RCBD & Latin Squares & BIBD

1. Four friends want to test *four different routes* for driving from the gym to the dorms. They decide that *time of day*, *driver*, and *automobile* may be important factors. How would you **design** a study to test for a route effect taking all these blocking variables into consideration? What would be the corresponding statistical **model** for the design that you propose?

   Solution:
   There are three blocking variables that should be taken into consideration besides the treatment. Then Graeco-Latin design would be a good fit if the blocking variables each has four distinct cases to test. That is, four drivers, four cars and four different time of day should be considered in the study. The design is as follows:

	Column			
Row	1	2	3	4
1	Aα	Bβ	Cγ	Dδ
2	Bδ	Aγ	Dβ	Cα
3	Cβ	Dα	Aδ	Bγ
4	Dγ	Cδ	Bα	Aβ

   Here, A, B, C, D represent the treatments, routes. The columns represents one blocking variable, drivers for example; the rows represents another blocking variable, time of day; and a third blocking variable is represented by $\alpha, \beta, \gamma, \delta$. For the study, conduct experiments each day according to the specific combinations in the table for four days, and observe the time spent along the way. The corresponding statistical model is:

$$Y_{ijkm} = \mu_{...} + \rho_i + \kappa_j + \tau_k + \psi_m + \epsilon_{ijkm}.$$

   Here, Y_{ijkm} denotes the observed data, τ_k denotes the treatment effect and ϵ_{ijkm} is the random error component. The other three are blocking effects.

2. Consider an $r \times r$ Latin square with rows ρ_i, columns κ_j, and treatments τ_k fixed. Obtain least squares estimated of the model parameters ρ_i, κ_j, and τ_k.

   Hint: You need to follow the procedure as described in pages 125-126 of the textbook.
   Solution:
   The statistic model is:

$$Y_{ijk} = \mu_{...} + \rho_i + \kappa_j + \tau_k + \epsilon_{ijk}.$$

   The error sum of squares is then:

$$\sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r}\epsilon_{ijk}^2 = \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r}(y_{ijk} - \mu_{...} - \rho_i - \kappa_j - \tau_k)^2$$

   The goal is then to find $\hat{\mu}, \hat{\rho}_i, \hat{\kappa}_j$ and $\hat{\tau}_k$ that minimizes F, that is, find the set of values such that the partial derivatives equal to 0:

$$-2\sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r}(Y_{ijk} - \hat{\mu} - \hat{\rho}_i - \hat{\kappa}_j - \hat{\tau}_k) = 0.$$

   First, note that $\sum \rho_i = \sum \kappa_j = \sum \tau_k = 0$ by assumption, then

$$N\hat{\mu} = \sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{k=1}^{r} Y_{ijk} = Y_{...}$$
$$\hat{\mu} = \bar{Y}_{....}$$

Then

$$\rho_i = r\hat{\mu} + r\hat{\rho}_i = Y_{i..}$$
$$\hat{\rho}_i = \bar{Y}_{i..} - \bar{Y}_{...},$$

and similarly we have

$$\hat{\kappa}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$$
$$\hat{\tau}_k = \bar{Y}_{..k} - \bar{Y}_{....}$$

3. A courier company is interested in deciding between five brands (A, B, C, D, and E) of car for its next purchase of fleet cars. The brands are all comparable in purchase price. The company wants to carry out a study that will enable them to compare the brands with respect to *operating costs.* For this purpose they select *five drivers.* In addition, the study will be carried out over a *five week period.* Each week a driver is assigned to a car using randomization and a Latin Square Design. The *average cost per mile* is recorded at the end of each week and is summarized below:

	Week				
Driver	1	2	3	4	5
1	5.83 (D)	6.22 (A)	7.67 (B)	9.43 (C)	6.57 (E)
2	4.80 (A)	7.56 (D)	10.34 (C)	5.82 (E)	9.86 (B)
3	7.43 (B)	11.29 (C)	7.01 (E)	10.48 (D)	9.27 (A)
4	6.60 (E)	9.54 (B)	11.11 (D)	10.84 (A)	15.05 (C)
5	11.24 (C)	6.34 (E)	11.30 (A)	12.58 (B)	16.04 (D)

(a) Write down the statistical model that corresponds to the experiment. Identify treatment and blocking factors. Make sure you state all model assumptions and that you clearly explain the notation used in the context of the problem.

(b) Is there a difference between the five brands? State the hypotheses, decision rule and conclusion in the context of the problem.

(c) Explain why interaction terms were not included in the statistical analysis.

(d) Perform all pairwise comparisons of the car brands using a Tukey **family** confidence coefficient of 90%. Interpret the results. Are there any grouping of the different categories?

Solution:

(a) The statistical model can be written as:

$$Y_{ijk} = \mu_{...} + \rho_i + \kappa_j + \tau_k + \epsilon_{ijk}$$

where
$\mu_{...}$ is a constant,
ρ_i are constants for the row blocking effects that have $\sum \rho_i = 0$. Here it represents the effect of different driver on operational cost,
κ_j are constants for the column blocking effects that have $\sum \kappa_j = 0$. Here it represents the effect of different time period on operational cost, τ_k are constants for the treatment effects that have $\sum \tau_j = 0$. Here it represents the effect of different car brand on operational cost,
ϵ_{ijk} are assumed to be independent $\mathcal{N}(0, \sigma^2)$ random error factor.
In addition, Y_{ijk} denotes the operation cost for the (i, j)th block and kth treatment. We assume that no interaction exists, and Y_{ijk} to be independent, normally distributed with mean $E(Y_{ijk}) = \mu_{..} + \rho_i + \kappa_j + \tau_k$ and constant variance $Var(Y_{ijk}) = \sigma^2$.

(b) The hypothesis for this test is:

$$\begin{cases} H_0 & : \text{all} \tau_j = 0 \\ H_1 & : \text{Otherwise.} \end{cases}$$

The test statistic is:

$$F^* = \frac{MSTR}{MSRem}$$

The decision rules are

- If $F^* \leq F(1-\alpha; r-1, (r-1)(r-2))$, conclude H_0.
- If $F^* > F(1-\alpha; r-1, (r-1)(r-2))$, conclude H_1.

Based on SAS result from (f), $p < 0.001$, we reject H_0 and conclude that there is a difference between treatments.

(c) Because each treatment level appears exactly once in each block level, there is no way to determine the interaction between treatments and blocks. Thus we can only use an additive model for this design, without explicit interaction terms in the model.

(d) Based on results from Tukey's pairwise comparison test, Brand C and D, B and D or A and B can be grouped together as the difference between them is not significant enough. In ascending order of operational cost, the car brands can be ordered as E, A, B, C, D. The grouping suggests that E is the single most cost efficient of the five, and thus should be the one purchased.

4. An automotive engineer wished to evaluate the effects of four rubber compounds on the life of automobile tires. The manufacturing process permitted the use of up to three different compounds used in each section. To do this, the tire is divided into three sections, and a different compound is used in each section. Because each segment of a tire would be subject to nearly identical road conditions, the investigator decided to use tires as blocks, with three of the four treatments (compounds) being applied to the three experimental units (tire segments) in each block. Four tires were tested. The response Y is a coded measure of wear. The experimental layout and response data are:

	Compound			
Tire	1	2	3	4
1	238	238	279	–
2	196	213	–	308
3	254	–	334	367
4	–	312	421	412

(a) Write the BIBD model. Fit the model, obtain the residuals, and plot them against the fitted values. Also, prepare a QQplot. What are your observations?

(b) Obtain estimates for the treatment means.

(c) Tests whether or not the type of compound affects tire wear; use $\alpha = 0.05$. State the hypotheses, decision rule, and conclusion.

(d) Test whether or not block effects are present; use $\alpha = 0.05$. State the hypotheses, decision rule, and conclusion.

(e) Give a 95% confidence interval for the mean wear for compound A.

(f) Analyze the nature of the treatment effects by making all pairwise comparisons among the treatment means. Use the Tukey procedure and a 95% family confidence coefficient.

Solution:

(a) The statistical model can be written as:

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

where
μ is a constant representing the overall mean,
τ_i are constants for the treatment effects that have $\sum_{i=1}^{a} \tau_i = 0$. Here it represents the effect of different compound on the life of automobile tire,

β_j are constants for the blocking effects that have $\sum_{j=1}^{b} \beta_j = 0$. Here it represents the effect of different time period on growth rate,
ϵ_{ij} are assumed to be independent $\mathcal{N}(0, \sigma^2)$ random error factor.
In addition, Y_{ij} denotes the response for the jth block and ith treatment. We assume that no interaction exists, and Y_{ij} to be independent, normally distributed with mean $E(Y_{ij}) = \mu + \tau_i + \beta_j$ and constant variance $Var(Y_{ij}) = \sigma^2$. There is no pattern or trend in the residual against fitted value plot, so the constant variance assumption holds. The residual fits the line well in QQ-plot, which means the normality assumption holds.

(b) The estimates for the treatment means is shown as follows:

(c) The hypothesis for this test is:

$$\begin{cases} H_0 & : \text{all} \tau_i = 0 \\ H_1 & : \text{Otherwise.} \end{cases}$$

The test statistic is:

$$F^* = \frac{MSTR}{MSRem}$$

The decision rules are

- If $F^* \leq F(1-\alpha; b-1, N-a-b-1)$, conclude H_0.
- If $F^* > F(1-\alpha; b-1, N-a-b-1)$, conclude H_1.

Based on SAS result, $p = 0.0034 < \alpha = 0.05$ (Type III SS), we reject H_0 and conclude that there is a difference between treatments.

(d) The hypothesis for this test is:

$$\begin{cases} H_0 & : \text{all} \beta_j = 0 \\ H_1 & : \text{Otherwise.} \end{cases}$$

The test statistic is:

$$F^* = \frac{MSBL}{MSRem}$$

The decision rules are

- If $F^* \leq F(1-\alpha; a-1, N-a-b-1)$, conclude H_0.
- If $F^* > F(1-\alpha; a-1, N-a-b-1)$, conclude H_1.

Based on SAS result from (c), $p = 0.0032 < \alpha = 0.05$ (Type III SS), we reject H_0 and conclude that there is a difference between blocks.

(e) See (b).

(f) Based on the pairwise comparison results, compound 1 and 2 are similar, compound 3 and 4 are similar, but there is significant difference between the two groups.