3 RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)

- 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 **homogeneous** 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.
- 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 **restricted randomization**.

3.1 RCBD Notation

- 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

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \text{and} \quad \epsilon_{ij} \sim IIDN(0, \sigma^2). \tag{6}$$

- 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 two constraints.
- Initially, we will assume the textbook constraints: $\sum_{i=1}^{a} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$.
- These are not the default SAS constraints $(\tau_a = 0, \beta_b = 0)$ or R constraints $(\tau_1 = 0, \beta_1 = 0)$.
- Applying these constraints, will yield least-squares estimates

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

where $\bar{y}_{i\cdot}$ is the mean for treatment i, and $\bar{y}_{\cdot j}$ is the mean for block j.

- Substitution of the estimates into the model yields:

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

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

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

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

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

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

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

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

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

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	$y_{\cdot 1}$=38.15
2	8.00	8.15	7.73	7.57	7.68	$y_{\cdot 2}$=39.13
3	7.93	7.87	7.74	7.80	7.21	$y_{\cdot 3}$=38.55
	$y_{1\cdot}$	$y_{2\cdot}$	$y_{3\cdot}$	$y_{4\cdot}$	$y_{5\cdot}$	
Totals	23.55	24.16	23.23	22.54	22.35	$y_{\cdot\cdot}$=115.83

Treatment Means	$\bar{y}_{1\cdot} = 7.850$	$\bar{y}_{2\cdot} = 8.053$	$\bar{y}_{3\cdot} = 7.743$	$\bar{y}_{4\cdot} = 7.513$	$\bar{y}_{5\cdot} = 7.450$
Block Means	$\bar{y}_{\cdot 1} = 7.630$	$\bar{y}_{\cdot 2} = 7.826$	$\bar{y}_{\cdot 3} = 7.710$		
Grand Mean	$\bar{y} = 7.723$				

Uncorrected Sum of Squares = $\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 \; =$

Correction factor = $y_{\cdot\cdot}^2/ab \; = \; 115.83^2/15 \; =$

$$\sum_{i=1}^{a} \frac{y_{i\cdot}^2}{b} = \frac{23.55^2 + 24.16^2 + 23.23^2 + 22.54^2 + 22.35^2}{3} = \frac{2685.5151}{3} =$$

$$\sum_{j=1}^{b} \frac{y_{\cdot j}^2}{a} = \frac{38.15^2 + 39.13^2 + 38.55^2}{5} = \frac{4472.6815}{5} =$$

$SS_{Total} \; = \; 895.6183 - 894.4393 \; =$

$SS_{Trt} \; = \; 895.1717 - 894.4393 \; =$

$SS_{Block} \; = \; 894.5364 - 894.4393 \; =$

$SS_E \; = \; 1.1790 - 0.7324 - 0.0971 \; =$

Analysis of Variance (ANOVA) Table

Source of Variation	Sum of Squares	d.f.	Mean Square	F Ratio	p-value
K_2O lbs/acre			.18311		.0404
Blocks			.04856	——	
Error			.043685	——	
Total		14	——	——	

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 ?
- The **conclusion** is to H$_0$ and conclude that

SAS Output for the RCBD Example

ANOVA RESULTS FOR STRENGTH BY TREATMENT

The GLM Procedure

Dependent Variable: strength

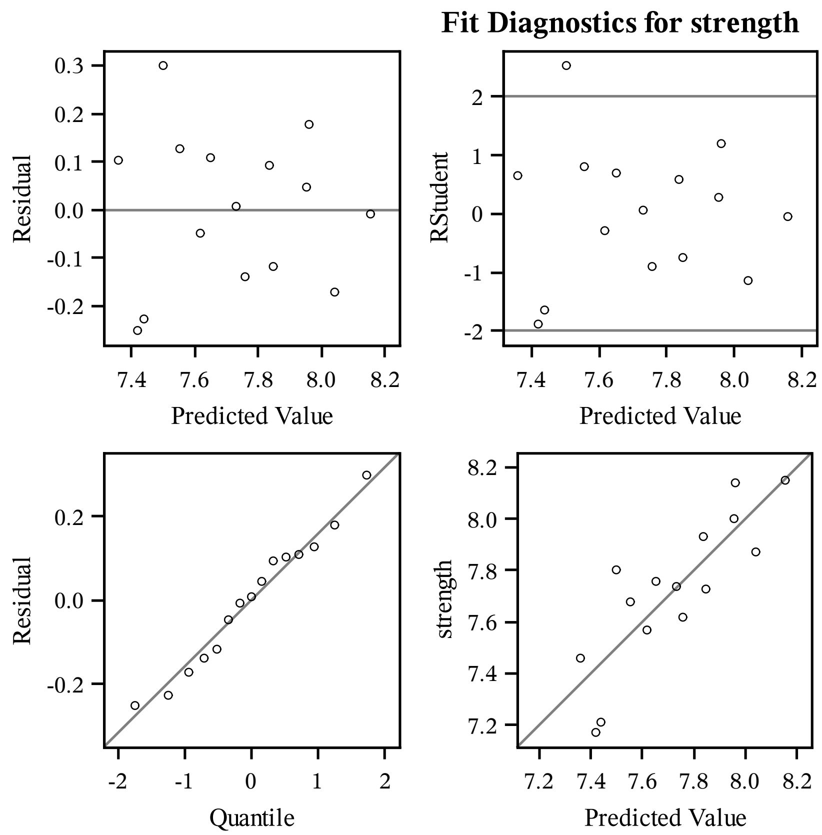
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.82956000	0.13826000	3.16	0.0677
Error	8	0.34948000	0.04368500		
Corrected Total	14	1.17904000			

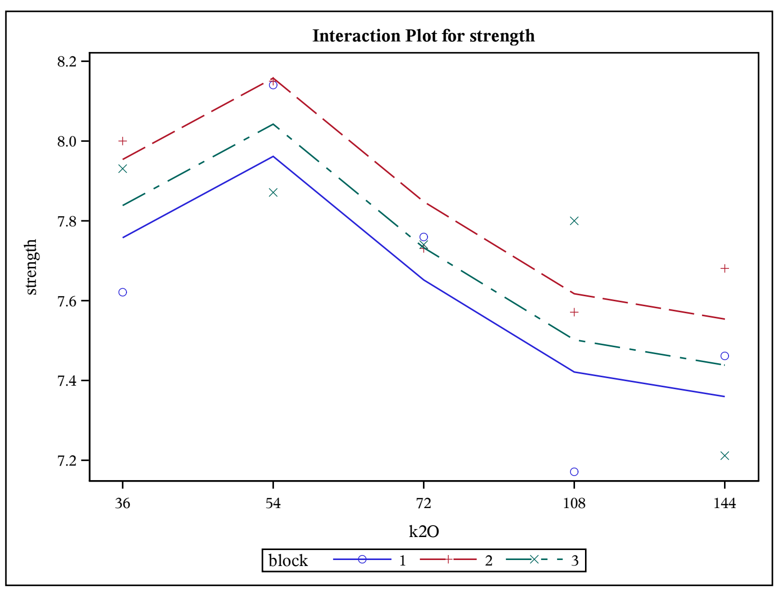
R-Square	Coeff Var	Root MSE	strength Mean
0.703589	2.706677	0.209010	7.722000

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

Parameter	Estimate		Standard Error	t Value	Pr > |t|
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
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
block 3	0.000000000	B	.	.	.

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	Std Dev
1	5	7.63000000	0.35972211
2	5	7.82600000	0.24047869
3	5	7.71000000	0.28853076

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

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

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	

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 **restriction on randomization**. 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 **restriction error** 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

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{7}$$

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

$$y_{ijk} = \mu + \tau_i + \beta_j + \delta_j + \epsilon_{ij} \tag{8}$$

where μ, τ_i, and β_j are the same in (8) as in (7), 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

$$\mathrm{H}_0:\ \tau_1 = \cdots = \tau_a = 0 \qquad \text{against} \qquad \mathrm{H}_1:\ \text{not all of the } \tau_i\text{s are equal} \tag{9}$$

- 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 (10) would be appropriate to test for a 'general' blocking effect.

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

$$\mathrm{H}_0:\ \sigma^2_\delta + \phi(\beta) = 0 \qquad \text{against} \qquad \mathrm{H}_1:\ \sigma^2_\delta + \phi(\beta) \neq 0 \tag{10}$$

  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 **there is no F test for blocks**. That is, there is no test for $\mathrm{H}_0:\ \sigma^2_\beta = 0$ if blocks are random and no test for $\mathrm{H}_0:\ \beta_1 = \beta_2 = \cdots = \beta_b$ 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. The SAS analysis for both models is on the next page. Here are important results:

	Without blocks	With blocks
R^2		
MS_E		
p-value		

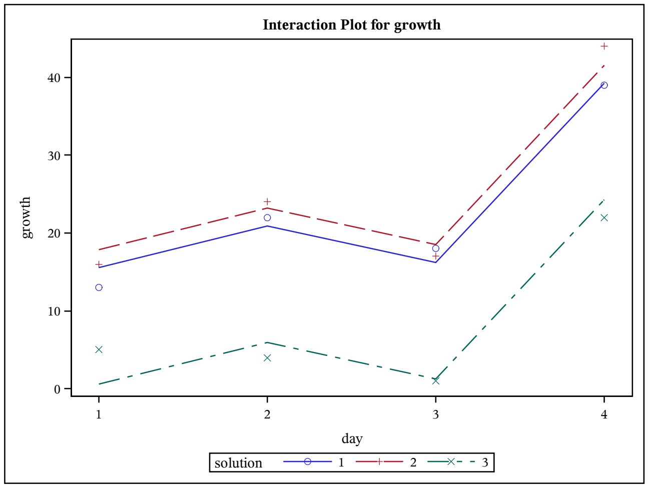
- Note that we would fail to reject H_0 if blocks were not in the model because there is large variability across blocks ($MS_{day} = 368.97$).
- If the $SS_{day} = 1106.92$ and $df_{day} = 3$ is pooled with the the $SS_E = 41.83$ and $df_E = 6$ in the model with days (blocks), then it forms the $SS_E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

SAS Code for RCBD Analyses With and Without Blocks

```
DM 'LOG;CLEAR;OUT;CLEAR';
ODS GRAPHICS ON;
* ODS PRINTER PDF file='C:\COURSES\ST541\RCBD2.PDF';
OPTIONS NODATE NONUMBER LS=76 PS=54;

********************************************;
*** RCBD ANALYSES WITH AND WITHOUT BLOCKS ***;
********************************************;
DATA IN;
 DO solution = 1 TO 3;
 DO day = 1 TO 4;
    INPUT growth @@; OUTPUT;
 END; END;
LINES;
13 22 18 39     16 24 17 44     5 4 1 22
;
******************************************************;
***  RUN AN ANOVA WITH SOLUTION ONLY, NO DAY BLOCKS ***;
******************************************************;
PROC GLM DATA=IN;
     CLASS solution;
     MODEL growth = solution / ss3;
TITLE 'RCBD WITHOUT DAYS (BLOCKS) IN THE MODEL';

****************************************;
***  RUN AN ANOVA WITH DAYS AS BLOCKS ***;
****************************************;
PROC GLM DATA=IN;
     CLASS day solution;
     MODEL growth = solution day / ss3;
TITLE 'RCBD WITH DAYS (BLOCKS) IN THE MODEL';
RUN;
```



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			

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

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

RCBD Without 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			

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

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

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 **sequential sum of squares method**.
- 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	
1	Total	μ	$N-1$	RSS_1	
2	Treatment	τ_i	$a-1$	$R(\tau\|\mu)$	$= RSS_1 - RSS_2$
3	Blocks	β_j	$b-1$	$R(\beta\|\tau,\mu)$	$= RSS_2 - RSS_3$
3	Error	ϵ_{ij}	$N-a-b+1$	RSS_3	

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

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

3.6.2 Type III Analysis

- The Type III analysis is referred to as the **marginal means** 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
1	Total	μ	$N-1$	RSS_1
2	Treatment	τ_i	$a-1$	$R(\tau\|\beta,\mu) \;=\; RSS_2^* - RSS_3$
3	Blocks	β_j	$b-1$	$R(\beta\|\tau,\mu) \;=\; RSS_2 - RSS_3$
1	Error	ϵ_{ij}	$N-a-b+1$	RSS_3

- 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)

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			

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

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

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

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			

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

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

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

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

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			

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

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

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

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			

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

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

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

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

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

*************************************;
*** 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;
```

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

$$\mathrm{H}_0 : \ \overline{\mu}_{1\cdot} \ = \ \overline{\mu}_{2\cdot} \ = \cdots = \overline{\mu}_{a\cdot}$$

$$\mathrm{H}_1 : \ \overline{\mu}_{i\cdot} \ \neq \ \overline{\mu}_{i^*\cdot} \quad \text{for some } i \neq i^* \quad \text{and} \quad \overline{\mu}_{i\cdot} = \left(\sum_{j=1}^{b} \mu_{ij} \right) /b.$$

Hypotheses for Type I (V1) Sum of Squares

$$\mathrm{H}_0 : \ \frac{1}{n_{1\cdot}} \sum_{j=1}^{b} n_{1j}\mu_{1j} \ = \ \frac{1}{n_{2\cdot}} \sum_{j=1}^{b} n_{2j}\mu_{2j} \ = \cdots = \ \frac{1}{n_{a\cdot}} \sum_{j=1}^{b} n_{aj}\mu_{aj}$$

$$\mathrm{H}_1 : \ \frac{1}{n_{i\cdot}} \sum_{j=1}^{b} n_{ij}\mu_{ij} \ \neq \ \frac{1}{n_{i^*\cdot}} \sum_{j=1}^{b} n_{i^*j}\mu_{i^*j} \quad \text{for some } i \neq i^*.$$

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 treamtment 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 $\widehat{\epsilon}_{ij} = e_{ij} = y_{ij} - \widehat{\mu} - \widehat{\tau}_i - \widehat{\beta}_j$.
- Goal: Find $\widehat{\mu}$, $\widehat{\tau}_i$, and $\widehat{\beta}_j$ that minimize L:

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

- Solution: Solve the normal equations

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

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

$$\begin{array}{llll} \text{(i)} & y_{\cdot\cdot} = & ab\widehat{\mu} + b\sum_{i=1}^{a}\widehat{\tau}_i + a\sum_{j=1}^{b}\widehat{\beta}_j & \\ \text{(ii)} & y_{i\cdot} = & b\widehat{\mu} + b\,\widehat{\tau}_i + \sum_{j=1}^{b}\widehat{\beta}_j & \text{for } i = 1,2,\ldots,a \\ \text{(iii)} & y_{\cdot j} = & a\widehat{\mu} + \sum_{i=1}^{a}\widehat{\tau}_i + a\widehat{\beta}_j & \text{for } j = 1,2,\ldots,b \end{array}$$

- 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 $\sum_{j=1}^{b}\beta_j = 0$.
- 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:

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

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

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

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			
Type of Coupon	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 $j = 1,2,3,4$

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

- 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).

$$X = \begin{array}{c} \begin{array}{ccccccc} \mu & \tau_1 & \tau_2 & \tau_3 & \beta_1 & \beta_2 & \beta_3 \end{array} \\ \left[\begin{array}{c|ccc|ccc} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 & -1 & -1 \\ \hline 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 & -1 \\ \hline 1 & -1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \end{array}\right] \end{array} \qquad y = \left[\begin{array}{c} -2 \\ -1 \\ 1 \\ 5 \\ \hline -1 \\ -2 \\ 3 \\ 4 \\ \hline -3 \\ -1 \\ 0 \\ 2 \\ \hline 2 \\ 1 \\ 5 \\ 7 \end{array}\right]$$

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

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

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

Thus, $\widehat{\tau}_4 = -\widehat{\tau}_1 - \widehat{\tau}_2 - \widehat{\tau}_3 = 10/4$ and $\widehat{\beta}_4 = -\widehat{\beta}_1 - \widehat{\beta}_2 - \widehat{\beta}_3 = 13/4$

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

$$X = \begin{array}{c} \begin{array}{ccccccccc} \mu & \tau_1 & \tau_2 & \tau_3 & \tau_4 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{array} \\ \left[\begin{array}{c|cccc|cccc}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\hline
0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{array}\right] \end{array} \qquad y = \left[\begin{array}{r}
-2 \\ -1 \\ 1 \\ 5 \\ -1 \\ -2 \\ 3 \\ 4 \\ -3 \\ -1 \\ 0 \\ 2 \\ 2 \\ 1 \\ 5 \\ 7 \\ \hline 0 \\ 0
\end{array}\right]$$

$$X'X = \left[\begin{array}{rrrrrrrrr}
16 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\
4 & 5 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
4 & 1 & 5 & 1 & 1 & 1 & 1 & 1 & 1 \\
4 & 1 & 1 & 5 & 1 & 1 & 1 & 1 & 1 \\
4 & 1 & 1 & 1 & 5 & 1 & 1 & 1 & 1 \\
4 & 1 & 1 & 1 & 1 & 5 & 1 & 1 & 1 \\
4 & 1 & 1 & 1 & 1 & 1 & 5 & 1 & 1 \\
4 & 1 & 1 & 1 & 1 & 1 & 1 & 5 & 1 \\
4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 5
\end{array}\right] \qquad (X'X)^{-1} = \frac{1}{16}\left[\begin{array}{rrrrrrrrr}
3 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4
\end{array}\right]$$

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