Topic 7: Randomized complete block design (RCBD)

Montgomery: chapter 4

Prof. Lingling An
University of Arizona

Outline

- Block design
- ANOVA for RCBD (one observation per cell)
- Model adequacy checking
  - Mentioned assumptions
  - Non-additivity
- Post-ANOVA comparisons
- Missing values (two approaches)

Nuisance Factor

Nuisance Factor (may be present in experiment)

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

Example: Penicillin Experiment

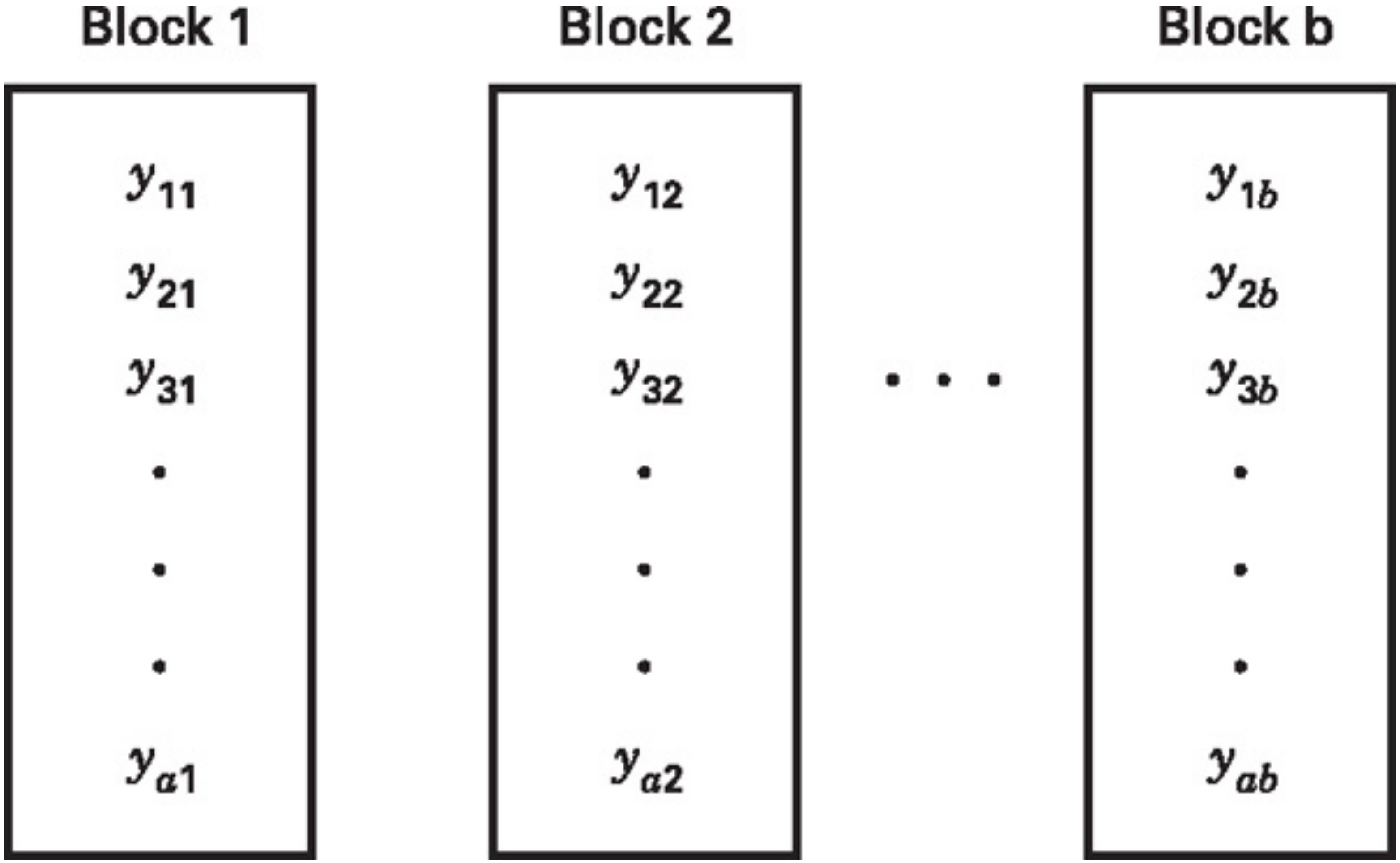
In this experiment, four penicillin manufacturing processes (A, B, C and D) were being investigated. Yield was the response. It was known that an important raw material, corn steep liquor, was quite variable. The experiment and its results were given below:

	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

Example: Penicillin Experiment -2

- Blend is a nuisance factor, treated as a block factor;
- (Complete) Blocking: all the treatments are applied within each block, and they are compared within blocks.
- Advantage: Eliminate blend-to-blend (between-block) variation from experimental error variance when comparing treatments.
- Cost: degree of freedom.

Randomized Complete Block Design (RCBD)



RCBD -2

- 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 move 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 related to each other. When $a = 2$, randomized complete block design becomes paired two sample case.

Statistical Model: two-way ANOVA

- b blocks and a treatments
- Statistical model is

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

μ - grand mean

τ_i - ith treatment effect

β_j - jth block effect

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

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

Estimates for Parameters

- Rewrite observation y_{ij} as:

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

- Compared with the model

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

- we have

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

Sum of Squares (SS)

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

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

$$\begin{array}{llll} \mathrm{SS_{Treatment}} & = b\sum(\overline{y}_{i.} - \overline{y}_{..})^2 & = b\sum\hat{\tau}_i^2 & \text{df} = a-1 \\ \mathrm{SS_{Block}} & = a\sum(\overline{y}_{.j} - \overline{y}_{..})^2 & = a\sum\hat{\beta}_j^2 & \text{df} = b-1 \\ \mathrm{SS_E} & = \sum\sum(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2 & = \sum\sum\hat{\epsilon}_{ij}^2 & \text{df} = (a-1)(b-1). \end{array}$$

Hence:

- $\mathrm{SS_T} = \mathrm{SS_{Treatment}} + \mathrm{SS_{Block}} + \mathrm{SS_E}$

- The Mean Squares are

$\mathrm{MS_{Treatment}} = \mathrm{SS_{Treatment}}/(a-1),\ \mathrm{MS_{Block}} = \mathrm{SS_{Block}}/(b-1),$
and $\mathrm{MS_E} = \mathrm{SS_E}/(a-1)(b-1).$

Testing Basic Hypotheses

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

  E(MS$_{\mathrm{E}}$)=σ^2

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

  E(MS$_{\mathrm{Block}}$) = $\sigma^2 + a\sum_{j=1}^{b} \beta_j^2/(b-1)$
- Use F-test to test H_0:

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

- Caution testing block effects
  - Usually not of interest.
  - Randomization is restricted: Differing opinions on F-test for testing blocking effects.
  - Can use ratio MS_{Block}/MSE to check if blocking successful.
  - Block effects can be random effects. (considered fixed effects in this chapter)

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	$\text{SS}_{\text{Treatment}}$	$a-1$	$\text{MS}_{\text{Treatment}}$	F_0
Error	SS_{E}	$(b-1)(a-1)$	MS_{E}	
Total	SS_{T}	$ab-1$		

$\text{SS}_{\text{T}} = \sum\sum y_{ij}^2 - y_{..}^2/N$

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

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

$\text{SS}_{\text{E}}=\text{SS}_{\text{T}} - \text{SS}_{\text{Treatment}} - \text{SS}_{\text{Block}}$

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

Another 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

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

Another example -2

$y_{1.} = 139$, $y_{2.} = 145$, $y_{3.} = 153$ and $y_{4.} = 128$;

$y_{.1} = 182$, $y_{.2} = 176$, and $y_{.3} = 207$

$$\mathrm{SS_T} = 26867 - 565^2/12 = 265$$

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

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

$$\mathrm{SS_E} = 265 - 111 - 135 = 19;$$

$$F_0 = (111/3)/(19/6) = 11.6;$$

P-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 Assumptions -2

- 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 - 3

- 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 transformation to eliminate interaction

SAS code

```
options nocenter ls=78;
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
;
run;

proc print data=wash;
run;

proc glm;
class stain soap;
model y = soap stain;
output out=diag r=res p=pred;
Run;
```

Obs	stain	soap	y
1	1	1	45
2	1	2	47
3	1	3	48
4	1	4	42
5	2	1	43
6	2	2	46
7	2	3	50
8	2	4	37
9	3	1	51
10	3	2	52
11	3	3	55
12	3	4	49

ANOVA tables

Overall ANOVA:

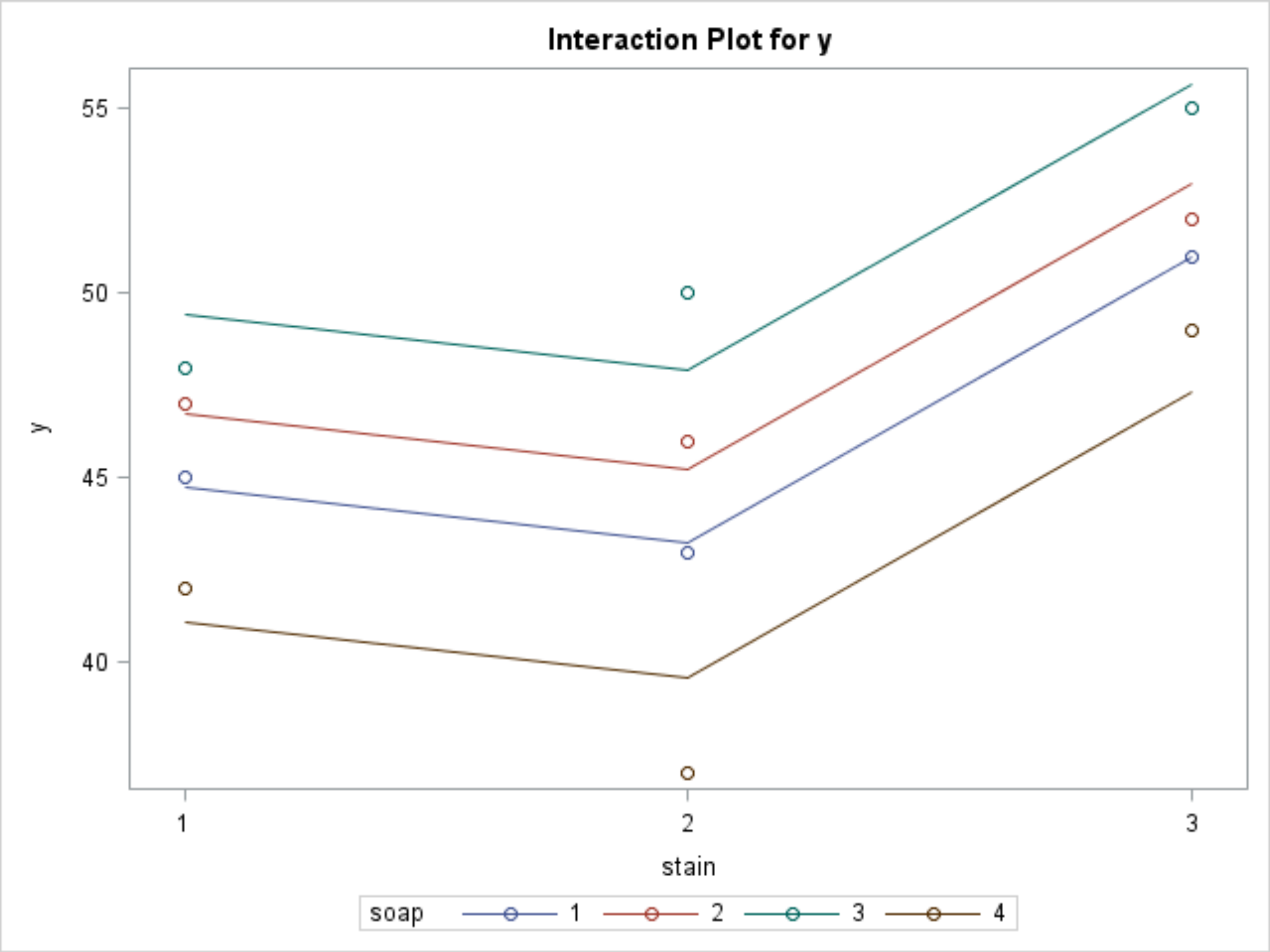
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	246.0833333	49.2166667	15.68	0.0022
Error	6	18.8333333	3.1388889		
Corrected Total	11	264.9166667			

Fit statistics:

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

Type III Model ANOVA

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


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

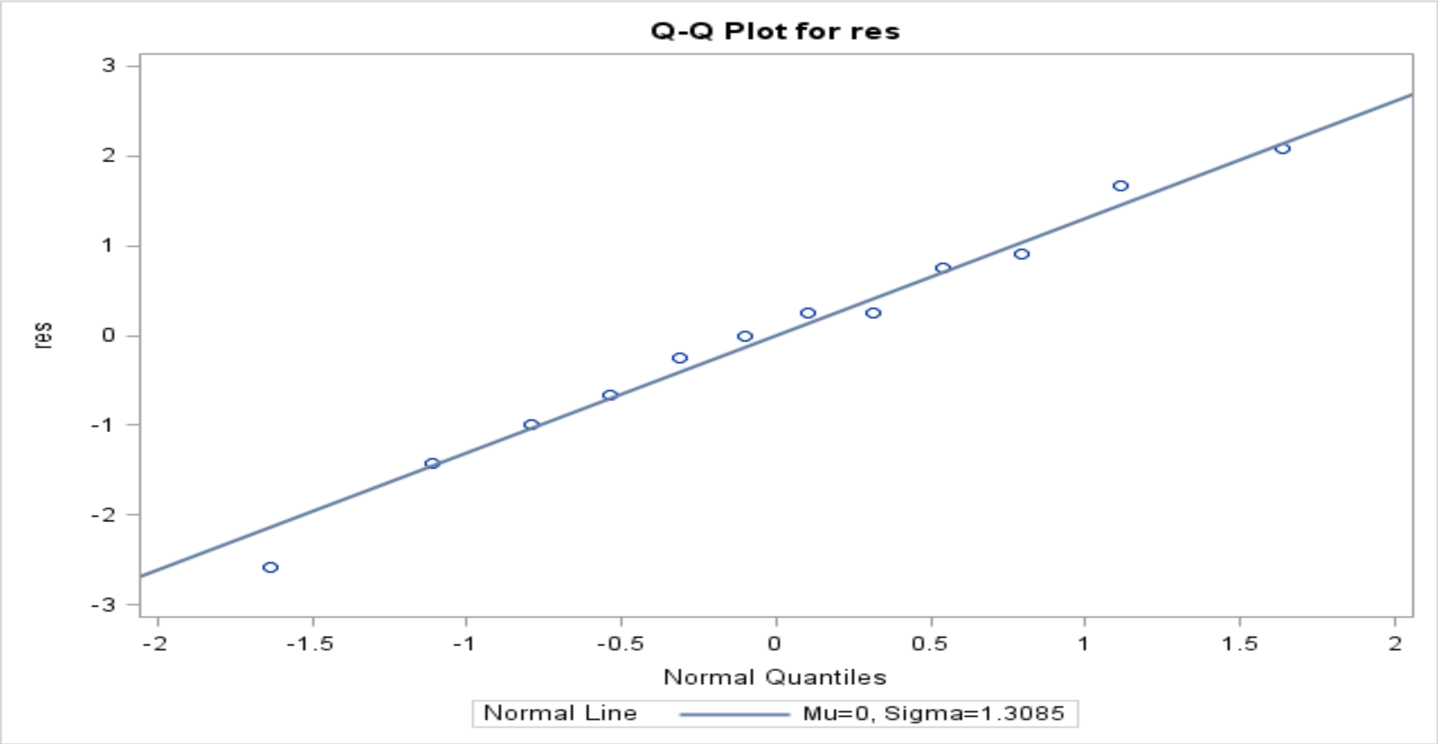
SAS code: model adequacy checking

```
/* check normality */
proc univariate data=diag normal;
var res;
qqplot res / normal (L=1 mu=est sigma=est);
Run;

/* check outliers */
data outlier;
set diag;
stdres=res/1.771691;
run;

proc print data=outlier;
run;
```

Output: normality checking



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.985667	Pr < W	0.9973
Kolmogorov-Smirnov	D	0.090905	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.017532	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.129122	Pr > A-Sq	>0.2500

Output: outlier checking

Obs	stain	soap	y	res	pred	stdres
1	1	1	45	0.25000	44.7500	**0.14111**
2	1	2	47	0.25000	46.7500	**0.14111**
3	1	3	48	-1.41667	49.4167	**-0.79961**
4	1	4	42	0.91667	41.0833	**0.51740**
5	2	1	43	-0.25000	43.2500	**-0.14111**
6	2	2	46	0.75000	45.2500	**0.42332**
7	2	3	50	2.08333	47.9167	**1.17590**
8	2	4	37	-2.58333	39.5833	**-1.45812**
9	3	1	51	0.00000	51.0000	**0.00000**
10	3	2	52	-1.00000	53.0000	**-0.56443**
11	3	3	55	-0.66667	55.6667	**-0.37629**
12	3	4	49	1.66667	47.3333	**0.94072**

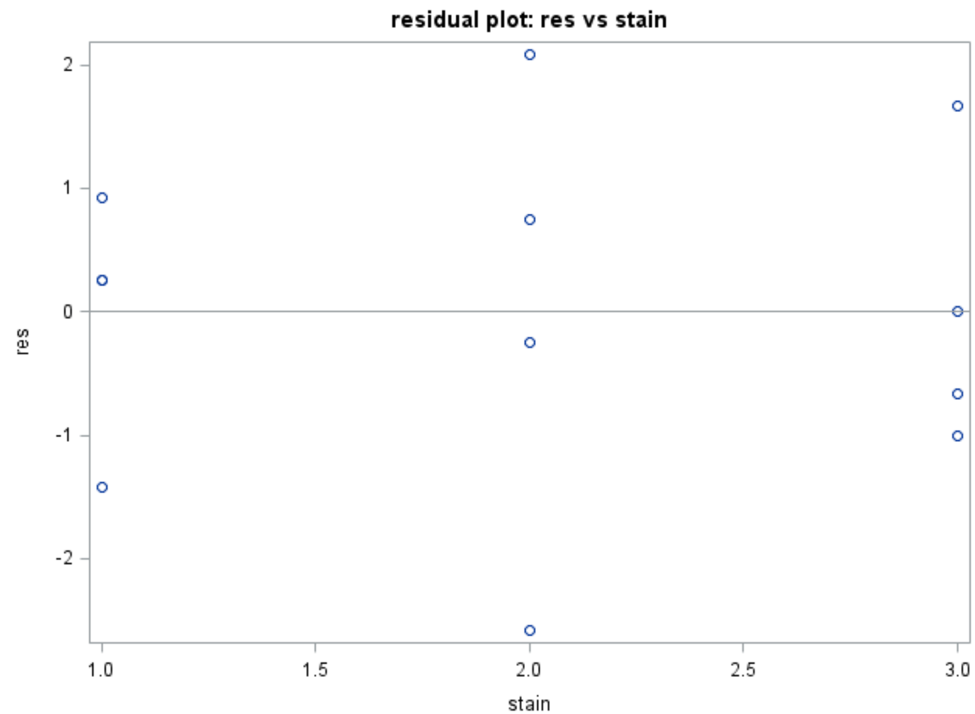
SAS code: model adequacy checking (cont’d)

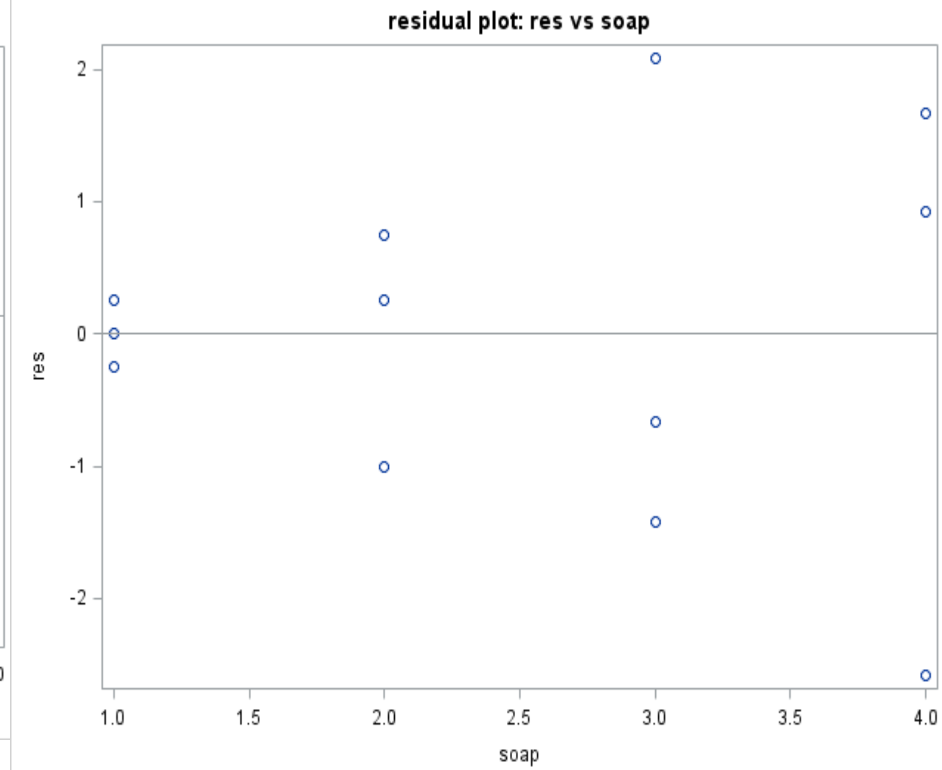
Constant variance checking

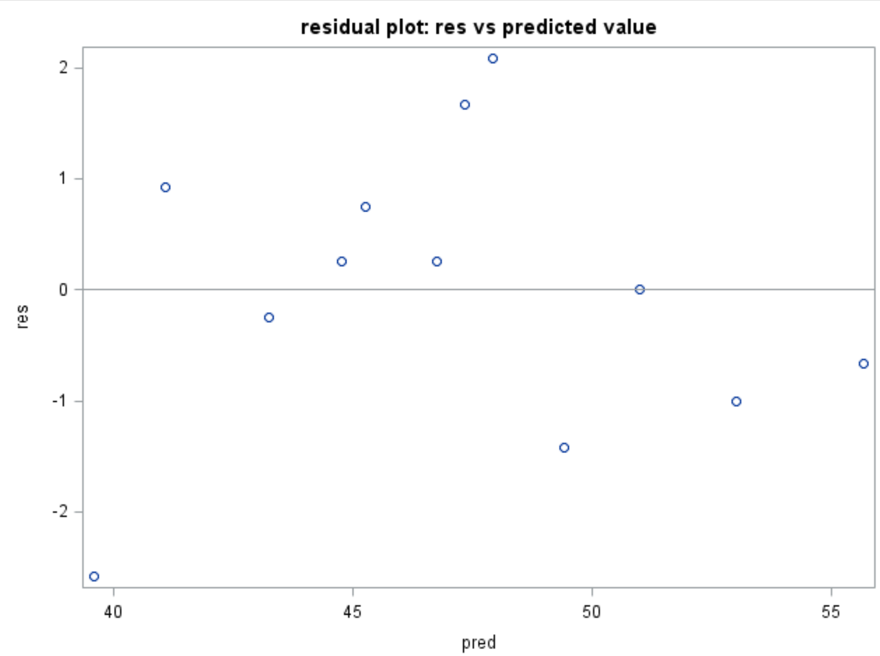
```
title 'residual plot: res vs soap ';
proc sgplot data=diag;
scatter x=soap y=res;
refline 0;
run;

title 'residual plot: res vs stain ';
proc sgplot data=diag;
scatter x=stain y=res;
refline 0;
run;

title 'residual plot: res vs predicted value ';
proc sgplot data=diag;
scatter x=pred y=res;
refline 0;
run;
```


residual plot: res vs stain
res
stain
2
1
0
-1
-2
1.0
1.5
2.0
2.5
3.0


residual plot: res vs soap
res
soap
2
1
0
-1
-2
1.0
1.5
2.0
2.5
3.0
3.5
4.0


residual plot: res vs predicted value
res
pred
2
1
0
-1
-2
40
45
50
55

Model adequacy checking: additivity

- We used a linear statistical model for RCBD:

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

- In other word, it is additive.
- The linear model is very useful, but in some situations it may be inadequate.
  - i.e. there may be an interaction between the treatment and block

Tukey's Test for Non-additivity

- Additivity assumption (or no interaction assumption) is crucial for block designs or experiments.
- 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

  (p204)

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

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

Sum of Squares caused by possible interaction:

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

Remaining error SS: $\mathsf{SS}'_{\mathsf{E}} = \mathsf{SS}_{\mathsf{E}} - \mathsf{SS}_{\mathsf{N}}$, $\; df = (a-1)(b-1) - 1$

Test Statistic:

$$F_0 = \frac{\mathsf{SS}_N/1}{\mathsf{SS}'_{\mathsf{E}}/[(a-1)(b-1)-1]} \sim F_{1,(a-1)(b-1)-1}$$

- 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.

SAS code for checking additivity

Still use the data “wash”

```
proc glm;
class stain soap;
model y = soap stain;
output out=diag r=res p=pred;
Run;

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

proc glm data=two;
class stain soap;
model y=stain soap q/ss3;
run;
```

Output: First “glm” procedure

Overall ANOVA table

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			

Type III model ANOVA table

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

Output: second “glm” procedure

Overall ANOVA table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	254.2775785	42.3795964	19.92	0.0024
Error	5	10.6390882	2.1278176		
Corrected Total	11	264.9166667			

Type III model ANOVA table

Source	DF	Type III SS	Mean Square	F Value	Pr > F
stain	2	12.91205848	6.45602924	3.03	~~0.1372~~
soap	3	13.11217219	4.37072406	2.05	~~0.2251~~
q	1	8.19424514	8.19424514	3.85	0.1070

X: not meaningful for testing blocks and treatments

Type I, III sum of squares

- $Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + e_i$
- **Type I (sequential) Sums of Squares**
  - The Type I Sums of Squares for b_1 are the Sums of Squares obtained from fitting b_1 over and above the mean;
  - The Type I Sums of Squares for b_2 are the Sums of Squares obtained from fitting b_2 after $b_{1.}$
  - etc.
- **Type III (marginal) Sums of Squares**
  - The Sums of Squares obtained by fitting each effect after all the other terms in the model,
  - The marginal (Type III) Sums of Squares do not depend upon the order in which effects are specified in the model.

More info: http://afni.nimh.nih.gov/sscc/gangc/SS.html

Post- ANOVA 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 $(\alpha = .05)$

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

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

Comparison of Treatment Means

Treatments

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

SAS code for pairwise comparisons

```
proc glm data=wash;
class stain soap;
model y = soap stain;
means soap / alpha=0.05 tukey lines;
Run;
```

Output

Tukey's Studentized Range (HSD) Test for y

Note: **This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.**

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

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

Missing Values

- When missing
  - Orthogonality lost
  - Design unbalanced
- Procedures
  1. Exact (Regression) approach

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

  2. Approximate approach : Estimate missing value

SAS code: regression approach for missing values

```
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 data=wash;
class stain soap;
model y = soap stain;
output out=diag r=res p=pred;
run;
```

output

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			

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

Use Type III SS when reporting with missing values

2 Approximate approach : Estimate missing value

Choose value to minimize SS_E

Take derivative and set equal to zero

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

$$
x = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}
$$

Missing value -- Example

- Consider detergent comparison example
- Suppose $y_{4,2} = 37$ is missing
- Estimate Approach

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

  - Estimate is

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

  - Do analysis but adjust error degrees of freedom

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

SAS code: approximate approach for missing values

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

proc glm;
class stain soap;
model y = soap stain;
output out=diag r=res p=pred;
run;
```

output

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			

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

Estimate - Must adjust F by hand

$$
\begin{aligned}
F_0 &= \frac{71.95/3}{5.49/5} \\
&= 21.84
\end{aligned}
$$

$$
\text{Pvalue} = 0.0027
$$

Remark:

Approximate approach produces a biased mean square for treatment.

Exact analysis approach is preferred.

Last slide

- Read Sections: 4.1