Randomized Complete Block Designs

In experiments where the available experimental units are not homogeneous, grouping the experimental units into blocks of homogeneous units will reduce the experimental error variance and also increase the range of validity for inferences about the treatment effects.

- **Examples of Randomized Complete Block Designs**

  1. In an experiment on the effects of four levels of newspaper advertising saturation on sales volume, 16 cities were included in the study. Size of city usually is highly correlated with the response variable, sales volume. Hence, it is desirable to block the 16 cities into four groups of four cities each, according to population size. Within each block, the four treatments are then assigned at random.

  2. In an experiment on the effects of three different incentive pay schemes on employee productivity of electronic assemblies, the experimental unit is an employee, and 30 employees are available for the study. Since productivity here is highly correlated with manual dexterity, it is desirable to block the 30 employees into 10 groups of three according to their manual dexterity. Thus, the three employees with the highest manual dexterity ratings are grouped into one block, and so on for the other employees. Within each block, the three incentive pay schemes are then assigned randomly to the three employees.

  3. A chemist is studying the reaction rate of five chemical agents. Only five agents can be analyzed effectively per day. Since day-to-day differences may affect the reaction rate, each day is used as a block, and all five chemical agents are tested each day in independently randomized orders.

- **Blocking Criteria**

  1. Characteristics associated with the unit
     - **a.** for persons - gender, age, income, intelligence, education, job experience, attitudes, etc.
     - **b.** for geographic areas - population size, average income, etc.
  2. Characteristics associate with the experimental setting – observer, time of processing (learning by observer, changes in equipment, environmental condition), machine, batch of material, measuring instrument, etc.

- **Advantages of Randomized Complete Block Design**

  1. It can, with effective grouping, provide substantially more precise results than a completely randomized design of comparable size.
  2. It can accommodate any number of treatments and replications.
  3. Different treatments need not have equal sample sizes. For instance, if the control is to have twice as large a sample size as each of three treatments, blocks of size five would be used: three units in a block are then assigned at random to the three treatments and two to the control.
  4. The statistical analysis is relatively simple.
  5. If an entire treatment or block needs to be dropped from the analysis for some reason, such as spoiled results, the analysis is not complicated thereby.
  6. Variability in experimental units can be deliberately introduced to widen the range of validity of the experimental results without sacrificing the precision of the results.

- **Disadvatages of Randomized Complete Block Design**

  1. If observations are missing within a block, a more complex analysis is required.
  2. The degrees of freedom for experimental error are not as large as with a completely randomized design. One degree of freedom is lost for each block after the first.
  3. More assumptions are required for the model (e.g., no interactions between treatments and blocks, constant variance from block to block) than for a completely randomized design model
  4. Because the blocking variable is an observational factor and not an experimental factor, cause-and-effect inferences concerning the relationship between the blocking variable and the response variable cannot be established.

- **Randomized Complete Block Design Model**

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \epsilon_{ij}, \qquad \text{for } i = 1, \ldots, n_b \text{ and } j = 1, \ldots, r$$

where:

1. Y_{ij} is the response in the ith block and jth treatment.
2. $\mu_{..}$ is a constant.
3. ρ_i are constants for the block effects, subject to the restriction $\sum_{i=1}^{n_b} \rho_i = 0$.
4. τ_j are constants for the treatment effects, subject to the restriction $\sum_{j=1}^{r} \tau_j = 0$.
5. ϵ_{ij} are independent $N(0, \sigma^2)$.

- **Estimators**

Parameter	Estimator
$\mu_{..}$	$\hat{\mu}_{..} = \bar{Y}_{..}$
ρ_i	$\hat{\rho}_i = \bar{Y}_{i.} - \bar{Y}_{..}$
τ_j	$\hat{\tau}_i = \bar{Y}_{.j} - \bar{Y}_{..}$

- **ANOVA Table**

ANOVA Table

Source	DF	Sum of Squares	Mean Square	F
Blocks	$n_b - 1$	$SSBL = \sum_{i=1} r(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$MSBL = \frac{SSBL}{n_b - 1}$	$F_{\text{BL}} = \frac{MSBL}{MSE}$
Treatments	$r - 1$	$SSTR = \sum_{j=1} n_b(\bar{Y}_{.j} - \bar{Y}_{..})^2$	$MSTR = \frac{SSTR}{r - 1}$	$F_{\text{TR}} = \frac{MSTR}{MSE}$
Error	$(n_b - 1)(r - 1)$	$SSE = \sum_{i=1} \sum_{j=1} (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$	$MSE = \frac{SSE}{(n_b - 1)(r - 1)}$	
Total	$n_b r - 1$	$SSTO = \sum_{i=1} \sum_{j=1} (Y_{ij} - \bar{Y}_{..})^2$		

Note:

1. $SSTO = SSTR + SSBL + SSE$.
2. $E(MSBL) = \sigma^2 + r\frac{\sum \rho_i^2}{n_b - 1}$.
3. $E(MSTR) = \sigma^2 + n_b\frac{\sum \tau_j^2}{r - 1}$.
4. $E(MSE) = \sigma^2$.

- **Example: Risk Premium.** In an experiment on decision making, executives were exposed to one of three methods of quantifying the maximum risk premium they would be willing to pay to avoid uncertainty in a business decision. The three methods are the utility method (U), the worry method (W), and the comparison method (C). After using the assigned method, the subjects were asked to state their degree of confidence in the method of quantifying the risk premium on a scale from 0 (no confidence) to 20 (highest confidence). Fifteen subjects were used in the study. They were grouped into five blocks of three executives, according to age. The table below contains the results of the experiment.

Age	Utility	Worry	Comparison
1 (oldest)	1	5	8
2	2	8	14
3	7	9	16
4	6	13	18
5 (youngest)	12	14	17

- **R commands:**

```
> response=c(1,5,8,2,8,14,7,9,16,6,13,18,12,14,17)
> block=gl(n=5,k=3) # n = no. of levels; k = no. of replications
> treatment=gl(3,1,length=15)
> data=data.frame(rating=response,age=block,method=treatment)
> interaction.plot(treatment,block,response)

> result=aov(response~block+treatment)
> anova(result)

Analysis of Variance Table
Response: response
          Df  Sum Sq Mean Sq F value    Pr(>F)
block      4 171.333  42.833  14.357 0.0010081 **
treatment  2 202.800 101.400  33.989 0.0001229 ***
Residuals  8  23.867   2.983
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> qqnorm(result$residuals)
> plot(result$fitted,result$residuals,xlab="Fitted values",ylab="Residuals")

> tapply(response,treatment,mean)
   1    2    3
 5.6  9.8 14.6

> TukeyHSD(result,"treatment",conf.level=.95)
$treatment
    diff      lwr       upr     p adj
2-1  4.2 1.078534  7.321466 0.0121268
3-1  9.0 5.878534 12.121466 0.0000920
3-2  4.8 1.678534  7.921466 0.0057757
```

- **Analysis of Treatment Effects**

  1. Single comparison: $t = qt(1-\alpha/2, df = (n_b-1)*(r-1))$
  2. Tukey procedure for all pairwise comparison: $T = qtukey(1-\alpha, r, df = (n_b-1)*(r-1))$
  3. Scheffe procedure: $S = sqrt((r-1)*qf(1-\alpha, df1 = r-1, df2 = (n_b-1)*(r-1))$
  4. Bonferroni procedure: $B = qt(1-\alpha/(2*g), df = (n_b-1)*(r-1))$