Randomized Complete Block Design

(RCBD)

Bruce A Craig

Department of Statistics
Purdue University

Blocking Design

- A **nuisance factor** is an factor that effects the response y but is not of interest to the researcher
- When planning an experiment must always consider the possibility of nuisance factors
- If unknown nuisance factor, randomization provides protection from bias but error variance will be inflated
- If known (and measurable) but uncontrollable use ANCOVA
- If known and controllable, we use a blocking design
- Extension of a paired t-test where pairs are the blocks

Randomized Complete Block Design (RCBD)

- Arrange b blocks, each containing a "similar" EUs
- Randomly assign a treatments to the EUs in block
- The linear statistical model is

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

τ_i - ith treatment effect
β_j - jth block effect
$\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$

- Model includes additional **additive** block effect

Partitioning the SS

- Rewrite observation as:

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

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

$$\begin{matrix} b\sum(\bar{y}_{i.} - \bar{y}_{..})^2 + & a\sum(\bar{y}_{.j} - \bar{y}_{..})^2 & + \sum\sum(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \\ \mathrm{SS}_{\mathrm{Treatment}} \quad + & \mathrm{SS}_{\mathrm{Block}} & + \quad \mathrm{SS}_{\mathrm{E}} \end{matrix}$$

- Under H_0, all SS/σ^2 independent χ^2
- Ratio of SS will be F distributed

Hypothesis Testing

- Can show (in the **fixed case**):

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

- Use F-test to test equality of treatment effects

$$F_0 = \frac{\mathrm{SS}_{\mathrm{Treatment}}/(a-1)}{\mathrm{SS}_{\mathrm{E}}/((a-1)(b-1))}$$

- Could also use F-test for inference on block effects but...
  - Usually not of interest (i.e., you chose to block for a reason)
  - Blocks not randomized to experimental units
  - Best to view F_0 and its P-value as a measure of blocking success

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}	$ba-1$		

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

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

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

Example - Consumer Testing

An experiment was designed to study the effectiveness of four different detergents to remove stains. Four white t-shirts were stained with one of three common stains and allowed to sit for a day. The shirts were then washed and the following "removal" readings (higher is better) were obtained with specially-designed equipment. 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

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

Constructing ANOVA Table

Using the earlier formulas...

$$\text{SS}_{\text{T}} = 26867 - 565^2/12 = 264.92$$
$$\text{SS}_{\text{Trt}} = (139^2 + 145^2 + 153^2 + 128^2)/3 - 565^2/12 = 110.92$$
$$\text{SS}_{\text{Block}} = (182^2 + 176^2 + 207^2)/4 - 565^2/12 = 135.17$$
$$\text{SS}_{\text{E}} = 265 - 111 - 135 = 18.83$$

$F_0 = (111/3)/(19/6) = 11.78$
P-value < 0.01 (Reject H_0 - At least one detergent effect is different from 0)

Diagnostics

- Assumptions / Model Conditions
  1. Model is correct (additive block effect assumption)
  2. Errors independent, Normally distributed, constant variance
- Assessing normality

  Histogram, normal probability plot of residuals
- Assessing constant variance

  Residuals vs blocks, treatments, and $\hat{y}_{ij}$
- Assessing additivity

  Is the block effect different for different treatments?

  Plot y vs block, connecting y from same treatment

  If roughly same pattern across treatments, additivity reasonable

  Tukey's Test of Non-additivity (formal test of specific alternative)

Comparisons of Treatments

- Multiple Comparisons/Contrasts
  - Similar procedures as before with CRD
  - n is replaced by b in all standard error formulas
  - Degrees of freedom error are $(b-1)(a-1)$
- Example: Comparison of detergents
  - Pairwise comparisons using Tukey's adjustment $(\alpha = .05)$

    6 degrees of freedom error $\rightarrow q_{0.05}(4,6) = 4.90$
    $s_{\overline{y}} = \sqrt{\mathrm{MSE}/3} = \sqrt{(18.83/6)/3} = 1.02$
    Least Significant difference is $4.90(1.02) = 5.01$.

Treatments

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

Using SAS

```
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 plots=all;
 class stain soap;  model y = soap stain;
 means soap / tukey lines; output out=diag r=res p=pred;
proc univariate noprint;
 qqplot res / normal (L=1 mu=0 sigma=est);
 hist res /normal (L=1 mu=0 sigma=est) kernel(L=2 K=quadratic);
proc gplot;
 plot res*soap/haxis=axis1; plot res*stain/haxis=axis1;
 plot res*pred;
run;
```

SAS Output

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		
Corr Total	11	264.9166667			

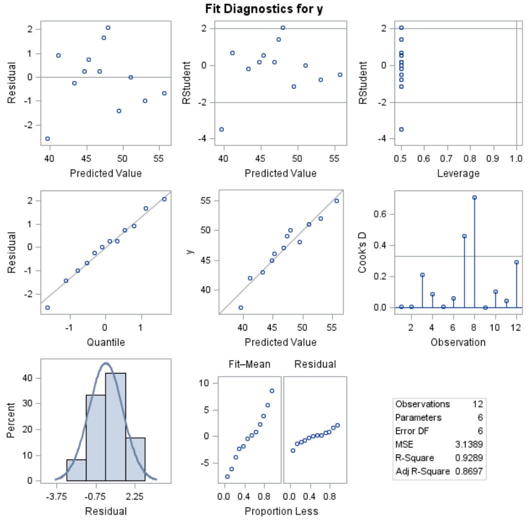
R-Square	Coeff Var	Root MSE	y Mean
0.928908	3.762883	1.771691	47.08333

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

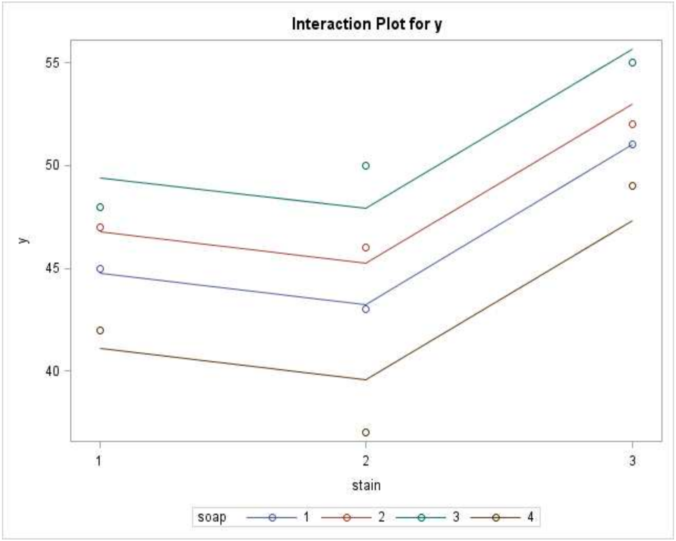
Source	DF	Type III SS	Mean Square	F Value	Pr > F
soap	3	110.9166667	36.9722222	11.78	0.0063
stain	2	135.1666667	67.5833333	21.53	0.0018

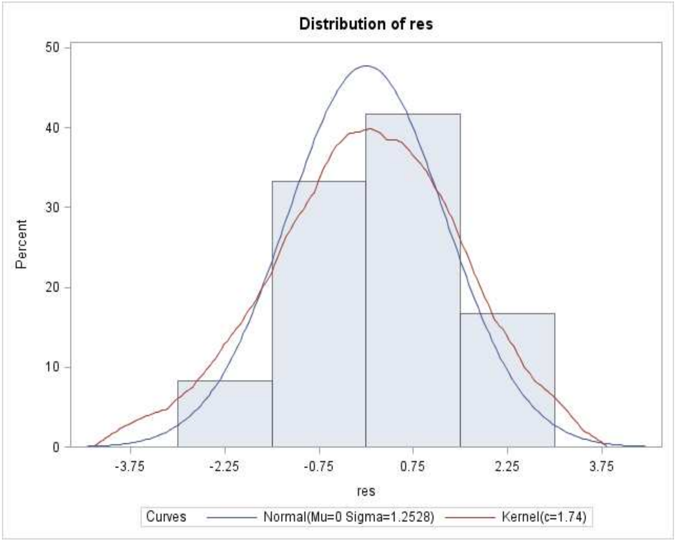
SAS Output

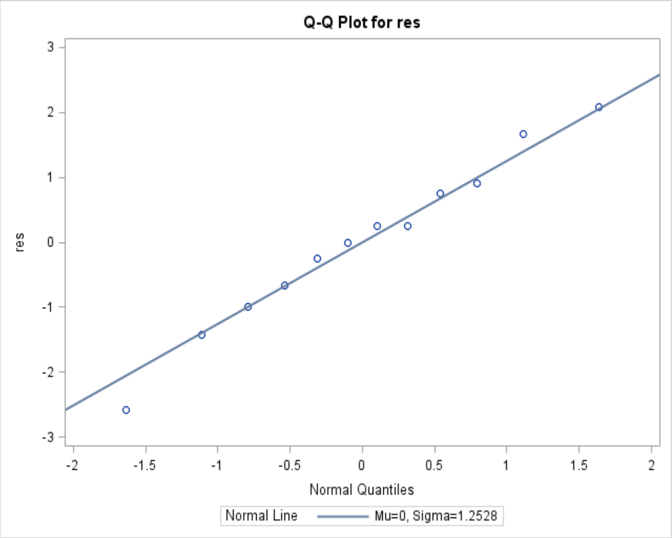
```
Tukey's Studentized Range (HSD) Test for y
Alpha                                          0.05
Error Degrees of Freedom                          6
Error Mean Square                          3.138889
Critical Value of Studentized Range         4.89559
Minimum Significant Difference               5.0076

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
```

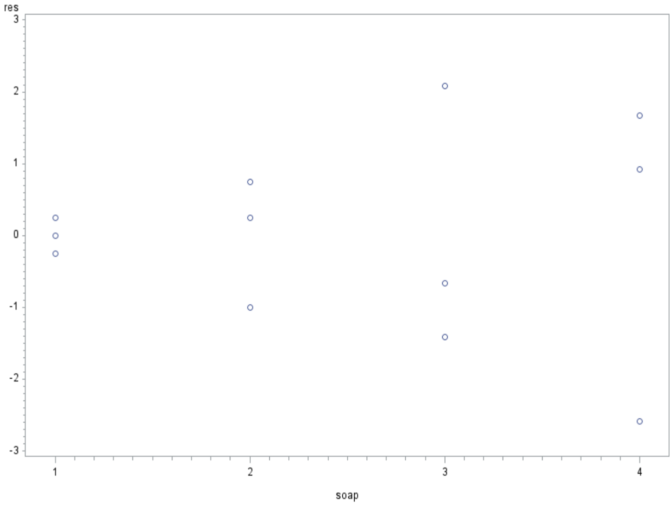


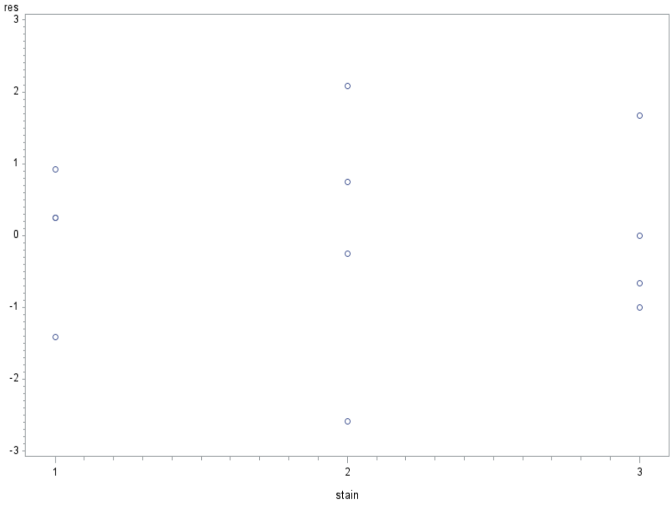
Observations	12
Parameters	6
Error DF	6
MSE	3.1389
R-Square	0.9289
Adj R-Square	0.8697

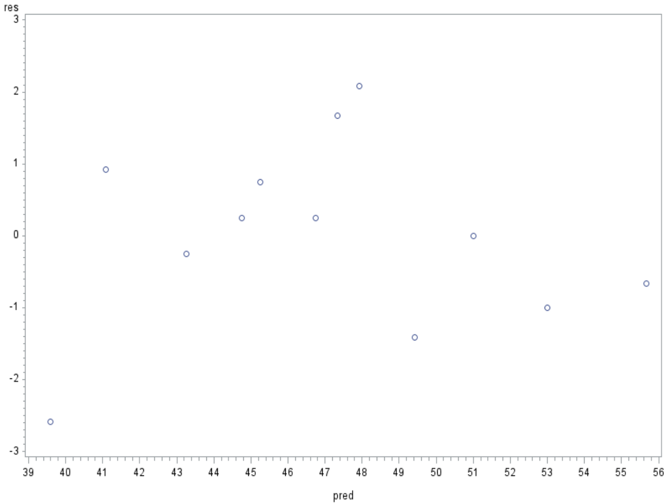

Interaction Plot for y
55
50
45
40
y
1
2
3
stain
soap
1
2
3
4


Distribution of res
Percent
50
40
30
20
10
0
-3.75
-2.25
-0.75
0.75
2.25
3.75
res
Curves
Normal(Mu=0 Sigma=1.2528)
Kernel(c=1.74)


Q-Q Plot for res
res
3
2
1
0
-1
-2
-3
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
Normal Quantiles
Normal Line
Mu=0, Sigma=1.2528


res
3
2
1
0
-1
-2
-3
1
2
3
4
soap


res
3
2
1
0
-1
-2
-3
1
2
3
stain


res
3
2
1
0
-1
-2
-3
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
pred

Summary

- Residuals appear relatively Normal but there may be a nonconstant variance issue or outlier
- Will consider removing to assess influence on inference
- Also considered a mixed model allowing the error variance to be different for soap groups
    - Model allowing different variance per soap does not converge
    - Model allowing different variance for Soaps #1 and #2 and for Soaps #3 and #4 suggests little difference in fit (BIC=31.8 versus 31.7)
    - Also, multiple comparison results do not change
- Therefore will use original model to draw conclusions

Underlying Regression Model

- Simple extension of CRD design matrix
- Add additional $b - 1$ columns to represent block
- Block columns orthogonal to treatment columns
- Thus, order of fit does not matter

Design Matrix for Detergent Study

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

Missing Values

- When missing observations (missing at random)

  Orthogonality lost - missing row in design matrix X

  Order of fit now important

- Procedures

  1. Regression approach

     Use Type III SS's (general regression significance test)

  2. Estimate missing value (single or multiple imputation)

     One option: Choose value that minimizes $\mathrm{SS_E}$ (minimize its contribution)

$$
\begin{aligned}
SS_{\mathrm{E}} &= \sum\sum y_{ij}^2 - y_{..}^2/ab - \frac{1}{b}\sum y_{i.}^2 + y_{..}^2/ab - \frac{1}{a}\sum y_{.j}^2 + y_{..}^2/ab \\
&= x^2 - \frac{1}{b}(y'_{i.} + x)^2 - \frac{1}{a}(y'_{.j} + x)^2 + \frac{1}{ab}(y'_{..} + x)^2 + R \\
x &= \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}
\end{aligned}
$$

Example - Detergent Study

- Suppose $y_{4,2} = 37$ was missing
- Estimation Approach

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

  - Estimate is

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

  - Plug this in and fit model but adjust error df!!!
- Regression: $\hat{\sigma}^2 = 1.097$
- Estimate: $\hat{\sigma}^2 = 1.097$ (must divide by 5 not 6)

```
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;
 classes stain soap;  model y = soap stain;
 output out=diag r=res p=pred;
 means soap / lsd lines; lsmeans soap / adjust=tukey lines;

data new1; set wash;
 if y=. then y=42.1666666666;

proc glm;
 classes stain soap; model y = soap stain;
 output out=diag r=res p=pred;
 means soap / tukey lines;  lsmeans soap / adjust=tukey lines;
run;
```

Regression - Type III

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		
C. Total	10	154.0000000			

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

Regression - Type III

```
Tukey's Studentized Range (HSD) Test for y
  Alpha                                        0.05
  Error Degrees of Freedom                        5
  Error Mean Square                        1.097222
  Critical Value of Studentized Range       5.21819
  Minimum Significant Difference             3.3472
  Harmonic Mean of Cell Sizes              2.666667
        NOTE: Cell sizes are not equal.

Tukey Grouping          Mean      N    soap
           A         51.0000      3    3
           A
      B    A         48.3333      3    2
      B
      B              46.3333      3    1
      B
      B              45.5000      2    4 ** Not correct
```

Regression - Type III

```
                     Least Squares Means
      Adjustment for Multiple Comparisons: Tukey-Kramer
                              Standard                LSMEAN
soap           y LSMEAN          Error    Pr > |t|    Number
1            46.3333333      0.6047650      <.0001         1
2            48.3333333      0.6047650      <.0001         2
3            51.0000000      0.6047650      <.0001         3
4            44.3888889      0.7807483      <.0001         4
                                                   LSMEAN
                                y LSMEAN    soap     Number

**Means based on model
  parameter estimates      A    51.00000    3             3
  adjust for missing       A
  block obs           B    A    48.33333    2             2
                      B
                      B    C    46.33333    1             1
                           C
                           C    44.38889    4             4
```

Estimate Approach

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

$$
\begin{aligned}
F_0 &= \frac{71.95/3}{5.49/5} \\
&= 21.84 \\
\mathrm{P-value} &= 0.0027
\end{aligned}
$$

Regression - Type III

```
Tukey's Studentized Range (HSD) Test for y
  Alpha                                       0.05
  Error Degrees of Freedom                       6
  Error Mean Square                       0.914352
  Critical Value of Studentized Range      4.89559
  Minimum Significant Difference            2.7027

Tukey Grouping         Mean      N     soap
           A        51.0000      3     3
           A
      B    A        48.3333      3     2    Same estimates as regr
      B                                     approach (lsmeans) but
      B    C        46.3333      3     1    not correct df and MSE
           C
           C        44.3889      2     4
```

Tukey's Test for Non-additivity

- Considers a special type of 1 df interaction
- Other types of interaction may also be considered
- Tukey assumes the following model (page 203-206)

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

- Use regression approach to test $H_0 : \gamma = 0$

Tukey's Test for Non-additivity

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

     Partitioning $\mathrm{SS_E}$ into $\mathrm{SS_N}$ and remainder

     The parameter $\hat{\gamma}$ is the slope estimate

     $\mathrm{SS}_{\mathrm{non-additivity}} = \hat{\gamma}^2 \sum\sum (q_{ij} - \hat{q}_{ij})^2$

$$F_0 = \frac{\mathrm{SS_N}/1}{(\mathrm{SS_E} - \mathrm{SS_N})/((a-1)(b-1)-1)}$$

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 are 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

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

$$F_0 = \frac{.0985/1}{1.9015/7} = .36$$

$F_0 < F_{1,7}$ - Do Not Reject.

SAS Procedures

```
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=resid1 r=res1 p=pred1;

data predsq; set resid1;
  predsq1 = pred1*pred1;

proc glm;
 class blk trt; model predsq1=blk trt;
 output out=resid2 r=res2 p=pred2;

proc glm;  model res1=res2; run;
```

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

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.09852217	0.09852217	0.67	0.4266
Error	13	1.90147783	0.14626753		
Corrected Total	14	2.00000000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
res2	1	0.09852217	0.09852217	0.67	0.4266

Parameter	Estimate	Std Error	t Value	Pr > |t|
Intercept	-.0000000000	0.09874800	-0.00	1.0000
res2	0.0369458128	0.04501655	0.82	0.4266

Random Block/Treatment Effects

- Could randomly select trts and/or blocks
- Do not need to worry about additivity
- Interaction considered random effect
- Interaction variance appears in all EMS
- Perform usual F-test (ratio of MS)
- Use Proc Mixed instead of Proc Glm
- Otherwise underestimate variability in trt means

```
data wash;
input stain soap y @@;                          ***Letting SAS compute EMS;
cards;                                          ***Adding soap*stain interaction;
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 soap*stain;
 random stain soap*stain / test;                ***EMS provided when random used;
run;
-------------------------------------------

Source          DF    Sum of Squares    Mean Square    F Value      Pr > F
Model           11      264.9166667      24.0833333        .           .
Error            0           .                .        ***With no df for error
Total           11      264.9166667                    ***Var(error) and Var(stain*soap)
                                                       ***confounded
Source                DF       Type III SS     Mean Square    F Value      Pr > F
SOAP                   3       110.9166667      36.9722222        .           .
STAIN                  2       135.1666667      67.5833333        .           .
STAIN*SOAP             6        18.8333333       3.1388889        .           .

Source          Type III Expected Mean Square
SOAP            Var(Error) + Var(STAIN*SOAP) + Q(SOAP)
STAIN           Var(Error) + Var(STAIN*SOAP) + 4 Var(STAIN)  *** The EMS;
STAIN*SOAP      Var(Error) + Var(STAIN*SOAP)
```

```
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 / test;          ***test option uses EMS as guide for F tests;
 lsmeans soap / stderr tdiff lines;

proc mixed;
 class stain soap; model y = soap;
 random stain; lsmeans soap / tdiff;
run;

proc glimmix;                  ***Mixed model procedure with lines option;
 class stain soap; model y = soap;
 random stain; lsmeans soap / lines;
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			

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

soap	y LSMEAN	Standard Error	Pr > |t|	LSMEAN Number	
1	46.3333333	1.0228863	<.0001	1	
2	48.3333333	1.0228863	<.0001	2	***Std errors
3	51.0000000	1.0228863	<.0001	3	***for indiv means
4	42.6666667	1.0228863	<.0001	4	***not correct

		y LSMEAN	soap	LSMEAN Number
	A	51.00000	3	3
	A			
B	A	48.33333	2	2
B				
B		46.33333	1	1
	C	42.66667	4	4

***Std errors for trt
difference is
correct

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	Std 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

***Std errors are correct

Differences of Least Squares Means

Effect	soap	_soap	Estimate	Std 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

The Glimmix Procedure

Covariance Parameter Estimates

Cov Parm	Estimate	StdError
stain	16.1111	16.9019
Residual	3.1389	1.8122

Type III Tests of Fixed Effects

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

soap Least Squares Means

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

soap	Estimate		
3	51.0000		A
			A
2	48.3333	B	A
		B	
1	46.3333	B	
4	42.6667		C

Choice of Sample Size

- Same as determining the number of blocks (b)
- Use same tables/procedures with b replacing n
- Can focus on overall F test for fixed or random trts
- Can also focus on specific contrast or contrasts

Example 4.2

```
data params;
  input a alpha d var;
  cards;
  4 .05 6 9
;
data new;
 set params;
 do b=2 to 15;
    df = (a-1)*(b-1);          ***df error now (a-1)(b-1);
    nc = b*d*d/(2*var);        ***replaced n by b in nc for CRD;
    fcut = finv(1-alpha,a-1,df);
    beta=probf(fcut,a-1,df,nc);
    power = 1- beta;
    output;
 end;
proc print;
run;
```

Example 4.2

Obs	a	alpha	d	var	b	df	nc	fcut	beta	power
3	4	0.05	6	9	4	9	8	3.86255	0.54011	0.45989
4	4	0.05	6	9	5	12	10	3.49029	0.39437	0.60563
5	4	0.05	6	9	6	15	12	3.28738	0.27616	0.72384
6	4	0.05	6	9	7	18	14	3.15991	0.18672	0.81328
7	4	0.05	6	9	8	21	16	3.07247	0.12254	0.87746
8	4	0.05	6	9	9	24	18	3.00879	0.07836	0.92164

Appears that for 80% power we need 7 blocks

RCBD with Replication

- What if multiple trt observations per block?
  - b blocks, a treatments, n replications/block

$$y_{ijl} = \mu + \tau_i + \beta_j + \epsilon_{ijl} \quad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ l = 1, 2, \ldots, n \end{cases}$$

- When would this occur?
  - Have large field with very gradual slope
  - Blocks expensive but observations cheap
- Increases df_{E} (or allows for interaction)

RCBD with Replication

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

RCBD with Replication

- Usual diagnostics checks
- Replace b by 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
Blk	SS_{Blk}	$b-1$	MS_{Blk}	
Treatment	$SS_{\mathrm{Treatment}}$	$a-1$	$MS_{\mathrm{Treatment}}$	F_0
Blk*Trt	$SS_{\mathrm{Blk*Trt}}$	$(b-1)(a-1)$	$MS_{\mathrm{Blk*Trt}}$	
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 at your disposal a field consisting of sixteen subplots (in a 4x4 grid). 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 subplots near the road or river will "behave" differently than subplots in the middle of the field, we might want to create $b = 3$ blocks. Block 1 consists of the four subplots along the road. Block 2 consists of the 4 subplots along the river and Block 3 consists of the eight subplots 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.

Example

- Statistical model is

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

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_0
Blocks	SS_{Block}	2	MS_{Block}	
Interaction	$SS_{\text{Trt}*\text{Blk}}$	6	$MS_{\text{Trt}*\text{Blk}}$	
Treatment	$SS_{\text{Treatment}}$	3	$MS_{\text{Treatment}}$	F_0
Error	SS_{E}	4	MS_{E}	
Total	SS_{T}	15		

Example

- If we used four blocks, we could not separate error and interaction
- In this analysis, SS_E based on observations within block 3 because it has replicates
- Only 4 df error so not a very powerful design
- Later, we will discuss the concept of pooling. In this case, we might test for interaction and if it is not significant, remove it thereby combining it with error. This increase the df from 4 to 10.

Background Reading

- Statistical analysis: Montgomery Section 4.1.1
- Checking model conditions: Montgomery Section 4.1.2
- Additivity assumption when blocks fixed: Montgomery Section 4.1.3
- Random block effects: Montgomery Section 4.1.3
- Block size determination : Montgomery Section 4.1.3
- Regression approach / Missing values: Montgomery Section 4.1.4