

Statistics 514: Design of Experiments

Topic 8

Topic Overview

This topic will cover

- RCB Designs
- Factorial Designs
- Latin Square Designs
- Balanced, Incomplete Block Designs

What Happens When There's More Than One Factor?

Naive Approach

- Vary one factor at a time
 - doesn't take into account *conditional expectations*
 - “Did I cover all possibilities?”
- Could work in theory, but **rarely** worth the trouble from a design perspective.

Situations

- *One Controllable Nuisance Factor*: Random Complete Block Design
- *More Than One Nuisance Factor*: Latin Squares Designs
- *Several Factors of Interest*: Factorial Design
- *Time is a Covariate*: Repeated Measures

Issues

- random vs fixed effects (Mixed Models)
- nesting
- recordable but uncontrollable factors (ANCOVA)
- budget (incomplete blocks/confounding)
- interactions
- balance

Randomized Complete Block Design (Section 4-1)

- Nuisance factor: a source of variation
 - Has effect on response but effect not of interest
 - If unknown → Protection through randomization; otherwise, bias.
 - If known but uncontrollable → Analysis of Covariance
 - If known and controllable → Blocking Design
- Randomized Complete Block Design (RCBD)
 - “most widely used design”
 - b blocks each consisting of a experimental units
 - a treatments are randomly assigned to EU’s within each block
 - “Complete” means every treatment is assigned to each block
 - Block for homogeneity – within blocks, units are homogeneous
 - results in “restriction on randomization”
 - extension of paired t -test where pairs = blocks
- Block
 - on batch
 - spatially
 - temporally
 - age, gender, history
- *Variance Reduction Design*: In addition to testing factor of interest, want to know if
 - process is “robust” to nuisance conditions
 - blocking is necessary in future experiments
- Sometimes need to follow randomization to determine design

Example

Four penicillin manufacturing processes – A , B , C , and D – are being investigated. Yield is the primary interest. It is known that an important raw material, corn steep liquor, is quite variable.

Corn Steep Liquor	A	B	C	D
Blend 1	89 ₁	88 ₃	97 ₂	94 ₄
Blend 2	84 ₄	77 ₂	92 ₃	79 ₁
Blend 3	81 ₂	87 ₁	87 ₄	85 ₃
Blend 4	87 ₁	92 ₃	89 ₂	84 ₄
Blend 5	79 ₃	81 ₄	80 ₁	88 ₂

- eliminates blend-to-blend variation
- provides wider inductive basis

Statistical Model

- b blocks and a treatments
- Statistical model is

$$y_{i,j} = \mu + \tau_i + \beta_j + \epsilon_{i,j} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases}$$

μ – grand mean

τ_i – i th treatment effect

β_j – j th block effect

$\epsilon_{i,j} \sim N(0, \sigma^2)$

Observation	Grand Mean	Conditions	Blocks	Residual Error																																										
<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">$y_{1,1}$</td><td style="padding: 2px 10px;">\dots</td><td style="padding: 2px 10px;">$y_{1,b}$</td></tr> <tr><td style="padding: 2px 10px;">$y_{2,1}$</td><td style="padding: 2px 10px;">\dots</td><td style="padding: 2px 10px;">$y_{2,b}$</td></tr> <tr><td style="padding: 2px 10px;">$y_{3,1}$</td><td style="padding: 2px 10px;">\dots</td><td style="padding: 2px 10px;">$y_{3,b}$</td></tr> <tr><td style="padding: 2px 10px;">\vdots</td><td></td><td style="padding: 2px 10px;">\vdots</td></tr> <tr><td style="padding: 2px 10px;">$y_{a,1}$</td><td style="padding: 2px 10px;">\dots</td><td style="padding: 2px 10px;">$y_{a,b}$</td></tr> </table>	$y_{1,1}$	\dots	$y_{1,b}$	$y_{2,1}$	\dots	$y_{2,b}$	$y_{3,1}$	\dots	$y_{3,b}$	\vdots		\vdots	$y_{a,1}$	\dots	$y_{a,b}$	=	<div style="border: 1px solid black; width: 60px; height: 60px; display: flex; align-items: center; justify-content: center; margin: 0 auto;"> μ </div>	+	<table style="border-collapse: collapse; width: 100%;"> <tr><td style="border: 1px solid black; padding: 2px 10px;">τ_1</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">τ_2</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">τ_3</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">\vdots</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">τ_a</td></tr> </table>	τ_1	τ_2	τ_3	\vdots	τ_a	+	<table style="border-collapse: collapse; width: 100%;"> <tr><td style="border: 1px solid black; padding: 2px 10px;">β_1</td><td style="border: 1px solid black; padding: 2px 10px;">\dots</td><td style="border: 1px solid black; padding: 2px 10px;">β_b</td></tr> </table>	β_1	\dots	β_b	+	<table style="border-collapse: collapse; width: 100%;"> <tr><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{1,1}$</td><td style="border: 1px solid black; padding: 2px 10px;">\dots</td><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{1,b}$</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{2,1}$</td><td style="border: 1px solid black; padding: 2px 10px;">\dots</td><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{2,b}$</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{3,1}$</td><td style="border: 1px solid black; padding: 2px 10px;">\dots</td><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{3,b}$</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">\vdots</td><td></td><td style="border: 1px solid black; padding: 2px 10px;">\vdots</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{a,1}$</td><td style="border: 1px solid black; padding: 2px 10px;">\dots</td><td style="border: 1px solid black; padding: 2px 10px;">$\epsilon_{a,b}$</td></tr> </table>	$\epsilon_{1,1}$	\dots	$\epsilon_{1,b}$	$\epsilon_{2,1}$	\dots	$\epsilon_{2,b}$	$\epsilon_{3,1}$	\dots	$\epsilon_{3,b}$	\vdots		\vdots	$\epsilon_{a,1}$	\dots	$\epsilon_{a,b}$
$y_{1,1}$	\dots	$y_{1,b}$																																												
$y_{2,1}$	\dots	$y_{2,b}$																																												
$y_{3,1}$	\dots	$y_{3,b}$																																												
\vdots		\vdots																																												
$y_{a,1}$	\dots	$y_{a,b}$																																												
τ_1																																														
τ_2																																														
τ_3																																														
\vdots																																														
τ_a																																														
β_1	\dots	β_b																																												
$\epsilon_{1,1}$	\dots	$\epsilon_{1,b}$																																												
$\epsilon_{2,1}$	\dots	$\epsilon_{2,b}$																																												
$\epsilon_{3,1}$	\dots	$\epsilon_{3,b}$																																												
\vdots		\vdots																																												
$\epsilon_{a,1}$	\dots	$\epsilon_{a,b}$																																												

- Assume additional additive effect due to block
Block effect the same for all EU's in block
- Treatment and block effects deviations from mean
Common parameter restrictions are then

$$\sum \tau_i = 0 \quad \sum \beta_j = 0$$

Partitioning the SS

- Rewrite observation as

$$y_{i,j} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{i,j} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

$$= \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \hat{\epsilon}_{i,j}$$
- Can partition $SS_{Tot} = \sum \sum (y_{i,j} - \bar{y}_{..})^2$ into

$$b \sum (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum \sum (y_{i,j} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

$$SS_{Treatment} + SS_{Block} + SS_{Error}$$
- Under H_0 , all SS/σ^2 independent χ^2 .
- Ratio of MS 's will be F distributed

Hypothesis Testing

- Can show (in the fixed case):

$$E(MS_E) = \sigma^2$$

$$E(MS_{Trt}) = \sigma^2 + b \sum \tau_i^2 / (a - 1) \text{ (no } \beta_j \text{'s)}$$

$$E(MS_{Blk}) = \sigma^2 + a \sum \beta_j^2 / (b - 1)$$

- Use F -test to test equality of treatment effects

$$F_0 = \frac{SS_{Trt} / (a - 1)}{SS_E / ((a - 1)(b - 1))}$$

- Caution when testing block effects

- Usually not of interest (blocked for a reason)
- Blocks not randomized to experimental units (restriction on randomization)
- Differing opinions on using F -test

Confounding may mean block effect not simply due to blocks.

This means a “significant” block effect does not show effect due to blocking.

- Can view ratio to see if blocking successful (what % explainable by blocking)

Analysis of Variance Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Blocks	SS_{Blk}	$b - 1$	MS_{Blk}	
Treatment	SS_{Trt}	$a - 1$	MS_{Trt}	F_0
Error	SS_E	$(b - 1)(a - 1)$	MS_E	
Total	SS_T	$ba - 1$		

- $SS_T = \sum \sum y_{i,j}^2 - y_{..}^2 / N$
- $SS_{Trt} = \frac{1}{b} \sum y_{i.}^2 - y_{..}^2 / N$
- $SS_{Blk} = \frac{1}{a} \sum y_{.j}^2 - y_{..}^2 / N$
- $SS_E = SS_T - SS_{Trt} - SS_{Blk}$

If $F_0 > F_{\alpha, a-1, (b-1)(a-1)}$, then reject H_0 .

Example

An experiment was designed to study the performance of four different detergents for cleaning clothes. The following “cleanness” readings (higher = cleaner) were obtained with specially designed equipment for three different types of common stains. Is there a difference among the detergents?

	Stain 1	Stain 2	Stain 3
Detergent 1	45	43	51
Detergent 2	47	46	52
Detergent 3	48	50	55
Detergent 4	42	37	49

$\sum \sum y_{i,j} = 565$ and $\sum \sum y_{i,j}^2 = 26867$
 $y_{1.} = 139, y_{2.} = 145, y_{3.} = 153,$ and $y_{4.} = 128$
 $y_{.1} = 182, y_{.2} = 176,$ and $y_{.3} = 207$

- $SS_T = \sum \sum y_{i,j}^2 - (\sum \sum y_{i,j})^2 / (ab) = 26867 - 565^2 / 12 \approx 265$
- $SS_{Trt} = (y_{1.}^2 + y_{2.}^2 + y_{3.}^2 + y_{4.}^2) / b - (\sum \sum y_{i,j})^2 / (ab) = (139^2 + 145^2 + 153^2 + 128^2) / 3 - 565^2 / 12 \approx 111$
- $SS_{Blk} = (y_{.1}^2 + y_{.2}^2 + y_{.3}^2) / a - (\sum \sum y_{i,j})^2 / (ab) = (182^2 + 176^2 + 207^2) / 4 - 565^2 / 12 \approx 135$
- $SS_E = SS_T - SS_{Trt} - SS_{Blk} \approx 265 - 111 - 135 = 19$
- $F_0 \approx (111/3) / (19/6) \approx 11.7$
- $p\text{-value} = 0.0063 < 0.01$ (Reject H_0 – There is a detergent difference)

Diagnostics

- Assumptions
 - Model is correct (additivity assumption)
 - Errors independent, normally distributed
 - Constant variance
- Normality
 - histogram, normal probability plot of residuals
- Variance and Unusual Observations
 - Residuals vs blocks
 - Residuals vs treatments
 - Residuals vs $\hat{y}_{i,j}$
- Additivity

- Residuals vs $\hat{y}_{i,j}$
- If curvilinear, interaction likely exists
- Interaction: Block effect different for different treatments
- Formal test for interaction: *Tukey's Test of Non-additivity*
- Usually transform to eliminate interaction

Example:

$$E(y_{i,j}) = \mu\tau_i\beta_j$$

$$\log(y_{i,j}) \approx \mu' + \tau'_i + \beta'_j + \epsilon_{i,j} \text{ (Additive Model)}$$

Comparisons of Treatments

- Multiple Comparisons/Contrasts
 - Similar procedures to 1-factor case
 - n is replaced by b in formulas
 - Degrees of freedom error is $(b - 1)(a - 1)$

- Example: Comparison of detergents

- Duncan's Multiple Range ($\alpha = 0.05$)

6 degrees of freed error $\rightarrow r_\alpha = (3.46, 3.58, 3.64)$

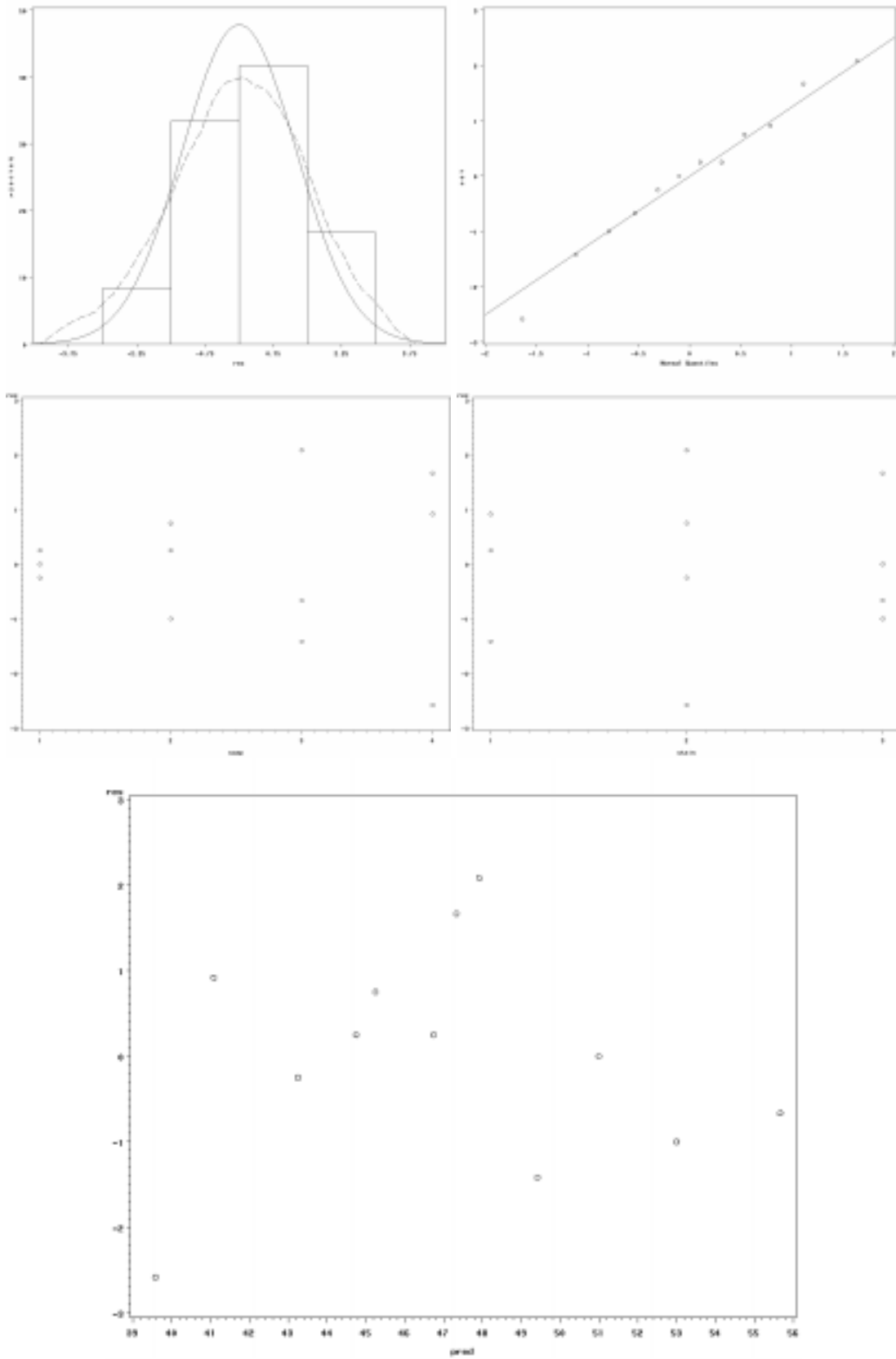
$$s_{\bar{y}} = \sqrt{MSE/3} = \sqrt{(19/6)/3} = 1.03$$

Least Significant ranges are $1.03r_\alpha = (3.56, 3.68, 3.74)$

Using SAS

```
options nocenter ls=78;
goptions colors=(none);
symbol1 v=circle; axis1 offset=(5);
data wash;
  input stain soap y @@;
  cards;
1 1 45 1 2 47 1 3 48 1 4 42
2 1 43 2 2 46 2 3 50 2 4 37
3 1 51 3 2 52 3 3 55 3 4 49
;

proc glm;
  class stain soap;
  model y = soap stain;
  means soap / duncan lines; /* Show Duncan Multiple Comparisons */
  output out=diag r=res p=pred;
```

Randomization/Hypothesis Tests

- Usually assume treatments are *assigned*, blocks are *allocated*
- Permute labels *within* blocks (see paired *t*-test example)
- Limiting results are same
- Permutation test different with interaction.

Regression Model

- Regression model simple extension of CRD
- Add additional $b - 1$ columns to represent block
- Block columns orthogonal to treatment columns
- Thus, order of fit does not matter

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 & -1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

Missing Values

- When missing
 - Orthogonality lost - missing row in X
 - Design unbalanced \rightarrow order of fit important
- Procedures
 1. Exact (Regression) approach
 - Use Type III SS's (general regression significance test)
 2. Approximate approach: Estimate missing value

Choose value to minimize SS_E (minimum contribution)
 Take derivative and set equal to zero

Suppose $y_{i_0, j_0} \equiv x$ is missing.

$$\begin{aligned} SS_E &= \sum \sum y_{i,j}^2 - y_{..}^2/ab - \frac{1}{b} \sum y_{i.}^2 + y_{..}^2/ab - \frac{1}{a} \sum y_{.j}^2 + y_{..}^2/ab \\ &= x^2 - \frac{1}{b}(y'_{i_0.} + x)^2 - \frac{1}{a}(y'_{.j_0} + x)^2 + \frac{1}{ab}(y'_{..} + x)^2 + R, \end{aligned}$$

where $y'_{i_0.}$ is the treatment sum with y_{i_0, j_0} missing, etc.

$$x_{min} = \frac{ay'_{i_0.} + by'_{.j_0} - y'_{..}}{(a-1)(b-1)}$$

Example

- Consider detergent comparison example
- Suppose $y_{4,2} = 37$ is missing
- Regression approach
 - Same procedure in SAS: look at Type III result
- Estimate approach

$$y'_{4.} = 91, y'_{.2} = 528, y'_{.2} = 139$$

– Estimate is

$$x = \frac{4(91) + 3(139) - 528}{6} = 42.17$$

– Do analysis but adjust error degrees of freedom

- Regression: $\hat{\sigma}^2 = 1.097$
- Estimate: $\hat{\sigma}^2 = 1.097$ (must divide by 5, not 6)

```
options nocenter ps=60 ls=78;
goptions colors=(none);
symbol1 v=circle; axis1 offset=(5);
```

```
data wash;
  input stain soap y @@;
  if y=37 then y=.;
  cards;
1 1 45 1 2 47 1 3 48 1 4 42
2 1 43 2 2 46 2 3 50 2 4 37
```

```

3 1 51 3 2 52 3 3 55 3 4 49
;

proc glm; /* Regression Approach */
  class stain soap;
  model y=soap stain;
  output out=diag p=pred r=res;
  means soap / lsd lines;
  lsmeans soap / stderr;

data new1; /* Input estimate */
  set wash;
  if y=. then y=42.166666666;

proc glm; /* Analysis of adjusted data*/
  class stain soap;
  model y=soap stain;
  output out=diag p=pred r=res;
  means soap / lsd lines;
run;

```

Regression – Type III

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	148.5138889	29.7027778	27.07	0.0013
Error	5	5.4861111	1.0972222		
Corrected Total	10	154.0000000			

R-Square Coeff Var Root MSE y Mean
0.964376 2.182258 1.047484 48.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
soap	3	48.1666667	16.0555556	14.63	0.0066
stain	2	100.3472222	50.1736111	45.73	0.0006

Source	DF	Type III SS	Mean Square	F Value	Pr > F
soap	3	58.9305556	19.6435185	17.90	0.0042
stain	2	100.3472222	50.1736111	45.73	0.0006

t

Grouping	Mean	N	soap
A	51.0000	3	3
B	48.3333	3	2
B			
C	46.3333	3	1
C			

C 45.5000 2 4

Least Squares Means

soap	y LSMEAN	Standard Error	Pr > t
1	46.3333333	0.6047650	<.0001
2	48.3333333	0.6047650	<.0001
3	51.0000000	0.6047650	<.0001
4	44.3888889	0.7807483	<.0001

Estimate – Must adjust F by hand

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	179.7060185	35.9412037	39.31	0.0002
Error	6	5.4861111	0.9143519		
Corrected Total	11	185.1921296			

R-Square Coeff Var Root MSE y Mean
 0.970376 2.012501 0.956217 47.51389

Source	DF	Type I SS	Mean Square	F Value	Pr > F
soap	3	71.9513889	23.9837963	26.23	0.0008
stain	2	107.7546296	53.8773148	58.92	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
soap	3	71.9513889	23.9837963	26.23	0.0008
stain	2	107.7546296	53.8773148	58.92	0.0001

LSD Test ** Must use correct error in comparisons

	Mean	N	soap
A	51.0000	3	3
B	48.3333	3	2
C	46.3333	3	1
D	44.3889	3	4 ** Really only two ‘‘observations’’ here

$$F_0 = \frac{71.95/3}{5.49/5}$$

$$= 21.84$$

$$p\text{-value} = 0.0027$$

Tukey’s Test for Non-additivity

- Consider possibility of interaction (if no replicates, then one df)
- Assume the following model

$$y_{i,j} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \epsilon_{i,j}$$

- Use regression approach to test significance
- Procedure to test if $\gamma = 0$
 1. Fit additive model $y_{i,j} = \mu + \tau_i + \beta_j + \epsilon_{i,j}$
 2. Obtain $\hat{y}_{i,j}$ and $r_{i,j} = y_{i,j} - \hat{y}_{i,j}$.
 3. Fit additive model $y_{i,j} = \gamma \hat{y}_{i,j}^2 / (2\bar{y}_{..}) + \mu + \tau_i + \beta_j + \epsilon_{i,j}$.
 4. Tukey sum of squares is $\sum r_{i,j}^2 - \sum r'_{i,j}{}^2$, where $r'_{i,j}$ are residuals from refitted model.
 5. Test MS (Tukey SS divided by MS_E of refitted model) against appropriate F .
 6. Use regression to estimate γ . Transformation based on the power $1 - \hat{\eta}$ (where $\eta = \gamma/\mu$) removes interaction.

Example 5-2 from Montgomery

- Impurity in chemical product is affected by temperature and pressure. We will assume temperature is the blocking factor. The data is shown below. We test for non-additivity.

	Pressure				
Temp	25	30	35	40	45
100	5	4	6	3	5
125	3	1	4	2	3
150	1	1	3	1	2

- Can use SAS to compute SS .
- Must divide by proper degrees of freedom

$$F_0 = \frac{0.0985/1}{1.9015/7} = 0.36$$

$F_0 < F_{1,7}$ – Do not reject

SAS Procedures

```
options nocenter ls=75;
```

```
data impurity;
input trt blk y @@;
cards;
1 1 5 1 2 3 1 3 1 2 1 4 2 2 1 2 3 1
3 1 6 3 2 4 3 3 3 4 1 3 4 2 2 4 3 1
5 1 5 5 2 3 5 3 2
```

```

;

proc univariate; /* Get mean of y for Step 2 */
  var y;

proc glm;
  class blk trt;
  model y = blk trt; /* Step 1 */
  output out=resid1 r=res1 p=pred1; /* Step 2 */

data resid1;
  set resid1;
  pred1sq = pred1*pred1/2/2.933333; /* 2.933 is mean from proc univariate */

proc glm;
  class blk trt;
  model y = blk trt pred1sq / solution;
  output out=resid2 r=res2 p=pred2; /* Steps 3-5 */

```

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	34.93333333	5.82222222	23.29	0.0001
Error	8	2.00000000	0.25000000		
Corrected Total	14	36.93333333			

R-Square	Coeff Var	Root MSE	y Mean
0.945848	17.04545	0.500000	2.933333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
blk	2	23.33333333	11.66666667	46.67	<.0001
trt	4	11.60000000	2.90000000	11.60	0.0021

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	35.03185550	5.00455079	18.42	0.0005
Error	7	1.90147783	0.27163969		
Corrected Total	14	36.93333333			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
blk	2	23.33333333	11.66666667	42.95	0.0001
trt	4	11.60000000	2.90000000	10.68	0.0042
pred1sq	1	0.09852217	0.09852217	0.36	0.5660

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	1.812807882 B	0.47262874	3.84	0.0064
pred1sq	0.216748744	0.35990363	0.60	0.5660

- p -value for testing $\gamma = 0$: 0.5660

- Can transform y 's to get additivity: $y^{1-\eta}$

$$1 - \hat{\eta} = 1 - 0.2167 = 0.783$$

Random Block/Treatment Effects

- Could randomly select treatments and/or blocks
- Do not need to worry about interaction
- Interaction considered random effect
- Interaction variance appears in all EMS 's.
- Perform usual F -test (ratio of MS 's)
- Use `proc mixed` instead of `proc glm`
- Otherwise, underestimate variability in treatment means

```
options nocenter ls=78; goptions colors=(none);
```

```
data wash;
```

```
input stain soap y @@;
```

```
cards;
```

```
1 1 45 1 2 47 1 3 48 1 4 42
```

```
2 1 43 2 2 46 2 3 50 2 4 37
```

```
3 1 51 3 2 52 3 3 55 3 4 49
```

```
;
```

```
proc glm;
```

```
class stain soap;
```

```
model y = soap stain;
```

```
random stain;
```

```
lsmeans soap / stderr tdiff;
```

```
proc mixed;
```

```
class stain soap;
```

```
model y = soap;
```

```
random stain;
```

```
lsmeans soap / diff;
```

```
run;
```

The GLM Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	246.0833333	49.2166667	15.68	0.0022
Error	6	18.8333333	3.1388889		
Corrected Total	11	264.9166667			

R-Square Coeff Var Root MSE y Mean

0.928908 3.762883 1.771691 47.08333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
soap	3	110.9166667	36.9722222	11.78	0.0063
stain	2	135.1666667	67.5833333	21.53	0.0018

Least Squares Means

soap	y LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	46.3333333	1.0228863	<.0001	1
2	48.3333333	1.0228863	<.0001	2
3	51.0000000	1.0228863	<.0001	3
4	42.6666667	1.0228863	<.0001	4

Least Squares Means for Effect soap
t for H0: LSmean(i)=LSmean(j) / Pr > |t|

Dependent Variable: y

i/j	1	2	3	4
1		-1.38257 0.2161	-3.226 0.0180	2.534715 0.0444
2	1.382572 0.2161		-1.84343 0.1148	3.917286 0.0078
3	3.226001 0.0180	1.843429 0.1148		5.760715 0.0012
4	-2.53471 0.0444	-3.91729 0.0078	-5.76072 0.0012	

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Estimate
stain	16.1111
Residual	3.1389

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
soap	3	6	11.78	0.0063

Least Squares Means

Effect	soap	Estimate	Standard Error	DF	t Value	Pr > t
soap	1	46.3333	2.5331	6	18.29	<.0001
soap	2	48.3333	2.5331	6	19.08	<.0001
soap	3	51.0000	2.5331	6	20.13	<.0001
soap	4	42.6667	2.5331	6	16.84	<.0001

Differences of Least Squares Means

Effect	soap	_soap	Estimate	Standard Error	DF	t Value	Pr > t
--------	------	-------	----------	----------------	----	---------	---------

soap	1	2	-2.0000	1.4466	6	-1.38	0.2161
soap	1	3	-4.6667	1.4466	6	-3.23	0.0180
soap	1	4	3.6667	1.4466	6	2.53	0.0444
soap	2	3	-2.6667	1.4466	6	-1.84	0.1148
soap	2	4	5.6667	1.4466	6	3.92	0.0078
soap	3	4	8.3333	1.4466	6	5.76	0.0012

Choice of Sample Size

- Same as determining number of blocks (b)
- Use same tables/procedures with b replacing n .

RCBD with Replicates

- What if multiple observations per treatment block?
 - b blocks
 - a treatments
 - n observations/treatment block

$$y_{i,j,k} = \mu + \tau_i + \beta_j + \epsilon_{i,j,k} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

- When would this occur?
 - Have large field with very gradual slope
 - Blocks expensive but observations cheap
- Increases df_E (or allows interaction)
- Usual diagnostics checks
- Replace b with bn in multiple comparisons or power
- Allows for easier assessment of additivity
 - More error degrees of freedom
 - Interaction and error not confounded
 - Can separate error and interaction SS

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Blocks	SS_{Blk}	$b - 1$	MS_{Blk}	
Treatment	SS_{Trt}	$a - 1$	MS_{Trt}	F_0
Error	SS_E	$abn - b - a + 1$	MS_E	
Total	SS_T	$abn - 1$		

or

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Blocks	SS_{Blk}	$b - 1$	MS_{Blk}	
Treatment	SS_{Trt}	$a - 1$	MS_{Trt}	F_0
Blk \times Trt	$SS_{Blk*Trt}$	$(b - 1)(a - 1)$	$MS_{Blk*Trt}$	
Error	SS_E	$ab(n - 1)$	MS_E	
Total	SS_T	$abn - 1$		

Example

You have been asked to design an experiment to compare four varieties of seed corn. You have a field consisting of sixteen subplots (in a 4×4 grid) at your disposal. If you were told that one side of the field is next to a highway and the side directly across from this one is next to a river, how would you design the experiment?

If we feel pretty certain that plots near the road or river will “behave” differently than plots in the middle, we might want to create $b = 3$ blocks. Block 1 consists of the four plots along the road. Block 2 consists of the 4 plots along the river, and Block 3 consists of the eight plots in the middle. Thus, we have two blocks which only have $n = 1$ observation per treatment and one block that has $n = 2$ observations per treatment.

- Statistical model is (with interaction)

$$y_{i,j,k} = \mu + \tau_i + \beta_j + (\tau\beta)_{i,j} + \epsilon_{i,j,k} \begin{cases} i = 1, 2, 3, 4 \\ j = 1, 2, 3 \\ k = 1, \dots, n_j, \end{cases}$$

$$\text{where } n_j = \begin{cases} 1 & \text{if } j = 1, 2 \\ 2 & \text{if } j = 3. \end{cases}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Blocks	SS_{Blk}	2	MS_{Blk}	
Treatment	SS_{Trt}	3	MS_{Trt}	F_0
Blk \times Trt	$SS_{Blk*Trt}$	6	$MS_{Blk*Trt}$	
Error	SS_E	4	MS_E	
Total	SS_T	15		

- If four blocks, cannot separate error and interaction
- SS_E based on observations within block 3 (has replicates)

How well did blocking work?

- F -test not valid since blocks not randomized
- Real Question: How much is saved using blocks as RCB, instead of CRD?
- Trade-off: reduction in variance vs loss in degrees of freedom

Relative efficiency

What sample size of CRD is needed to get same amount of information as RCBD?

$$RE_{RCB:CRD} = \frac{(\nu_{RCB} + 1)(\nu_{CRD} + 3)\sigma_{CRD}^2}{(\nu_{RCB} + 3)(\nu_{CRD} + 1)\sigma_{RCB}^2},$$

where

- $\sigma_{CRD}^2, \sigma_{RCB}^2$ – error variances
- $\nu_{RCB} = (a - 1)(b - 1)$ – error degrees of freedom for RCB
- $\nu_{CRD} = a(b - 1)$ – error degrees of freedom of same size

$$\hat{\sigma}_{CRD}^2 = \frac{(b - 1)MS_{Blk} + b(a - 1)MS_E}{ab - 1}$$

Ultimately, the loss in degrees of freedom will have little effect as long as a moderate number of error degrees of freedom are available.

Detergent Comparison Example

$$\hat{\sigma}_{CRD}^2 = \frac{2 \times 67.583 + 3 \times 3 \times 3.139}{11} \approx 14.9$$
$$RE = \frac{(6 + 1)(8 + 3)14.9}{(6 + 3)(8 + 1)3.139} \approx 4.50$$

... implies that approximately four and a half times as many replicates would be needed with a completely randomized design to obtain the same sensitivity as is obtained by blocking on stain types.

Factorial Designs

- Treatments often different levels of one factor
- What if interested in combinations of two factors
 - Temperature and Pressure

- Seed variety and Fertilizer
- Diet and Exercise Regime
- Could treat each combination as treatment and do ANOVA
 - A levels of factor A and b levels of factor B
 - ab total treatments, each with n observations
 - Use contrasts to study specific effects

Experiment

An experiment is conducted to study the effect of hormones injected into test rats. There are two distinct hormones – A, B – each with two distinct levels. For purposes here, we will consider this to be four different treatments labeled $\{A, a, B, b\}$. Each treatment is applied to six rats with the response being the amount of glycogen (in mg) in the liver.

Treatment	Responses					
A	106	101	120	86	132	97
a	51	98	85	50	111	72
B	103	84	100	83	110	91
b	50	66	61	72	85	60

Three contrasts are of interest. They are

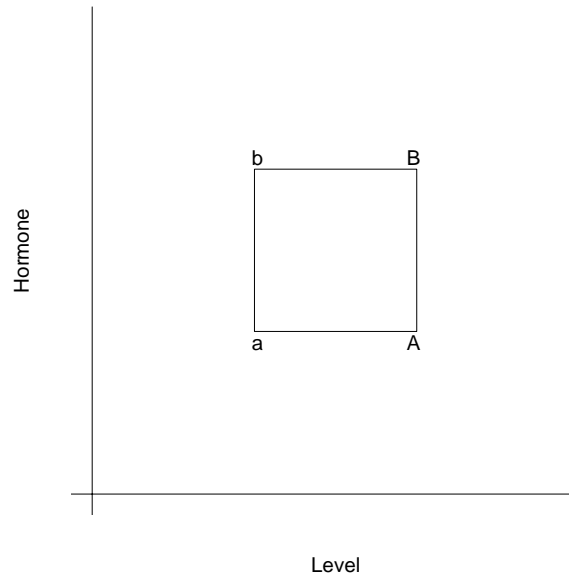
Comparison	A	a	B	b
Hormone A vs Hormone B	1	1	-1	-1
Low level vs High level	1	-1	1	-1
Equivalence of level effect	1	-1	-1	1

Can we redo the experiment in such a way that the contrast sums of squares are separated already?

Factorial Experiment

- Break up treatments into the two factors (two levels each)
- Known as 2^2 factorial
 - Factor A: Hormone (A or B)
 - Factor B: Level (L or H)
- Factorial – investigates all combinations of factors
- Single replicate of factorial involves ab trials
- Design illustrated as table or graphically

	Level	
Hormone	Low	High
A	XXXXXX	XXXXXX
B	XXXXXX	XXXXXX



Statistical Model (Two Factors)

- Statistical model is

$$y_{i,j,k} = \mu + \tau_i + \beta_j + (\tau\beta)_{i,j} + \epsilon_{i,j,k} \begin{cases} i = 1, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

μ – grand mean

τ_i – i th level effect of factor A (ignores B)

β_j – j th level effect of factor B (ignores A)

$(\tau\beta)_{i,j}$ – interaction effect of combination (i, j)

Explains variation not described by mean effects

$\epsilon_{i,j,k} \sim N(0, \sigma^2)$

- Over-parametrized model
- Must include $a + b + 1$ model constraints. Typically

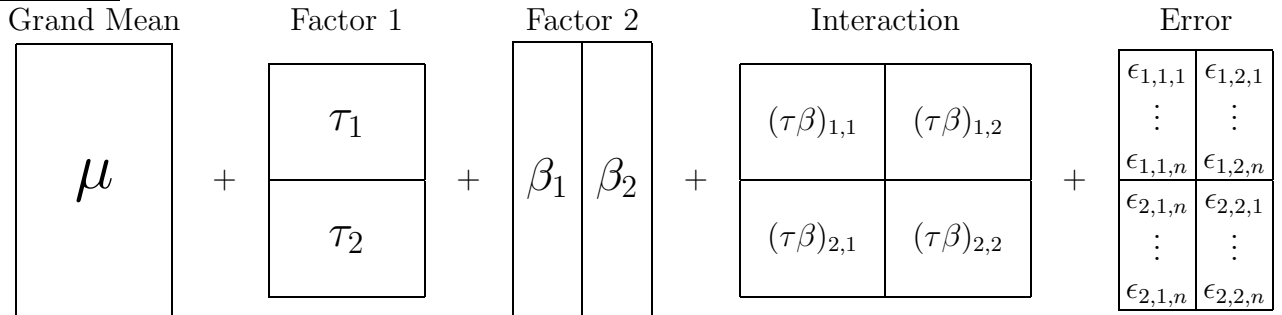
$$\sum_i \tau_i = 0, \sum_j \beta_j = 0, \sum_i (\tau\beta)_{i,j} = 0, \sum_j (\tau\beta)_{i,j} = 0$$

Picture of 2^2 Model

Observation

$y_{1,1,1}$	$y_{1,2,1}$
\vdots	\vdots
$y_{1,1,n}$	$y_{1,2,n}$
$y_{2,1,1}$	$y_{2,2,1}$
\vdots	\vdots
$y_{2,1,n}$	$y_{2,2,n}$

=



Model Estimates

- Previous constraints result in estimates

$$\begin{aligned}\hat{\mu} &= \bar{y}_{...} \\ \hat{\tau}_i &= \bar{y}_{i..} - \bar{y}_{...} \\ \hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...} \\ (\hat{\tau\beta})_{i,j} &= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\end{aligned}$$

- The predicted values is (i, j) combination average, so $\hat{y}_{i,j,k} = \bar{y}_{ij.}$ and $\hat{\epsilon}_{i,j,k} = y_{i,j,k} - \bar{y}_{ij.}$
- Note: From model (ignoring parameter constraints)

$$\begin{aligned}\mathbb{E}(\bar{y}_{i..}) &= \mu + \tau_i + \bar{\beta}_{.} + (\bar{\tau\beta})_{i.} \\ \mathbb{E}(\bar{y}_{i..} - \bar{y}_{i'..}) &= \tau_i - \tau_{i'} + (\bar{\tau\beta})_{i.} - (\bar{\tau\beta})_{i'..}\end{aligned}$$

Difference in treatment effects depends on $(\tau\beta)$ constraints
 $\tau_i - \tau_{i'}$ is non-estimable (if interaction included in model)

- Using typical constraints

$$\begin{aligned}\mathbb{E}(\bar{y}_{i..}) &= \mu + \tau_i \\ \mathbb{E}(\bar{y}_{i..} - \bar{y}_{i'..}) &= \tau_i - \tau_{i'}\end{aligned}$$

- Caution if interaction present
- Should always test interaction first

Partitioning the Sum of Squares

- Rewrite observation as:

$$y_{i,j,k} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (y_{i,j,k} - \bar{y}_{i,j.})$$

- Can look at $\sum (y_{i,j,k} - \bar{y}_{...})^2$
- RHS simplifies to

$$\begin{aligned} & bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + \\ & an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 + \\ & n \sum_i \sum_j (\bar{y}_{i,j.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \dots \end{aligned}$$

- $SS_A + SS_B + SS_{AB} + SS_E$
- Under normality, all SS/σ^2 independent

Hypothesis Testing

- Can show (fixed case)

$$\begin{aligned} E(MS_E) &= \sigma^2 \\ E(MS_A) &= \sigma^2 + bn \sum \tau_i^2 / (a - 1) \\ E(MS_B) &= \sigma^2 + an \sum \beta_j^2 / (b - 1) \\ E(MS_{AB}) &= \sigma^2 + n \sum (\tau\beta)_{i,j}^2 / (a - 1)(b - 1) \end{aligned}$$

- Use F -test to test equality of A, B, and AB effects

$$\begin{aligned} F_0 &= \frac{SS_A / (a - 1)}{SS_E / (ab(n - 1))} \\ F_0 &= \frac{SS_B / (b - 1)}{SS_E / (ab(n - 1))} \\ F_0 &= \frac{SS_{AB} / (a - 1)(b - 1)}{SS_E / (ab(n - 1))} \end{aligned}$$

Calculations

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Factor A	SS_A	$a - 1$	MS_A	F_0
Factor B	SS_B	$b - 1$	MS_B	F_0
Interaction	SS_{AB}	$(a - 1)(b - 1)$	MS_{AB}	F_0
Error	SS_E	$ab(n - 1)$	MS_E	
Total	SS_T	$abn - 1$		

$$\begin{aligned}
SS_T &= \sum \sum y_{i,j,k}^2 - y_{...}^2/abn \\
SS_A &= \frac{1}{bn} \sum y_{i..}^2 - y_{...}^2/abn \\
SS_B &= \frac{1}{an} \sum y_{.j.}^2 - y_{...}^2/abn \\
SS_{Submodel} &= \frac{1}{n} \sum y_{ij.}^2 - y_{...}^2/abn \\
SS_{AB} &= SS_{Sub} - SS_A - SS_B \\
SS_E &= \text{Subtraction}
\end{aligned}$$

$df_E > 0$ only if $n > 1$. When $n = 1$, cannot separate interaction from error (confounded). Recall typically RCBD uses $n = 1$. Assuming no interaction allows us to estimate error and test for treatment differences.

Rat Hormone Example

$$\begin{aligned}
\sum \sum y_{i,j,k} &= 2074 \text{ and } \sum \sum \sum y_{i,j,k}^2 = 191022 \\
y_{1,2,.} &= 642, y_{2,2,.} = 571, y_{1,1,.} = 467, \text{ and } y_{2,1,.} = 394
\end{aligned}$$

$$\begin{aligned}
SS_T &= 191022 - 2074^2/24 = 11793.83 \\
SS_A &= (1109^2 + 965^2)/12 - 2074^2/24 = 864.00 \\
SS_B &= (861^2 + 1213^2)/12 - 2074^2/24 = 5162.67 \\
SS_{Sub} &= (642^2 + 467^2 + 571^2 + 394^2)/6 - 2074^2/24 = 6026.83 \\
SS_{AB} &= 6026.83 - 5162.67 - 864.00 = 0.016 \\
SS_E &= 11793.83 - 6026.83 = 5767.00 \\
F_0^{AB} &= (0.16/1)/(5767/20) \approx 0 \text{ (Not Significant)} \\
F_0^A &= (864/1)/(5767/20) \approx 3.0 \text{ (Not Significant)} \\
F_0^B &= (5162.67/1)/(5767/20) \approx 17.9 \text{ (Significant)}
\end{aligned}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Hormone	864.00	1	864	2.99
Level	5162.67	1	5163	17.9
Interaction	0.016	1	0.016	5.5×10^{-5}
Error	5767.00	20	288.35	
Total	11793.83	23		

Example: Comparing Factors

- The mean at each combination is

	Hormone	
Level	A	B
L	77.83	65.67
H	107.00	95.17

- Can look at one factor by averaging out other
- Only interpretable if no interaction
- Comparing Hormone A to Hormone B
 - Average out the level

$$\frac{(77.83 + 107.00)}{2} - \frac{(65.677 + 95.17)}{2} \approx 12.0$$

- Comparing Low Level to High Level
 - Average out the hormone

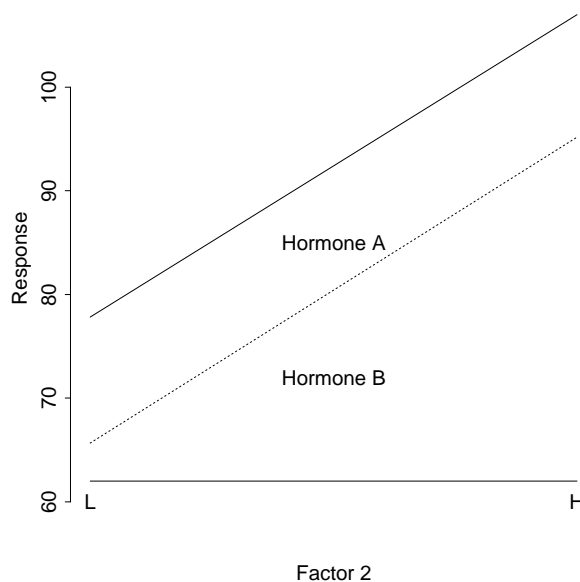
$$\frac{(77.83 + 65.67)}{2} - \frac{(107.00 + 95.17)}{2} \approx -29.335$$

Two Factor Experiment

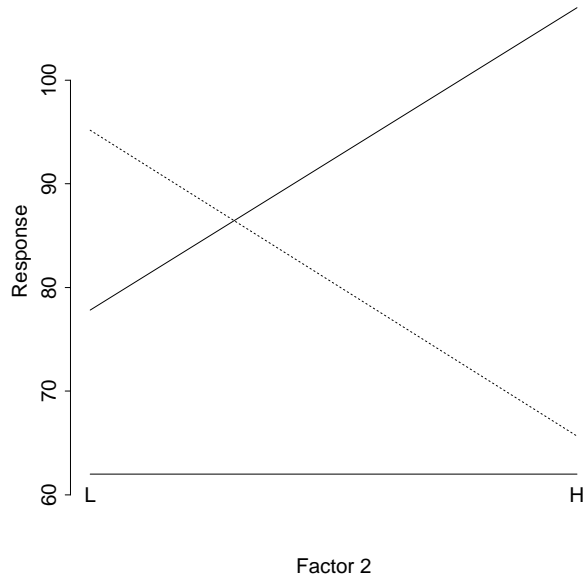
- Interaction
 - Difference in response of one factor not constant over other factor levels.
 - Difference in level of Hormone A vs Difference in level of Hormone B

$$(107.00 - 77.83) - (95.17 - 65.67) \approx -0.33$$

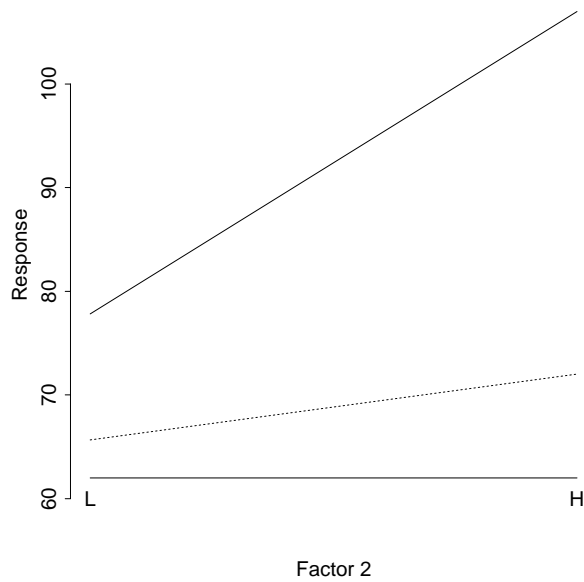
- Interaction Plot
 - Can view changes in response difference
- Are lines parallel or not?



- Completely opposite behavior (no Factor 2 effect?)



- Increase but not same amount (still Factor 2 effect?)



Regression/Response Surface Approach

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2 + \epsilon$$

- $x_1 = 1$ if Hormone A and $x_1 = -1$ if Hormone B

- $x_2 = 1$ if High Level and $x_2 = -1$ if Low Level

- $x_1x_2 = \begin{cases} 1 & \text{Hormone A/High Level} \\ & \text{Hormone B/Low Level} \\ -1 & \text{Hormone A/Low Level} \\ & \text{Hormone B/High Level} \end{cases} \quad \begin{matrix} \text{or} \\ \\ \text{or} \end{matrix}$

- Trt A: $E(y) = \beta_0 + \beta_1 + \beta_2 + \beta_{1,2}$

- Trt a: $E(y) = \beta_0 + \beta_1 - \beta_2 - \beta_{1,2}$

- Trt B: $E(y) = \beta_0 - \beta_1 + \beta_2 - \beta_{1,2}$

- Trt b: $E(y) = \beta_0 - \beta_1 - \beta_2 + \beta_{1,2}$

- β_1 estimates Hormone effect

- β_2 estimates Level effect

- $\beta_{1,2}$ estimates Interaction

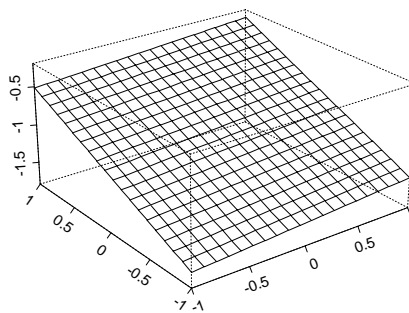
- Response surface described by $-1 \leq x_i \leq 1$

- If $\beta_{1,2} = 0$, then surface is a plane (additive model)

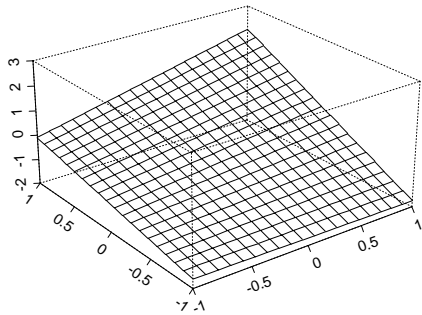
- If $\beta_{1,2} \neq 0$, then surface “curved” (pages 162-163)

- Recall non-additivity residual plot diagnostic

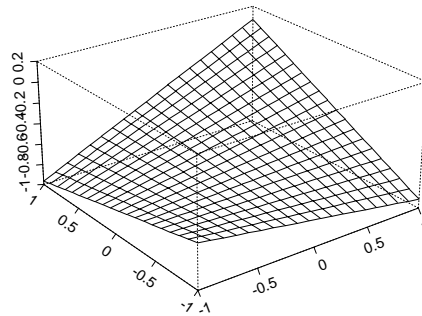
Response Surface without Interaction



Response Surface with Positive Interaction

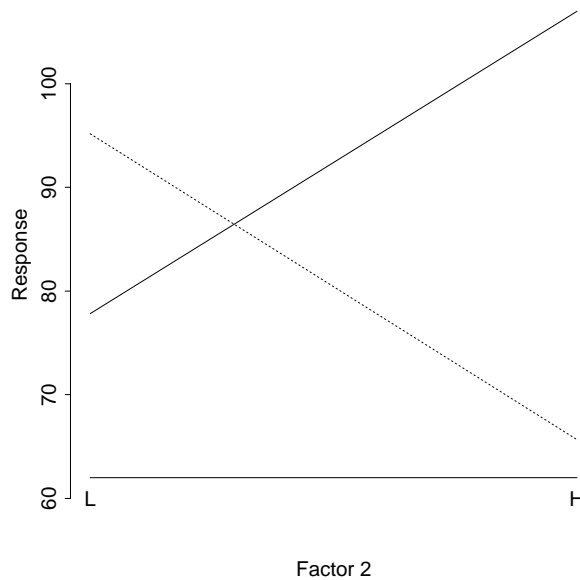


Response Surface with Negative Interaction



Interaction in Two Factor Experiment

- If interaction, how to interpret main effects
 - In regression, if $\beta_{1,2} \neq 0$, leave x_1 and x_2 in model
 - Main effect result depends on level of other factor
 - Cannot average over other factor
 - Compare $\bar{y}_{i,j}$'s instead
 - For example, if opposite behavior, effects will cancel out



- May still be able to discuss main effects (but not estimate)
- Sometimes interaction due to only a few combos

Hidden Replication Feature: Main Effects

- One factor approach
 - Can estimate main effects using 3 combinations (A, a, b)
 - Main effect Level: $y_A - y_a$, Main effect Hormone: $y_a - y_b$
 - Must replicate to have variance estimate
 - Cannot estimate interaction without 4th combination
 - If $n = 2$ so $N = 6$, $\text{Var}(\text{Effect}) = 2\sigma^2/n = \sigma^2$
- Factorial approach
 - Estimate effects using $N = 4$ observations (A, a, B, b)
 - Have variance estimate if no interaction
 - Main effect Hormone: $0.5(y_A + y_a) - 0.5(y_B + y_b)$
 - $\text{Var}(\text{Effect}) = \sigma^2$
 - Replication provide ability to estimate interaction

Factorial gives same accuracy with less ($N = 6$ vs $N = 4$)

Diagnostics

Assumptions

1. Error independent, normally distributed
2. Constant variance
3. No interaction (if $n = 1$)

Tukey's Test of Nonadditivity

Normality

Histogram or QQplot of residuals

Variance and Unusual Observations

Residuals vs factor A

Residuals vs factor B

Residuals vs $\hat{y}_{i,j}$.

Multiple Comparisons

- If no interaction
 - Compare means of each factor (average over other factor)
 - Similar procedures as CRD
 - an or bn observations used in calculations
 - Degrees of freedom error is $ab(n - 1)$
 - Use `proc means` or `lsmeans` in SAS
- If interaction significant
 - Construct interaction plot for visual description
 - Can compare means of A at fixed level of B and vice-versa
 - n observations used in calculations
 - Degrees of freedom error is $ab(n - 1)$
 - `means` statement only does main effect comparisons
 - `lsmeans` statement has `slice` option
 - Can also do comparison by hand

Using SAS

```
options nocenter ls=78;
goptions colors=(none);

data new;
infile 'h:\System\Desktop\glycogen.dat';
input trt $ resp; /* $ */

/* Transform data into two factors */
if trt="a" then horm=1; if trt="a" then level=1;
if trt="A" then horm=1; if trt="A" then level=2;
if trt="b" then horm=2; if trt="b" then level=1;
if trt="B" then horm=2; if trt="B" then level=2;

/* Break down one-way analysis by contrasts */
proc glm;
class trt;
model resp = trt;
contrast 'level' trt 1 1 -1 -1;
contrast 'hormone' trt 1 -1 1 -1;
contrast 'interaction' trt 1 -1 -1 1;
```

```

proc glm;
  class level horm;
  /* Do anova with interactions */
  model resp = level horm level*horm / solution;
  /* Gives (adjusted) t-tests */
  means level|horm / tukey lines;
  /* Shows means at each level */
  lsmeans level horm level*horm / adjust=tukey tdiff;
  output out=new1 r=res p=pred;

  /* Normal diagnostics */
proc univariate noprint;
  hist res / normal (L=1 mu=0 sigma=est) kernel (L=2);

  /* Residuals diagnostics */
symbol1 v=circle; axis1 offset=(5);
proc gplot;
  plot res*pred res*level res*horm / haxis=axis1;

run;

```

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6026.83333	2008.94444	6.97	0.0022
Error	20	5767.00000	288.35000		
Corrected Total	23	11793.83333			

R-Square	Coeff Var	Root MSE	resp Mean
0.511016	19.65000	16.98087	86.41667

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
level	1	5162.666667	5162.666667	17.90	0.0004
hormone	1	864.000000	864.000000	3.00	0.0989
interaction	1	0.166667	0.166667	0.00	0.9811

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6026.83333	2008.94444	6.97	0.0022
Error	20	5767.00000	288.35000		
Corrected Total	23	11793.83333			

R-Square	Coeff Var	Root MSE	resp Mean
0.511016	19.65000	16.98087	86.41667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
level	1	5162.666667	5162.666667	17.90	0.0004
horm	1	864.000000	864.000000	3.00	0.0989

level*horm	1	0.166667	0.166667	0.00	0.9811
------------	---	----------	----------	------	--------

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	288.35
Critical Value of Studentized Range	2.95000
Minimum Significant Difference	14.461

Means with the same letter are not significantly different.

	Mean	N	level
A	101.083	12	2
B	71.750	12	1
	Mean	N	horm
A	92.417	12	1
A	80.417	12	2

Level of	Level of		-----resp-----	
level	horm	N	Mean	Std Dev
1	1	6	77.833333	24.8468643
1	2	6	65.666667	11.9443152
2	1	6	107.000000	16.5650234
2	2	6	95.166667	10.9071842

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

		H0:LSMean1=LSMean2	
level	resp LSMEAN	t Value	Pr > t
1	71.750000	-4.23	0.0004
2	101.083333		

		H0:LSMean1=LSMean2	
horm	resp LSMEAN	t Value	Pr > t
1	92.4166667	1.73	0.0989
2	80.4166667		

		LSMEAN	
level	horm	resp LSMEAN	Number
1	1	77.833333	1
1	2	65.666667	2
2	1	107.000000	3
2	2	95.166667	4

Least Squares Means for Effect level*horn
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

i/j	Dependent Variable: resp			
	1	2	3	4
1		1.241001 0.6091	-2.975 0.0347	-1.768 0.3171
2	-1.241 0.6091		-4.216 0.0022	-3.009 0.0323
3	2.975003 0.0347	4.216005 0.0022		1.207001 0.6297
4	1.768002 0.3171	3.009003 0.0323	-1.207 0.6297	

What if Interaction Present?

A grape grower is interested in maximizing the number of bushels/acre on her winery. She limits her study to the combinations of 3 varieties and 4 pesticides (12 combinations). For each combination, two replicates will be obtained.

$$y_{i,j,k} = \mu + v_i + p_j + (vp)_{i,j} + \epsilon_{i,j,k}$$

If there is an interaction, the grower wants to compare the pesticides for each fixed variety of grape. In other words, the grower wants to compare $\bar{y}_{i,1}$, $\bar{y}_{i,2}$, $\bar{y}_{i,3}$, and $\bar{y}_{i,4}$. Must compute appropriate standard error.

$$\begin{aligned} \bar{y}_{i,j} &= \mu + v_i + p_j + (vp)_{i,j} + \bar{\epsilon}_{i,j}. \\ \bar{y}_{i,1} - \bar{y}_{i,2} &= p_1 - p_2 + (vp)_{i,1} - (vp)_{i,2} + \bar{\epsilon}_{i,1} - \bar{\epsilon}_{i,2}. \\ \text{Var}(\bar{y}_{i,1} - \bar{y}_{i,2}) &= \sigma^2/n + \sigma^2/n = 2\sigma^2/n \end{aligned}$$

Could do calculations by hand, use `lsmean`, or `lsmean slice` option.

```
options nocenter ls=78;
goptions colors=(none);

data new;
input varty pest resp @@;
cards;
1 1 49 1 1 39 1 2 50 1 2 55 1 3 43 1 3 38 1 4 85 1 4 73
2 1 55 2 1 41 2 2 67 2 2 58 2 3 53 2 3 42 2 4 53 2 4 48
3 1 66 3 1 68 3 2 85 3 2 92 3 3 69 3 3 62 3 4 85 3 4 99
;

proc sort; by pest varty;
proc means noprint;
var resp; by pest varty;
output out=new1 mean=mn;
```

```

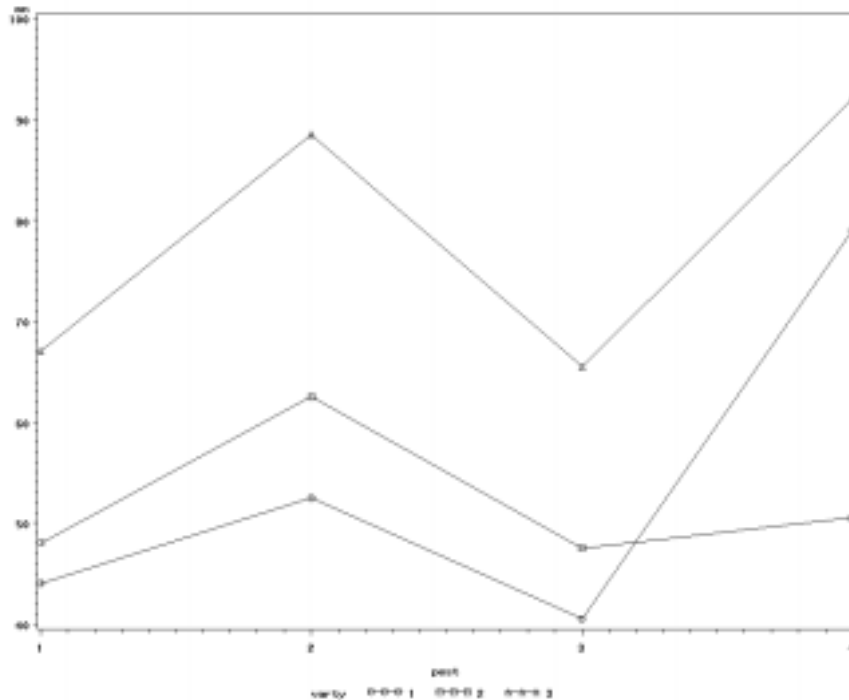
symbol1 v=circle i=join; symbol2 v=square i=join; symbol3 v=triangle i=join;
proc gplot;
  plot mn*pest=varty;

proc glm data=new;
  class varty pest;
  model resp = varty|pest;
  lsmeans varty*pest / slice=varty adjust=tukey stderr tdiff;
  output out=new1 r=res p=pred;

symbol1 v=circle i=none;
proc gplot;
  plot res*pred;
run;

```

Interaction Plot



Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	6680.458333	607.314394	14.36	<.0001
Error	12	507.500000	42.291667		
Corrected Total	23	7187.958333			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
varty	2	3397.583333	1698.791667	40.17	<.0001

pest	3	2227.458333	742.486111	17.56	0.0001
varty*pest	6	1055.416667	175.902778	4.16	0.0171

Least Squares Means

Adjustment for Multiple Comparisons: Tukey

varty	resp LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	54.0000000	2.2992299	<.0001	1
2	52.1250000	2.2992299	<.0001	2
3	78.2500000	2.2992299	<.0001	3

pest	resp LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	53.0000000	2.6549221	<.0001	1
2	67.8333333	2.6549221	<.0001	2
3	51.1666667	2.6549221	<.0001	3
4	73.8333333	2.6549221	<.0001	4

Least Squares Means

Adjustment for Multiple Comparisons: Tukey

varty	pest	resp LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	1	44.0000000	4.5984599	<.0001	1
1	2	52.5000000	4.5984599	<.0001	2
1	3	40.5000000	4.5984599	<.0001	3
1	4	79.0000000	4.5984599	<.0001	4
2	1	48.0000000	4.5984599	<.0001	5
2	2	62.5000000	4.5984599	<.0001	6
2	3	47.5000000	4.5984599	<.0001	7
2	4	50.5000000	4.5984599	<.0001	8
3	1	67.0000000	4.5984599	<.0001	9
3	2	88.5000000	4.5984599	<.0001	10
3	3	65.5000000	4.5984599	<.0001	11
3	4	92.0000000	4.5984599	<.0001	12

Least Squares Means for Effect varty*pest
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: resp

i/j	1	2	3	4	5	6
1		-1.30705	0.538196	-5.38196	-0.61508	-2.84475
		0.9623	1.0000	0.0055	0.9999	0.2695
2	1.307048		1.845244	-4.07491	0.691967	-1.5377
	0.9623		0.7699	0.0424	0.9998	0.9024
3	-0.5382	-1.84524		-5.92016	-1.15328	-3.38295

	1.0000	0.7699		0.0024	0.9838	0.1242
4	5.381962	4.074914	5.920158		4.766881	2.537211
	0.0055	0.0424	0.0024		0.0142	0.3992
5	0.615081	-0.69197	1.153278	-4.76688		-2.22967
	0.9999	0.9998	0.9838	0.0142		0.5595
6	2.844751	1.537703	3.382948	-2.53721	2.22967	
	0.2695	0.9024	0.1242	0.3992	0.5595	
7	0.538196	-0.76885	1.076392	-4.84377	-0.07689	-2.30656
	1.0000	0.9994	0.9902	0.0126	1.0000	0.5174
8	0.999507	-0.30754	1.537703	-4.38245	0.384426	-1.84524
	0.9945	1.0000	0.9024	0.0261	1.0000	0.7699
9	3.536718	2.22967	4.074914	-1.84524	2.921637	0.691967
	0.0983	0.5595	0.0424	0.7699	0.2427	0.9998
10	6.842781	5.535733	7.380977	1.460818	6.227699	3.998029
	0.0007	0.0043	0.0003	0.9264	0.0016	0.0478
11	3.306063	1.999015	3.844259	-2.0759	2.690981	0.461311
	0.1394	0.6881	0.0609	0.6454	0.3299	1.0000
12	7.380977	6.073929	7.919173	1.999015	6.765895	4.536225
	0.0003	0.0020	0.0002	0.6881	0.0007	0.0204
i/j	7	8	9	10	11	12
1	-0.5382	-0.99951	-3.53672	-6.84278	-3.30606	-7.38098
	1.0000	0.9945	0.0983	0.0007	0.1394	0.0003
2	0.768852	0.307541	-2.22967	-5.53573	-1.99901	-6.07393
	0.9994	1.0000	0.5595	0.0043	0.6881	0.0020
3	-1.07639	-1.5377	-4.07491	-7.38098	-3.84426	-7.91917
	0.9902	0.9024	0.0424	0.0003	0.0609	0.0002
4	4.843766	4.382455	1.845244	-1.46082	2.0759	-1.99901
	0.0126	0.0261	0.7699	0.9264	0.6454	0.6881
5	0.076885	-0.38443	-2.92164	-6.2277	-2.69098	-6.7659
	1.0000	1.0000	0.2427	0.0016	0.3299	0.0007
6	2.306555	1.845244	-0.69197	-3.99803	-0.46131	-4.53623
	0.5174	0.7699	0.9998	0.0478	1.0000	0.0204
7		-0.46131	-2.99852	-6.30458	-2.76787	-6.84278
		1.0000	0.2180	0.0014	0.2986	0.0007
8	0.461311		-2.53721	-5.84327	-2.30656	-6.38147
	1.0000		0.3992	0.0027	0.5174	0.0013
9	2.998522	2.537211		-3.30606	0.230656	-3.84426
	0.2180	0.3992		0.1394	1.0000	0.0609
10	6.304584	5.843273	3.306063		3.536718	-0.5382
	0.0014	0.0027	0.1394		0.0983	1.0000
11	2.767866	2.306555	-0.23066	-3.53672		-4.07491
	0.2986	0.5174	1.0000	0.0983		0.0424
12	6.842781	6.381469	3.844259	0.538196	4.074914	
	0.0007	0.0013	0.0609	1.0000	0.0424	

varty*pest Effect Sliced by varty for resp

varty	DF	Sum of Squares	Mean Square	F Value	Pr > F
1	3	1819.000000	606.333333	14.34	0.0003
2	3	297.375000	99.125000	2.34	0.1245
3	3	1166.500000	388.833333	9.19	0.0020

Using SAS Functions

In the previous example, SAS, when making the Tukey correction, assumed we would compare all 66 pairs of means, instead of the $6 \times 3 = 18$ of interest. (That is, only comparisons between levels of pesticide are of interest.) Here is a way to adjust for this in SAS. It involves using built-in SAS functions.

1. There are three sets of comparisons (i.e., contrasts). Each set is orthogonal to another set. Think of this experiment as 12 treatments. Each set of contrasts involves a different set of four treatments.
2. For each set, we're looking at all pairwise comparisons.

Because the sets are orthogonal, consider breaking up the overall α level among the three groups using orthogonality principles and then compute Tukey's minimum significant difference for each pairwise comparison within a group. In this case, the critical value of the test statistic is 3.56411 instead of 3.97016 .

```
data test;
  input FWE alev blev df mse;
/* First determine alpha level for orthogonal contrast groups */
  cut1 = probmc('maxmod',.,1-FWE,df,alev);
  group_alpha = 2*(1-probt(cut1,df));
/* Now compute Tukey's minimum significant difference */
  qval = probmc('range',.,1-group_alpha,df,blev)/sqrt(2);
  sig_dist = qval*sqrt(2*mse/(df/(alev*blev)+1));
  output;
datalines;
0.05 3 4 12 42.29166718
;
proc print;
run;
```

Sample Size/Power

- Assume both factors are fixed
- Have three different Φ^2
 - Factor A: $\Phi^2 = nbD_A^2/2a\sigma^2$
 - Factor B: $\Phi^2 = naD_B^2/2b\sigma^2$
 - Interaction: $\Phi^2 = nD_{AB}^2/2((a-1)(b-1)+1)\sigma^2$
- $df_E = ab(n-1)$
- df_{Num} depends on factor
- Compute n for each factor and choose largest n
- If $D_A = D_B = D_{AB}$, will be interaction Φ

Example: Hormone Study

Want to find any difference larger than 25 with 80% probability ($\alpha = 0.05$ and $\hat{\sigma}^2 = 288.35$)

$$\Phi_A^2 = 1.084n \quad \Phi_B^2 = 1.084n \quad \Phi_{AB}^2 = 0.542n$$

$$df_E = 4(n - 1) \text{ and } df_{Num} = 1$$

For $df_{Num} = 1 \rightarrow n = 4 < 0.80$ and $n = 5 \approx 80\%$.

Example: Fruit Yield

Want to find any difference larger than 5 bushels with 80% probability ($\alpha = 0.05$ and $\hat{\sigma}^2 = 42.29$)

$$\Phi_A^2 = 0.388n \quad \Phi_B^2 = 0.222n \quad \Phi_{AB}^2 = 0.042n$$

$$df_E = 12(n - 1) \text{ and } df_{Num} = 2, 3, 6$$

For $df_{Num} = 2 \rightarrow n = 8 \approx 80\%$

For $df_{Num} = 3 \rightarrow n = 8 \approx 55\%$, $n = 13 \approx 80\%$

For $df_{Num} = 6 \rightarrow n = 45 \approx 80\%$

No Interaction Assumption

- Only one replicate, error and interaction confounded
- Experimenter may feel no interaction exists
- Allows tests of main effects
- Similar to additivity assumptions – RCB design
- Check residual plot of violations – curvilinear
- Tukey's Test for Non-Additivity

General Factorial Model

- Factorial Design - observations at all possible combinations
- a levels of Factor A , b levels of Factor B , ...
- Straightforward construction if all **fixed effects**.
- In 3-factor model $\rightarrow nabc$ observations
- Need $n > 1$ to test for all possible interactions

Statistical Model (3 factor)

$$y_{i,j,k,\ell} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{i,j} + (\beta\gamma)_{j,k} + (\tau\gamma)_{i,k} + (\tau\beta\gamma)_{i,j,k} + \epsilon_{i,j,k,\ell}$$

$$\left\{ \begin{array}{l} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ \ell = 1, 2, \dots, n \end{array} \right.$$

Analysis of Variance Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Factor A	SS_A	$a - 1$	MS_A	F_0
Factor B	SS_B	$b - 1$	MS_B	F_0
Factor C	SS_C	$c - 1$	MS_C	F_0
AB	SS_{AB}	$(a - 1)(b - 1)$	MS_{AB}	F_0
AC	SS_{AC}	$(a - 1)(c - 1)$	MS_{AC}	F_0
BC	SS_{BC}	$(b - 1)(c - 1)$	MS_{BC}	F_0
ABC	SS_{ABC}	$(a - 1)(b - 1)(c - 1)$	MS_{ABC}	F_0
Error	SS_E	$abc(n - 1)$	MS_E	
Total	SS_T	$abcn - 1$		

Analysis of 2^k Design (Example 6-1)

```

data bottle;
input A B C devi;
datalines;
-1 -1 -1 -3
-1 -1 -1 -1
 1 -1 -1  0
 1 -1 -1  1
-1  1 -1 -1
-1  1 -1  0
 1  1 -1  2
 1  1 -1  3
-1 -1  1 -1
-1 -1  1  0
 1 -1  1  2
 1 -1  1  1
-1  1  1  1
-1  1  1  1
 1  1  1  6
 1  1  1  5
;

```

```

proc glm; /* Full ANOVA */
  class A B C; model devi = A|B|C;
  output out = botone r = res p = pred;
run;

proc glm; /* ANOVA with no 3-way interaction */
  class A B C; model devi = A B C A*B A*C B*C;
  output out = botone r = res p = pred;
run;

proc glm; /* ANOVA table with no interaction */
  class A B C; model devi = A B C ;
  output out = botone r = res p = pred;
run;

```

Dependent Variable: devi

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	73.00000000	10.42857143	16.69	0.0003
Error	8	5.00000000	0.62500000		
Corrected Total	15	78.00000000			

R-Square Coeff Var Root MSE devi Mean
0.935897 79.05694 0.790569 1.000000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	36.00000000	36.00000000	57.60	<.0001
B	1	20.25000000	20.25000000	32.40	0.0005
A*B	1	2.25000000	2.25000000	3.60	0.0943
C	1	12.25000000	12.25000000	19.60	0.0022
A*C	1	0.25000000	0.25000000	0.40	0.5447
B*C	1	1.00000000	1.00000000	1.60	0.2415
A*B*C	1	1.00000000	1.00000000	1.60	0.2415

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	72.00000000	12.00000000	18.00	0.0002
Error	9	6.00000000	0.66666667		
Corrected Total	15	78.00000000			

R-Square Coeff Var Root MSE devi Mean
0.923077 81.64966 0.816497 1.000000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	36.00000000	36.00000000	54.00	<.0001
B	1	20.25000000	20.25000000	30.38	0.0004
C	1	12.25000000	12.25000000	18.38	0.0020
A*B	1	2.25000000	2.25000000	3.38	0.0994
A*C	1	0.25000000	0.25000000	0.38	0.5554
B*C	1	1.00000000	1.00000000	1.50	0.2518

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	68.50000000	22.83333333	28.84	<.0001
Error	12	9.50000000	0.79166667		
Corrected Total	15	78.00000000			

R-Square	Coeff Var	Root MSE	devi Mean
0.878205	88.97565	0.889757	1.000000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	36.00000000	36.00000000	45.47	<.0001
B	1	20.25000000	20.25000000	25.58	0.0003
C	1	12.25000000	12.25000000	15.47	0.0020

What happens when there are no replicates?

General Factorial Model

- Usual assumptions and diagnostics
- Multiple comparisons simple extension of two factor
- Often higher order interactions ignored
- Beyond three-way interactions difficult to picture
- Pooled together with error (increase df_E)
- If all factors random
 - For example, a two factor model...
 - $E(MS_E) = \sigma^2$
 - $E(MS_A) = \sigma^2 + bn\sigma_\tau^2 + n\sigma_{\tau\beta}^2$
 - $E(MS_B) = \sigma^2 + an\sigma_\beta^2 + n\sigma_{\tau\beta}^2$
 - $E(MS_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2$
- Don't necessarily test over MS_E
- Mixture of fixed/random effect \rightarrow mixed models (Chapter 12)

Response Curves and Surfaces

When factors are quantitative (distinct ordering of levels of factor), may want to look at response curve or response surface to interpolate (i.e., predict the response for levels between those actually used). This is frequently done when the goal of the experiment is to find the “best” combination of factors (highest average response).

Example 5-4: An experimenter is interested in the effective life of a battery made of $a = 3$ different types of material and stored at $b = 3$ different temperatures. Since temperature is a quantitative factor, the experimenter is interested in not only seeing if either of the factors (or combination) affect battery life but also predicting what the battery life would be for other temperature values.

One can get crude representations of the response curves by creating an interaction plot (Figure 5-9). While there is an overall decrease in the battery life as temperature increases, the rate at which it decreases is quite different across materials. Notice that the life using Material 1 dramatically drops between the lowest and middle temperatures. This suggests an interaction, and this interaction is found significant in Table 5.5 ($p = 0.0186$).

Montgomery presents a response surface analysis using Design-Expert. This response surface analysis is a combination of ANOVA and regression. Orthogonal polynomial contrasts (Table X) separate the appropriate SS into linear and quadratic components (both for the main effect and interaction). From this analysis, you see the difference in response curves is primarily due to a differing quadratic component ($p = 0.0106$). This is pretty obvious from the interaction plot.

Below are the commands to do this same analysis using SAS. The temperature variable is treated as if it were a covariate (i.e., considered continuous). The parameter estimates can then be used to estimate the response curve for each material.

```
options nocenter;

data machine;
  input mat temp y @@;
  temp2 = temp*temp;
cards;
1 15 130 1 15 155 1 70 34 1 70 40 1 125 20 1 125 70
1 15 74 1 15 180 1 70 80 1 70 75 1 125 82 1 125 58
2 15 150 2 15 188 2 70 136 2 70 122 2 125 25 2 125 70
2 15 159 2 15 126 2 70 106 2 70 115 2 125 58 2 125 45
3 15 138 3 15 110 3 70 174 3 70 120 3 125 96 3 125 104
3 15 168 3 15 160 3 70 150 3 70 139 3 125 82 3 125 60
;

proc glm;
  class mat;
  model y=temp mat temp2 mat*temp mat*temp2/ solution;
run;
```

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	59416.22222	7427.02778	11.00	<.0001
Error	27	18230.75000	675.21296		

Corrected Total 35 77646.97222

Source	DF	Type I SS	Mean Square	F Value	Pr > F
temp	1	39042.66667	39042.66667	57.82	<.0001
mat	2	10683.72222	5341.86111	7.91	0.0020
temp2	1	76.05556	76.05556	0.11	0.7398
temp*mat	2	2315.08333	1157.54167	1.71	0.1991
temp2*mat	2	7298.69444	3649.34722	5.40	0.0106

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temp	1	1239.179404	1239.179404	1.84	0.1867
mat	2	1147.938218	573.969109	0.85	0.4385
temp2	1	76.055556	76.055556	0.11	0.7398
temp*mat	2	7170.660365	3585.330183	5.31	0.0114
temp2*mat	2	7298.694444	3649.347222	5.40	0.0106

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	132.7623967 B	20.56765571	6.45	<.0001
temp	0.9028926 B	0.75514800	1.20	0.2422
mat 1	36.6177686 B	29.08705766	1.26	0.2188
mat 2	26.8615702 B	29.08705766	0.92	0.3639
mat 3	0.0000000 B	.	.	.
temp2	-0.0102479 B	0.00526030	-1.95	0.0619
temp*mat 1	-3.4043388 B	1.06794054	-3.19	0.0036
temp*mat 2	-1.0762397 B	1.06794054	-1.01	0.3225
temp*mat 3	0.0000000 B	.	.	.
temp2*mat 1	0.0230992 B	0.00743919	3.11	0.0044
temp2*mat 2	0.0045868 B	0.00743919	0.62	0.5427
temp2*mat 3	0.0000000 B	.	.	.

Response Curves and Surfaces

The following model is used

$$y_{i,j,k} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3t + \beta_4x_1t + \beta_5x_2t + \beta_6t^2 + \beta_7x_1t^2 + \beta_8x_2t^2 + \epsilon_{i,j,k}$$

where

$$x_1 = \begin{cases} 1 & \text{if } mat = 1 \\ 0 & \text{if } mat = 2 \\ -1 & \text{if } mat = 3 \end{cases} \quad x_2 = \begin{cases} 0 & \text{if } mat = 1 \\ 1 & \text{if } mat = 2 \\ -1 & \text{if } mat = 3 \end{cases}$$

and results in the following three response curves.

$$\begin{aligned} E(y_{1,t}) &= 169.38 - 2.5014t + 0.0129t^2 \\ E(y_{2,t}) &= 159.62 - 0.1733t - 0.0057t^2 \\ E(y_{3,t}) &= 132.76 + 0.9029t - 0.0102t^2 \end{aligned}$$

These curves could then be used to predict other lives for different temperatures. Interestingly, material 1 has a curve that suggests the best temperature is a plus or minus infinity. One needs to be careful with interpolation.

Blocking in Factorial Model

- For randomized complete block
 - Each treatment appears in each block.
 - CRD replicate in each block
- For factorial design
 - All combinations appear in each block.
 - Factorial replicate in each block
- Blocking on two factors – Latin Square
 - Need p treatments run in $p \times p$ square
 - If $ab_{..} = p$, combinations can be run in Latin Square.
- Additional additivity assumptions
- Interaction with blocks reduces df for error

Using SAS (Example 5-6 – 2fblk.sas)

```
options nocenter ps=50 ls=80;
```

```
data new;  
input oper filter clutter resp;  
cards;  
1 1 1 90
```

```
...
```

```
4 2 3 83  
;
```

```
proc glm;  
class oper filter clutter;  
model resp = oper filter|clutter;  
output out=new1 r=res p=pred;  
run;
```

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	1881.500000	235.187500	21.21	<.0001
Error	15	166.333333	11.088889		
Corrected Total	23	2047.833333			

R-Square	Coeff Var	Root MSE	resp Mean
0.918776	3.508339	3.329998	94.91667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
oper	3	402.166667	134.055556	12.09	0.0003
filter	1	1066.666667	1066.666667	96.19	<.0001
clutter	2	335.583333	167.791667	15.13	0.0003
filter*clutter	2	77.083333	38.541667	3.48	0.0575

Source	DF	Type III SS	Mean Square	F Value	Pr > F
oper	3	402.166667	134.055556	12.09	0.0003
filter	1	1066.666667	1066.666667	96.19	<.0001
clutter	2	335.583333	167.791667	15.13	0.0003
filter*clutter	2	77.083333	38.541667	3.48	0.0575

Unbalanced Factorial Design

- Sometimes n_i varies for each combination.
- Could be result of missing observations
- Could be designed to be unbalanced

Certain combinations of more importance

- Use same approaches as before
 1. Regression Approach
 2. Estimate missing value(s)
 3. Remove data so balanced
- SAS Type I-Type IV SS
 - Type I – fit in order of `model` statement
 - Type II – fits only unrelated terms or lower order first
 - Type III – fits all other terms first
 - Type IV – used when some $n_i = 0$.
- Should report/discuss all types.
- Think statistical definition of causation.

Example

A factorial experiment is performed to investigate the effects of genetic strain and environment on the number of errors a rat makes in a maze. Due to shortage of rats, the number of replicates per combination varied.

Environment	Genetic Strain					
	1		2		3	
Free	92	85	76	51	61	
	89	100	72	92	47	
Controlled	98	80		73	82	
	106	72	92	59	77	

From SAS

Level of ENV	Level of TRAIT	N	Mean	SD
1	1	3	93.666667	5.6862407
1	2	4	81.250000	8.9953692
1	3	3	53.000000	7.2111026
2	1	2	102.000000	5.6568542
2	2	3	81.333333	10.0664459
2	3	4	75.250000	5.5602758

SS_E is the sum of squared differences between the cell observations and the cell mean.

```

data rat;
  input env trait num @@;
  cards;
  1 1 92  1 1 100  1 1 89
  2 1 106  2 1 98
  1 2 85  1 2 76  1 2 72  1 2 92
  1 3 51  1 3 61  1 3 47
  2 2 80  2 2 72  2 2 92
  2 3 73  2 3 82  2 3 77  2 3 69
  ;

proc glm;
  class env trait;
  model num = env|trait / ss1 ss2 ss3 ss4;
  means env|trait;
  lsmeans env|trait / stderr;
  output out = new r = res p = pred;

proc plot;
  plot res*pred;
run;

```

Dependent Variable: num

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	3815.271930	763.054386	13.43	<.0001
Error	13	738.833333	56.833333		
Corrected Total	18	4554.105263			

R-Square Coeff Var Root MSE num Mean

0.837765 9.460831 7.538789 79.68421

Source	DF	Type I SS	Mean Square	F Value	Pr > F
env	1	214.049708	214.049708	3.77	0.0743
trait	2	3172.919753	1586.459877	27.91	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Source	DF	Type II SS	Mean Square	F Value	Pr > F
env	1	503.721340	503.721340	8.86	0.0107
trait	2	3172.919753	1586.459877	27.91	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Source	DF	Type III SS	Mean Square	F Value	Pr > F
env	1	470.222222	470.222222	8.27	0.0130
trait	2	3254.880952	1627.440476	28.64	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
env	1	470.222222	470.222222	8.27	0.0130
trait	2	3254.880952	1627.440476	28.64	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Level of env	N	Mean	Std Dev
1	10	76.5000000	18.3862872
2	9	83.2222222	12.7355583

Level of trait	N	Mean	Std Dev
1	5	97.0000000	6.7082039
2	7	81.2857143	8.6161532
3	7	65.7142857	13.1999278

Level of env	Level of trait	N	Mean	Std Dev
1	1	3	93.666667	5.6862407
1	2	4	81.250000	8.9953692
1	3	3	53.000000	7.2111026
2	1	2	102.000000	5.6568542
2	2	3	81.333333	10.0664459
2	3	4	75.250000	5.5602758

Least Squares Means

env	num LSMEAN	Standard Error	Pr > t
1	75.9722222	2.4059469	<.0001
2	86.1944444	2.6155400	<.0001

trait	num LSMEAN	Standard Error	Pr > t
1	97.8333333	3.4409705	<.0001
2	81.2916667	2.8789224	<.0001
3	64.1250000	2.8789224	<.0001

env	trait	num LSMEAN	Standard Error	Pr > t
1	1	93.666667	4.352522	<.0001
1	2	81.250000	3.769394	<.0001
1	3	53.000000	4.352522	<.0001
2	1	102.000000	5.330729	<.0001
2	2	81.333333	4.352522	<.0001
2	3	75.250000	3.769394	<.0001

Blocking (Review)

- Used when controlling for nuisance factors
- Units are *grouped*, not randomly assigned.
- Usually need to “recruit” to fill with each factor level

Question

- **NOT** “Is there a block effect?”, but rather ...
- “Did blocking help me detect a treatment effect?” (Relative Efficiency)

Often there is no replication within blocks

- restricts inference to additive effects
- use *Tukey’s Test for Nonadditivity* to check for interaction

Latin Square Design

- Block on two nuisance factors
- One treatment observation per block 1
- One treatment observation per block 2
- Must have same number of blocks and treatments
- Two restrictions on randomization

$$y_{i,j,k} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{i,j,k} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{cases}$$

μ -- grand mean
 α_i -- i th block 1 effect (row)
 τ_j -- j th treatment effect
 β_k -- k th block 2 effect (column)
 $\epsilon_{i,j,k} \sim N(0, \sigma^2)$

- Completely additive model (no interaction)
- Design represented in $p \times p$ grid
- Randomization restrictions
 - one treatment per row
 - one treatment per column
- Shuffle rows and columns of standard square

Examples

C	B	A	
B	A	C	
A	C	B	
B	A	C	
A	C	B	
C	B	A	
D	B	C	A
C	D	A	B
B	A	D	C
A	C	B	D

```

proc plan
title 'Latin Square Design';
proc plan seed = 12;
  factors rows = 4 ordered cols = 4 ordered / NOPRINT;
  treatments tmts = 4 cyclic;
  output out = g
    rows cvals = ('Day 1' 'Day 2' 'Day 3' 'Day 4') random
    cols cvals = ('Lab 1' 'Lab 2' 'Lab 3' 'Lab 4') random
    tmts nvals = ( 0      100      250      450 ) random;
proc tabulate;
  class rows cols;
  var tmts;
  table rows, cols*(tmts*f = 6.) / rts = 8;
run;

```

	cols			
	Lab 1	Lab 2	Lab 3	Lab 4
	tmts	tmts	tmts	tmts
	Sum	Sum	Sum	Sum
rows				
Day 1	450	100	0	250
Day 2	100	0	250	450
Day 3	0	250	450	100
Day 4	250	450	100	0

Partitioning the SS

- Rewrite observation as

$$\begin{aligned}
 y_{i,j,k} &= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...}) + (\bar{y}_{i,j,k} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}) \\
 &= \hat{\mu} + \hat{\alpha}_i + \hat{\tau}_j + \hat{\beta}_k + \hat{\epsilon}_{i,j,k}
 \end{aligned}$$

- Partition SS_T into

$$p \sum (\bar{y}_{i..} - \bar{y}_{...})^2 + p \sum (\bar{y}_{.j.} - \bar{y}_{...})^2 + p \sum (\bar{y}_{..k} - \bar{y}_{...})^2 + \sum \sum \hat{\epsilon}_{i,j,k}^2$$

$SS_{Row} + SS_{Treatment} + SS_{Col} + SS_E$

- Under H_0 , all SS/σ^2 independent chi-squared RV's
- Usual F -test analysis
- Caution testing column and row effects

Analysis of Variance Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F_0
Rows	SS_{Row}	$p - 1$	MS_{Row}	
Treatment	SS_{Trt}	$p - 1$	MS_{Trt}	F_0
Column	SS_{Col}	$p - 1$	MS_{Col}	
Error	SS_E	$(p - 2)(p - 1)$	MS_E	
Total	SS_T	$p^2 - 1$		

$$\begin{aligned}
SS_T &= \sum \sum \sum y_{i,j,k}^2 - y_{...}^2/p^2 \\
SS_{Row} &= \frac{1}{p} \sum y_{i..}^2 - y_{...}^2/p^2 \\
SS_{Trt} &= \frac{1}{p} \sum y_{.j.}^2 - y_{...}^2/p^2 \\
SS_{Col} &= \frac{1}{p} \sum y_{..k}^2 - y_{...}^2/p^2 \\
SS_E &= \text{Use subtraction}
\end{aligned}$$

If $F_0 > F_{\alpha, p-1, (p-2)(p-1)}$, then H_0 .

Using SAS

Consider an experiment to investigate the effect of 4 diets on milk production. There are 4 cows. Each lactation period, the cows receive a different diet. Assume there is a washout period. So previous diet does not affect future results. Will block on lactation period and cow.

```

options nocenter ls=75;goptions colors=(none);

data new;
  input cow period trt resp @@;
cards;
  1 1 1 38 1 2 2 32 1 3 3 35 1 4 4 33
  2 1 2 39 2 2 3 37 2 3 4 36 2 4 1 30
  3 1 3 45 3 2 4 38 3 3 1 37 3 4 2 35
  4 1 4 41 4 2 1 30 4 3 2 32 4 4 3 33
;

proc glm;
class cow trt period;
model resp=trt period cow;
means trt/ lines tukey;
means period cow;
output out=new1 r=res p=pred;

symbol1 v=circle;
proc gplot; plot res*pred;

proc univariate noprint;
histogram res / normal (L=1 mu=0 sigma=est) kernel (L=2);
run;

```

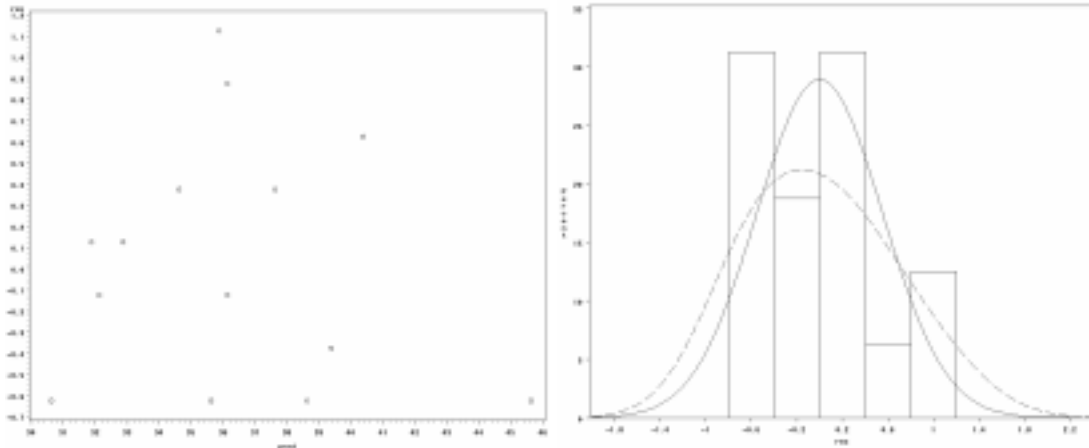
Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	242.5625000	26.9513889	33.17	0.0002

Error	6	4.8750000	0.8125000		
Corrected Total	15	247.4375000			
Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	3	40.6875000	13.5625000	16.69	0.0026
period	3	147.1875000	49.0625000	60.38	<.0001
cow	3	54.6875000	18.2291667	22.44	0.0012

Tukey's Studentized Range (HSD) Test for resp
Alpha 0.05
Error Degrees of Freedom 6
Error Mean Square 0.8125
Critical Value of Studentized Range 4.89559
Minimum Significant Difference 2.2064

	Mean	N	trt
A	37.5000	4	3
A	37.0000	4	4
B	34.5000	4	2
B	33.7500	4	1



Missing Values

- When missing
 - Design unbalanced
 - Orthogonality lost
 - Order of fit important
- Procedures
 1. Regression approach – Use Type III sum of squares
 2. Estimate missing value

Choose value to minimize SS_E
Take derivative and set equal to zero

$$y_{i,j,k} = \frac{p(y'_{i..} + y'_{.j.} + y'_{..k}) - 2y'_{...}}{(p-2)(p-1)}$$

Using proc mixed

- Sometimes, block should be considered random (i.e., every possible block effect is not taken into the model)

Example: 4 cows randomly chosen from large herd
Want the inference to extend to the herd
Treat cow as a random blocking factor

- Sometimes measuring same EU over time/period

Example: Cow measured over 4 lactation periods
Are lactation periods closer together more similar?
Will treat as crossover design later

- `proc mixed` can incorporate both concepts into model

Random Effects

- Similar results as with RCBD
- Standard error for a mean: `proc glm` incorrect
- Standard error for a contrast: `proc glm` correct

```
proc glm;
class cow trt period;
model resp = cow trt period;
random cow;
lsmeans trt/stderr tdiff;
```

```
proc mixed;
class cow trt period;
model resp = trt period;
random cow;
lsmeans trt/ diff;
run;
```

The GLM Procedure

trt	resp LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	33.7500000	0.4506939	<.0001	1
2	34.5000000	0.4506939	<.0001	2
3	37.5000000	0.4506939	<.0001	3
4	37.0000000	0.4506939	<.0001	4

Least Squares Means for Effect trt
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

i/j	1	2	3	4
1		-1.1767 0.2839	-5.88348 0.0011	-5.09902 0.0022
2	1.176697 0.2839		-4.70679 0.0033	-3.92232 0.0078
3	5.883484 0.0011	4.706787 0.0033		0.784465 0.4626
4	5.09902 0.0022	3.922323 0.0078	-0.78446 0.4626	

The Mixed Procedure

Least Squares Means

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	33.7500	1.1365	6	29.70	<.0001
trt	2	34.5000	1.1365	6	30.36	<.0001
trt	3	37.5000	1.1365	6	33.00	<.0001
trt	4	37.0000	1.1365	6	32.56	<.0001

Differences of Least Squares Means

Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	2	-0.7500	0.6374	6	-1.18	0.2839
trt	1	3	-3.7500	0.6374	6	-5.88	0.0011
trt	1	4	-3.2500	0.6374	6	-5.10	0.0022
trt	2	3	-3.0000	0.6374	6	-4.71	0.0033
trt	2	4	-2.5000	0.6374	6	-3.92	0.0078
trt	3	4	0.5000	0.6374	6	0.78	0.4626

Correlated Observations

- Residuals within an EU may be correlated.
- Residuals closer in time may be more similar
- Can incorporate various correlation structures
- Represented as $p \times p$ covariance matrix
- Main diagonal contains the variances

- Off-diagonal elements represent covariances
- Uncorrelated residuals (for $p = 4$ obs/cow)

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- 1st order autoregressive (Generalized LS)

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

SAS Code

```
/* Fit a simple correlation structure */
/* Similar to standard mixed analysis */

proc mixed covtest cl;
class cow trt period;
model resp = trt period / ddfm = kr outp = diago;
random cow;
repeated period / subject = cow type = simple;
lsmeans trt / diff;
run;

/* Fit an AR(1) correlation structure */

proc mixed covtest cl;
class cow trt period;
model resp = trt period / ddfm = kr outp = diago;
random cow;
repeated period / type = ar(1) subject = cow;
lsmeans trt / diff;
run;
```

Dimensions	
Covariance Parameters	2
Columns in X	9
Columns in Z	4
Subjects	1
Max Obs Per Subject	16
Observations Used	16
Observations Not Used	0
Total Observations	16

Covariance Parameter Estimates							
Cov		Standard	Z				
Parm	Subject	Estimate	Error	Value	Pr Z	Alpha	Lower
cow		4.3542	3.7229	1.17	0.1211	0.05	1.3461
period	cow	0.8125	0.4691	1.73	0.0416	0.05	0.3374

Fit Statistics	
-2 Res Log Likelihood	41.3
AIC (smaller is better)	45.3
AICC (smaller is better)	47.3
BIC (smaller is better)	44.1

Type 3 Tests of Fixed Effects				
Effect	Num	Den	F Value	Pr > F
trt	3	6	16.69	0.0026
period	3	6	60.38	<.0001

Least Squares Means						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	33.7500	1.1365	3.82	29.70	<.0001
trt	2	34.5000	1.1365	3.82	30.36	<.0001
trt	3	37.5000	1.1365	3.82	33.00	<.0001
trt	4	37.0000	1.1365	3.82	32.56	<.0001

Differences of Least Squares Means							
Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	2	-0.7500	0.6374	6	-1.18	0.2839
trt	1	3	-3.7500	0.6374	6	-5.88	0.0011
trt	1	4	-3.2500	0.6374	6	-5.10	0.0022
trt	2	3	-3.0000	0.6374	6	-4.71	0.0033
trt	2	4	-2.5000	0.6374	6	-3.92	0.0078
trt	3	4	0.5000	0.6374	6	0.78	0.4626

Dimensions	
Covariance Parameters	3
Columns in X	9
Columns in Z	4
Subjects	1
Max Obs Per Subject	16
Observations Used	16
Observations Not Used	0
Total Observations	16

Covariance Parameter Estimates							
Cov Parm	Subject	Estimate	Standard Error	Z	Value	Pr Z	Alpha
cow		4.1459	3.7154	1.12	0.1322	0.05	1.2331
AR(1)	cow	0.2184	0.5794	0.38	0.7062	0.05	-0.9171
Residual		0.9292	0.7343	1.27	0.1029	0.05	0.3061

Fit Statistics	
-2 Res Log Likelihood	41.2
AIC (smaller is better)	47.2
AICC (smaller is better)	52.0
BIC (smaller is better)	45.3

Type 3 Tests of Fixed Effects				
Effect	Num	Den	F Value	Pr > F
trt	3	5.04	8.65	0.0197
period	3	4.06	41.27	0.0017

Least Squares Means						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	33.7306	1.1424	3.81	29.53	<.0001
trt	2	34.4506	1.1424	3.81	30.16	<.0001
trt	3	37.5563	1.1424	3.81	32.87	<.0001
trt	4	37.0125	1.1424	3.81	32.40	<.0001

Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	2	-0.7200	0.7831	5.82	-0.92	0.3943
trt	1	3	-3.8257	0.8834	5.51	-4.33	0.0060
trt	1	4	-3.2819	0.7831	5.82	-4.19	0.0061
trt	2	3	-3.1057	0.7831	5.82	-3.97	0.0079
trt	2	4	-2.5619	0.8834	5.51	-2.90	0.0301
trt	3	4	0.5438	0.7831	5.82	0.69	0.5141

Replicated Latin Square

- Latin Square Design $\rightarrow df_E$ small
 - If 3 treatments \rightarrow 2 df error
 - if 4 treatments \rightarrow 6 df error
- Replication increases df_E without increasing treatments
- Method of replication
 - Same row and column blocks
 - New rows and same columns
 - Same rows and new columns
 - New rows and new columns
- Degrees of freedom depend on what is “new”/randomized
- Often include additional block- “replicate” effect

Replicate Square

- Same row/column blocks used in additional squares
- Usually includes replicate (e.g., time) effect

$$y_{i,j,k,\ell} = \mu + \alpha_i + \tau_j + \beta_k + \delta_\ell + \epsilon_{i,j,k,\ell} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \\ \ell = 1, 2, \dots, n \end{cases}$$

Example: same rows and same columns

	1	2	3	Data			Replication
1	A	B	C	7.0	8.0	9.0	1
2	B	C	A	4.0	5.0	6.0	
3	C	A	B	6.0	3.0	4.0	
	1	2	3	Data			Replication
1	A	B	C	8.0	4.0	7.0	2
2	B	C	A	6.0	3.0	6.0	
3	C	A	B	5.0	8.0	7.0	
	1	2	3	Data			Replication
1	A	B	C	9.0	6.0	8.0	3
2	B	C	A	5.0	7.0	6.0	
3	C	A	B	9.0	3.0	7.0	

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Rows	SS_{Row}	$p - 1$		
Columns	SS_{Col}	$p - 1$		
Replicate	SS_{Rep}	$n - 1$		
Treatment	SS_{Trt}	$p - 1$	MS_{Trt}	F_0
Error	SS_E	$(p - 1)(n(p + 1) - 3)$	MS_E	
Total	SS_T	$np^2 - 1$		

Replicated Rows (or Columns)

- Different rows only (p^2 subjects, n replications)
- Row effect often **nested** within square
- β_k can be different for each square

$$\sum \beta_{k(\ell)} = 0 \text{ instead of } \sum \beta_k = 0$$

$$y_{i,j,k,\ell} = \mu + \alpha_{i(\ell)} + \tau_j + \beta_k + \delta_\ell + \epsilon_{i,j,k,\ell} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \\ \ell = 1, 2, \dots, n \end{cases}$$

Example: different rows and same columns

	1	2	3	Data			Replication
1	A	C	B	7.0	8.0	9.0	1
2	C	B	A	4.0	5.0	6.0	
3	B	A	C	6.0	3.0	4.0	
	1	2	3	Data			Replication
4	C	B	A	8.0	4.0	7.0	2
5	B	A	C	6.0	3.0	6.0	
6	A	C	B	5.0	8.0	7.0	
	1	2	3	Data			Replication
7	B	A	C	9.0	6.0	8.0	3
8	A	C	B	5.0	7.0	6.0	
9	C	B	A	9.0	3.0	7.0	

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Rows	SS_{Row}	$n(p - 1)$		
Columns	SS_{Col}	$p - 1$		
Replicate	SS_{Rep}	$n - 1$		
Treatment	SS_{Trt}	$p - 1$	MS_{Trt}	F_0
Error	SS_E	$(p - 1)(np - 1)$	MS_E	
Total	SS_T	$np^2 - 1$		

Replicated Rows (or Columns) without nesting

- Considered Latin Rectangle: “No replicate effect”
- np separate rows (i.e., np^2 experimental units)
- Row not nested within square

$$y_{i,j,k} = \mu + \alpha_i + \tau_j + \beta_k + \epsilon_{i,j,k} \begin{cases} i = 1, 2, \dots, np \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{cases}$$

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Rows	SS_{Row}	$np - 1$		
Columns	SS_{Col}	$p - 1$		
Treatment	SS_{Trt}	$p - 1$	MS_{Trt}	F_0
Error	SS_E	$(p - 1)(np - 2)$	MS_E	
Total	SS_T	$np^2 - 1$		

Replicated Rows and Columns

- Have completely separate squares
- Usually row and column effect **nested** within square

$$y_{i,j,k,\ell} = \mu + \alpha_{i(\ell)} + \tau_j + \beta_{k(\ell)} + \delta_\ell + \epsilon_{i,j,k,\ell} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \\ \ell = 1, 2, \dots, n \end{cases}$$

Example: different rows and different columns

	1	2	3	Data			Replication
1	A	B	C	7.0	8.0	9.0	1
2	B	C	A	4.0	5.0	6.0	
3	C	A	B	6.0	3.0	4.0	
	4	5	6	Data			Replication
4	C	B	A	8.0	4.0	7.0	2
5	B	A	C	6.0	3.0	6.0	
6	A	C	B	5.0	8.0	7.0	
	7	8	9	Data			Replication
7	B	A	C	9.0	6.0	8.0	3
8	A	C	B	5.0	7.0	6.0	
9	C	B	A	9.0	3.0	7.0	

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Rows	SS_{Row}	$n(p - 1)$		
Columns	SS_{Col}	$n(p - 1)$		
Replicate	SS_{Rep}	$n - 1$		
Treatment	SS_{Trt}	$p - 1$	MS_{Trt}	F_0
Error	SS_E	$(p - 1)(n(p - 1) - 1)$	MS_E	
Total	SS_T	$np^2 - 1$		

Example: Carbon Monoxide Emissions

Four gasoline blends are of interest, two levels each of factors A and B . Testing will take place in “nonlaboratory” conditions on city streets. Blocking factors are car-to-car variation and route differences. There will be also three replicates taken. The following designs use different types of replications schemes:

- If only four cars and four routes are used, they will appear in all three replications. Thus, row and column factors are being replicated across squares.

$$df_E = (p - 1)(n(p + 1) - 3) = 36.$$

Every blend/car/route combination appears once.

- Four cars are used on twelve different routes. Then, row blocks (cars) are being reused, but not column blocks (routes). (Note there is no nesting of route within square.)

$$df_E = (p - 1)(np - 2) = 30.$$

More routes used. Not every blend/car combination appears on every route.

- Four cars are used on twelve different routes divided between three different days (in potentially three different conditions). Routes on each day do not correspond. Thus, routes are *nested* within day.

$$df_E = (p - 1)(np - 1) = 33.$$

More routes used. No ability to check route-by-day interaction. Can check for day-to-day effect. However, not every treatment/driver combination appears each day.

- Twelve cars on twelve routes are divided into three sets of four each to create squares. Thus, neither rows nor columns are reused.

$$df_E = (p - 1)(np - n - 1) = 24.$$

More cars and routes used. No ability to check route-by-square or car-by-square interactions. Square acts as another (incomplete) block effect.

Having more block levels does not necessarily lead to more precision if number of experimental units is not also increased.

Extensions

- Crossover Design
 - p treatments and p periods
 - np subjects (experimental units)
 - Analysis similar to replicated column Latin Square

- Used in drug comparisons/physiology experiments
- Delay between periods to remove residual effect
- Model can handle residual effects ($p > 2$)
- Graeco-Latin Square Design (Section 4.3)
 - Superimpose two Latin Squares onto each other
 - Can block on three factors ($p \geq 3$ and $p \neq 6$)

Graeco-Latin Square: An Example

An experiment is conducted to compare four gasoline additives by testing them on four cars with four drivers over four days. Only four runs can be conducted in each day. The response is the amount of automobile emission.

- Treatment factor: gasoline additive, denoted by $A, B, C,$ and $D.$
- Block factor 1: driver, denoted by 1, 2, 3, 4.
- Block factor 2: day, denoted by 1, 2, 3, 4.
- Block factor 3: car, denoted by $\alpha, \beta, \gamma, \delta.$

drivers	days			
	1	2	3	4
1	$A\alpha = 32$	$B\beta = 25$	$C\gamma = 31$	$D\delta = 27$
2	$B\delta = 24$	$A\gamma = 36$	$D\beta = 20$	$C\alpha = 25$
3	$C\beta = 28$	$D\alpha = 30$	$A\delta = 23$	$B\gamma = 31$
4	$D\gamma = 34$	$C\delta = 35$	$B\alpha = 29$	$A\beta = 33$

(Completely Additive) Model

$$y_{i,j,k,\ell} = \mu + \alpha_i + \tau_j + \beta_k + \eta_\ell + \epsilon_{i,j,k,\ell} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, p \end{cases}$$

- μ – grand mean
- α_i – i th block 1 effect (row)
- τ_j – j th treatment effect
- β_k – k th block 2 effect (column)
- η_ℓ – ℓ th block 3 effect (Greek letters)
- $\epsilon_{i,j,k,\ell} \sim N(0, \sigma^2)$

Graeco-Latin Square

Design Matrix

driver	day	additive	car
1	1	A	α
1	2	B	β
1	3	C	γ
1	4	D	δ
\vdots	\vdots	\vdots	\vdots
4	1	D	γ
4	2	C	δ
4	3	B	α
4	4	A	β

- Two $p \times p$ Latin squares are said to be **orthogonal** if the two squares – when superimposed – have the property that each pair of letters appears once.
- The superimposed square is called a Graeco-Latin square.
- Tables available for $p \leq 8$
- 6×6 Graeco-Latin square does not exist
- Can be extended to considering $p - 1$ collection of orthogonal Latin Squares (with no error degrees of freedom). (*Latin Hypersquare*)

Example:

drivers	days			
	1	2	3	4
1	$A\alpha a = 32$	$B\beta b = 25$	$C\gamma c = 31$	$D\delta d = 27$
2	$B\delta c = 24$	$A\gamma d = 36$	$D\beta a = 20$	$C\alpha b = 25$
3	$C\beta d = 28$	$D\alpha c = 30$	$A\delta b = 23$	$B\gamma a = 31$
4	$D\gamma b = 34$	$C\delta a = 35$	$B\alpha d = 29$	$A\beta c = 33$

Crossover Design

- Time is a blocking factor (usually called period)
- np subjects (S_k) receive p treatments τ_j in p periods (P_i)
- Anticipate high level of variability between subjects
- Improve precision with several observations from each subject
- Common used for only 2, 3, or 4 periods
- Should consider problem of subsequent use

- Analysis similar to Latin Rectangle/Replicated Columns

$$y_{i,j,k} = \mu + P_i + \tau_j + S_k + \epsilon_{i,j,k}$$

Advantages

- Ethical
 - No one receives *just* a placebo.

Provisos

- *Wash-out period*, in addition to...
- Time periods of sufficient duration for factor to have full effect.
- Factors effects do not change over time (no time/factor interaction)
- Treatment doesn't cure
- May need extra period
 - Instead of: $a \rightarrow b \rightarrow c$,
 - use $a \rightarrow b \rightarrow b \rightarrow c \rightarrow c$
 - or

$$\begin{aligned} a &\rightarrow b \rightarrow b \rightarrow c \\ b &\rightarrow c \rightarrow c \rightarrow a \\ c &\rightarrow a \rightarrow a \rightarrow b \end{aligned}$$

Consider 2 treatment, 2 period experiment with n subjects. Based on the model, the difference in treatments for the two groups can be written

$$\text{Received Trt 1 first: } \text{diff}_{1,k} = (\tau_1 - \tau_2) + (P_1 - P_2) + (\epsilon_{1,1,k} - \epsilon_{2,2,k})$$

$$\text{Received Trt 2 first: } \text{diff}_{2,k} = (\tau_2 - \tau_1) + (P_1 - P_2) + (\epsilon_{2,1,k} - \epsilon_{1,2,k})$$

Thus, $\bar{\text{diff}}_1 - \bar{\text{diff}}_2$ estimates $2(\tau_1 - \tau_2)$ with variance $2\hat{\sigma}^2/n$, where $\hat{\sigma}^2$ estimates the variance of $\epsilon_{2,1,k} - \epsilon_{1,2,k}$ and $\epsilon_{2,1,k} - \epsilon_{1,2,k}$.

Residual Effects

But what if there is a residual effect? In other words, the treatment effect is different for the different periods. The model can then be written

$$\text{Treatment 1 first: } \text{diff}_{1,k} = (\tau_1 - \tau'_2) + (P_1 - P_2) + \epsilon_{1,k}$$

$$\text{Treatment 2 first: } \text{diff}_{2,k} = (\tau_2 - \tau'_1) + (P_1 - P_2) + \epsilon_{2,k}$$

Thus, $\bar{\text{diff}}_1 - \bar{\text{diff}}_2$ estimates $(\tau_1 - \tau_2) + (\tau'_1 - \tau'_2)$. Cannot yield inference about both differences (i.e., confounded).

Can test for residual effect by looking at sums instead of differences. Subject variability incorporated into error ($\delta_{i,j,k} = \epsilon_{i,j,k} + S_k$)

Treatment 1 first: $\text{sum}_{1,k} = 2\mu + (\tau_1 + \tau'_2) + (P_1 + P_2) + \delta_{1,k}$

Treatment 2 first: $\text{sum}_{2,k} = 2\mu + (\tau_2 + \tau'_1) + (P_1 + P_2) + \delta_{2,k}$

Thus, $\bar{\text{sum}}_1 - \bar{\text{sum}}_2$ estimates $(\tau_1 - \tau_2) - (\tau'_1 - \tau'_2)$. Can check to see if different from zero. Problem if that occurs. Low power test because it incorporates subject to subject variability.

If residual effects stay constant...

- Having more than 2 periods allows incorporation of residual effect in model
- Not orthogonal, order of fit important (Type III)

$$y_{i,j,k} = \mu + P_i + \tau_j + S_k + r_{i,j'} + \epsilon_{i,j,k} \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, p \\ k = 1, 2, \dots, np, \end{cases}$$

where $r_{i,j'}$ only occurs when $i \neq 1$ and j' references the treatment used in the previous model.

SAS Code (latin2.sas)

```
options nocenter ls=75;options colors=(none);

data new;
  input cow period trt resp @@;
  if period = 1 then resid = 0;
  else resid = a;
  resid1 = 0; resid2 = 0; resid3 = 0;
  if resid = 1 then resid1 = 1; if resid = 4 then resid1 = -1;
  if resid = 2 then resid2 = 1; if resid = 4 then resid2 = -1;
  if resid = 3 then resid3 = 1; if resid = 4 then resid3 = -1;
  a = trt;
  retain a;
cards;
  1 1 1 38 1 2 2 32 1 3 3 35 1 4 4 33
  2 1 2 39 2 2 3 37 2 3 4 36 2 4 1 30
  3 1 3 45 3 2 4 38 3 3 1 37 3 4 2 35
  4 1 4 41 4 2 1 30 4 3 2 32 4 4 3 33
;

proc print;
```

```

proc glm;
class cow period trt;
model resp = cow period trt resid1 resid2 resid3 / solution;
lsmeans trt / stderr pdiff cl;
run;

```

Obs	cow	period	trt	resp	resid	a	resid1	resid2	resid3
1	1	1	1	38	0	1	0	0	0
2	1	2	2	32	1	2	1	0	0
3	1	3	3	35	2	3	0	1	0
4	1	4	4	33	3	4	0	0	1
5	2	1	2	39	0	2	0	0	0
6	2	2	3	37	2	3	0	1	0
7	2	3	4	36	3	4	0	0	1
8	2	4	1	30	4	1	-1	-1	-1
9	3	1	3	45	0	3	0	0	0
10	3	2	4	38	3	4	0	0	1
11	3	3	1	37	4	1	-1	-1	-1
12	3	4	2	35	1	2	1	0	0
13	4	1	4	41	0	4	0	0	0
14	4	2	1	30	4	1	-1	-1	-1
15	4	3	2	32	1	2	1	0	0
16	4	4	3	33	2	3	0	1	0

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	244.6875000	20.3906250	22.24	0.0133
Error	3	2.7500000	0.9166667		
Corrected Total	15	247.4375000			

R-Square Coeff Var Root MSE resp Mean
0.988886 2.682808 0.957427 35.68750

Source	DF	Type I SS	Mean Square	F Value	Pr > F
cow	3	54.6875000	18.2291667	19.89	0.0175
period	3	147.1875000	49.0625000	53.52	0.0042
trt	3	40.6875000	13.5625000	14.80	0.0265
resid1	1	0.5625000	0.5625000	0.61	0.4906
resid2	1	0.5208333	0.5208333	0.57	0.5057
resid3	1	1.0416667	1.0416667	1.14	0.3646

Source	DF	Type III SS	Mean Square	F Value	Pr > F
cow	3	46.0833333	15.3611111	16.76	0.0223
period	3	147.1875000	49.0625000	53.52	0.0042
trt	3	7.8409091	2.6136364	2.85	0.2062
resid1	1	0.3750000	0.3750000	0.41	0.5679
resid2	1	1.0416667	1.0416667	1.14	0.3646
resid3	1	1.0416667	1.0416667	1.14	0.3646

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	3	7.84090909	2.61363636	2.85	0.2062
resid	3	2.12500000	0.70833333	0.77	0.5814

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		33.0000000 B	0.95742711	34.47	<.0001
cow	1	0.6250000 B	0.82915620	0.75	0.5057
cow	2	2.0000000 B	0.82915620	2.41	0.0948
cow	3	5.3750000 B	0.82915620	6.48	0.0075
cow	4	0.0000000 B	.	.	.
period	1	8.0000000 B	0.67700320	11.82	0.0013
period	2	1.5000000 B	0.67700320	2.22	0.1135
period	3	2.2500000 B	0.67700320	3.32	0.0449
period	4	0.0000000 B	.	.	.
trt	1	-3.6250000 B	1.58771324	-2.28	0.1066
trt	2	-4.0000000 B	1.58771324	-2.52	0.0862
trt	3	-1.3750000 B	1.58771324	-0.87	0.4502
trt	4	0.0000000 B	.	.	.
resid1		0.75000000	1.17260394	0.64	0.5679
resid2		1.25000000	1.17260394	1.07	0.3646
resid3		-1.25000000	1.17260394	-1.07	0.3646

Residual Effect of *A* increases response 0.75 units
Residual Effect of *B* increases response 1.25 units
Residual Effect of *C* decreases response 1.25 units
Residual Effect of *D* decreases response 0.75 units

$$\begin{aligned} \text{LSMEANS treatment 1} &= 35.6875 + (-3.625 - 0.25(-3.625 - 4.000 - 1.375)) = 34.3125 \\ &= \text{grand mean} + \text{treatment effect} \end{aligned}$$

$$\begin{aligned} \text{LSMEANS treatment 2} &= 35.6875 + (-4.000 - 0.25(-3.625 - 4.000 - 1.375)) = 33.9375 \\ &= \text{grand mean} + \text{treatment effect} \end{aligned}$$

Least Squares Means

trt	resp LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	34.3125000	1.0013012	<.0001	1
2	33.9375000	1.0013012	<.0001	2
3	36.5625000	1.0013012	<.0001	3
4	37.9375000	1.0013012	<.0001	4

Least Squares Means for effect trt
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: resp				
i/j	1	2	3	4
1		0.8285	0.2514	0.1066
2	0.8285		0.1968	0.0862
3	0.2514	0.1968		0.4502
4	0.1066	0.0862	0.4502	

Least Squares Means for Effect trt				
Difference				
		Between	95% Confidence Limits for	
i	j	Means	LSMean(i)-LSMean(j)	
1	2	0.375000	-4.677812	5.427812
1	3	-2.250000	-7.302812	2.802812
1	4	-3.625000	-8.677812	1.427812
2	3	-2.625000	-7.677812	2.427812
2	4	-4.000000	-9.052812	1.052812
3	4	-1.375000	-6.427812	3.677812

Designs Balanced for Residual Effects

- Consider the following two Latin Squares
- Suppose row = period and column = subject

D	C	B	A	D	C	B	A
C	D	A	B	C	A	D	B
B	A	D	C	B	D	A	C
A	B	C	D	A	B	C	D

- (Left) $C \rightarrow D$ twice, $A \rightarrow D$ once, $B \rightarrow D$ never
- (Right) Each treatment follows each other treatment once
- Right square balanced for residual effects
- If p even, can be balanced using p subjects
- If p odd, need multiple of $2p$ subjects.

Relative Efficiency of Latin Squares

What sample size is needed in order for a simpler design to have as much info as LS design?

$$E_{LS:RCB} = \left(\frac{\nu_{LS} + 1}{\nu_{LS} + 3} \right) \left(\frac{\nu_{RCB} + 3}{\nu_{RCB} + 1} \right) \frac{\sigma_{RCB}^2}{\sigma_{LS}^2}$$

$$E_{LS:CRD} = \left(\frac{\nu_{LS} + 1}{\nu_{LS} + 3} \right) \left(\frac{\nu_{CRD} + 3}{\nu_{CRD} + 1} \right) \frac{\sigma_{CRD}^2}{\sigma_{LS}^2},$$

where

$$\begin{aligned} \nu_{LS} &= (p-1)(p-2) \\ \nu_{RCB} &= (p-1)(p-1) \\ \nu_{CRD} &= (p-1)p \end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}_{RCB}^2 &= \frac{MS_{Rows} + (p-1)MS_E}{p} && \text{(row blocks eliminated)} \\ \hat{\sigma}_{RCB}^2 &= \frac{MS_{Cols} + (p-1)MS_E}{p} && \text{(column blocks eliminated)} \\ \hat{\sigma}_{CRD}^2 &= \frac{MS_{Rows} + MS_{Cols} + (p-1)MS_E}{p+1} && \text{(both eliminated)}\end{aligned}$$

Example of Diminishing Returns: Tree Plots

- 5 treatments over 25 subplots (no residual effects assumed)

III	IV	V	II	I
I	II	III	V	IV
V	I	II	IV	III
II	III	IV	I	V
IV	V	I	III	II

- 3 species in columns *within* blocks:

A	B	C	B	C	A	C	A	B	A	C	B	C	B	A
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- Question: Would it be better to change order for each row?

ABC	BCA	CAB	ACB	CBA
BAC	ACB	CBA	BAC	ACB
CBA	BAC	ACB	CBA	BAC
ACB	CBA	BAC	ACB	CBA
BAC	ACB	CBA	BAC	ACB

- Practical Answer: Farmers wouldn't do it.
- Statistical Answer: Diminishing returns – such careful blocking probably will not return much.

Youden Squares

- “Incomplete” Latin Squares Designs
 - #Cols ≠ #Rows = #Trts
 - #Rows ≠ #Cols = #Trts

- #Trts \neq #Cols = #Rows
- Youden Square is a Latin square from which at least one columns (or row or diagonal) is missing
 - A Youden Square is formed by taking away rows or columns or diagonals.
 - Every LS with rows/cols/diags missing *is not* a Youden Square
 - Youden Squares are *balanced*

Example:

	1	2	3	4
1	A	B	C	D
2	B	C	D	E
3	C	D	E	A
4	D	E	A	B
5	E	A	B	C

This is an example of a **Balanced Incomplete Block Design**.

Balanced Incomplete Blocks

- *Incomplete*: cannot fit all treatments into each block
 - *Balanced*: each **pair** of treatments occur together λ times
 - *Balanced*: $\text{Var}(\hat{\tau}_i - \hat{\tau}_j)$ is constant
- a treatments, b blocks, r replicates, and k treatments per blocks
 Total number of observations is $kb = ar = N$.

So treatment i occurs in r blocks. To have balance, each other treatment is equally likely to be with treatment i in a block. Since there are $k - 1$ other units in a block and $a - 1$ other treatments, the number of times each pair occurs together is

$$\lambda = r(k - 1)/(a - 1),$$

where λ is an integer. One way to generate this is

- Select $\binom{a}{k}$ blocks and assign each a different k treatment combination.
- The number of replicates is $r = \binom{a - 1}{k - 1}$.
- $\lambda = \binom{a - 2}{k - 2}$
- Sometimes can do this in less than $\binom{a}{k}$ blocks.

Extensive list of BIB designs found in Fisher and Yates (1963) and Cochran and Cox (1957).

Examples

$$a = 3, b = 3, k = 2, \rightarrow r = 2, \lambda = 1$$

1	2	3
A	B	A
B	C	C

$$a = 4, k = 2, b = 6, \rightarrow r = 3, \lambda = 1$$

1	2	3	4	5	6
A	A	A	B	B	C
B	C	D	C	D	D

$$a = 4, k = 3, b = 4, \rightarrow r = 3, \lambda = 2$$

1	2	3	4
A	A	A	B
B	B	C	C
C	D	D	D

Balanced Incomplete Block

- Similar construction as RCBD
- Statistical Model

$$y_{i,j} = \mu + \tau_i + \beta_j + \epsilon_{i,j} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases}$$

- Not all $y_{i,j}$ exist because of incompleteness
- Additive effect due to block / No Interaction
- Usual treatment and block restrictions

$$\sum \tau_i = 0 \quad \sum \beta_j = 0$$

- Nonorthogonality of treatments and blocks

Use Type III Sums of Squares and lsmeans

Analysis of Variance Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Blocks	SS_{Blk}	$b - 1$	MS_{Blk}	
Treatment	SS_{Trt}	$a - 1$	MS_{Trt}	F_0
Error	SS_E	$N - a - b + 1$	MS_E	
Total	SS_T	$N - 1$		

- $SS_T = \sum \sum y_{i,j}^2 - y_{..}^2/N$

- $SS_{Blk} = \frac{1}{k} \sum y_{.j}^2 - y_{..}^2/N$
- SS_{Trt} needs adjust for incompleteness

$$Q_i = y_{i.} - \frac{1}{k} \sum_{j=1}^k n_{i,j} y_{.j}, \text{ where } n_{i,j} = \begin{cases} 1 & \text{if treatment } i \text{ in block } j \\ 0 & \text{otherwise} \end{cases}$$

Treatment i 's **total** minus treatment i 's block averages

$$\sum Q_i = 0$$

Cannot consider n equals r (Q_i 's correlated)

Can consider $n = \lambda a/k < r$.

- $SS_{Trt(adj)} = k \sum Q_i^2 / \lambda a = n \sum \hat{\tau}_i^2$
- If $F_0 > F_{\alpha, a-1, N-a-b+1}$, then reject H_0

Model Estimates

- Design matrix X is RCBD with certain rows missing
- Can form normal equations to solve for $\hat{\mu}$, etc.

$$\begin{aligned} \hat{\mu} &= y_{..}/N \\ \hat{\tau}_i &= kQ_i/\lambda a \\ \hat{\beta}_j &= rQ'_j/\lambda b \end{aligned}$$

where

$$\begin{aligned} Q_i &= y_{i.} - \frac{1}{k} \sum n_{i,j} y_{.j} \\ Q'_j &= y_{.j} - \frac{1}{r} \sum n_{i,j} y_{i.} \end{aligned}$$

$$\begin{aligned} \text{Var}(Q_i) &= \text{Var}(y_{i.}) + \text{Var}\left(\frac{1}{k} \sum n_{i,j} y_{.j}\right) - 2\text{Cov}\left(y_{i.}, \frac{1}{k} \sum n_{i,j} y_{.j}\right) \\ &= r\sigma^2 + \frac{r}{k^2} k\sigma^2 - \frac{2}{k} r\sigma^2 \\ &= \frac{(k-1)r}{k} \sigma^2 \\ \text{Var}(\hat{\tau}_i) &= \left(\frac{k}{\lambda a}\right)^2 \text{Var}(Q_i) \\ &= \left(\frac{k}{\lambda a}\right)^2 \frac{(k-1)r}{k} \sigma^2 \\ &= \frac{k(a-1)}{\lambda a^2} \sigma^2 \end{aligned}$$

Power and Multiple Comparisons

- Power Calculations

- Assume a and k are known
- Limited to values of b such that λ an integer
 Non-centrality parameter $\delta = \lambda a \sum \tau_i^2 / k\sigma^2$
 Use integer values of λ and solve for b to get df
- Can also use confidence interval estimation method
 Can show $\text{Var}(\hat{\tau}_i - \hat{\tau}_j) = 2k\sigma^2 / \lambda a$
 Want α level CI to be no larger than $2D$
 $t_{\alpha/2, N-a-b+1} \sqrt{2k\sigma^2 / a\lambda} = D$

- Multiple Comparisons and Contrasts

- Must compute adjusted means (`lsmeans`)
- Adjusted mean is $\hat{\mu} + \hat{\tau}_i$
- Standard error of adjusted mean is $\sqrt{\sigma^2 \left(\frac{k(a-1)}{\lambda a^2} + \frac{1}{N} \right)}$
- Contrasts based on adjusted treatment totals

SAS Example (bib.sas)

```
options nocenter ps=60 ls=75;
options reset=all;
```

```
data example;
  input trt block resp @@;
  cards;
  1 1 73 1 2 74 1 4 71
  2 2 75 2 3 67 2 4 72
  3 1 73 3 2 75 3 3 68
  4 1 75 4 3 72 4 4 75
  ;
```

```
proc glm;
  class block trt;
  model resp = block trt;
  lsmeans trt / pdiff tdiff adjust=bon stderr;
  contrast 'a' trt 1 -1 0 0;
  estimate 'b' trt 0 0 1 -1;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	77.75000000	12.95833333	19.94	0.0024

Error	5	3.25000000	0.65000000
Corrected Total	11	81.00000000	

R-Square	Coeff Var	Root MSE	resp Mean
0.959877	1.112036	0.806226	72.50000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
block	3	55.00000000	18.33333333	28.21	0.0015
trt	3	22.75000000	7.58333333	11.67	0.0107

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	3	66.08333333	22.02777778	33.89	0.0010
trt	3	22.75000000	7.58333333	11.67	0.0107

Adjustment for Multiple Comparisons: Bonferroni

		Standard		LSMEAN
trt	resp LSMEAN	Error	Pr > t	Number
1	71.3750000	0.4868051	<.0001	1
2	71.6250000	0.4868051	<.0001	2
3	72.0000000	0.4868051	<.0001	3
4	75.0000000	0.4868051	<.0001	4

Least Squares Means for Effect trt
t for H0: LSMean(i)=LSMean(j) / Pr > |t|
Dependent Variable: resp

i/j	1	2	3	4
1		-0.35806	-0.89514	-5.19183
		1.0000	1.0000	0.0209
2	0.358057		-0.53709	-4.83378
	1.0000		1.0000	0.0284
3	0.895144	0.537086		-4.29669
	1.0000	1.0000		0.0464
4	5.191833	4.833775	4.296689	
	0.0209	0.0284	0.0464	

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
a	1	0.08333333	0.08333333	0.13	0.7349

**** Contrast a compares trt 1 and trt 2. 0.13 = (-0.35806)*(-0.35806)
**** Different p-values because Bonferroni used in first comparison

		Standard		
Parameter	Estimate	Error	t Value	Pr > t
b	-3.00000000	0.69821200	-4.30	0.0077

Interblock Analysis

Fixed effects analysis is known as *intrablock analysis*. If blocks are random, we can obtain additional information about τ 's by considering the information between block totals. Based

on our model,

$$y_{i,j} = \mu + \tau_i + (\beta_j + \epsilon_{i,j})$$

we can write the block totals as

$$y_{.j} = k\mu + \sum n_{i,j}\tau_i + \text{error}_i$$

and compute the least squares estimates of μ and τ_i .

- The least square estimate for τ_i is

$$\tilde{\tau}_i = \frac{\sum n_{i,j}y_{.j} - kr\bar{y}_{..}}{r - \lambda}$$

- Two estimates are uncorrelated $\rightarrow \text{Cov}(\hat{\tau}_i, \tilde{\tau}_i) = 0$.
- Use weighted combinations of estimates
- Weights based on the variances of the two estimates

$$\hat{\sigma}^2 = MS_E \text{ and } \hat{\sigma}_\beta^2 = \frac{(b-1)[MS_{Blk(adj)} - MS_E]}{a(r-1)}$$

$$\tau_i^* = \begin{cases} \frac{kQ_i(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2) + (\sum n_{i,j}y_{.j} - kr\bar{y}_{..})\hat{\sigma}^2}{(r-\lambda)\hat{\sigma}^2 + a\lambda(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2)} & \text{if } \hat{\sigma}_\beta > 0 \\ \frac{y_{i.} - (1/a)y_{..}}{r} & \text{if } \hat{\sigma}_\beta = 0 \end{cases}$$

- `proc mixed` computes combined estimates
- More information used \rightarrow more precise estimates
- Not much benefit if $\sigma_\beta^2 \gg \sigma^2$
- Can be worse in certain situations
- Combined estimate does not account for uncertainty in variance estimates
- Kenward-Rogers df correction also adjusts std error

SAS Commands

```
proc mixed;
class block trt;
model resp=trt / ddfm=kr;
random block;
lsmeans trt / diff adjust=bon;
contrast 'a' trt 1 -1 0 0;
estimate 'b' trt 0 0 1 -1;
run;
```

The Mixed Procedure

Model Information

Data Set WORK.EXAMPLE
 Dependent Variable resp
 Covariance Structure Variance Components
 Estimation Method REML
 Residual Variance Method Profile
 Fixed Effects SE Method Prasad-Rao-Jeske-
 Kackar-Harville
 Degrees of Freedom Method Kenward-Roger

Dimensions

Covariance Parameters 2
 Columns in X 5
 Columns in Z 4
 Subjects 1
 Max Obs Per Subject 12
 Observations Used 12
 Observations Not Used 0
 Total Observations 12

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	44.37333968	
1	1	34.22046396	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
block	8.0167
Residual	0.6500

Fit Statistics

-2 Res Log Likelihood	34.2
AIC (smaller is better)	38.2
AICC (smaller is better)	40.6
BIC (smaller is better)	37.0

Type 3 Tests of Fixed Effects

Effect	Num		Den		F Value	Pr > F
	DF	DF	DF	DF		
trt	3	5.03	11.33	0.0112		

Estimates

Label	Standard		DF	t Value	Pr > t
	Estimate	Error			
b	-2.9705	0.6995	5.03	-4.25	0.0080

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
a	1	5.03	0.08	0.7829

Least Squares Means						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	71.4131	1.4973	3.51	47.70	<.0001
trt	2	71.6164	1.4973	3.51	47.83	<.0001
trt	3	72.0000	1.4973	3.51	48.09	<.0001
trt	4	74.9705	1.4973	3.51	50.07	<.0001

Differences of Least Squares Means							
Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	1	2	-0.2033	0.6995	5.03	-0.29	0.7829
trt	1	3	-0.5869	0.6995	5.03	-0.84	0.4395
trt	1	4	-3.5574	0.6995	5.03	-5.09	0.0037
trt	2	3	-0.3836	0.6995	5.03	-0.55	0.6069
trt	2	4	-3.3541	0.6995	5.03	-4.79	0.0048
trt	3	4	-2.9705	0.6995	5.03	-4.25	0.0080

Differences of Least Squares Means				
Effect	trt	_trt	Adjustment	Adj P
trt	1	2	Bonferroni	1.0000
trt	1	3	Bonferroni	1.0000
trt	1	4	Bonferroni	0.0225
trt	2	3	Bonferroni	1.0000
trt	2	4	Bonferroni	0.0289
trt	3	4	Bonferroni	0.0480

Other Incomplete Designs

- Youden Square
 - Latin Square with one row (col) deleted
 - Each treatment occurs same number of times in each row (col)
 - Columns (rows) for BIBD
 - Analysis combination of Latin Square and BIBD
- Partially Balanced Incomplete Block Design
 - Doesn't require each pair to occur together λ times
 - Pair in associate class i appears together λ_i times
 - All treatments have same # of i th associates
 - Plus additional restrictions on # of associates
 - Extensive list in Bose, Clatworthy, and Shrikhande (1954)

Blocks					
1	2	3	4	5	6
2	3	4	5	6	1
4	5	6	1	2	3
3	4	5	6	1	2

- Cyclic Designs

- Includes some BIB and PBIB designs
- Consider situation where $r = mk$ and $b = ma$
- Determine m initial blocks, generate others by cycling
- PBIB example is cyclic design with initial block (1243)
- If $k = a$ and rows also blocks, get Latin Square

- Square, Cubic, and Rectangular Lattices

- Square: $a = k^2$; Cubic: $a = k^3$; Rect: $a = k(k + 1)$
- Square Example: Consider 9 treatments and blocks of size three

1	2	3	A	B	C
4	5	6	B	C	A
7	8	9	C	A	B

- Rep 1 blocks: (123)(456)(789) using Rows
- Rep 2 blocks: (147)(258)(369) using Columns
- Rep 3 blocks: (168)(249)(357) using Latin Squares
- Additional Reps obtained from orthogonal squares