

Lecture 6: Block Designs

Montgomery: Chapter 4

Nuisance Factor (may be present in experiment)

- Has effect on response but its effect is not of interest
- If unknown → Protecting experiment through randomization
- If known (measurable) but uncontrollable → Analysis of Covariance (Chapter 15 or 14 Section 3)
- If known and controllable → Blocking

Penicillin Experiment

In this experiment, four penicillin manufacturing processes were being investigated. They were denoted by A , B , C , and D . Yield was the response of primary interest. It was known that an important raw material, corn steep liquor, was quite variable. The experiment and its results were given below:

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

- eliminate blend-to-blend variation.
- provide wider inductive basis.

Randomized Complete Block Design

- b blocks each consisting of (partitioned into) a experimental units
- a treatments are randomly assigned to the experimental units within each block
- Typically after the runs in one block have been conducted, then moves to another block.
- Typical blocking factors: day, batch of raw material etc.
- results in restriction on randomization because randomization is only within blocks.
- data within a block are dependent on each other. When $a = 2$, randomized complete block design becomes paired two sample case.

Statistical Model

- b blocks and a treatments
- Statistical model is

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

μ - grand mean

τ_i - i th treatment effect

β_j - j th block effect

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

- The model is additive because within a fixed block, the block effect is fixed; for a fixed treatment, the treatment effect is fixed. In other words, blocks and treatments do not interact.
- parameter constraints: $\sum_{i=1}^a \tau_i = 0$; $\sum_{j=1}^b \beta_j = 0$

Estimates for Parameters

- Rewrite observation y_{ij} as:

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

- Compared with the model, we have

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

$$\hat{\epsilon}_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$$

Sum of Squares (SS)

- Can partition $SS_T = \sum \sum (y_{ij} - \bar{y}_{..})^2$ into

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

$$SS_{\text{Treatment}} = b \sum (\bar{y}_{i.} - \bar{y}_{..})^2 \text{ with df} = a - 1$$

$$SS_{\text{Block}} = a \sum (\bar{y}_{.j} - \bar{y}_{..})^2 \text{ with df} = b - 1$$

$$SS_E = \sum \sum (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \text{ with df} = (a - 1)(b - 1).$$

Hence:

- $SS_T = SS_{\text{Treatment}} + SS_{\text{Block}} + SS_E$
- The Mean Squares are SS divided by df, respectively, we have $MS_{\text{Treatment}}$, MS_{Block} and MS_E .

Hypothesis Testing

- $H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0$ vs H_1 : at least one is not
- Can show:

$$E(\text{MS}_E) = \sigma^2$$

$$E(\text{MS}_{\text{Treatment}}) = \sigma^2 + b \sum_{i=1}^a \tau_i^2 / (a - 1)$$

$$E(\text{MS}_{\text{Block}}) = \sigma^2 + a \sum_{j=1}^b \beta_j^2 / (b - 1)$$

- Use F-test to test H_0 :

$$F_0 = \frac{\text{MS}_{\text{Treatment}}}{\text{MS}_E} = \frac{\text{SS}_{\text{Treatment}} / (a - 1)}{\text{SS}_E / ((a - 1)(b - 1))}$$

- Caution testing block effects
 - Usually not of interest.
 - Randomization is restricted and within block runs are dependent.
 - Differing opinions on using F-test for testing blocking effects.
 - Can use ratio $\text{MS}_{\text{Block}}/\text{MSE}$ to check if blocking successful.

Analysis of Variance Table

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

$$SS_T = \sum \sum y_{ij}^2 - y_{..}^2 / N$$

$$SS_{\text{Treatment}} = \frac{1}{b} \sum y_{i.}^2 - y_{..}^2 / N$$

$$SS_{\text{Block}} = \frac{1}{a} \sum y_{.j}^2 - y_{..}^2 / N$$

$$SS_E = SS_T - SS_{\text{Treatment}} - SS_{\text{Block}}$$

Decision Rule: 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 in 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 between 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

$$\sum \sum y_{ij} = 565 \text{ and } \sum \sum y_{ij}^2 = 26867$$

$$y_{1.} = 139, y_{2.} = 145, y_{3.} = 153 \text{ and } y_{4.} = 128; y_{.1} = 182, y_{.2} = 176, \text{ and } y_{.3} = 207$$

$$SS_T = 26867 - 565^2/12 = 265$$

$$SS_{Trt} = (139^2 + 145^2 + 153^2 + 128^2)/3 - 565^2/12 = 111$$

$$SS_{Block} = (182^2 + 176^2 + 207^2)/4 - 565^2/12 = 135$$

$$SS_E = 265 - 111 - 135 = 19; F_0 = (111/3)/(19/6) = 11.6; P\text{-value} < 0.01$$

Checking Assumptions (Diagnostics)

- Assumptions
 - Model is additive (no interaction between treatment effects and block effects) (additivity assumption)
 - Errors are independent and normally distributed
 - Constant variance
- Checking normality:
 - Histogram, QQ plot of residuals, Shapiro-Wilk Test.
- Checking constant variance
 - Residual Plot: Residuals vs \hat{y}_{ij}
 - Residuals vs blocks
 - Residuals vs treatments

Checking Assumptions (Continued)

- Additivity
 - Residual Plot: residuals vs \hat{y}_{ij}
 - If residual plot shows curvilinear pattern, interaction between treatment and block likely exists
 - Interaction: block effects can be different for different treatments
- Formal test: Tukey's One-degree Freedom Test of Non-additivity
- If interaction exists, usually try to use transformation to eliminate interaction

Treatments Comparison

- Multiple Comparisons/Contrasts
 - procedures (methods) are similar to those for Completely Randomized Design (CRD)
 - n is replaced by b in all formulas
 - Degrees of freedom error is $(b - 1)(a - 1)$
- Example : Comparison of Detergents
 - Tukey's Method ($\alpha = .05$)

$$q_{\alpha}(a, df) = q_{\alpha}(4, 6) = 4.896.$$

$$CD = \frac{q_{\alpha}(4,6)}{\sqrt{2}} \sqrt{\text{MSE}(\frac{1}{b} + \frac{1}{b})} = 4.896 \sqrt{\frac{19}{6*3}} = 5.001$$

Comparison of Means

Treatments

4	1	2	3
42.67	46.33	48.33	51.00
A	A	A	
		B	B

Using SAS

```

options nocenter ls=78;
options 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 / alpha=0.05 tukey lines;
  output out=diag r=res p=pred;

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

```

```
run;
```

```
proc gplot;
```

```
  plot res*soap / haxis=axis1;
```

```
  plot res*stain / haxis=axis1;
```

```
  plot res*pred;
```

```
run;
```

Output

Dependent Variable: y

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

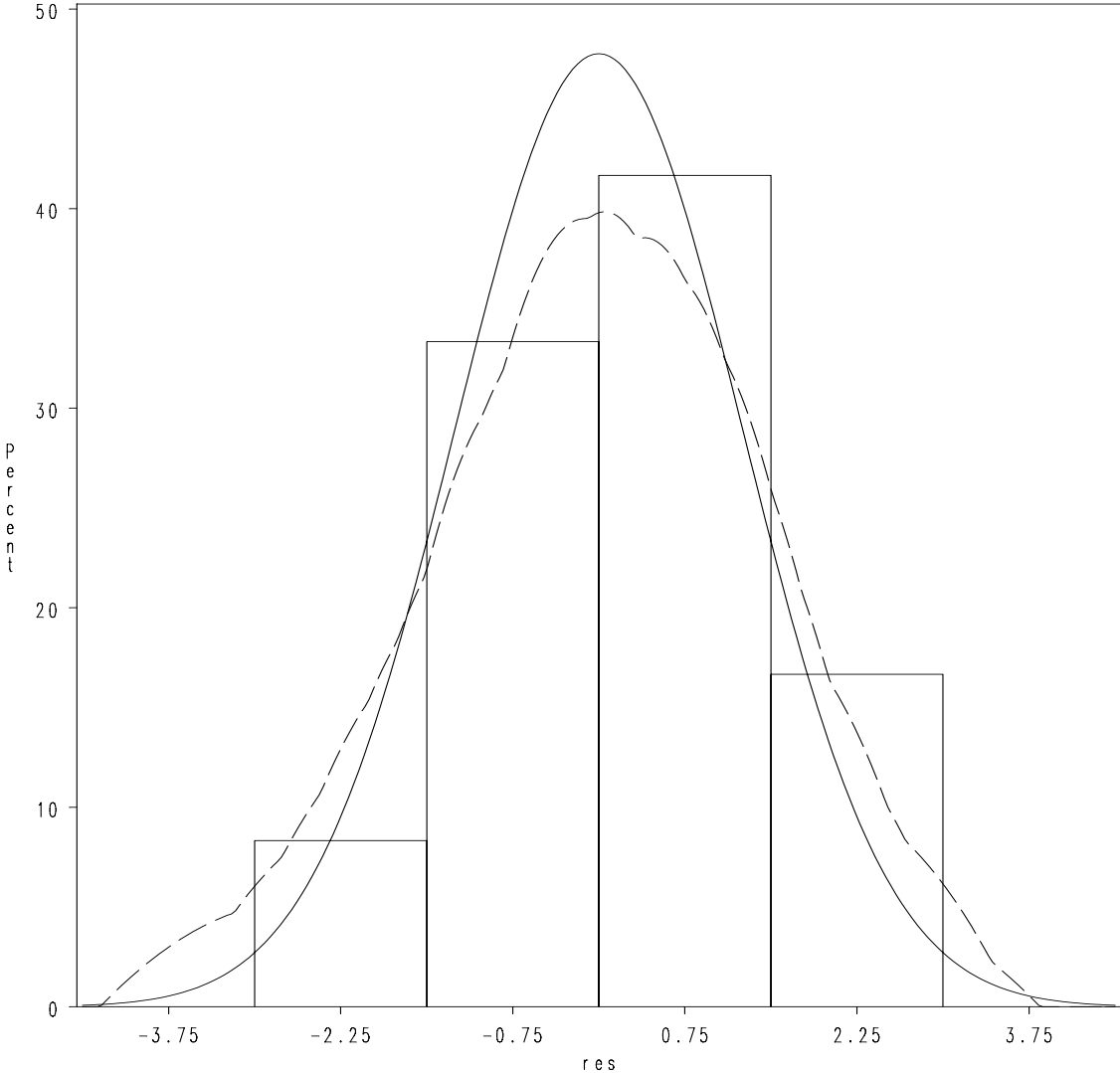
Source	DF	Type III SS	Mean Square	F Value	Pr > F
soap	3	110.9166667	36.9722222	11.78	0.0063
stain	2	135.1666667	67.5833333	21.53	0.0018

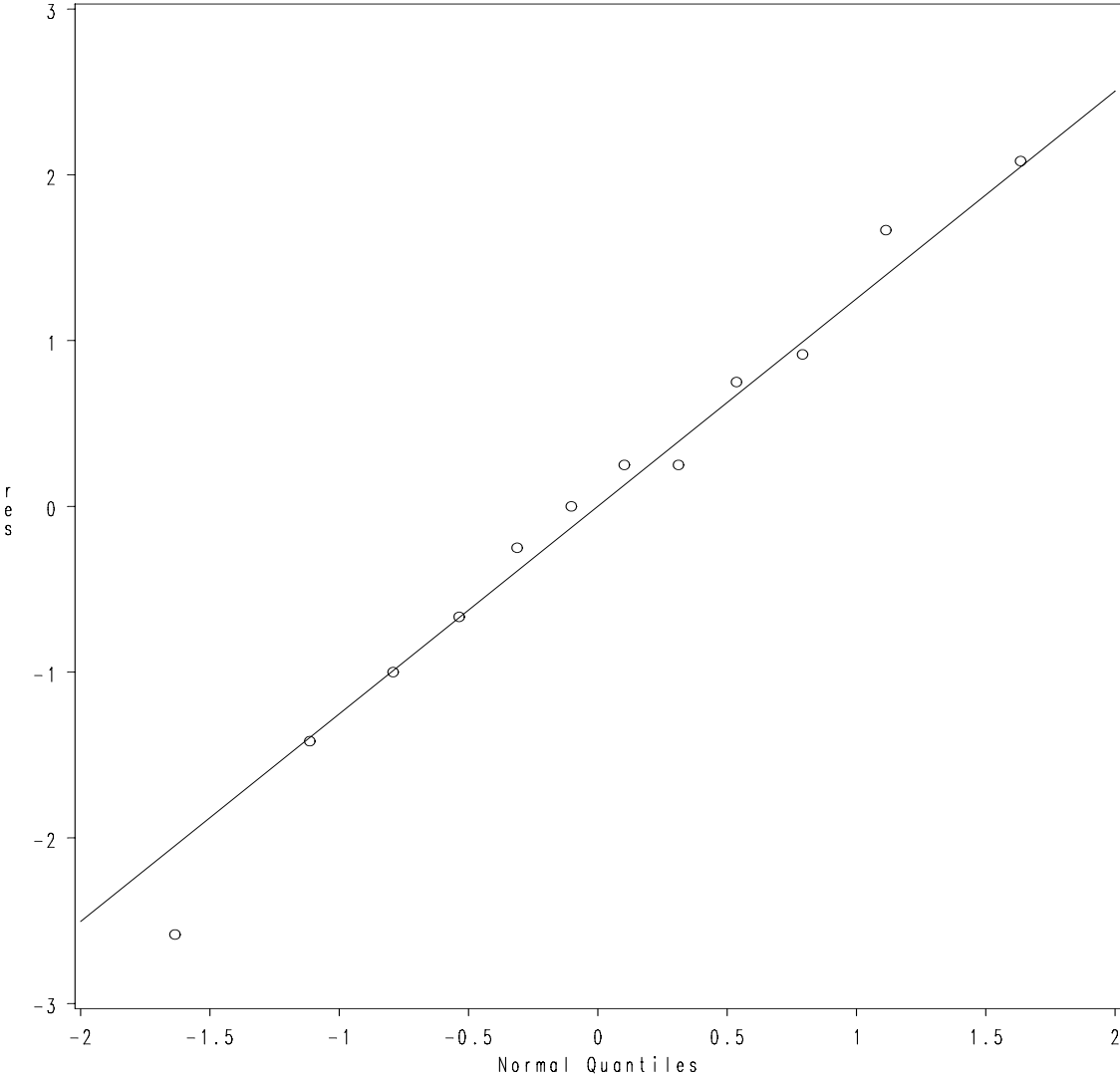
Tukey's Studentized Range (HSD) Test for res

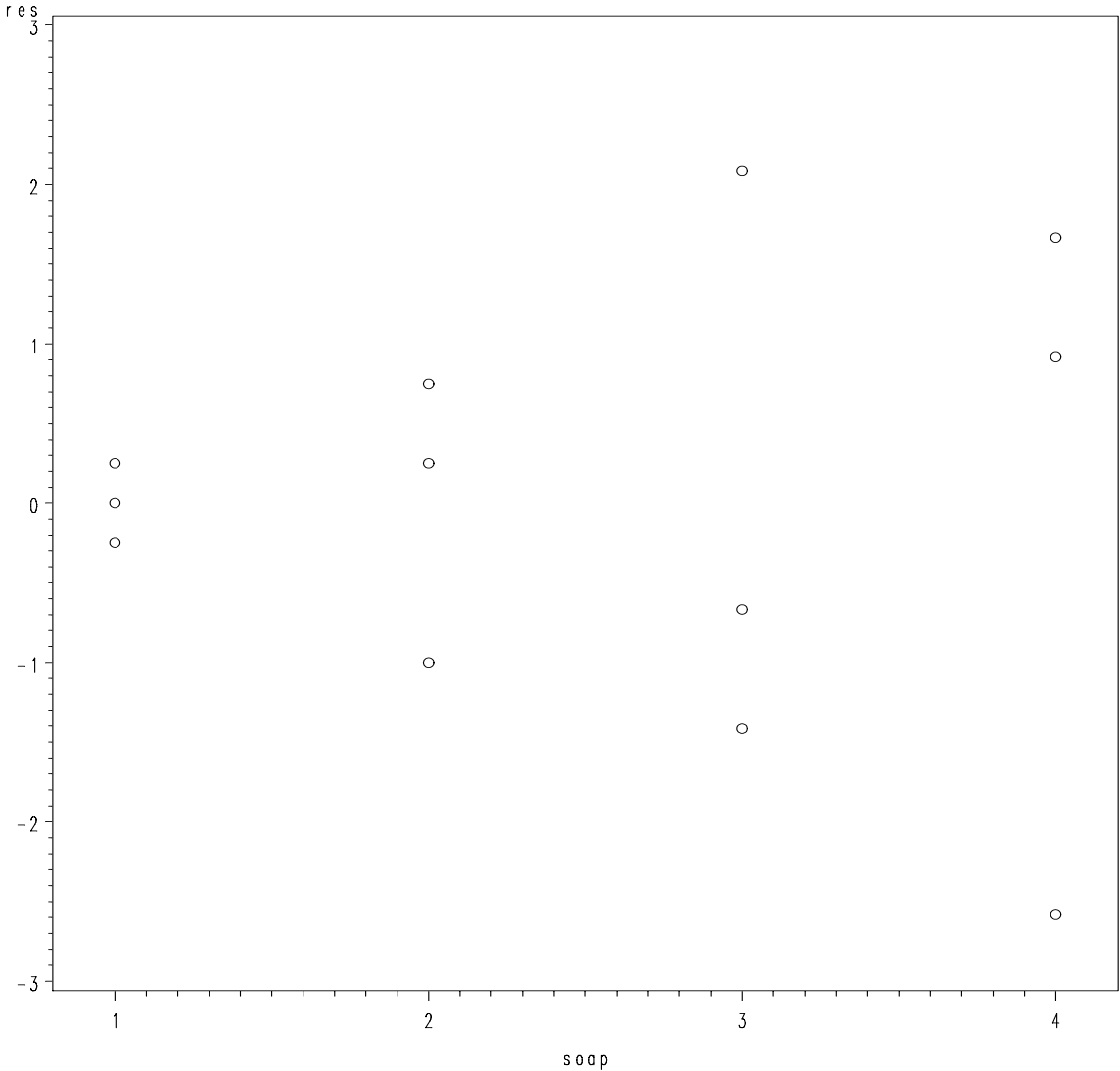
Alpha	0.05
Error Degrees of Freedom	6
Error Mean Square	3.138889
Critical Value of Studentized Range	4.89559
Minimum Significant Difference	5.007

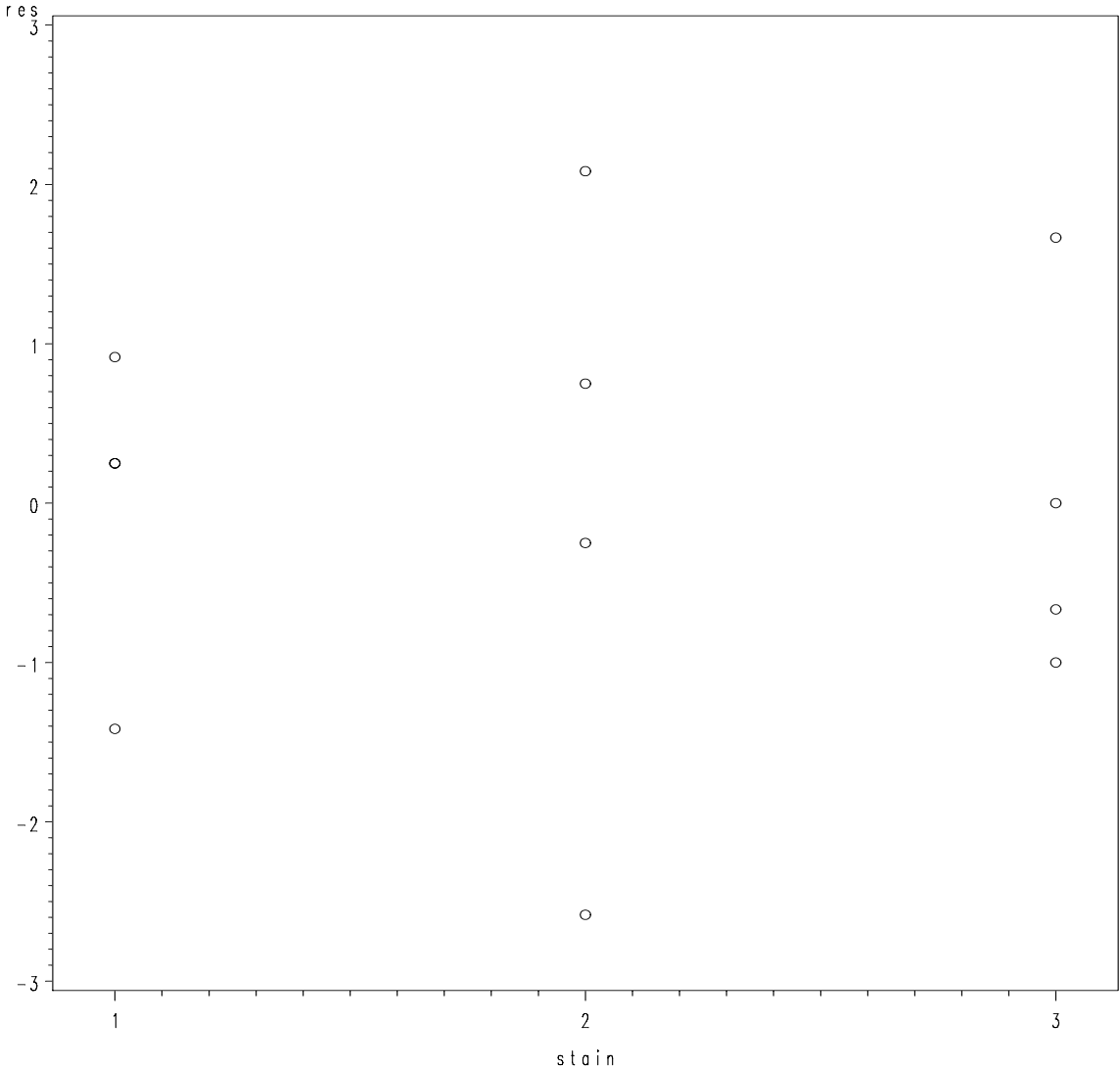
Means with the same letter are not significantly different.

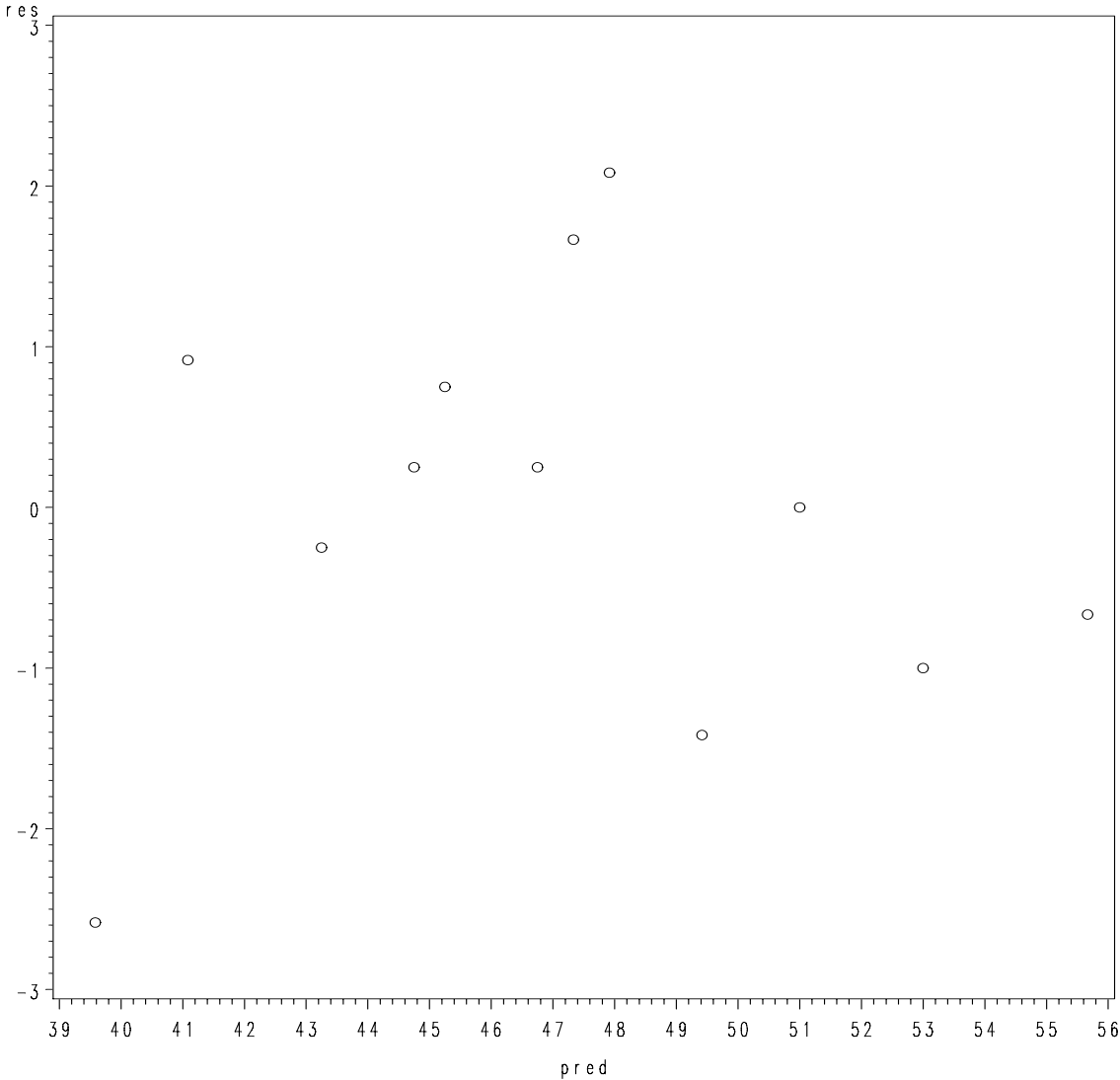
	Tukey Grouping	Mean	N	soap
	A	51.000	3	3
	A			
	A	48.333	3	2
	A			
B	A	46.333	3	1
B				
B		42.667	3	4











Tukey's Test for Non-additivity

- Additivity assumption (or no interaction assumption) is crucial for block design without replicates.
- To check the interaction between block and treatment **fully** needs $(a - 1)(b - 1)$ degree of freedom. It is not affordable when without replicates.
- Instead consider a special type of interaction. Assume following model (pages 190-193 or pages 178-181)

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \epsilon_{ij}$$

- $H_0 : \gamma = 0$ vs $H_1 : \gamma \neq 0$

Sum of Squares caused by possible interaction:

$$SS_N = \frac{\left[\sum_i \sum_j y_{ij} y_{i.} y_{.j} - y_{..} (SS_{\text{Trt}} + SS_{\text{Blk}} + y_{..}^2/ab) \right]^2}{ab SS_{\text{Trt}} SS_{\text{Blk}}} \quad df = 1.$$

Remaining error SS: $SS'_E = SS_E - SS_N$, $df = (a - 1)(b - 1) - 1$

Test Statistic:

$$F_0 = \frac{SS_N/1}{SS'_E/[(a - 1)(b - 1) - 1]} \sim F_{1, (a-1)(b-1)-1}$$

- Decision rule: Reject H_0 if $F_0 > F_{\alpha, 1(a-1)(b-1)-1}$.

A Convenient Procedure to Calculate SS_N , SS'_E and F_0

- 1 Fit additive model $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$
- 2 Obtain \hat{y}_{ij} and $q_{ij} = \hat{y}_{ij}^2$
- 3 Fit the model $y_{ij} = \mu + \tau_i + \beta_j + q_{ij} + \epsilon_{ij}$

Use the test for q_{ij} in the ANOVA table with type III SS and ignore the tests for the treatment and block factors.

Example 5-2 from Montgomery

- Impurity in chemical product is affected by temperature and pressure. We will assume temperature is a blocking factor. The data is shown below. We will 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

$$SS_N = 0.0985, SS'_E = 1.9015, F_0 = .36, P - \text{value} = 0.566$$

Do Not Reject, there appears to be no interaction between block and treatment.

SAS Code

```

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 glm;
class blk trt;
model y=blk trt;
output out=one r=res p=pred;

data two;
set one;
q=pred*pred;

proc glm data=two;
class blk trt;
model y=blk trt q/ss3; run;

```

Output

From the first model statement:

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			

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

Output

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 III SS	Mean Square	F Value	Pr > F
blk	2	1.25864083	0.62932041	2.32	0.1690 XXX
trt	4	1.09624963	0.27406241	1.01	0.4634 XXX
q	1	0.09852217	0.09852217	0.36	0.5660

XXX: not meaningful for testing blocks and treatments

RCBD with Replicates

- a treatments ($i = 1, 2, \dots, a$)
- b blocks ($j = 1, 2, \dots, b$)
- n observations for each treatment in each block ($l = 1, 2, \dots, n$)

$$y_{ijl} = \mu + \tau_i + \beta_j + \epsilon_{ijl} \quad \left\{ \begin{array}{l} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ l = 1, 2, \dots, n \end{array} \right.$$

- Similar assumptions as before. $N = abn$ and many more degree of freedom to get around. It allows interaction (but we have to be really careful about the possible interaction)

Assuming no interaction between Ttr and Blk: ANOVA

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

For multiple comparison, df_E becomes $abn - a - b + 1$ and the number of replicates for a fixed treatment now is bn instead of n . Hence, the formulas need to be modified accordingly.

Not Assuming no interaction between Trt and Blk: ANOVA

- Assess additivity (no interaction) by Sum of Squares due to interaction $SS_{\text{Trt*Blk}}$.
- Interaction and error are not confounded; their SS's are separated

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

- Usually, we do not want to see large (significant) $SS_{\text{Blk*Trt}}$ because this will (first) affect the interpretation of the results and (second) the prediction of the fitted model. Sometimes, transformation is considered to eliminate interaction.