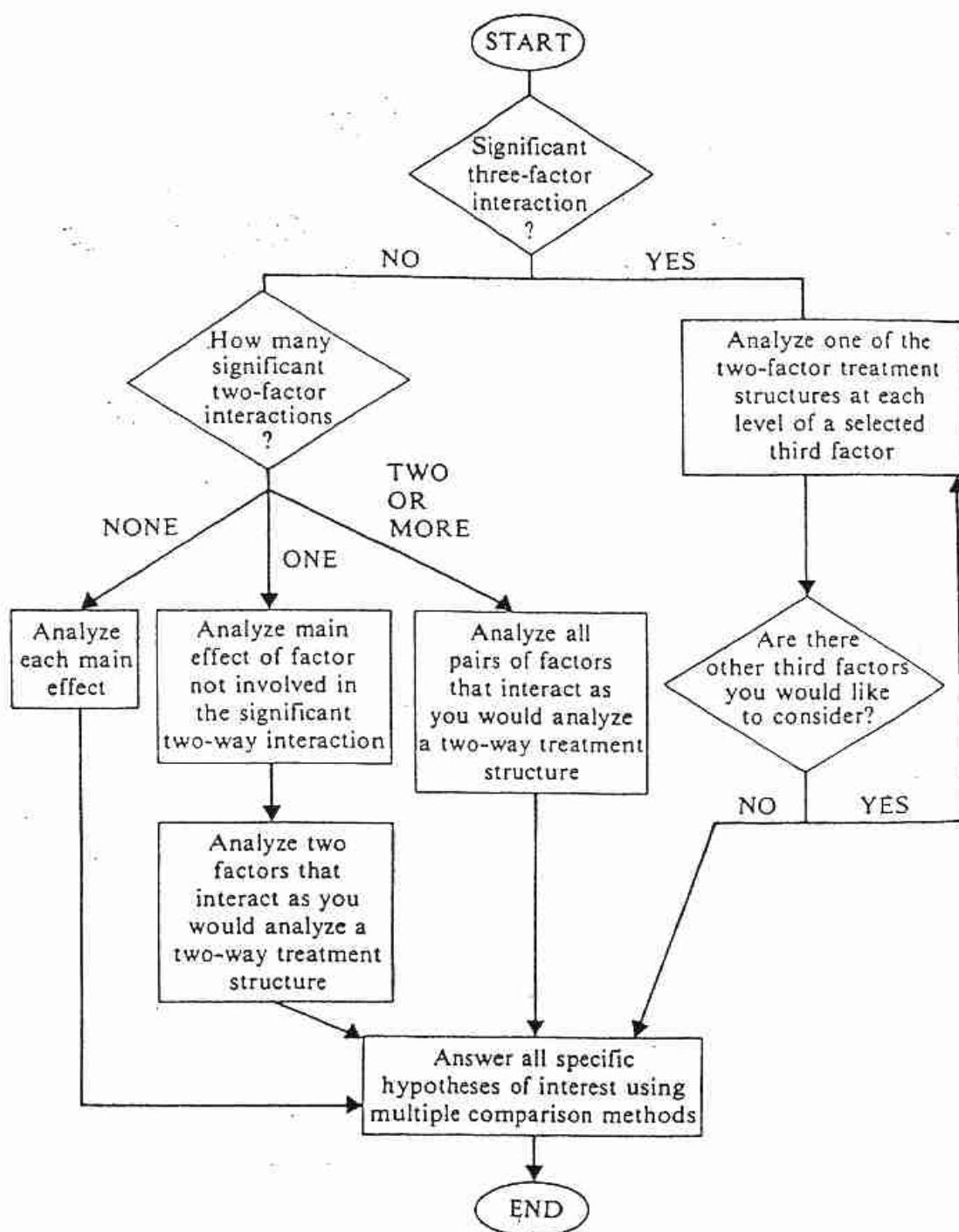
Under the restrictions, the treatment means μ_{ijk} are given by

1. $\mu_{abc} = \mu$
2. $\mu_{abk} = \mu + \gamma_k$ when $k \neq c$
3. $\mu_{ajc} = \mu + \beta_j$ when $j \neq b$
4. $\mu_{ibc} = \mu + \tau_i$ when $i \neq a$
5. $\mu_{ajk} = \mu + \beta_j + \gamma_k + (\beta\gamma)_{jk}$ when $j \neq b, k \neq c$
6. $\mu_{ibk} = \mu + \tau_i + \gamma_k + (\tau\gamma)_{ik}$ when $i \neq a, k \neq c$
7. $\mu_{ijc} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$ when $i \neq a, j \neq b$
8. $\mu_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk}$ when $i \neq a, j \neq b, k \neq c$

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 be analyzed.

~~K~~ 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 i th level of S, j th level of F, and the k th level of P.

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

Linear

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

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

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

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

2. A 2-way interaction between S and P would consist 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,

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

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

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

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

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

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

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

~~X~~ 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}
$C_1 = \text{Main P}$	1	-2	1	1	-2	1	1	-2	1	1	-2	1
$C_2 = \text{Main P}$	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 significant 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 significant 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 they 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 temperature 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 ($25^\circ, 35^\circ$); 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.

In our case
 $y = \text{average weight gain}$

$$EU = MU = \text{aquarium}$$

we were interested in individual weight gain, we would have observed

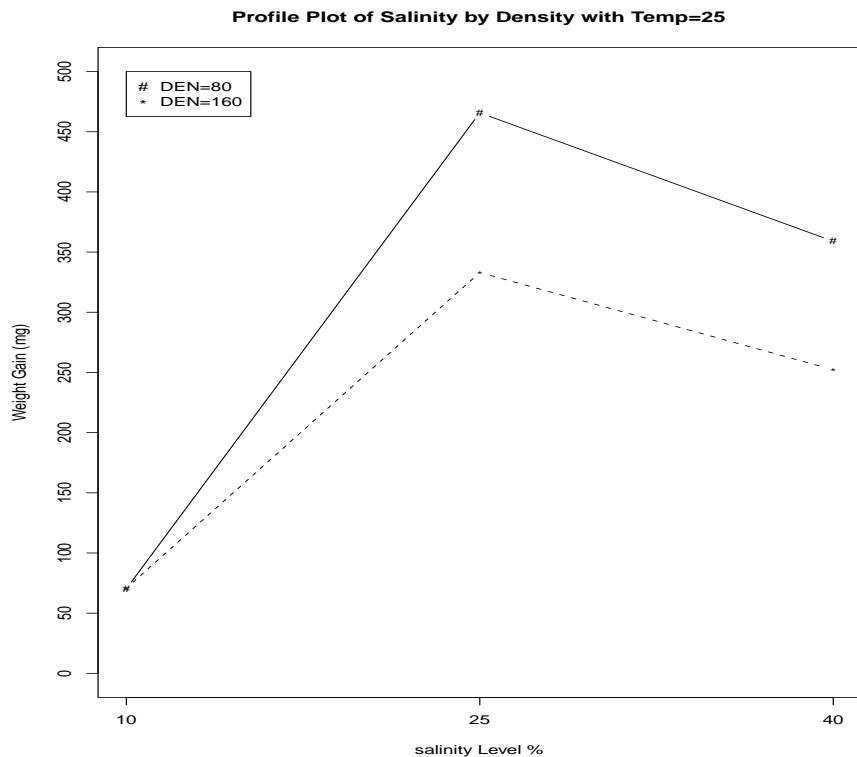
$y = \text{individual weight gain}$

$$EU = \text{shrimp}, MU = \text{shrimp}$$

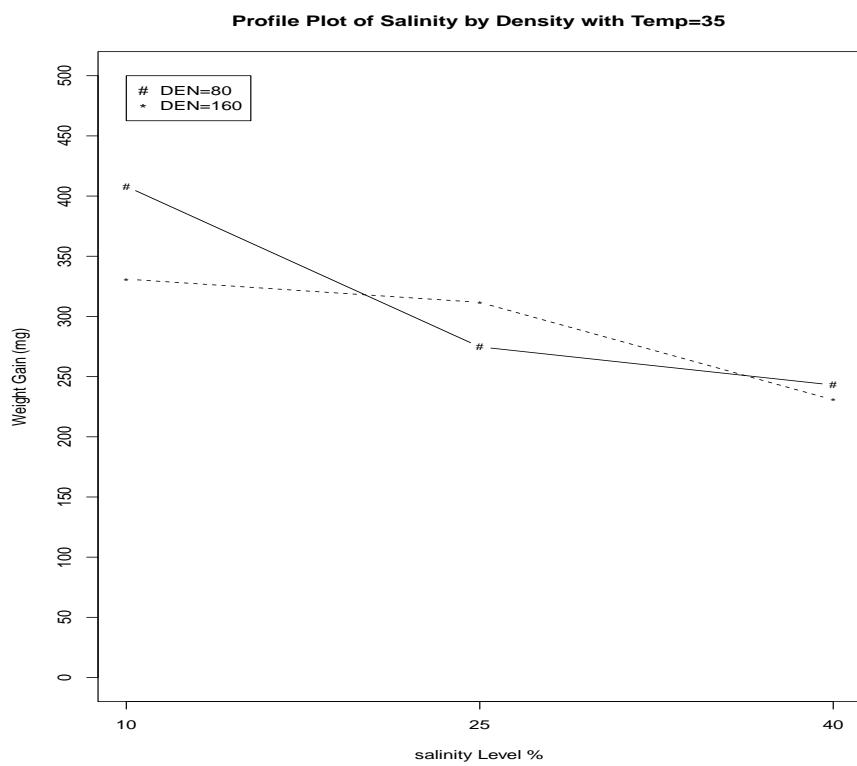
Table 1. Four-week Weight Gain of Shrimp Cultured in Aquaria at Different Levels of Temperature (T), Water Salinity (S) and Density of Shrimp Populations (D)

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

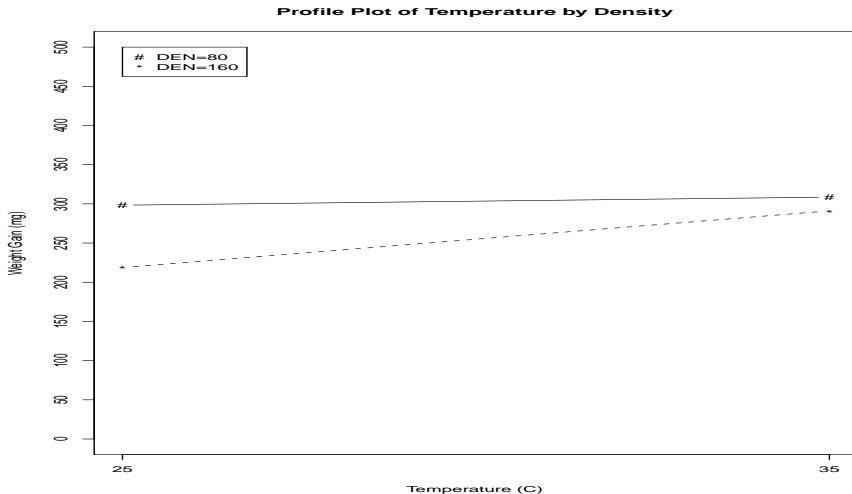


implies possible
consistent
interaction b/w
Salinity & Density
at Temp = 25

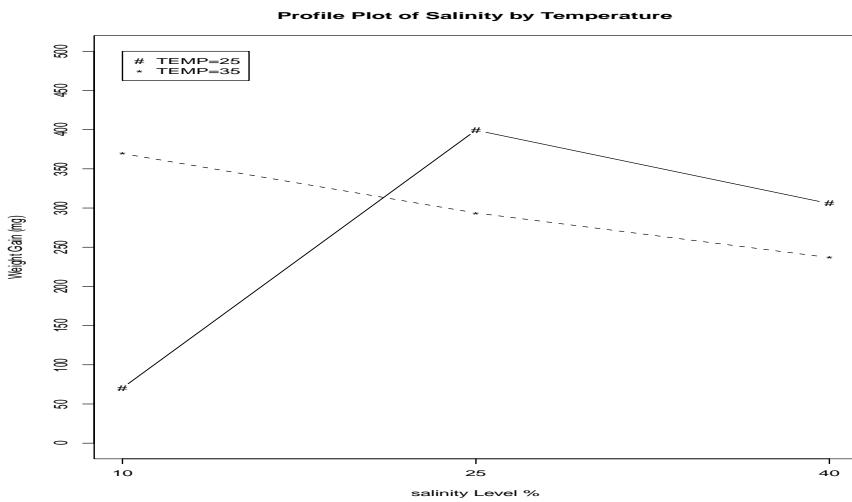


implies possible
nonconstant interaction
b/w Salinity & Density
at Temp = 35

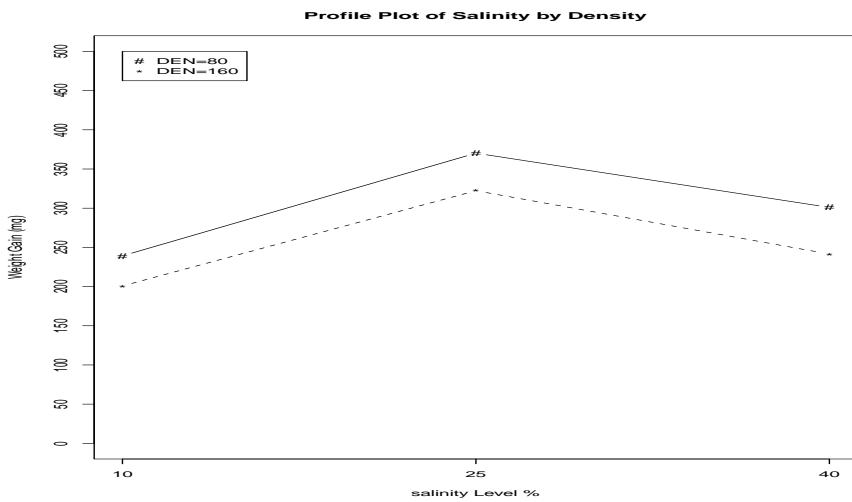
Profile Plots Depicting Two-Way Interactions



implies possible
-consistent interaction
when temp & density



implies possible
means consistent interaction
between salinity and temp for both



⇒ possibly no interaction
between salinity & density.

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;
```

Homogeneity of variances.

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

(include all interactions when we write libtst.)

*T D S T*D T*S D*S T*D*S*

compute means at each two-way interaction.

*means S*D S*T T*D;*

*lsmeans S*T*D/stderr pdiff adjust=tukey;* *want to compare means for 3-way (for comparing all differences)*

*lsmeans S*T*D/stderr pdiff;* *want to compare means for 3-way (for comparing main and diff.)*

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;

if interaction is significant were not interested in not different.

Homogeneity of variances.

get fancier plots.

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

→ Model is significant.

- when looking at ANOVA table, always use & (adj. needed) b/c we first check only if (main int. is significant). → interaction is significant.

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	Standard		Pr > t	LSMEAN Number
				LSMEAN	Error		
25	80	10	70.333333	31.111508	0.0331	<.0001	1
25	80	25	465.666667	31.111508	0.0331	<.0001	2
25	80	40	359.000000	31.111508	0.0331	<.0001	3
25	160	10	70.666667	31.111508	0.0324	<.0001	4
25	160	25	333.000000	31.111508	0.0324	<.0001	5
25	160	40	252.333333	31.111508	0.0324	<.0001	6
35	80	10	408.000000	31.111508	0.0317	<.0001	7
35	80	25	274.666667	31.111508	0.0317	<.0001	8
35	80	40	243.000000	31.111508	0.0317	<.0001	9
35	160	10	331.000000	31.111508	0.0317	<.0001	10
35	160	25	311.666667	31.111508	0.0317	<.0001	11
35	160	40	230.666667	31.111508	0.0317	<.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):

B1

$$\begin{aligned} T &= 25, D = 80 \\ G1 &= \{10\}, G2 = \{25, 40\} \end{aligned}$$

$$\begin{aligned} T &= 35, D = 80 \\ G1 &= \{10, 25\}, G2 = \{25, 40\} \end{aligned}$$

B2

$$\begin{aligned} T &= 25, D = 160 \\ G1 &= \{10\}, G2 = \{25, 40\} \end{aligned}$$

$$\begin{aligned} T &= 35, D = 160 \\ G1 &= \{10, 25, 40\} \end{aligned}$$

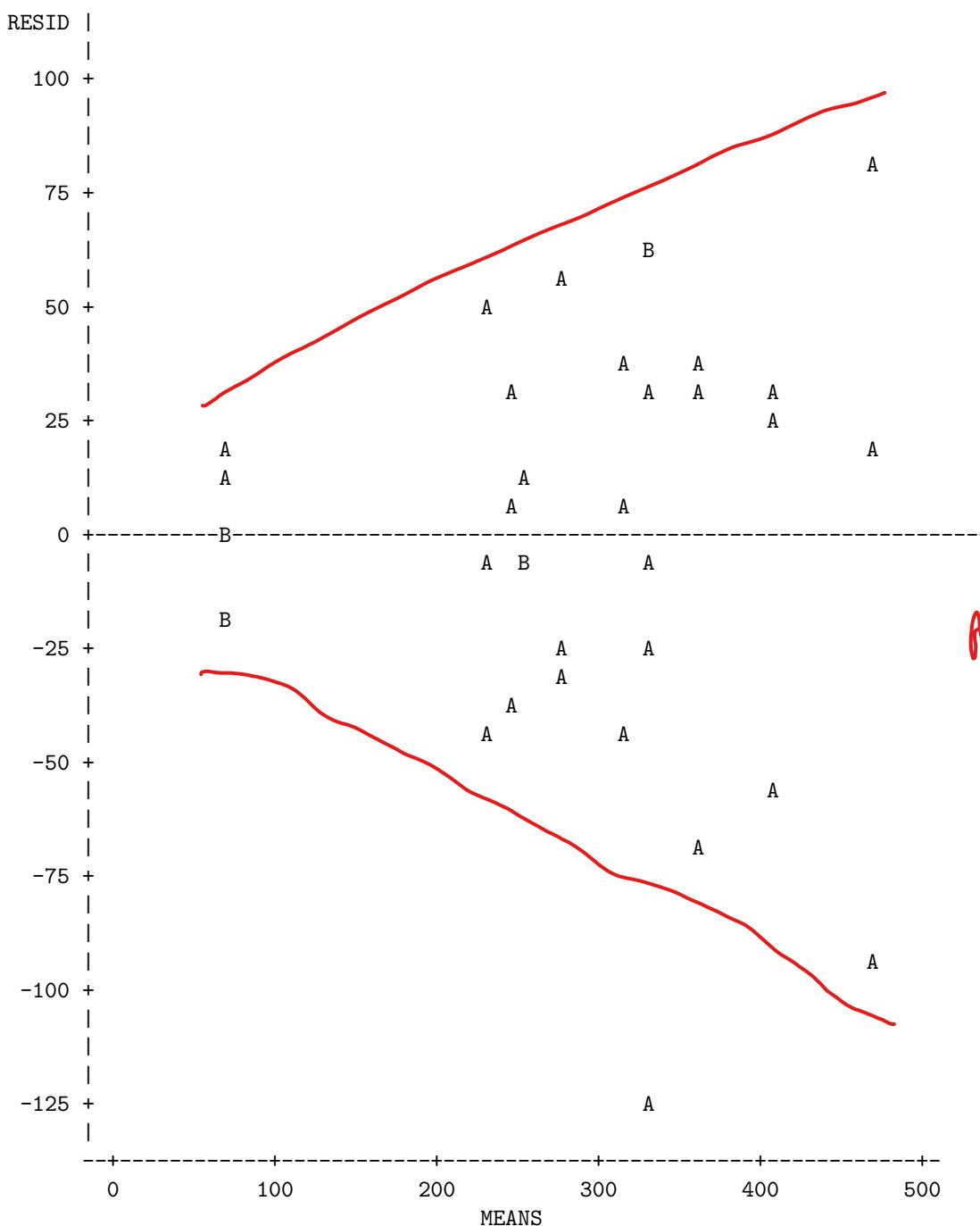
$$\begin{aligned} (3) &= 12 \\ \alpha_{pc} &= \frac{0.05}{12} \\ &= 0.004167 \end{aligned}$$

For each
unit of
T-D-S
looking at
first column

STOP Friday 3/25/22 (Week 9, lecture 24)

START Monday 3/28/22 (Week 10, Lecture 25)

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

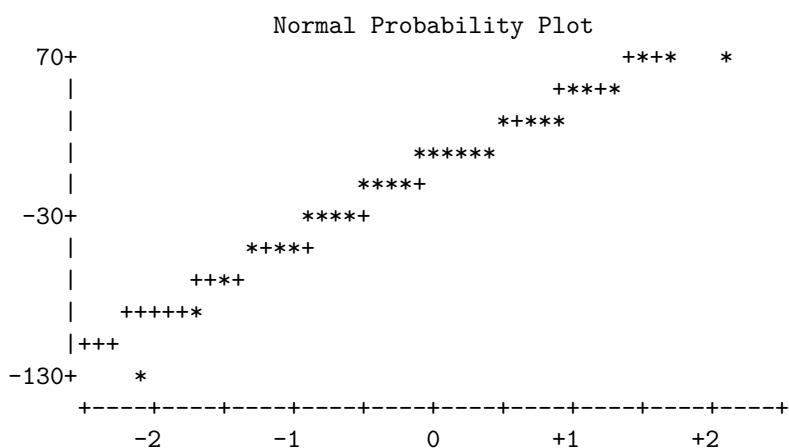


for constant error
variable.
heteroskedastic
errors.

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



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

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
TRT	11	9805.6	891.4	0.44	0.9233
Error	24	48977.3	2040.7		

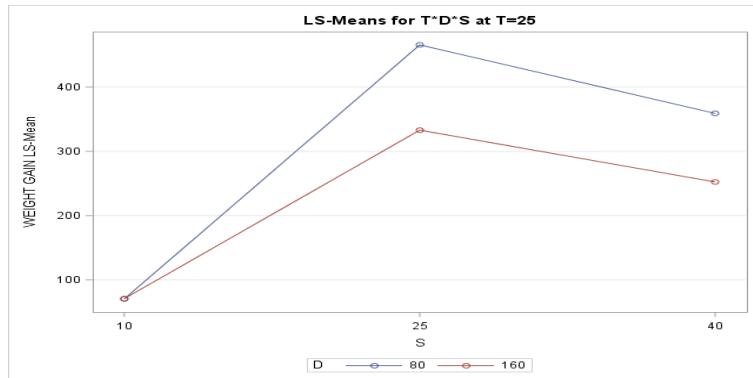
$\gamma \alpha = 0.05$. No evidence
for deviation of normality.

$$H_0: \sigma_1 = \sigma_2 = \dots = \sigma_{12}$$

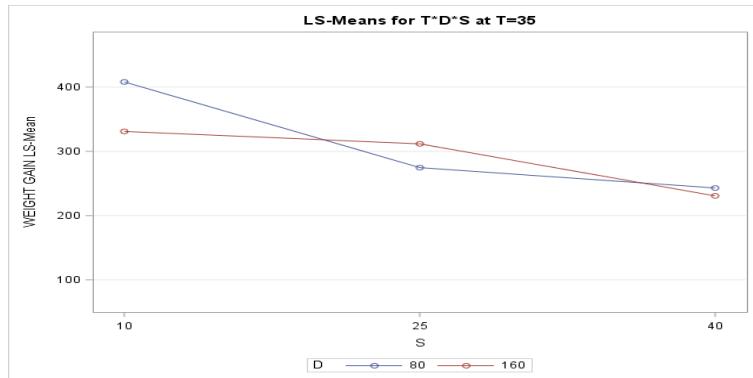
- no violation
of homogeneity.
 \Rightarrow plot of residual was misleading.

Profile Plot Depicting Three- Way Interaction-SAS Output

Page 1 of 1



file:///C:/Users/longneck/AppData/Local/Temp/SAS%20Temporary%20Files/_TD11212_... 3/27/2019
Page 1 of 1



file:///C:/Users/longneck/AppData/Local/Temp/SAS%20Temporary%20Files/_TD11212_... 3/27/2019

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 correction of $\alpha = \frac{.05}{30} = .0017$. We divide by 30 because there are 6 combinations of Salinity and Density which yields $\binom{6}{2} = 15$ pairs of means and the comparison is done for each of the 2 Temperatures. Thus, $(2)(15) = 30$ pairs of treatment means are being compared. This yields:

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

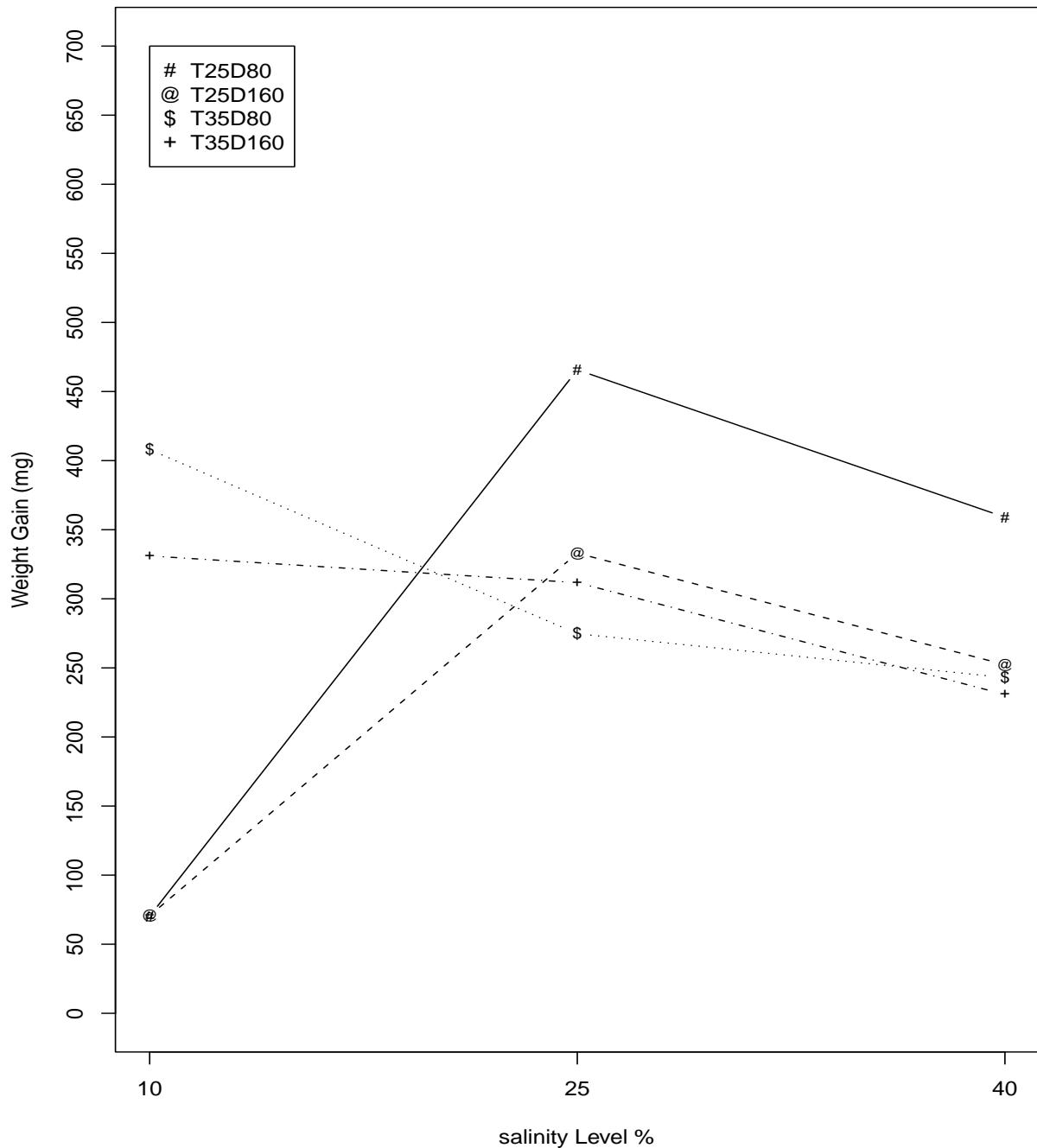
At TEMP=35, (S,D)=(10%,80), (10%,160) and (25%,160) have the highest mean weight gain and they are not significantly different.

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



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 0 0 0 0 0;
ESTIMATE 'SLIN25160' TRT 0 0 0 -1 0 1 0 0 0 0 0 0 0 0 0 0 0;
ESTIMATE 'SLIN3580' TRT 0 0 0 0 0 0 -1 0 1 0 0 0 0 0 0 0 0;
ESTIMATE 'SLIN35160' TRT 0 0 0 0 0 0 0 0 0 0 0 -1 0 1 0 0 0;
ESTIMATE 'SQUAD2580' TRT 1 -2 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0;
ESTIMATE 'SQUAD25160' TRT 0 0 0 1 -2 1 0 0 0 0 0 0 0 0 0 0 0;
ESTIMATE 'SQUAD3580' TRT 0 0 0 0 0 0 1 -2 1 0 0 0 0 0 0 0 0;
ESTIMATE 'SQUAD35160' TRT 0 0 0 0 0 0 0 0 0 1 -2 1 0 0 0 0 0;
run;
```

Cell means model.

$$\alpha_{pc} = \frac{0.05}{6} = 0.00625$$

Parameter	Estimate	Standard 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: $\bar{y}_{i\cdot\cdot\cdot}$, $\bar{y}_{\cdot j\cdot\cdot}$, and $\bar{y}_{\cdot\cdot k\cdot}$. 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	173.83	Yes
25	80	40	359.00	-106.67	9.49	Yes
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

$(Temp = 25^\circ, Density = 80, Salinity = 25\%)$ ~~and~~

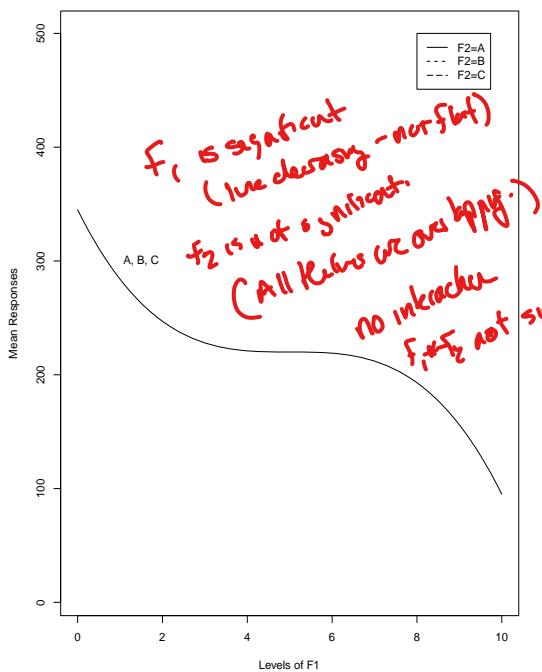
$(Temp = 35^\circ, Density = 80, Salinity = 10\%)$ ~~;~~

$(25^\circ, 80, 40\%)$

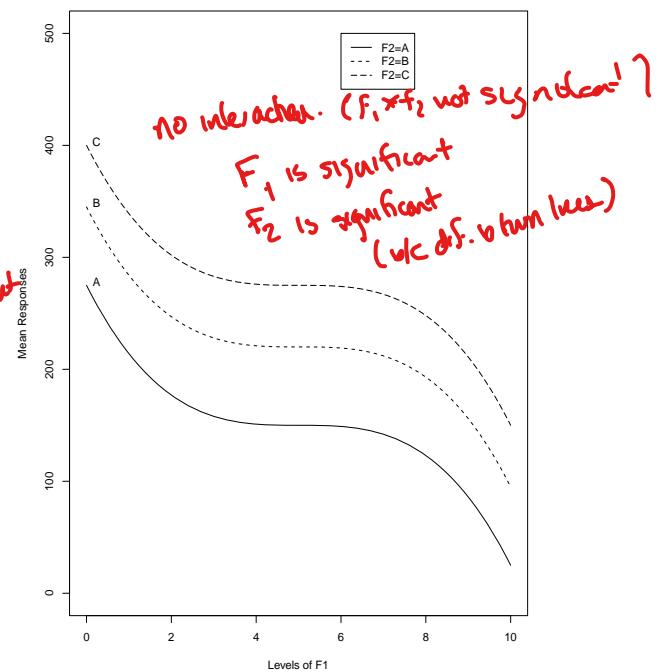
Factorial Experiment With Quantitative and Qualitative Factors

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.

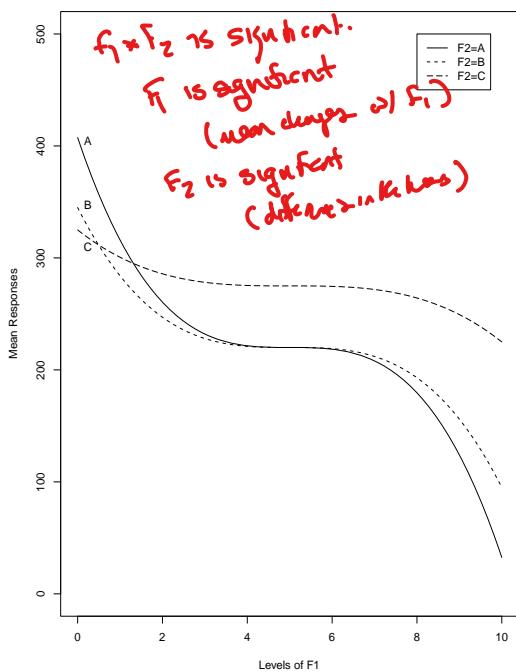
Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)



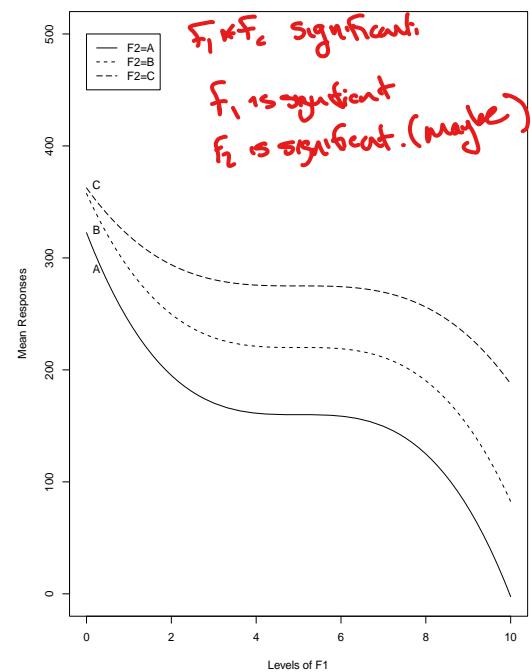
Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)



Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)



Profile Plot of Quant. Factor (F1) and Qual. Factor (F2)



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 qualitative factor, "City" with three fixed levels and one quantitative factor, "Rate" with three fixed levels.

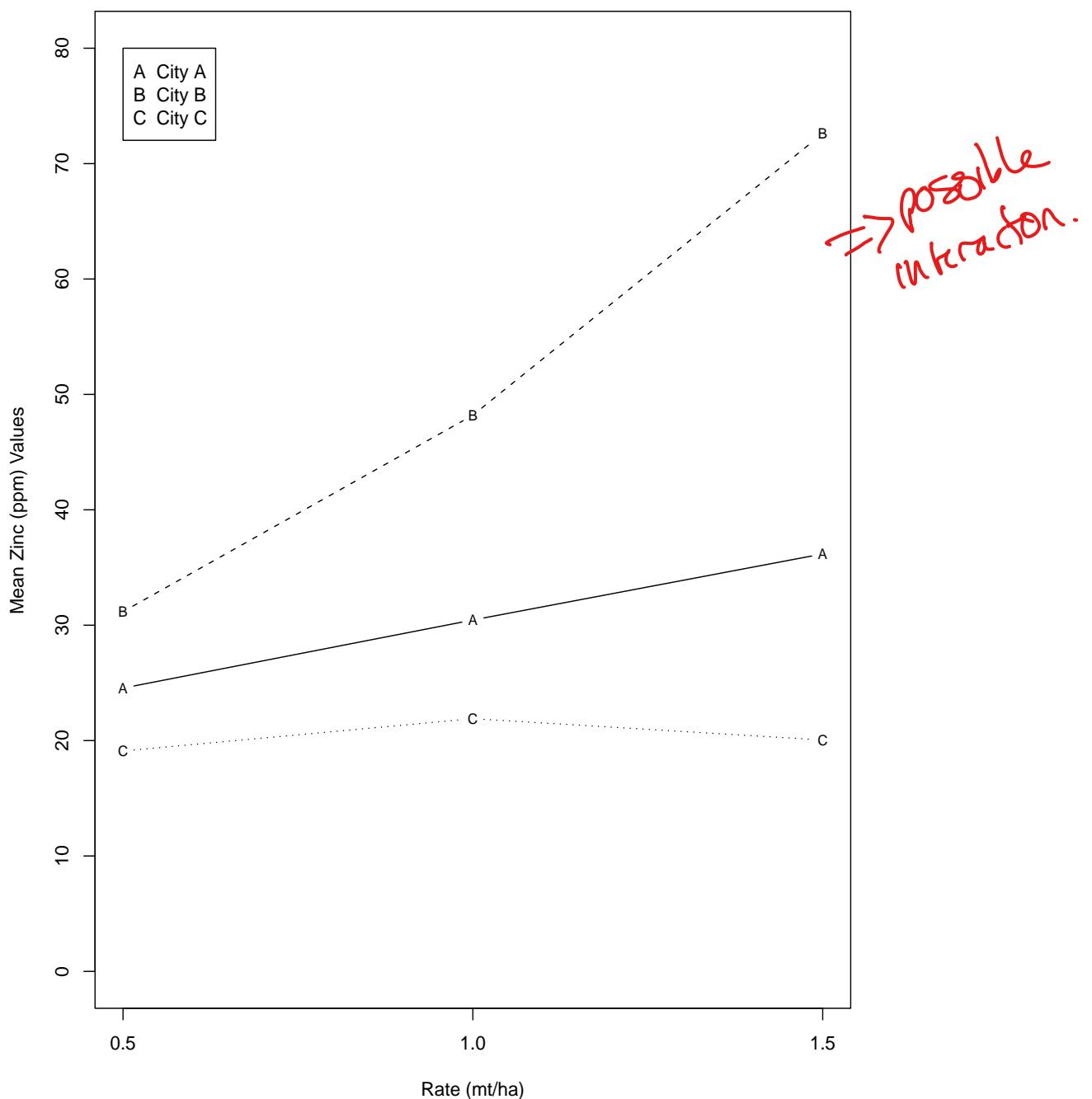
Experiment Design: Each of the nine treatment combinations was randomly assigned to four replicate containers in a completely randomized design. The containers were arranged completely at random in a growth chamber. At a certain stage of growth the zinc content in parts per million was determined for 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	

*Note - we have
1 qualitative factor
City (w/ 3 levels)
; 1 quant. factor
Rate (w/ 3 levels)
. Levels for both
factors are fixed.*

Profile Plots of Rate by City



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 $\mu_{i1}, \mu_{i2}, \mu_{i3}$, for each City i :

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

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 quadratic 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 quadratic.

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

The significant 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.

$$\begin{aligned}\alpha_3 &= 0 \\ \beta_3 &= 0 \\ (\alpha\beta)_{i3} &= 0 \quad \text{if } i \\ (\alpha\beta)_{sj} &= 0 \quad \text{if } j \\ \text{Interaction is sig.} & \\ \text{Rate trend at rate.} & \\ \text{City independently.} & \\ \text{Need to look at} &\end{aligned}$$

SAS code and output:

```
* Factorial_sludges.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;
```

- resid
- normality

*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;
```

+ for yet to include ranadjuated me2, so we will look at them all together.

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

Level of	X		
F2	N	Mean	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	Level of	X		
F1	F2	N	Mean	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

Model is significant.

Pr > F <.0001

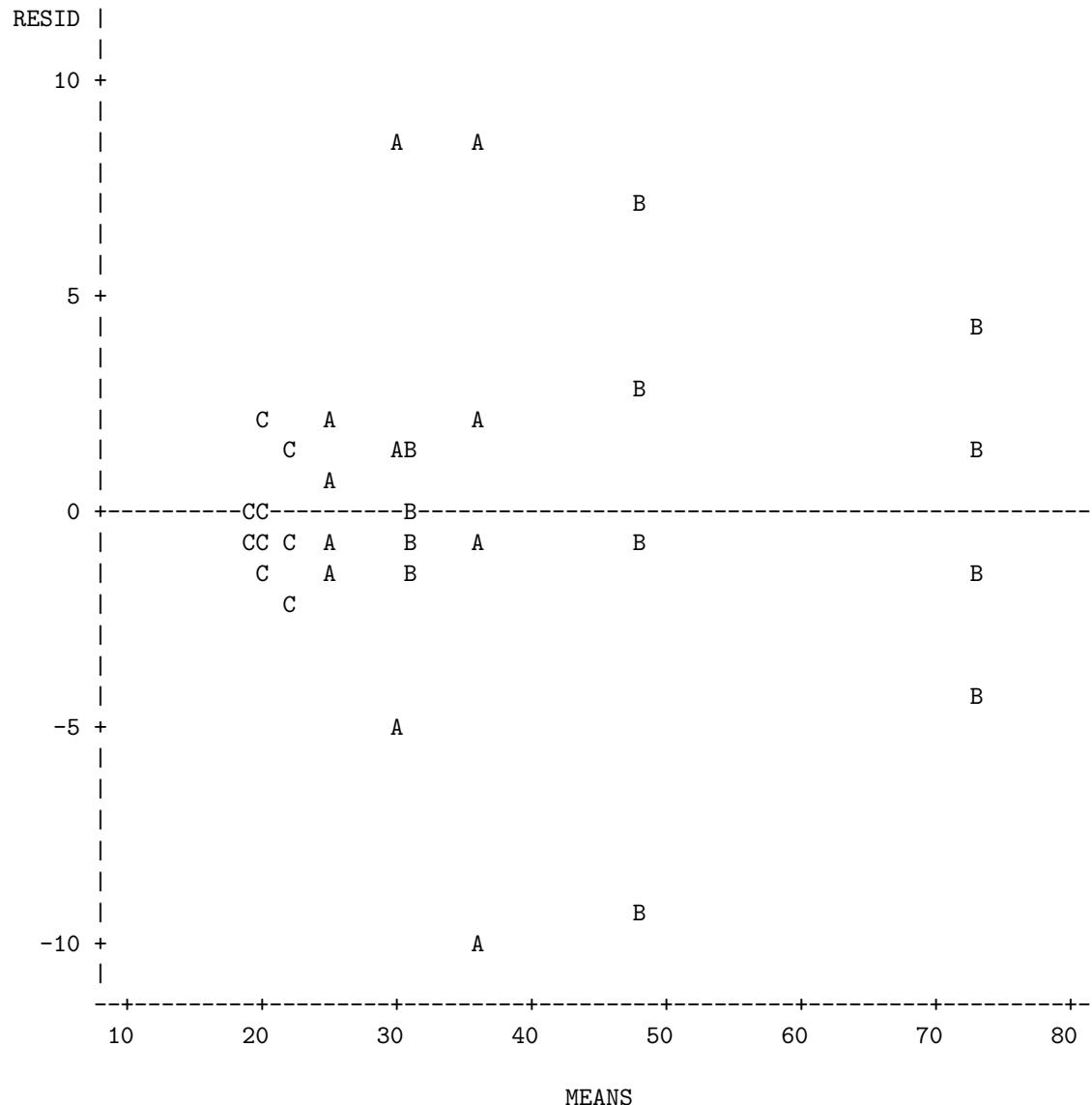
Interaction is significant
=> can't look at main effects alone

$$\bar{y}_{i\cdot}$$

$$\bar{y}_{\cdot j}$$

$$\bar{y}_{ij}$$

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



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		

→ Evidence of Nonconstant variance
(Heteroskedasticity)
(Homogeneity violated)

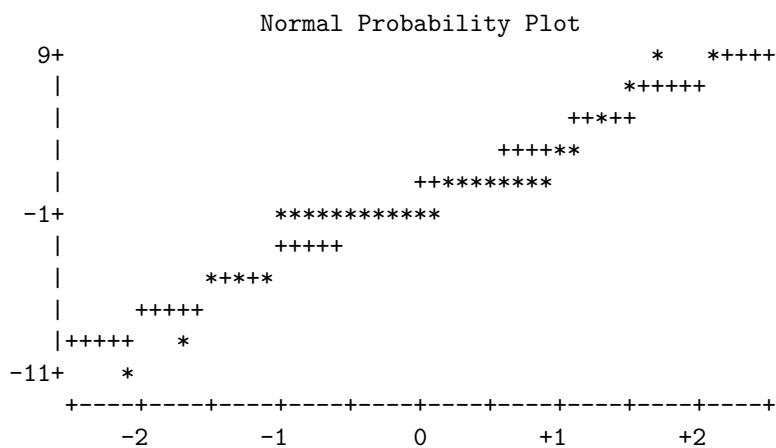
Tests for Normality

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	*
-----+-----+-----+		

Heavy tails

Variable: RESID



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

Least squares

$$k_{pc} = \frac{0.05}{6} = 0.0083$$

CITY-RATE	Estimate	Standard Error		DF	Alpha	Lower	Upper	LSMEAN Number
		Estimate	Standard Error					
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	

a_{ij}

only testing $a_{ij} = \mu_0$

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	
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	4	A10	
D	F	E	4	A05	
D	F	E	4	C10	
	F	E	4	C15	
	F		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}

$\binom{9}{2} = 36 \text{ d.f.s.}$
 $\alpha = 0.05$
 + not actually interested in the 10% interaction
 is significant.
 • actually interested in trying
 if far enough apart
 city is the difference
 start to sales.
 • only interested in
 $\binom{3}{2} = 3 \text{ differences.}$

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 $X = Y^{-0.83} \approx 1/Y$. 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

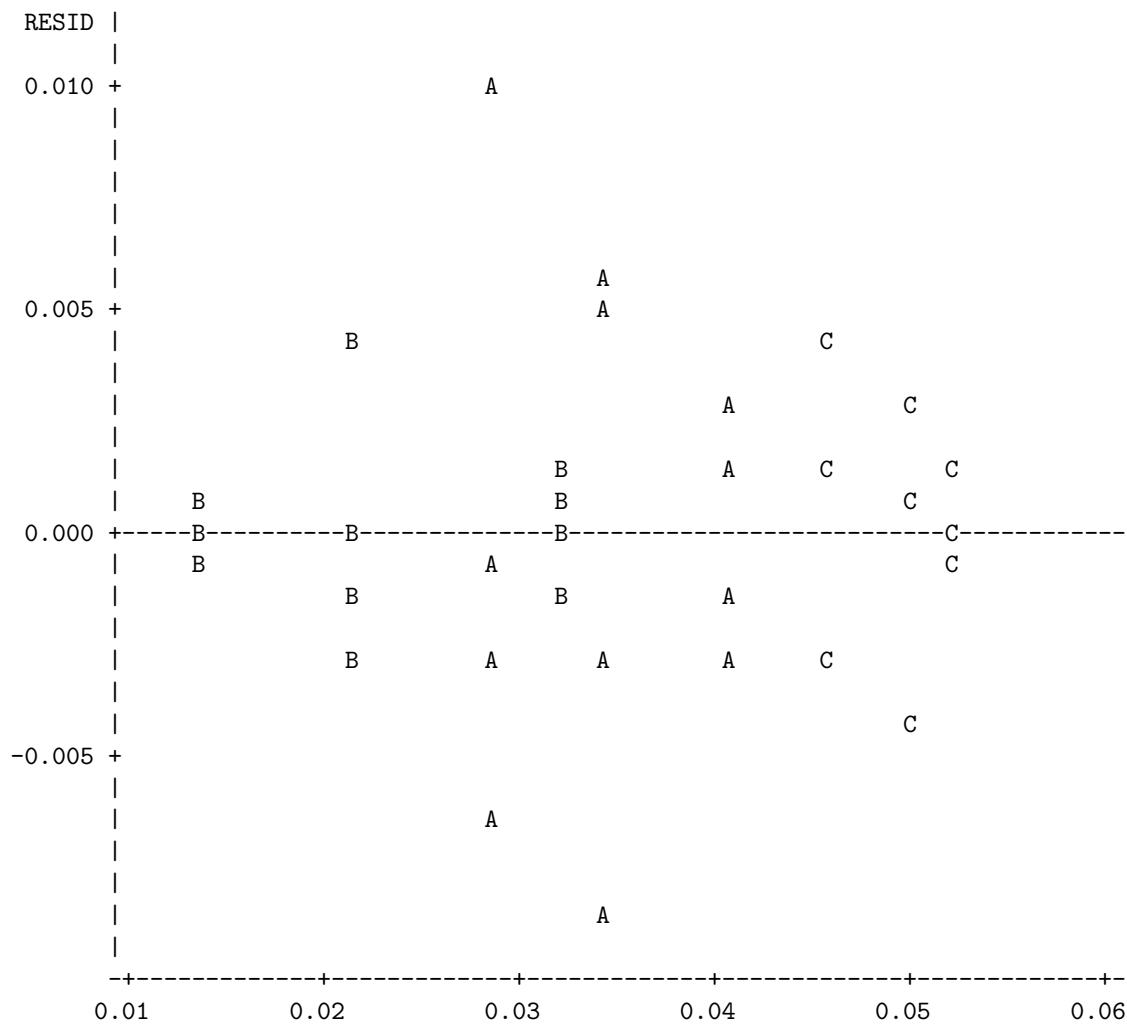
Source	DF	Sum of		F Value	Pr > F
		Squares	Mean Square		
Model	8	0.00547830	0.00068479	45.51	<.0001
Error	27	0.00040631	0.00001505		
Corrected Total	35	0.00588462			

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

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
		F1*F2			
	4	0.00028869	0.00007217	4.80	0.0047

model significant.
Interaction significant.

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



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		

→ St. ll don't have constant error variance.

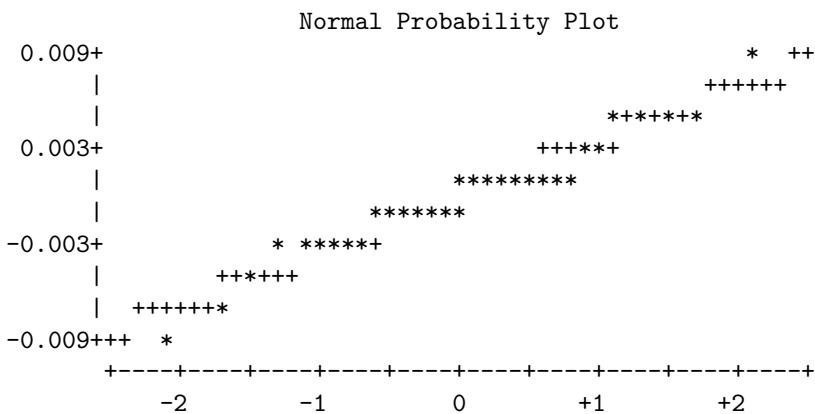
Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W	0.964985
Kolmogorov-Smirnov	D	0.149674
Cramer-von Mises	W-Sq	0.097865
Anderson-Darling	A-Sq	0.556962

↑ normality not violated.

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



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	Sum of Squares		
			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 <i>or</i>
LIN-B	1	0.00067035	0.00067035	44.55	<.0001 <i>or</i>
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

$$\Delta_{QC} = \frac{0.05}{6} - \frac{-0.06}{3}$$

Least Squares Means

Adjustment for Multiple Comparisons: Tukey

TRT	X LSMEAN	Standard		LSMEAN	
		Error	Pr > t	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.

Finished Monday 3/28/22 (week 19 lecture 5)