3 BLOCK DESIGNS

- The experimenter is concerned with studying the effects of a single factor on a response of interest. However, variability from another factor that is not of interest is expected.
- The goal is to control the effects of a variable not of interest by bringing experimental units that are similar into a group called a "block". The treatments are then randomly applied to the experimental units within each block. The experimental units are assumed to be within each block.
- By using blocks to control a source of variability, the mean square error (MSE) will be reduced. A smaller MSE makes it easier to detect significant results for the factor of interest.

3.1 Randomized Complete Block Designs (RCBDs)

- Assume there are a treatments and b blocks. If we have one observation per treatment within each block, and if treatments are randomized to the experimental units within each block, then we have a **randomized complete block design (RCBD)**. Because randomization only occurs within blocks, this is an example of .
- Assume μ is the baseline mean, τ_i is the i^{th} treatment effect, β_j is the j^{th} block effect, and ϵ_{ij} is the random error of the observation. The statistical model for a RCBD is

$$y_{ij} = \qquad \text{and} \quad \epsilon_{ij} \sim IIDN(0, \sigma^2). \tag{8}$$

- μ, τ_i $(i = 1, 2, \ldots, a)$, and β_j $(j = 1, 2, \ldots, b)$ are not uniquely estimable. Constraints must be imposed. To be able to calculate estimates $\widehat{\mu}$, $\widehat{\tau}_i$, and $\widehat{\beta}_j$, we need to impose <u>two</u> constraints.
- Initially, we will assume the textbook constraints: $\sum_{i=1}^{a} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$.
- The default SAS constraints are $\tau_a = 0, \beta_b = 0$. The default R constraints are $\tau_1 = 0, \beta_1 = 0$.
- Applying these constraints, will yield least-squares estimates

$$\widehat{\mu} = \qquad \widehat{\tau}_i = \qquad \text{and} \quad \widehat{\beta}_j =$$

  where $\overline{y}_{i\cdot}$ is the mean for treatment i, and $\overline{y}_{\cdot j}$ is the mean for block j.
- Substitution of the estimates into the model yields:

$$\begin{aligned} y_{ij} &= \widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j + e_{ij} \\ &= \overline{y}_{\cdot\cdot} + \end{aligned}$$

  where $e_{ij} = \widehat{\epsilon}_{ij}$ is the residual of an observation y_{ij} from a RCBD. The value of e_{ij} is

$$e_{ij} = y_{ij} - (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}) - (\overline{y}_{\cdot j} - \overline{y}_{\cdot\cdot}) - \overline{y}_{\cdot\cdot} =$$

- The total sum of squares (SS_{total}) for the RCBD is partitioned into 3 components:

$$\begin{aligned} \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{\cdot\cdot})^2 &= \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2 + \sum_{j=1}^{b}\sum_{i=1}^{a}(\overline{y}_{\cdot j} - \overline{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \overline{y}_{i\cdot} - \overline{y}_{\cdot j} + \overline{y}_{\cdot\cdot}\right)^2 \\ &= b\sum_{i=1}^{a}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2 + a\sum_{j=1}^{b}(\overline{y}_{\cdot b} - \overline{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \overline{y}_{i\cdot} - \overline{y}_{\cdot j} + \overline{y}_{\cdot\cdot}\right)^2 \\ &= b\sum_{i=1}^{a} \quad + \quad a\sum_{j=1}^{b} \quad + \quad \sum_{i=1}^{a}\sum_{j=1}^{b} \end{aligned}$$

  OR $SS_{Total} = SS_{Trt} + SS_{Block} + SS_E$

- Alternate formulas to calculate SS_{Total}, SS_{Trt} and SS_{Block}.

$$SS_{Total} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab} \qquad SS_{Trt} = \sum_{i=1}^{a} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab} \qquad SS_{Block} = \sum_{j=1}^{b} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$$

$SS_E = SS_{Total} - SS_{Trt} - SS_{Block}$ where $\frac{y_{..}^2}{ab}$ is the **correction factor**.

3.2 Cotton Fiber Breaking Strength Experiment

An agricultural experiment considered the effects of K_2O (potash) on the breaking strength of cotton fibers. Five K_2O levels were used (36, 54, 72, 108, 144 lbs/acre). A sample of cotton was taken from each plot, and a strength measurement was taken. The experiment was arranged in 3 blocks of 5 plots each.

	K_2O lbs/acre (treatment)					
Block	36	54	72	108	144	Totals
1	7.62	8.14	7.76	7.17	7.46	$y_{.1}$=38.15
2	8.00	8.15	7.73	7.57	7.68	$y_{.2}$=39.13
3	7.93	7.87	7.74	7.80	7.21	$y_{.3}$=38.55
	$y_{1.}$	$y_{2.}$	$y_{3.}$	$y_{4.}$	$y_{5.}$	
Totals	23.55	24.16	23.23	22.54	22.35	$y_{..}$=115.83

Treatment Means	$\overline{y}_{1.} = 7.850$	$\overline{y}_{2.} = 8.053$	$\overline{y}_{3.} = 7.743$	$\overline{y}_{4.} = 7.513$	$\overline{y}_{5.} = 7.450$
Block Means	$\overline{y}_{.1} = 7.630$	$\overline{y}_{.2} = 7.826$	$\overline{y}_{.3} = 7.710$		
Grand Mean	$\overline{y} = 7.722$				

SS_{Total} = SS_{Trt} =

SS_{Block} = SS_E =

Analysis of Variance (ANOVA) Table

Source of Variation	Sum of Squares	d.f.	Mean Square	F Ratio	p-value
K_2O lbs/acre					
Blocks				___	
Error				____	
Total			____	____	

Test the hypotheses $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$ versus $H_1: \tau_i \neq 0$ *for some* i.

- The **test statistic** is $F_0 = 4.1916$.
- The **reference distribution** is $F(a-1, (a-1)(b-1)) = F(4, 8)$.
- The **critical value** is $F_{.05}(4, 8) =$.
- The **decision rule** is to reject H_0 if the test statistic F_0 is greater than $F_{.05}(4, 8)$.

  Is $F_0 > F_{.05}(4, 8)$? Is ? Is p-value $< .5$?
- The **conclusion** is to H_0 and conclude that

ANOVA RESULTS FOR STRENGTH BY TREATMENT

The GLM Procedure

Dependent Variable: strength

$SS_{TRT} + SS_{BLOCKS}$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.82956000	0.13826000	3.16	0.0677
Error	8	0.34948000	0.04368500		
Corrected Total	14	1.17904000			

SS_E

SS_{TOTAL}

R-Square	Coeff Var	Root MSE	strength Mean
0.703589	2.706677	0.209010	7.722000

SS_{TRT}

$F_0 = 4.19$ FOR TESTING

$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$ vs $H_1: \tau_i \neq 0$ FOR SOME i

P-VALUE = .0404 $\Rightarrow$ REJECT H_0

(WEAK EVIDENCE TO REJECT H_0)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TREATMENT → k2O	4	0.73244000	0.18311000	4.19	0.0404
block	2	0.09712000	0.04856000	1.11	0.3750

SS_{BLOCKS}

Parameter		Estimate		Standard Error	t Value	Pr > |t|
$\hat{y}_{5,3}$ → Intercept		7.438000000	B	0.14278072	52.09	<.0001
k2O	36	0.400000000	B	0.17065560	2.34	0.0471
k2O	54	0.603333333	B	0.17065560	3.54	0.0077
k2O	72	0.293333333	B	0.17065560	1.72	0.1240
k2O	108	0.063333333	B	0.17065560	0.37	0.7202
$\hat{\tau}_5 = 0$ → k2O	144	0.000000000	B	.	.	.
block	1	-0.080000000	B	0.13218926	-0.61	0.5618
block	2	0.116000000	B	0.13218926	0.88	0.4058
$\hat{\beta}_3 = 0$ → block	3	0.000000000	B	.	.	.

SAS DEFAULT $\hat{\tau}_a = 0$ $\hat{\beta}_b = 0$

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

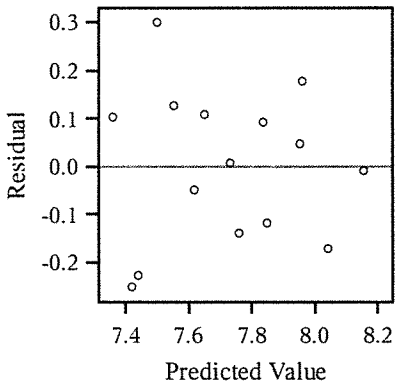
Level of block	N	strength Mean	strength Std Dev
1	5	7.63000000	0.35972211
2	5	7.82600000	0.24047869
3	5	7.71000000	0.28853076

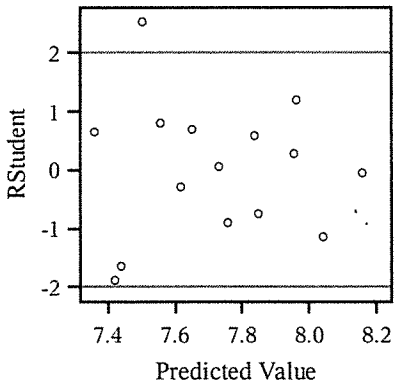
$\hat{\tau}_i$

Parameter	Estimate	Standard Error	t Value	Pr > |t|
K2O=36	0.12800000	0.10793208	1.19	0.2697
K2O=54	0.33133333	0.10793208	3.07	0.0154 (*)
K2O=72	0.02133333	0.10793208	0.20	0.8482
K2O=108	-0.20866667	0.10793208	-1.93	0.0893
K2O=144	-0.27200000	0.10793208	-2.52	0.0358

$\sum_{i=1}^{5} \hat{\tau}_i = 0$ CONSTRAINT

Fit Diagnostics for strength

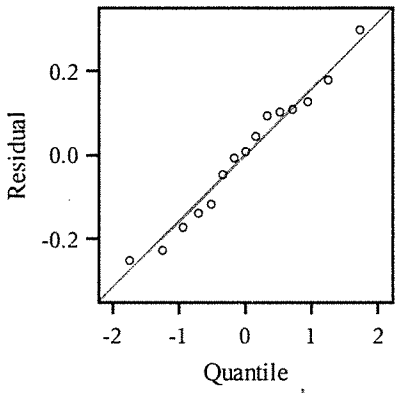


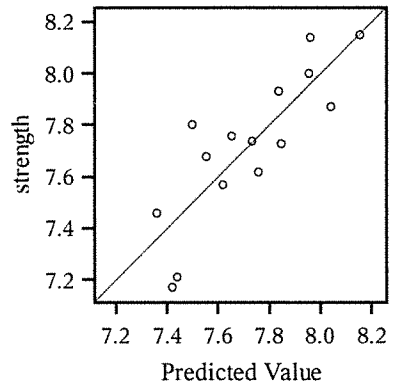


NO SERIOUS VIOLATIONS OF MODEL ASSUMPTIONS.

$$\bar{\mu}_{i\cdot} = \frac{\sum_{j=1}^{b} \mu_{ij}}{b}$$

(MEAN AVERAGED OVER BLOCKS)





YOU CANNOT RUN LEVENE'S TEST OR BROWN-FORSYTHE TEST BECAUSE THERE ARE NO REPLICATIONS ($n = 1$)

REJECT $H_0: \bar{\mu}_{i\cdot} = \bar{\mu}_{i'\cdot}$ IF $|\bar{y}_{i\cdot} - \bar{y}_{i'\cdot}| > .5896$

OR

REJECT $H_0: \tau_i = \tau_{i'}$ IF $|\bar{y}_{i\cdot} - \bar{y}_{i'\cdot}| > .5896$

TUKEY'S TEST

Alpha	0.05
Error Degrees of Freedom	8
Error Mean Square	0.043685
Critical Value of Studentized Range	4.88569
Minimum Significant Difference	0.5896

Means with the same letter are not significantly different.

Tukey Grouping		Mean	N	k2O
	A	8.0533	3	54
	A			
B	A	7.8500	3	36
B	A			
B	A	7.7433	3	72
B	A			
B	A	7.5133	3	108
B				
B		7.4500	3	144

TRT

2
1
3
4
5

Comparisons significant at the 0.05 level are indicated by *.**

k2O Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
54 - 36	0.2033	-0.3862	0.7929	
54 - 72	0.3100	-0.2796	0.8996	
54 - 108	0.5400	-0.0496	1.1296	
54 - 144	0.6033	0.0138	1.1929	***
36 - 54	-0.2033	-0.7929	0.3862	
36 - 72	0.1067	-0.4829	0.6962	
36 - 108	0.3367	-0.2529	0.9262	
36 - 144	0.4000	-0.1896	0.9896	
72 - 54	-0.3100	-0.8996	0.2796	
72 - 36	-0.1067	-0.6962	0.4829	
72 - 108	0.2300	-0.3596	0.8196	
72 - 144	0.2933	-0.2962	0.8829	
108 - 54	-0.5400	-1.1296	0.0496	
108 - 36	-0.3367	-0.9262	0.2529	
108 - 72	-0.2300	-0.8196	0.3596	
108 - 144	0.0633	-0.5262	0.6529	
144 - 54	-0.6033	-1.1929	-0.0138	***
144 - 36	-0.4000	-0.9896	0.1896	
144 - 72	-0.2933	-0.8829	0.2962	
144 - 108	-0.0633	-0.6529	0.5262	

0 IS NOT IN THE CONFIDENCE INTERVAL

REJECT ONLY $H_0: \bar{\mu}_{2\cdot} = \bar{\mu}_{5\cdot}$

$k_2O = 54$ $k_2O = 144$

OR

REJECT $H_0: \tau_2 = \tau_5$

TRT: 54, 144 NOT CONNECTED

$\Rightarrow$ REJECT $H_0: \bar{\mu}_{2\cdot} = \bar{\mu}_{5\cdot}$ FAIL TO REJECT ALL OTHER COMPARISONS

3.3 SAS Code for Cotton Fiber Breaking Strength RCBD

```
DM 'LOG; CLEAR; OUT; CLEAR;';
OPTIONS NODATE NONUMBER LS=76;
ODS GRAPHICS ON;
ODS PRINTER PDF file='C:\COURSES\ST541\RCBD.PDF';

****************************************;
*** A RANDOMIZED COMPLETE BLOCK DESIGN ***;
****************************************;

DATA in; INPUT k2O block strength @@; CARDS;
 36 1 7.62    36 2 8.00    36 3 7.93
 54 1 8.14    54 2 8.15    54 3 7.87
 72 1 7.76    72 2 7.73    72 3 7.74
108 1 7.17   108 2 7.57   108 3 7.80
144 1 7.46   144 2 7.68   144 3 7.21

PROC GLM DATA=in PLOTS = (ALL);
   CLASS k2O block;
   MODEL strength = k2O block / SS3 SOLUTION;
   MEANS block;
   MEANS k2O / TUKEY CLDIFF LINES;
   ESTIMATE 'K2O=36'  K2O  4 -1 -1 -1 -1 / DIVISOR=5;
   ESTIMATE 'K2O=54'  K2O -1  4 -1 -1 -1 / DIVISOR=5;
   ESTIMATE 'K2O=72'  K2O -1 -1  4 -1 -1 / DIVISOR=5;
   ESTIMATE 'K2O=108' K2O -1 -1 -1  4 -1 / DIVISOR=5;
   ESTIMATE 'K2O=144' K2O -1 -1 -1 -1  4 / DIVISOR=5;
TITLE 'ANOVA RESULTS FOR STRENGTH BY TREATMENT';
RUN;
```

3.4 Restrictions on Randomization

- Two common reasons for blocking:

  1. The experimenter has multiple sets of experimental units that are homogeneous within sets but are heterogeneous across sets. This typically occurs when there is not a sufficient number of homogeneous experimental units available to run a CRD leading the experimenter to form groups of units that are as homogeneous as possible.
  2. The experimenter has time constraints that do not allow a CRD to be run within a continuous period of time that would ensure uniformity of experimental conditions. Under these circumstances, blocks take the form of a time unit (such as a day).

- For a RCBD, there is one . Randomization is restricted to randomly assigning the a treatments to the a experimental units within each block.

- In their *Design of Experiments* text, Anderson and McLean (A&M) introduce a random component called a into the traditional RCBD model to present a more realistic picture of the experimental situation. This approach will be useful later when we have multiple restrictions on randomizations (e.g., split-plot designs).

- Essentially, we're saying there must be a different error structure between a completely randomized design and a design that has a restriction on randomization. And, because there is a different error structure, there must be differences in the model and the analysis.

- Thus, A&M suggest that the traditional model equation

$$y_{ij} \;=\; \mu \;+\; \tau_i \;+\; \beta_j \;+\; \epsilon_{ij} \tag{9}$$

should include a term indicating where the restriction on randomization occurred. That is:

$$y_{ijk} \;=\; \mu \;+\; \tau_i \;+\; \beta_j \;+\; \delta_j \;+\; \epsilon_{ij} \tag{10}$$

where μ, τ_i, and β_j are the same in (10) as in (9), y_{ij} is the response from the i^{th} treatment in block j for the k^{th} randomization, and δ_j is the restriction error associated with the j^{th} block.

- We also assume $\delta_j \sim N(0, \sigma^2_\delta)$, and each δ_j is completely confounded with the j^{th} block effect.

Comparison of CRD and RCBD ANOVA Tables

CRD with 2 model effects

Source	term	d.f.	EMS
Blocks	β_j	$b-1$	$\sigma^2 + a\phi(\beta)$
Treatments	τ_i	$a-1$	$\sigma^2 + b\left(\sum_{i=1}^{a} \tau_i^2\right)/(a-1)$
Error	ϵ_{ij}	$(a-1)(b-1)$	σ^2

RCBD from A&M

Source	term	d.f.	EMS
Blocks	β_j	$b-1$	$\sigma^2 + a\sigma^2_\delta + a\phi(\beta)$
Restriction Error	$\delta_{j(k)}$	0	$\sigma^2 + a\sigma^2_\delta$
Treatments	τ_i	$a-1$	$\sigma^2 + b\left(\sum_{i=1}^{a} \tau_i^2\right)/(a-1)$
Error	ϵ_{ijk}	$(a-1)(b-1)$	σ^2

where $\phi(\beta)$ is a function of $\beta_1, \ldots, \beta_b$ if blocks are fixed or $\phi(\beta) = \sigma^2_\beta$ if blocks are random.

- In both the fixed and random block cases, the ANOVA F-tests associated with treatment effects are identical. You use $F_0 = MS_{trt}/MS_E$ to test

$$\mathrm{H}_0 : \tau_1 = \cdots = \tau_a = 0 \qquad \text{against} \qquad \mathrm{H}_1 : \text{not all of the } \tau_i\text{s are equal} \tag{11}$$

- The EMS for the RCBD indicates that the correct denominator EMS for testing for a significant block effect (either fixed or random) is the EMS for the restriction error. The problem is that this is not estimable from the data.

- Under these circumstances, the test of the hypothesis involving the combination of the block effects and the restriction error in (12) would be appropriate to test for a 'general' blocking effect.

- The statistic $F = MS_{blocks}/MS_E$ is actually a test of

$$\mathrm{H}_0 : \qquad\qquad \text{against} \qquad \mathrm{H}_1 : \tag{12}$$

Note that even if $\beta_1 = \beta_2 = \cdots = \beta_b = 0$ (fixed) or $\sigma^2_\beta = 0$ (random) is true, we still have the restriction error in the EMS which prevents it from matching the error EMS $= \sigma^2$.

- Because of the restriction on randomization, A&M claim that . That is, there is no test for H_0 : if blocks are random and no test for H_0 : if blocks are fixed.

- Fortunately this is not a problem because most of the time the experimenter is only interested in whether or not blocking had been effective in reducing the MS_E for improved testing of the effects of the treatment of interest.

3.5 Example of an Analysis With and Without Blocks

Three different disinfecting solutions are being compared to study their effectiveness in stopping the growth of bacteria in milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design with days as blocks. Observations are taken for four days. The inside of the milk containers are covered with a certain amount of bacteria. The response is the percentage of bacteria remaining after rinsing the container with a disinfecting solution.

	Day			
Solution	1	2	3	4
1	13	22	18	39
2	16	24	17	44
3	5	4	1	22

- The data were analyzed assuming two different models. The first model does not include blocks. The second model includes blocks.

SSE FOR MODEL WITHOUT BLOCKS = 1158.75 = $SS_{SOLUTION}$ (TRT) + SS_{DAYS} (BLOCKS) FOR MODEL WITH BLOCKS

RCBD Without Days as Blocks

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	703.500000	351.750000	2.73	0.1182
Error	9	1158.750000	128.750000		
Corrected Total	11	1862.250000			

MSE

R^2

R-Square	Coeff Var	Root MSE	growth Mean
0.377769	60.51630	11.34681	18.75000

TREATMENTS ONLY (NO BLOCKS)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
solution	2	703.5000000	351.7500000	2.73	0.1182

FAIL TO REJECT $H_0: \tau_1 = \tau_2 = \tau_3$ (NO EVIDENCE TO REJECT H_0)

RCBD With Days as Blocks

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1810.416667	362.083333	41.91	0.0001
Error	6	51.833333	8.638889		
Corrected Total	11	1862.250000			

MSE

R^2

R-Square	Coeff Var	Root MSE	growth Mean
0.972166	15.67573	2.939199	18.75000

TREATMENTS AND BLOCKS

Source	DF	Type III SS	Mean Square	F Value	Pr > F
solution	2	703.500000	351.750000	40.72	0.0003
day	3	1106.916667	368.972222	42.71	0.0002

REJECT $H_0: \tau_1 = \tau_2 = \tau_3$ (VERY STRONG EVIDENCE TO REJECT H_0)

Here are important results:

	Without blocks	With blocks
R^2		
MS_E		
p-value		

- Note that we would fail to reject H_0 if blocks were not in the model because there is large variability across blocks ($MS_{day} = 368.97$).
- If the $SS_{day} = 1106.92$ and $df_{day} = 3$ is pooled with the the $SS_E = 41.83$ and $df_E = 6$ in the model with days (blocks), then it forms the $SS_E = 1158.75$ and $df_E = 9$ for the model without days (blocks).

3.6 Type I vs Type III Analyses

- Without the **/ss3** option in the MODEL statement, *SAS* will contain two ANOVA tables: ANOVA for Type I sum of squares and ANOVA for Type III sum of squares.
- If there are no missing observations, the Type I and Type III analyses are identical.
- If there are missing observations, the Type I and Type III analyses are different. To see how they differ we will first look at the Type I analysis.

3.6.1 Type I Analysis

- The Type I analysis is based on sequentially fitting the data to the model one factor at a time. It is often referred to as the .
- For the RCBD there are two possibilities that I will refer to as
  - Version 1 (V1) when fitting treatments before blocks.
  - Version 2 (V2) when fitting blocks before treatments.
- Let RSS_i be the error sum of squares (SS_E) after fitting the model in the i^{th} step.
- The steps for determining the ANOVA SS for V1 are:
  1. Fit $y_{ij} = \mu + \epsilon_{ij}$ and obtain $RSS_1 = SS_{total}$.
  2. Fit $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ and obtain $RSS_2 = SS_E$ for the model with treatments only.
  3. Fit $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ and obtain $RSS_3 = SS_E$ for the model with treatments and blocks.
- The steps for determining the ANOVA SS for V2 are:
  1. Fit $y_{ij} = \mu + \epsilon_{ij}$ and obtain $RSS_1 = SS_{total}$.
  2′. Fit $y_{ij} = \mu + \beta_j + \epsilon_{ij}$ and obtain $RSS_2^* = SS_E$ for the model with blocks only.
  3. Fit $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ and obtain $RSS_3 = SS_E$ for the model with blocks and treatments..
- The ANOVA sum of squares for V1 and V2 are summarized in the following table:

Step	V1 Source	Fit	*df*	Type I SS for V1	Fitted Model
1	Total	μ	$N-1$	RSS_1	$y_{ij} = \widehat{\mu} + e_{ij}$
2	Treatment	τ_i	$a-1$	$R(\tau\|\mu) \quad = RSS_1 - RSS_2$	$y_{ij} = \widehat{\mu} + \widehat{\tau}_i + e_{ij}$
3	Blocks	β_j	$b-1$	$R(\beta\|\tau,\mu) = RSS_2 - RSS_3$	$y_{ij} = \widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j + e_{ij}$
3	Error	ϵ_{ij}	$N-a-b+1$	RSS_3	

Step	V2 Source	Fit	*df*	Type I SS for V2	Fitted Model
1	Total	μ	$N-1$	RSS_1	$y_{ij} = \widehat{\mu} + e_{ij}$
$2'$	Blocks	β_j	$b-1$	$R(\beta\|\mu) = RSS_1 - RSS_2^*$	$y_{ij} = \widehat{\mu} + \widehat{\beta}_j + e_{ij}$
3	Treatment	τ_i	$a-1$	$R(\tau\|\beta,\mu) = RSS_2^* - RSS_3^*$	$y_{ij} = \widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j + e_{ij}$
3	Error	ϵ_{ij}	$N-a-b+1$	RSS_3	

- In V1, the quantity $R(\tau|\mu)$ is called the **due to τ adjusted for μ** and $R(\beta|\tau,\mu)$ is called the **for β adjusted for τ and μ.**
- In V2, the quantity $R(\beta|\mu)$ is called the due to β adjusted for μ and $R(\tau|\beta,\mu)$ is called the for τ adjusted for β and μ.

3.6.2 Type III Analysis

- The Type III analysis is referred to as the or the **Yates weighted squares of means** analysis.
- For a RCBD, the Type III SS_{trt} and SS_{blocks} are computed using the following procedure:
  1. Fit the model with treatments only: $y_{ij} = \mu + \tau_i + \epsilon_{ij}$. Then $RSS_2 = SS_E$ for this model.
  2. Fit the model with blocks only: $y_{ij} = \mu + \beta_j + \epsilon_{ij}$. Then $RSS_2^* = SS_E$ for this model.
  3. Fit the model $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$. Then $RSS_3 = SS_E$ and $RSS_1 = SS_{total}$ for the model with both treatments and blocks.

Step	Source	Fit	*df*	Type III SS	Fitted Model
1	Total	μ	$N-1$	RSS_1	$y_{ij} = \widehat{\mu} + e_{ij}$
2	Treatment	τ_i	$a-1$	$R(\tau\|\beta,\mu) \;=\; RSS_2^* - RSS_3$	$y_{ij} = \widehat{\mu} + \widehat{\tau}_j + e_{ij}$
3	Blocks	β_j	$b-1$	$R(\beta\|\tau,\mu) \;=\; RSS_2 - RSS_3$	$y_{ij} = \widehat{\mu} + \widehat{\tau}_i + e_{ij}$
1	Error	ϵ_{ij}	$N-a-b+1$	RSS_3	$y_{ij} = \widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j + e_{ij}$

- If any y_{ij} values are missing, then $SS_{trt} + SS_{blocks} + SS_E \neq SS_{total}$ for a Type III analysis.

3.6.3 RCBD Analysis with a Missing Observation

See the example in Section 3.5 for the description of the experiment. Suppose y_{23} was missing from the RCBD. The RCBD data table is:

	Day			
Solution	1	2	3	4
1	13	22	18	39
2	16	24	.	44
3	5	4	1	22

- Let us examine the Type I and Type III sums of squares. The next page contains the SAS output.
- The top of the next page contains the Type I (V1) sum of squares and the bottom of the page contains the Type I (V2) sum of squares. Note the difference in sums of squares, mean squares, F-statistics, and p-values for the Type I analyses.

- The reason for the difference between the V1 and V2 Type I sum of squares is that a Type I analysis is sequential so the order in which terms enter the model is important.
- The Type III analysis is the same for both analyses Type III sums of squares are not calculated sequentially. That is. the order in which terms enter the model is not important.
- The following page contains the two analyses with only one effect in each model. I included these analyses so you can see how RSS_2 and RSS_2^* are calculated.

ANOVA RESULTS: (MODEL WITH SOLUTION THEN DAY)

DIFFERENT ANALYSES FOR TYPE I SS ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1811.575758	362.315152	38.27	0.0005
Error	5	47.333333	9.466667		
Corrected Total	10	1858.909091			

$RSS_3 = 47.\overline{3}$

$RSS_1 = 1858.\overline{90}$ (SS_{TOTAL})

R-Square	Coeff Var	Root MSE	growth Mean
0.974537	16.27151	3.076795	18.90909

(V1)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
solution	2	790.909091	395.454545	41.77	0.0008
day	3	1020.666667	340.222222	35.94	0.0008

$R(\tau|\mu) = 790.\overline{90}$

$R(\beta|\tau,\mu) = RSS_2 - RSS_3 = 1020.\overline{6}$

(*)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
solution	2	670.500000	335.250000	35.41	0.0011
day	3	1020.666667	340.222222	35.94	0.0008

$R(\tau|\beta,\mu) = RSS_2^* - RSS_3 = 670.5$

$R(\beta|\tau,\mu) = 1020.\overline{6}$

ANOVA RESULTS: (MODEL WITH DAY THEN SOLUTION)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1811.575758	362.315152	38.27	0.0005
Error	5	47.333333	9.466667		
Corrected Total	10	1858.909091			

$RSS_3 = 47.\overline{3}$

$RSS_1 = 1858.\overline{90}$

R-Square	Coeff Var	Root MSE	growth Mean
0.974537	16.27151	3.076795	18.90909

(V2)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
day	3	1141.075758	380.358586	40.18	0.0006
solution	2	670.500000	335.250000	35.41	0.0011

$R(\beta|\mu) = 1141.\overline{075}$

$R(\tau|\beta,\mu) = 670.5$

RECOMMENDED

Source	DF	Type III SS	Mean Square	F Value	Pr > F
day	3	1020.666667	340.222222	35.94	0.0008
solution	2	670.500000	335.250000	35.41	0.0011

SAME ANALYSIS AS (*). ORDER DOES NOT MATTER FOR TYPE III SS.

So where did RSS_2 and RSS_2^* come from?

RSS_2 is the SSE for the model with only treatments and no blocks.

ANOVA RESULTS FOR THE MODEL WITH SOLUTION (TREATMENTS) ONLY

VERIFICATION ANOVAS FOR V1, V2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	790.909091	395.454545	2.96	0.1090
Error	8	1068.000000	133.500000		
Corrected Total	10	1858.909091			

$RSS_2 = 1068.0$
$RSS_1 = 1858.\overline{90}$
FOR (V1)

R-Square	Coeff Var	Root MSE	growth Mean
0.425469	61.10405	11.55422	18.90909

FOR MODEL 2

Source	DF	Type I SS	Mean Square	F Value	Pr > F
solution	2	790.9090909	395.4545455	2.96	0.1090

$R(\tau|\mu) = RSS_1 - RSS_2 = 790.\overline{90}$ (STEP 2)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
solution	2	790.9090909	395.4545455	2.96	0.1090

(SEE PAGE 75)

RSS_2^* is the SSE for the model with only blocks and no treatments.

ANOVA RESULTS FOR THE MODEL WITH DAYS (BLOCKS) ONLY

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1141.075758	380.358586	3.71	0.0696
Error	7	717.833333	102.547619		
Corrected Total	10	1858.909091			

$RSS_2^* = 717.8\overline{3}$
$RSS_1 = 1858.\overline{90}$
FOR (V_2)

R-Square	Coeff Var	Root MSE	growth Mean
0.613842	53.55403	10.12658	18.90909

FOR MODEL 2'

Source	DF	Type I SS	Mean Square	F Value	Pr > F
day	3	1141.075758	380.358586	3.71	0.0696

$R(\beta|\mu) = RSS_1 - RSS_2^* = 1141.\overline{075}$

Source	DF	Type III SS	Mean Square	F Value	Pr > F
day	3	1141.075758	380.358586	3.71	0.0696

All of these calculations are done automatically in the RCBD analyses for the two models on the previous page.

Type I SS (V1) Summary

$RSS_1 = 1858.91$	$R(\mu) = RSS_1$	$= 1858.91$
$RSS_2 = 1068.00$	$R(\tau\|\mu) = RSS_1 - RSS_2$	$= 790.91$
$RSS_3 = 47.33$	$R(\beta\|\tau,\mu) = RSS_2 - RSS_3$	$= 1020.67$

Type I SS (V2) Summary

$RSS_1 = 1858.91$	$R(\mu) = RSS_1$	$= 1858.91$
$RSS_2^* = 717.83$	$R(\beta\|\mu) = RSS_1 - RSS_2^*$	$= 1141.08$
$RSS_3 = 47.33$	$R(\tau\|\beta,\mu) = RSS_2^* - RSS_3$	$= 670.50$

Type III SS Summary

$RSS_1 = 1858.91$	$R(\mu) = RSS_1$	$= 1858.91$
$RSS_3 = 47.33$		
$RSS_2^* = 717.83$	$R(\beta\|\tau,\mu) = RSS_2^* - RSS_1$	$= 1020.67$
$RSS_2 = 1068.00$	$R(\tau\|\beta,\mu) = RSS_2 - RSS_1$	$= 670.50$

```
****************************************;
*** RCBD WITH A MISSING OBSERVATION ***;
****************************************;
DATA IN;
 DO solution = 1 TO 3;
 DO day = 1 TO 4;
    INPUT growth @@; OUTPUT;
 END; END;
CARDS;
13 22 18 39     16 24 . 44     5 4 1 22
;
************************************************;
***  RUN AN ANOVA WITH SOLUTION APPEARING FIRST ***;
************************************************;
PROC GLM DATA=IN;
     CLASS solution day;
     MODEL growth = solution day;
TITLE 'ANOVA RESULTS (SOLUTION THEN DAY)';

**********************************************;
***  RUN AN ANOVA WITH DAY APPEARING FIRST ***;
**********************************************;
PROC GLM DATA=IN;
     CLASS day solution;
     MODEL growth = day solution;
TITLE 'ANOVA RESULTS (DAY THEN SOLUTION)';

**************************************;
***  RUN AN ANOVA WITH SOLUTION ONLY ***;
**************************************;
PROC GLM DATA=IN;
     CLASS solution;
     MODEL growth = solution;
TITLE 'ANOVA RESULTS (SOLUTION ONLY)';

**********************************;
***  RUN AN ANOVA WITH DAY ONLY ***;
**********************************;
PROC GLM DATA=IN;
     CLASS day;
     MODEL growth = day;
TITLE 'ANOVA RESULTS (DAY ONLY)';

RUN;
```

- **Warning: The default in R is Type I SS. So be careful to enter blocks before treatments in the model.**

R code for RCBD with missing value

```
# ANOVA for RCBD with missing observation

strength <- c(13,22,18,39,16,24,NA,44,5,4,1,22)
solution <- c(1,1,1,1,2,2,2,2,3,3,3,3)
day      <- c(1,2,3,4,1,2,3,4,1,2,3,4)

f1 <- aov(strength~factor(day)+factor(solution))
summary (f1)
f2 <-  lm(strength~factor(day)+factor(solution))
summary(f2)
```

R output for RCBD with missing value

```
> summary (f1)
                 Df Sum Sq Mean Sq F value   Pr(>F)
factor(day)       3 1141.1   380.4   40.18 0.000636 ***
factor(solution)  2  670.5   335.2   35.41 0.001116 **
Residuals         5   47.3     9.5
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
1 observation deleted due to missingness


> summary(f2)

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         15.333      2.206   6.952 0.000946 ***
factor(day)2         5.333      2.512   2.123 0.087176 .
factor(day)3         1.667      2.901   0.575 0.590479
factor(day)4        23.667      2.512   9.421 0.000227 ***
factor(solution)2    3.000      2.432   1.233 0.272264
factor(solution)3  -15.000      2.176  -6.895 0.000983 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.077 on 5 degrees of freedom
  (1 observation deleted due to missingness)

Multiple R-squared: 0.9745,     Adjusted R-squared: 0.9491

F-statistic: 38.27 on 5 and 5 DF,  p-value: 0.0005468
```

3.6.4 Type I vs Type III Hypotheses

- Because of differences between Type I and Type III SS, there will be differences in the hypotheses associated with the F-tests (assuming the restriction on randomization is ignored).
- Let $\mu_{ij} = \mu + \tau_i + \beta_j$ be the i^{th} treatment, j^{th} block mean.

Hypotheses for Type III and Type I (V2) Sum of Squares

H_0 :

H_1 : for some $i \neq i^*$ and $\overline{\mu}_{i\cdot} = \left(\sum_{j=1}^{b} \mu_{ij}\right)/b.$

Hypotheses for Type I (V1) Sum of Squares

$$H_0 : \frac{1}{n_{1\cdot}}\sum_{j=1}^{b} n_{1j}\mu_{1j} = \frac{1}{n_{2\cdot}}\sum_{j=1}^{b} n_{2j}\mu_{2j} = \cdots = \frac{1}{n_{a\cdot}}\sum_{j=1}^{b} n_{aj}\mu_{aj}$$

$$H_1 : \frac{1}{n_{i\cdot}}\sum_{j=1}^{b} n_{ij}\mu_{ij} \neq \frac{1}{n_{i^*\cdot}}\sum_{j=1}^{b} n_{i^*j}\mu_{i^*j} \quad \text{for some } i \neq i^*.$$

where $n_{i\cdot}$ = the number of nonmissing y_{ij} values for the i^{th} treatment, and $n_{ij} = 1$ if y_{ij} is not missing and $n_{ij} = 0$ if y_{ij} is missing.

- The Type III hypotheses are comparing the treatment means average across the blocks (and are the ones I want to test.) Therefore I recommend using the p-values from a Type III analysis.
- If there are no missing y_{ij} values, the Type I and Type III hypotheses are the same.

3.7 RCBD Normal Equations

- For model $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$, the error is $\epsilon_{ij} = y_{ij} - \mu - \tau_i - \beta_j$
- Substituting in estimates produces the residual $\widehat{\epsilon}_{ij} = e_{ij} =$
- Goal: Find $\widehat{\mu}$, $\widehat{\tau}_i$, and $\widehat{\beta}_j$ that minimize L:

$$L = \sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{\epsilon}_{ij}^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}$$

- Solution: Solve the normal equations

$$\frac{\partial L}{\partial \widehat{\mu}} = -2\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j) = 0$$

$$\frac{\partial L}{\partial \widehat{\tau}_i} = -2\sum_{j=1}^{b}(y_{ij} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j) = 0 \quad \text{for } i = 1, 2, \ldots, a$$

$$\frac{\partial L}{\partial \widehat{\beta}_j} = -2\sum_{i=1}^{a}(y_{ij} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j) = 0 \quad \text{for } j = 1, 2, \ldots, b$$

- After distributing the sum and then simplifying, we get:

$$\text{(i)} \quad y_{\cdot\cdot} \;=\; ab\widehat{\mu} \;+\; b\sum_{i=1}^{a}\widehat{\tau}_i + \;a\sum_{j=1}^{b}\widehat{\beta}_j$$

$$\text{(ii)} \quad y_{i\cdot} \;=\; b\widehat{\mu} \;+\; b\,\widehat{\tau}_i \;+\; \sum_{j=1}^{b}\widehat{\beta}_j \qquad \text{for } i = 1, 2, \ldots, a$$

$$\text{(iii)} \quad y_{\cdot j} \;=\; a\widehat{\mu} \;+\; \sum_{i=1}^{a}\widehat{\tau}_i \;+\; a\widehat{\beta}_j \qquad \text{for } j = 1, 2, \ldots, b$$

- For (i), (ii), and (iii), there is a total of $1 + a + b$ equations. If you sum the a equations in (ii), you get (i). If you sum the b equations in (iii), you also get (i). Thus, the rank is $a + b - 1$ which implies that μ and each τ_i and β_j are not uniquely estimable. To get estimates of μ and each τ_i and β_j, we must impose 2 constraints. We will use $\sum_{i=1}^{a}\tau_i = 0$ and $\sum_{j=1}^{b}\beta_j = 0$ (textbook constraints).

- Substitution of these constraints into (i), (ii), and (iii) yields

$$(1) \quad ab\widehat{\mu} \;=\; y_{\cdot\cdot} \qquad (2) \quad b\widehat{\mu} + b\widehat{\tau}_i \;=\; y_{i\cdot} \qquad (3) \quad a\widehat{\mu} + a\widehat{\beta}_j \;=\; y_{\cdot j}$$

- Then, from (1), we have

$$\widehat{\mu} \;=\; \frac{y_{\cdot\cdot}}{ab} \;=$$

- Substitution of $\widehat{\mu} = \overline{y}_{\cdot\cdot}$ in (2) yields:

$$b\overline{y}_{\cdot\cdot} + b\widehat{\tau}_i = y_{i\cdot} \quad\longrightarrow\quad \overline{y}_{\cdot\cdot} + \widehat{\tau}_i = \overline{y}_{i\cdot} \quad\longrightarrow\quad \widehat{\tau}_i =$$

- Substitution of $\widehat{\mu} = \overline{y}_{\cdot\cdot}$ in (3) yields:

$$a\overline{y}_{\cdot\cdot} + a\widehat{\beta}_j = y_{\cdot j} \quad\longrightarrow\quad \overline{y}_{\cdot\cdot} + \widehat{\beta}_j = \overline{y}_{\cdot j} \quad\longrightarrow\quad \widehat{\beta}_j =$$

3.8 Matrix Forms for the RCBD

Example: The goal is to determine whether or not four different tips produce different readings on a hardness testing machine. The machine operates by pressing the tip into a metal test coupon, and from the depth of the resulting depression, the hardness of the coupon can be determined. The experimenter decides to obtain four observations for each tip. Four randomly selected coupons (blocks) were used and each tip (treatment) was tested on each coupon. The data represent deviations from a desired depth in 0.1 mm units:

	Type of Tip **(Treatment)**			
Coupon (**Block**)	1	2	3	4
1	−2	−1	−3	2
2	−1	−2	−1	1
3	1	3	0	5
4	5	4	2	7

- Model: $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ for $i = 1, 2, 3, 4$ and $j = 1, 2, 3, 4$

$$\epsilon_{ij} \sim N(0, \sigma^2) \qquad \beta_j \sim N(0, \sigma_\beta^2)$$

- Assume (i) $\sum_{i=1}^4 \tau_i = 0$ and (ii) $\sum_{j=1}^4 \beta_j = 0$. If we estimate $[\ \mu, \tau_1, \tau_2, \tau_3, \beta_1, \beta_2, \beta_3\ ]$, we can then estimate $\tau_4 = -\tau_1 - \tau_2 - \tau_3$ from (i) and $\beta_4 = -\beta_1 - \beta_2 - \beta_3$ from (ii).

$$X = \left[\begin{array}{c|ccc|ccc} \mu & \tau_1 & \tau_2 & \tau_3 & \beta_1 & \beta_2 & \beta_3 \\ \hline 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 & -1 & -1 \\ \hline 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 & -1 \\ \hline 1 & -1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{array}\right] \qquad \theta = \left[\begin{array}{c} \mu \\ \hline \tau_1 \\ \tau_2 \\ \tau_3 \\ \hline \beta_1 \\ \beta_2 \\ \beta_3 \end{array}\right] \qquad \epsilon = \left[\begin{array}{c} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \hline \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \hline \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \\ \hline \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \\ \epsilon_{44} \\ \hline 0 \\ 0 \end{array}\right] \qquad y = \left[\begin{array}{c} \text{-2} \\ \text{-1} \\ 1 \\ 5 \\ \hline \text{-1} \\ \text{-2} \\ 3 \\ 4 \\ \hline \text{-3} \\ \text{-1} \\ 0 \\ 2 \\ \hline 2 \\ 1 \\ 5 \\ 7 \end{array}\right]$$

$$X'X = \begin{bmatrix} 16 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 4 & 4 & 0 & 0 & 0 \\ 0 & 4 & 8 & 4 & 0 & 0 & 0 \\ 0 & 4 & 4 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 4 & 4 \\ 0 & 0 & 0 & 0 & 4 & 8 & 4 \\ 0 & 0 & 0 & 0 & 4 & 4 & 8 \end{bmatrix} \qquad (X'X)^{-1} = \frac{1}{16} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & \text{-1} & \text{-1} & 0 & 0 & 0 \\ 0 & \text{-1} & 3 & \text{-1} & 0 & 0 & 0 \\ 0 & \text{-1} & \text{-1} & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & \text{-1} & \text{-1} \\ 0 & 0 & 0 & 0 & \text{-1} & 3 & \text{-1} \\ 0 & 0 & 0 & 0 & \text{-1} & \text{-1} & 3 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 20 \\ \text{-12} \\ \text{-11} \\ \text{-17} \\ \text{-22} \\ \text{-21} \\ \text{-9} \end{bmatrix} \qquad (X'X)^{-1}X'y = \frac{1}{16} \begin{bmatrix} 20 \\ \text{-36 + 11 + 17} \\ \text{12 - 33 + 17} \\ \text{12 + 11 -51} \\ \text{-66 + 21 + 9} \\ \text{22 - 63 + 9} \\ \text{22 + 21 -27} \end{bmatrix}$$

$$= \frac{1}{16} \begin{bmatrix} 20 \\ \text{-8} \\ \text{-4} \\ \text{-28} \\ \text{-36} \\ \text{-32} \\ 16 \end{bmatrix} = \begin{bmatrix} 5/4 \\ \text{-2/4} \\ \text{-1/4} \\ \text{-7/4} \\ \text{-9/4} \\ \text{-8/4} \\ 4/4 \end{bmatrix} = \begin{bmatrix} \overline{y}_{..} \\ \overline{y}_{1\cdot} - \overline{y}_{..} \\ \overline{y}_{2\cdot} - \overline{y}_{..} \\ \overline{y}_{3\cdot} - \overline{y}_{..} \\ \overline{y}_{\cdot 1} - \overline{y}_{..} \\ \overline{y}_{\cdot 2} - \overline{y}_{..} \\ \overline{y}_{\cdot 3} - \overline{y}_{..} \end{bmatrix} = \begin{bmatrix} \widehat{\mu} \\ \widehat{\tau}_1 \\ \widehat{\tau}_2 \\ \widehat{\tau}_3 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix}$$

Thus, $\widehat{\tau}_4 = -\widehat{\tau}_1 - \widehat{\tau}_2 - \widehat{\tau}_3 =$ and $\widehat{\beta}_4 = -\widehat{\beta}_1 - \widehat{\beta}_2 - \widehat{\beta}_3 =$

Alternate Approach: Keeping $a + b + 1$ Columns

- Because we have 2 constraints, we have to add 2 rows to the X matrix.

$$X = \begin{array}{c} \begin{matrix} \mu & \tau_1 & \tau_2 & \tau_3 & \tau_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{matrix} \\ \left[\begin{array}{c|cccc|cccc} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right] \end{array} \quad \theta = \begin{bmatrix} \mu \\ \hline \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \hline \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \hline \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \hline \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \\ \hline \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \\ \epsilon_{44} \\ \hline 0 \\ 0 \end{bmatrix} \quad y = \begin{bmatrix} -2 \\ -1 \\ 1 \\ 5 \\ \hline -1 \\ -2 \\ 3 \\ 4 \\ \hline -3 \\ -1 \\ 0 \\ 2 \\ \hline 2 \\ 1 \\ 5 \\ 7 \\ \hline 0 \\ 0 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 16 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ 4 & 5 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 5 & 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 1 & 5 & 1 & 1 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 & 5 & 1 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 & 1 & 5 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 & 1 & 1 & 5 & 1 & 1 \\ 4 & 1 & 1 & 1 & 1 & 1 & 1 & 5 & 1 \\ 4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 5 \end{bmatrix} \qquad (X'X)^{-1} = \frac{1}{16} \begin{bmatrix} 3 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 20 \\ 3 \\ 4 \\ -2 \\ 15 \\ -4 \\ -3 \\ 9 \\ 18 \end{bmatrix} \qquad (X'X)^{-1}X'y = \begin{bmatrix} 5/4 \\ -2/4 \\ -1/4 \\ -7/4 \\ 10/4 \\ -9/4 \\ -8/4 \\ 4/4 \\ 13/4 \end{bmatrix} = \begin{bmatrix} \widehat{\mu} \\ \widehat{\tau}_1 \\ \widehat{\tau}_2 \\ \widehat{\tau}_3 \\ \widehat{\tau}_4 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \\ \widehat{\beta}_4 \end{bmatrix} = \widehat{\theta}$$

3.9 SAS Power Analysis for RCBD

SAS Code for determining power for a 4 treatment, 10 block RCBD for three α levels (.01,.05.10) and various σ values

```
ODS LISTING;
DM 'LOG;CLEAR;OUT;CLEAR;';
OPTIONS LS=72 PS=54 NONUMBER NODATE;

DATA rcbd;
b=10;                              ** enter number of blocks;
DO Trt = 'A', 'B', 'C', 'D';
DO Blocks = 1 to b;
   IF Trt = 'A' THEN tau = 3;      ** enter pattern for;
   IF Trt = 'B' THEN tau = 0;      ** the tau_i or mu_i;
   IF Trt = 'C' THEN tau = 0;
   IF Trt = 'D' THEN tau = 0;
OUTPUT;
END;
END;
;

DATA rcbd2; SET rcbd; BY Trt;
     IF last.Trt THEN OUTPUT;
DROP b taudiff;
PROC PRINT DATA=rcbd2;

PROC GLMPOWER DATA = rcbd;
      CLASS Trt Blocks;
      MODEL tau = Trt Blocks;
      POWER
         STDDEV = 1.5 2.0 2.5 3.0
         ALPHA  = 0.01 0.05 0.10
         NTOTAL = 40
         POWER  = .;
TITLE 'Determine power for an RCBD with 4 treatment and 10 blocks';
TITLE2 'for three alpha levels and various sigma values';
RUN;
```

SAS Output

Determine power for an RCBD with 4 treatment and 10 blocks
for three alpha levels and sigma=2

Obs	Trt	Blocks	tau
1	A	10	3
2	B	10	0
3	C	10	0
4	D	10	0

The GLMPOWER Procedure

Fixed Scenario Elements

Dependent Variable	tau
Total Sample Size	40
Error Degrees of Freedom	27

Computed Power

Index	Source	Alpha	Std Dev	Test DF	Power
1	Trt	0.01	1.5	3	0.963
2	Trt	0.01	2.0	3	0.730
3	Trt	0.01	2.5	3	0.465
4	Trt	0.01	3.0	3	0.292
5	Trt	0.05	1.5	3	0.995
6	Trt	0.05	2.0	3	0.909
7	Trt	0.05	2.5	3	0.729
8	Trt	0.05	3.0	3	0.554
9	Trt	0.10	1.5	3	0.998
10	Trt	0.10	2.0	3	0.955
11	Trt	0.10	2.5	3	0.833
12	Trt	0.10	3.0	3	0.688

3.10 Simple Repeated Measures Designs

The following comments are based on *Statistical Principles in Experimental Design* by B. Winer and from your text.

- In experimental work in biomedical, pharmaceutical, social, behavioral sciences, and occasionally in physical sciences and engineering, the experimental units are people (or animals).
- Because subjects vary (e.g., with respect to physical characteristics, health history, life experiences, training, etc.), the responses of subjects who receive the same treatment can vary greatly. If it is not controlled or accounted for in the analysis, it will can greatly inflate the experimental error making it difficult to detect real differences between treatments .
- This is similar to the case when blocking is ignored in a RCBD when there is large variability among the blocks.
- If this source of subject-to-subject variability can be separated from the treatment effects and the random experimental error, then the sensitivity of the experiment to detect real differences between treatments in increased ( ).
- It may be possible to control for the subject-to-subject variability by running a design in which each subject receives all a treatments. Such a design is called a .
- We will look at the simplest RMD in which each subject receives all a treatments in a random order.
- One potential problem is the in which the effect of receiving one treatment may influence the effect of (or carry-over to) the next treatment received by the subject.
- The goal is to prevent (or, at least minimize) any carry-over effects when designing the study. For example, in drug studies, this involves waiting a sufficient amount of time until the drug is out of the subject's system before administering the next drug.
- In our analysis, we will be assuming that there are no carry-over effects associated with a treatment. That is, we are assuming that the subjects unique characteristics remain constant (uniform) at those times when the treatments are administered.

Notation:

- Assume there are a treatments and n subjects.
- If we have one observation per treatment level for each subject, and the order in which the treatments are run within each subject is determined randomly, then we have a
- Variability due to differences in the average responses of the subjects to treatments is removed from the experimental error (assuming the SRMD model is appropriate).
- Like the RCBD, because randomization occurs within each subject, this represents an example of restricted randomization.
- The SRMD data can be summarized as:

	Treatment (i)				
Subject (j)	1	2	$\cdots$	a	Totals
1	y_{11}	y_{21}	$\cdots$	y_{a1}	$y_{\cdot 1}$
2	y_{12}	y_{22}	$\cdots$	y_{a2}	$y_{\cdot 2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{1n}	y_{2n}	$\cdots$	y_{an}	$y_{\cdot n}$
Totals	$y_{1\cdot}$	$y_{2\cdot}$	$\cdots$	$y_{a\cdot}$	$y_{\cdot\cdot}$

- Like all designs, we have for total variation ($df = an - 1$):

$$SS_{total} = \sum_{i=1}^{a}\sum_{j=1}^{n}$$

- The part of total variation attributable to differences among the means of the subjects is

$$SS_{between\ subj} = a\sum_{j=1}^{n} \tag{13}$$

- The variation within subject j having $df = a - 1$ is

$$SS_{within\ subj\ j} = \sum_{i=1}^{a} \tag{14}$$

- The within-subject variation in (14) pooled over the n subjects having $df = n(a-1)$ is

$$SS_{within\ subj} = \sum_{j=1}^{n}\sum_{i=1}^{a} \tag{15}$$

- It can be shown that SS_{total} partitions as

$$SS_{total} = SS_{between\ subj} + SS_{within\ subj} \tag{16}$$

- Note that the variation of responses within a subject has two sources: one part depends on differences due to the treatment received and the other is random residual variation. The part which depends on differences due to the treatments (like the RCBD) is defined as

$$SS_{trt} = n\sum_{i=1}^{a} \tag{17}$$

and the part that is residual error variation is

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \overline{y}_{i\cdot} - \overline{y}_{\cdot j} + \overline{y}_{\cdot\cdot}\right)^2 = \tag{18}$$

- It can be shown that $SS_{within\ subj}$ partitions as

$$SS_{within\ subj} = SS_{trt} + SS_E \tag{19}$$

- Combining (16) and (19) we have

$$SS_{total} =$$

- Numerically, these are exactly the same sums of squares for the RCBD with $SS_{between\ subjects}$ for the SRMD replacing the SS_{blocks} for the RCBD notationally.
- If μ is the baseline mean, τ_i is the i^{th} treatment effect, β_j is the j^{th} subject effect, and ϵ_{ij} is the random error of the observation, then the statistical model for a SRMD is:

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

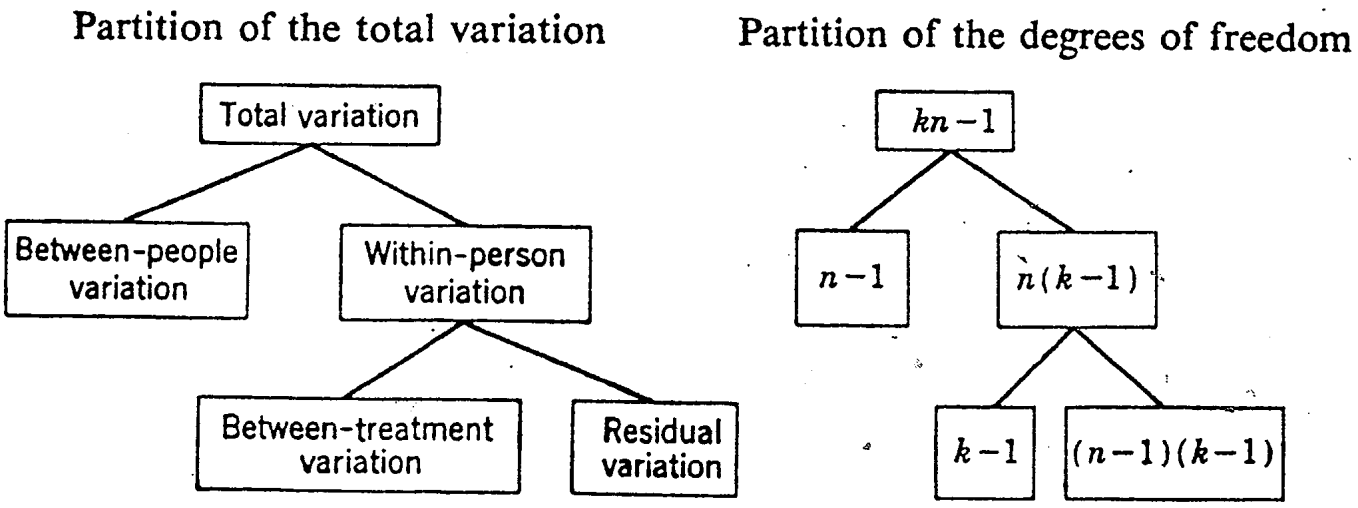
with random error $\epsilon \sim N(0, \sigma^2)$ and random subject effects $\beta_j \sim N(0, \sigma^2_\beta)$.

- Like the RBCD, we use $F = \dfrac{MS_{trt}}{MS_e}$ to test

$$\text{H}_0 : \tau_1 = \tau_2 = \cdots = \tau_a \qquad \text{vs} \qquad \text{H}_1 : \tau_i \neq \tau_j \text{ for some } i \neq 1'$$

or

$$\text{H}_0 : \overline{\mu}_{1\cdot} = \overline{\mu}_{2\cdot} = \cdots = \overline{\mu}_{a\cdot} \qquad \text{vs} \qquad \text{H}_1 : \overline{\mu}_{i\cdot} = \overline{\mu}_{i'\cdot} \text{ for some } i \neq i'$$



EXAMPLE: The purpose of this experiment was to study the effects of 4 drugs upon reaction time to a series of standardized tasks. The 10 subjects were randomly selected and had been given extensive training on these tasks prior to the experiment. Each of the subjects was observed under each of the 4 drugs and the order of administering the drug was randomized. A sufficient amount of time was allowed between administration of the drugs to avoid a carryover effect of one drug upon the effects of subsequent drugs (known as a drug interaction). The following table summarizes the reaction times.

Drug	Subject									
Treatment	1	2	3	4	5	6	7	8	9	10
1	30	14	24	38	26	28	20	27	37	29
2	28	18	20	34	28	26	23	24	35	32
3	16	10	18	20	14	19	17	22	22	18
4	34	22	30	44	30	31	30	32	43	32

SAS Code for Simple Repeated Measures Design

```
DM 'LOG; CLEAR; OUT; CLEAR;';

ODS GRAPHICS ON;
ODS PRINTER PDF file='C:\COURSES\ST541\REPMEAS.PDF';
OPTIONS NODATE NONUMBER;

DATA IN;
   DO drug =1 TO 4;
   DO subject = 1 TO 10;
      INPUT reaction @@; OUTPUT;
   END; END;
LINES;
30 14 24 38 26  28 20 27 37 29
28 18 20 34 28  26 23 24 35 32
16 10 18 20 14  19 17 22 22 18
34 22 30 44 30  31 30 32 43 32
;
PROC GLM DATA=IN PLOTS=(ALL);
     CLASS subject drug;
     MODEL reaction = drug subject / SS3;
     MEANS drug / TUKEY;
     MEANS subject;
TITLE 'SINGLE FACTOR REPEATED MEASURES DESIGN';
RUN;
```

SINGLE FACTOR REPEATED MEASURES DESIGN

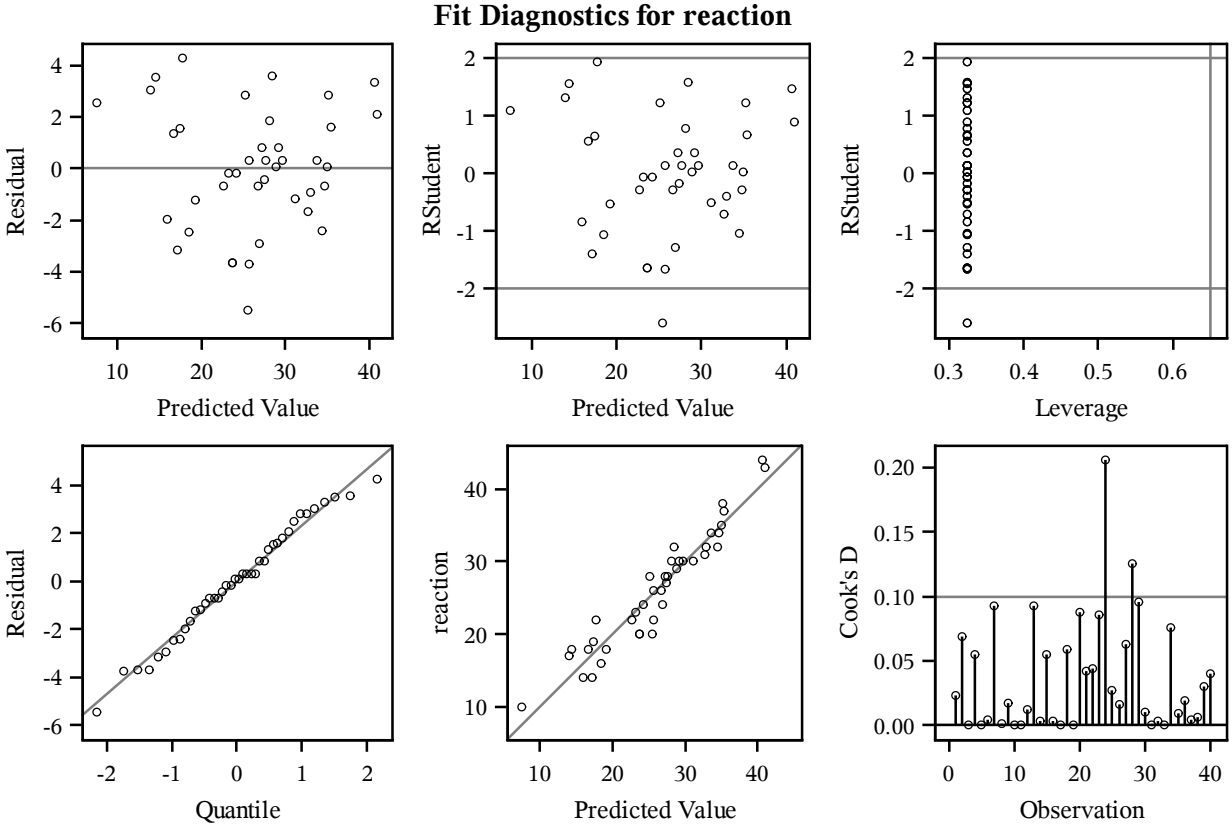
The GLM Procedure

Variable: reaction

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	2228.800000	185.733333	23.26	<.0001
Error	27	215.575000	7.984259		
Corrected Total	39	2444.375000			

R-Square	Coeff Var	Root MSE	reaction Mean
0.911808	10.81586	2.825643	26.12500

Source	DF	Type III SS	Mean Square	F Value	Pr > F
drug	3	1190.675000	396.891667	49.71	<.0001
subject	9	1038.125000	115.347222	14.45	<.0001



Tukey's Studentized Range (HSD) Test for reaction

Alpha	0.05
Error Degrees of Freedom	27
Error Mean Square	7.984259
Critical Value of Studentized Range	3.87009
Minimum Significant Difference	3.4581

Means with the same letter are not significantly different.			
Tukey Grouping	**Mean**	**N**	**drug**
A	32.800	10	4
B	27.300	10	1
B			
B	26.800	10	2
C	17.600	10	3

		reaction	
Level of subject	**N**	**Mean**	**Std Dev**
1	4	27.0000000	7.7459667
2	4	16.0000000	5.1639778
3	4	23.0000000	5.2915026
4	4	34.0000000	10.1980390
5	4	24.5000000	7.1879529
6	4	26.0000000	5.0990195
7	4	22.5000000	5.5677644
8	4	26.2500000	4.3493295
9	4	34.2500000	8.8459030
10	4	27.7500000	6.6520673