

STAT 500

Randomized Complete Block Design (RCBD)

Randomized Complete Block Design (RCBD)

- **Block:** a group of experimental units that, prior to treatment, are expected to be more like one another (with respect to one or more response variables) than experimental units in general. (in simple words, groups of similar experimental units)
- **Randomized Complete Block Design (RCBD):** experimental design in which separate and completely randomized treatment assignments are made for each of multiple blocks in such a way that all treatments have at least one experimental unit in each block.

Typical RCBDs

- J treatments
- n blocks with J units in each block
 - Units within each block are similar
 - Within each block, randomly assign J treatments to the units so that one experimental unit for each treatment
 - Each block is essentially a repetition of the experiment

Remarks about Block Designs

- If blocks have too few units for full repetition, we could use an **incomplete** block design
 - Balanced Incomplete Block Design (BIBD).
 - * Each treatment is represented by the same number of experimental units.
 - * Each pair of treatments appears together in the same block an equal number of times.
- If blocks have many units (some multiple of J), then we can apply each treatment more than once

RCBD: Model

Model for experiments with one unit per treatment per block

$$Y_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}$$

- $i = 1, \dots, n$ indexes blocks
- $j = 1, \dots, J$ indexes treatments
- τ_j are fixed treatment effects (with $\tau_J = 0$)
- β_i are block effects
 - Could be fixed effects with $\beta_n = 0$
 - Could be random effects with $\beta_i \sim N(0, \sigma_B^2)$
- Additive model (same treatment effects in each block)
- $\epsilon_{ij} \sim N(0, \sigma_e^2)$

RCBD: ANOVA table

source of variation	degrees of freedom	sums of squares
blocks	$n - 1$	$J \sum_{i=1}^n (\bar{Y}_{i.} - \bar{Y}_{..})^2$
treatments	$J - 1$	$n \sum_{j=1}^J (\bar{Y}_{.j} - \bar{Y}_{..})^2$
error	$(n - 1)(J - 1)$	$\sum_{i=1}^n \sum_{j=1}^J (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$
total	$nJ - 1$	$\sum_{i=1}^n \sum_{j=1}^J (Y_{ij} - \bar{Y}_{..})^2$

RCBD: Expectations of Mean Squares

Residual (error) Mean Square: $E(MS_{error}) = \sigma_e^2$

Fixed Treatment Effects (with $\bar{\tau} = \sum_{j=1}^J \tau_j$)

$$E(MS_{treatments}) = \sigma_e^2 + \frac{n}{J-1} \sum_{j=1}^J (\tau_j - \bar{\tau})^2$$

Fixed Blocks (with $\bar{\beta} = \sum_{i=1}^n \beta_i$)

$$E(MS_{blocks}) = \sigma_e^2 + \frac{J}{n-1} \sum_{i=1}^n (\beta_i - \bar{\beta})^2$$

Random Blocks

$$E(MS_{blocks}) = \sigma_e^2 + J\sigma_{\beta}^2$$

RCBD: Tests of Treatment Effects

Test the null hypothesis of no treatment effects:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_J$$

against the alternative that at least two population means are different.

Reject H_0 if

$$F = \frac{MS_{treatments}}{MS_{error}} \geq F_{(J-1, (n-1)(J-1)), 1-\alpha}$$

Note that

$$\frac{E(MS_{treatments})}{E(MS_{error})} = \frac{\sigma_e^2 + \frac{n}{J-1} \sum_{j=1}^J (\tau_j - \bar{\tau})^2}{\sigma_e^2}$$

RCBD: Test of Block Effects

Test the null hypothesis of no block effects:

$H_o : \beta_1 = \beta_2 = \dots = \beta_n$ for fixed block effects
or $H_o : \sigma_{\beta}^2 = 0$ for random block effects

Reject H_o if

$$F = \frac{MS_{blocks}}{MS_{error}} \geq F_{(n-1, (n-1)(J-1)), 1-\alpha}$$

Usually this test is of little interest, because variation among blocks is anticipated.

RCBD: Differences in Estimated Treatment Means

- $Y_{ij} - Y_{ik} = (\mu + \beta_i + \tau_j + \epsilon_{ij}) - (\mu + \beta_i + \tau_k + \epsilon_{ik})$
 $= \tau_j - \tau_k + \epsilon_{ij} - \epsilon_{ik}$
- $Var(Y_{ij} - Y_{ik}) = Var(\epsilon_{ij} - \epsilon_{ik}) = 2\sigma_e^2$
- Similarly,

$$\begin{aligned} Var(\bar{Y}_{.j} - \bar{Y}_{.k}) &= Var\left(\frac{1}{n} \sum_{i=1}^n Y_{ij} - \frac{1}{n} \sum_{i=1}^n Y_{ik}\right) \\ &= Var\left(\frac{1}{n} \sum_{i=1}^n (Y_{ij} - Y_{ik})\right) \\ &= Var\left(\frac{1}{n} \sum_{i=1}^n (\tau_j - \tau_k + \epsilon_{ij} - \epsilon_{ik})\right) \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n Var(\tau_j - \tau_k + \epsilon_{ij} - \epsilon_{ik}) \right) = \frac{1}{n^2} \left(\sum_{i=1}^n 2\sigma_e^2 \right) = \frac{2\sigma_e^2}{n} \end{aligned}$$

RCBD: Contrasts

- Variances for contrasts among the treatment means are computed in a similar manner:
 - The design is balanced (each treatment occurs the same number of times in each block)
 - Block effects cancel out because $\sum_j c_j = 0$
 - variance for contrast estimate: $Var(\sum_j c_j \bar{Y}_{.j}) = \sigma_e^2 \sum_j c_j^2 / n$, estimated by $MS_{error} \sum_j c_j^2 / n$
- Can use orthogonal contrasts to partition treatment sums of squares

RCBD: Penicillin Experiment

$\left. \begin{array}{l} \textit{Process A} \\ \textit{Process B} \\ \textit{Process C} \\ \textit{Process D} \end{array} \right\}$ Levels of a “fixed” treatment factor

Blocks correspond to different batches of a basic raw material

- Random sample of five batches from all possible batches
- Split each batch into four parts:
 - run each process on one randomly selected part
 - randomize the order in which the processes are run within each batch

Random Block (Batch) Effects

- Batches are sampled from a population of many possible batches
- To repeat this experiment you would need to use a different set of batches of raw material

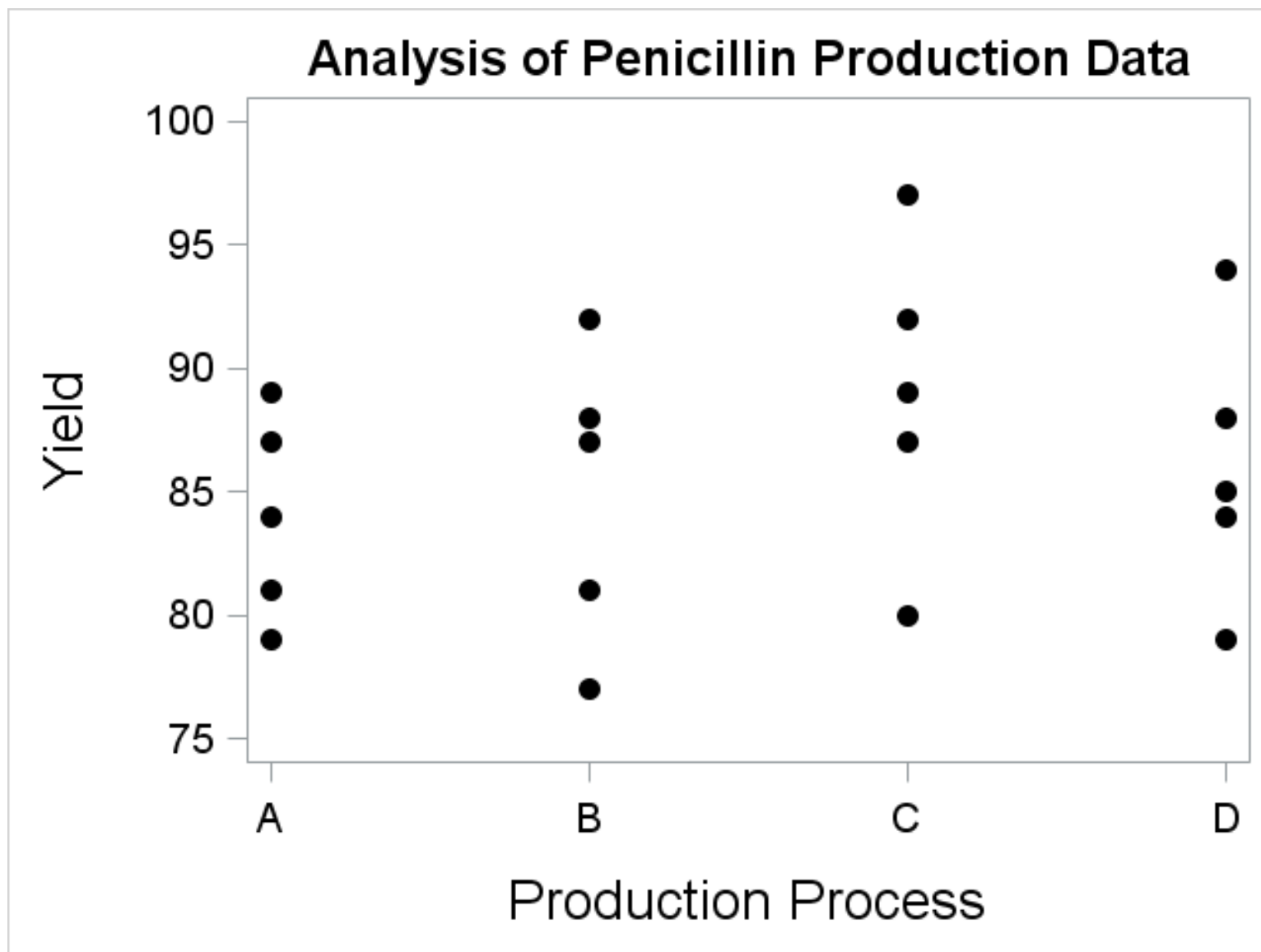
Data Source: Box, Hunter & Hunter (1978),
Statistics for Experimenters,
Wiley & Sons, New York.

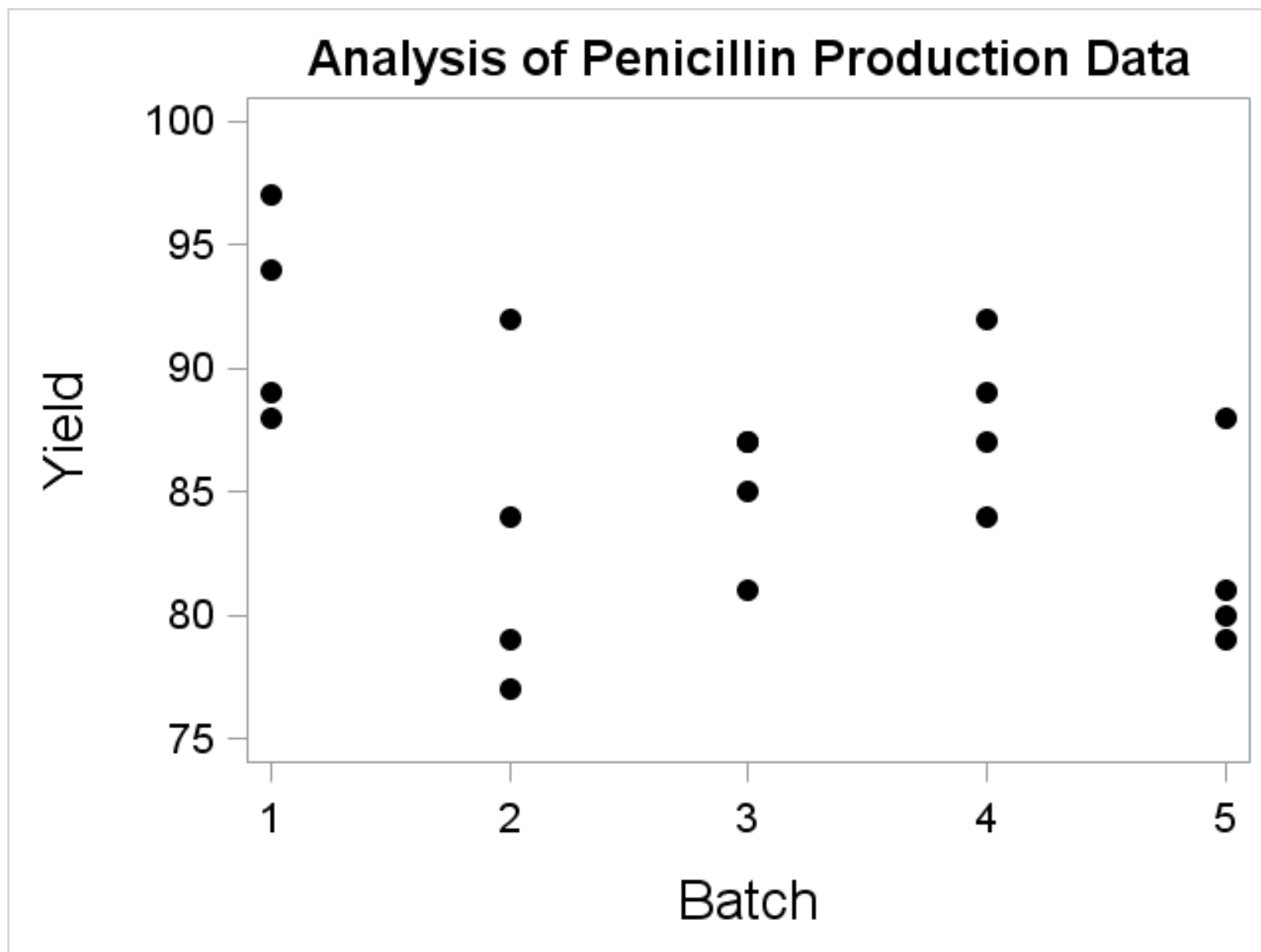
SAS code: penicillin.sas

Penicillin Experiment

Data: Yields

Batch	Processes			
	A	B	C	D
1	89	88	97	94
2	84	77	92	79
3	81	87	87	85
4	87	92	89	84
5	79	81	80	88
Means	84	85	86	89





Model

$$Y_{ij} = \mu + \tau_j + \beta_i + \epsilon_{ij}$$

↑	↑	↑	↑
Yield for the j -th process applied to the i -th batch	mean yield for the j -th process, averaging across the entire population of possible batches	random batch effect	random error

where $\beta_i \sim N(0, \sigma_\beta^2)$, $\epsilon_{ij} \sim N(0, \sigma_e^2)$ and any ϵ_{ij} is independent of any β_i .

Model

$$\begin{aligned}\mu_j = E(Y_{ij}) &= E(\mu + \tau_j + \beta_i + e_{ij}) \\ &= \mu + \tau_j + E(\beta_i) + E(e_{ij}) \\ &= \mu + \tau_j \quad i = 1, 2, 3, 4\end{aligned}$$

represents the mean yield for the j -th process, averaging across all possible batches.

PROC GLM and PROC MIXED in SAS would fit a restricted model with $\tau_4 = 0$. Then

- $\mu = \mu_4$ is the mean yield for process D
- $\tau_j = \mu_j - \mu_4 \quad j = 1, 2, 3, 4.$

Variance-Covariance Structure

$$\begin{aligned} \text{Var}(Y_{ij}) &= \text{Var}(\mu + \tau_j + \beta_i + e_{ij}) \\ &= \text{Var}(\beta_i + e_{ij}) \\ &= \text{Var}(\beta_i) + \text{Var}(e_{ij}) \\ &= \sigma_{\beta}^2 + \sigma_e^2 \quad \text{for all } (i, j) \end{aligned}$$

is the variance of a measurement of yield for one run of a process with a random sample of a batch

Variance-Covariance Structure

For runs of different processes on the same batch:

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{ik}) &= \text{Cov}(\mu + \tau_j + \beta_i + e_{ij}, \mu + \tau_k + \beta_i + e_{ik}) \\ &= \text{Cov}(\beta_i + e_{ij}, \beta_i + e_{ik}) \\ &= \text{Cov}(\beta_i, \beta_i) + \text{Cov}(\beta_i, e_{ij}) + \text{Cov}(e_{ik}, \beta_i) \\ &\quad + \text{Cov}(e_{ij}, e_{ik}) \\ &= \text{Var}(\beta_i) = \sigma_\beta^2 \quad \text{for all } j \neq k \end{aligned}$$

Correlation Among Yields for Runs on the Same Batch

$$\rho = \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{Var(Y_{ij})Var(Y_{ik})}}$$

$$= \frac{\sigma_{\beta}^2}{\sigma_{\beta}^2 + \sigma_e^2} \text{ for } i \neq k$$

Results for runs on different batches are independent:

$$Cov(Y_{ij}, Y_{\ell k}) = 0 \quad \text{for } i \neq \ell$$

Variance-Covariance Matrix for the Four Runs on the Same Batch

$$\text{Var} \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix} = \begin{bmatrix} \sigma_{\beta}^2 + \sigma_e^2 & \sigma_{\beta}^2 & \sigma_{\beta}^2 & \sigma_{\beta}^2 \\ \sigma_{\beta}^2 & \sigma_{\beta}^2 + \sigma_e^2 & \sigma_{\beta}^2 & \sigma_{\beta}^2 \\ \sigma_{\beta}^2 & \sigma_{\beta}^2 & \sigma_{\beta}^2 + \sigma_e^2 & \sigma_{\beta}^2 \\ \sigma_{\beta}^2 & \sigma_{\beta}^2 & \sigma_{\beta}^2 & \sigma_{\beta}^2 + \sigma_e^2 \end{bmatrix}$$

$$= \underset{\substack{\uparrow \\ \text{matrix} \\ \text{of ones}}}{\sigma_{\beta}^2 J} + \underset{\substack{\uparrow \\ \text{identity} \\ \text{matrix}}}{\sigma_e^2 I}$$

This special type of covariance structure is called *compound symmetry*

RCBD model for Random Blocks

Write this model in the form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \\ Y_{41} \\ Y_{42} \\ Y_{43} \\ Y_{44} \\ Y_{51} \\ Y_{52} \\ Y_{53} \\ Y_{54} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{34} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{44} \\ e_{51} \\ e_{52} \\ e_{53} \\ e_{54} \end{bmatrix}$$

RCBD model for Random Blocks

Impose the SAS constraint $\tau_4 = 0$

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \\ Y_{41} \\ Y_{42} \\ Y_{43} \\ Y_{44} \\ Y_{51} \\ Y_{52} \\ Y_{53} \\ Y_{54} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \\ e_{31} \\ e_{32} \\ e_{33} \\ e_{34} \\ e_{41} \\ e_{42} \\ e_{43} \\ e_{44} \\ e_{51} \\ e_{52} \\ e_{53} \\ e_{54} \end{bmatrix}$$

$$Y = X\beta + Zu + \epsilon$$

ANOVA

For this restricted model, the least squares estimator for

$$\beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}$$

is

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

ANOVA

Source	d.f.	SS	MS	F	p-value
Blocks	4	264	66.000	3.50	0.0407
Processes	3	70	23.333	1.24	0.3387
Error	12	226	18.833		
Total	19	560			

- $\hat{\sigma}_e^2 = MS_{error} = 18.833$
- $S_{\bar{Y}_{.j} - \bar{Y}_{.k}} = \sqrt{MS_{error} \frac{2}{n}} = \sqrt{(18.833) \frac{2}{5}} = 2.745$
- MS_{blocks} is an estimate of $E(MS_{blocks}) = \sigma_e^2 + 4\sigma_\beta^2$
- An estimate of σ_β^2 is

$$\frac{MS_{blocks} - MS_{error}}{4} = \frac{66 - 18.833}{4} = 11.792$$

RCBD: Multiple Comparisons

Tukey-Kramer (HSD, *Honest Significant Difference*)

- Declare a significant difference in treatment means if

$$|\bar{Y}_i - \bar{Y}_j| \geq \frac{1}{\sqrt{2}} q(r, df_{error}, 1-\alpha) \sqrt{MS_{error} \left(\frac{1}{n} + \frac{1}{n} \right)}$$

- For the penicillin study

$$HSD = \frac{1}{\sqrt{2}} (4.199) \sqrt{18.833 \left(\frac{1}{5} + \frac{1}{5} \right)} = 8.15$$

- Order sample means from smallest to largest:

<u>Process A</u>	<u>Process B</u>	<u>Process D</u>	<u>Process C</u>
84	85	86	89

Fixed or Random Block?

- F-test for treatment effects is the same if blocks are fixed or random
- $Var(\bar{Y}_{.j} - \bar{Y}_{.k}) = \sigma_e^2(\frac{2}{n})$ is the same if blocks are fixed or random. Use

$$S_{\bar{Y}_{.j} - \bar{Y}_{.k}} = \sqrt{MS_{error} \frac{2}{n}}$$

Fixed or Random Block?

- Standard error of a single treatment mean

- Fixed blocks: $Var(\bar{Y}_{.j}) = \sigma_e^2/n$

Use $S_{\bar{Y}_{.j}} = \sqrt{MS_{error} \frac{1}{n}}$ with df_{error}

- Random blocks:

$Var(\bar{Y}_{.j}) = (\sigma_e^2 + \sigma_\beta^2)/n$ Use

$$S_{\bar{Y}_{.j}} = \sqrt{\frac{(J-1)MS_{error} + MS_{blocks}}{Jn}}$$

$$\text{with } df = \frac{[\frac{J-1}{J}MS_{error} + \frac{1}{J}MS_{blocks}]^2}{\frac{[\frac{J-1}{J}MS_{error}]^2}{df_{error}} + \frac{[\frac{1}{J}MS_{blocks}]^2}{df_{blocks}}}$$