Latin Square Design (LSD)

In this session we will introduce Latin Square Design (LSD). The entire topic is subdivided in the following subdivisions.

1. Introduction and Layout of LSD:
2. Model and hypothesis
3. Least square estimates of the parameters
4. Decomposition of the total sum of squares
5. Statistical Analysis
6. ANOVA Table
7. Advantages and disadvantages
8. Example
9. conclusion

1. Introduction and Layout of LSD:

LSD used to split the experiment in two directions in order to further increase power or decrease sample size. This design removes two possible sources of heterogeneity (i) row wise and (ii) column wise. i.e. Latin square designs allow for two blocking factors. For instance, if you had a plot of land the fertility of this land might change in both directions, North -- South and East -- West due to soil or moisture gradients. So, both rows and columns can be used as blocking factors. Whenever, you have more than one blocking factor a Latin square design will allow you to remove the variation for these two sources from the error variation.

Here:

- The treatments are assigned at random within rows and columns, with each treatment appears once per row and once per column.
- There are equal number of rows, columns, and treatments.
- Latin Square Design Useful where the experimenter desires to control variation in two different directions

Suppose a 5 x 5 Latin Square and it looks like this:

Raw material batch	Operator				
	1	2	3	4	5
1	**A**	**B**	**C**	**D**	**E**
2	**B**	**C**	**D**	**E**	**A**
3	**C**	**D**	**E**	**A**	**B**
4	**D**	**E**	**A**	**B**	**C**
5	**E**	**A**	**B**	**C**	**D**

A Latin square of order p is an arrangement of p symbols in p^2 cells arranged in p rows and p columns such that each symbol occurs once and only once in each row and in each column.

Thus in LSD the

1. treatments are grouped into replications in two ways
   Once in rows and in columns
2. Row and column variations are eliminated from the within treatment variation:

   The experimental units are grouped according to two factors, hence two effects like two block effects are removed from the experimental error.

   Hence the error variance can be considerable reduced in LSD

The LSD is an incomplete three way layout in which each of the three factors row, column and treatment are at p levels each and observations on the possible treatment combinations are taken

2. Model and Hypothesis

Let $y_{i\,jk}$ represent observation from k^{th} row, j^{th}column and receiving i^{th} treatment, i,j,k= 1….p

We define the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \qquad (4)$$

Where μ general effect
α_i – i^{th} treatment effect
βj – j^{th} block effect;
γk – k^{th} row effect;
ϵijk- error term , which are identically and independently distributed normal random variables with mean 0 and variance σ^2.

The null hypothesis under consideration are

1. Related to the treatment effects

H_A: $\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$
And the alternative hypothesis is
H_{A1}: atleast one $\alpha_i \neq \alpha_j$ for all i,j

2. Related to the block effects

H_B: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$
And the alternative hypothesis is
H_{B1}: atleast one$\beta_i \neq \beta_j$ for all i,j

3. Related to the row effects

H_C: $\gamma_1 = \gamma_2 = \ldots = \gamma_p = 0$

And the alternative hypothesis is
H_{C1}: atleast one $\gamma_i \neq \gamma_j$ for all i,j

3 .Least square estimates of the parameters:

The parameters μ, α_i, β_j and γ_k are estimated by the method of least squares. i.e. by minimising error sum of squares. L = $\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \varepsilon_{ijk}^2$ =

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_k)^2$$

solving the normal equations $\frac{\partial L}{\partial \mu} = 0$, $\frac{\partial L}{\partial \alpha_i} = 0$, $\frac{\partial L}{\partial \beta_j} = 0$, $\frac{\partial L}{\partial \gamma_k} = 0$, i ,j,k=1,2…p

we obtain 3p+1 normal equations as ,

$$p^2\mu + \sum_{i=1}^{p} p\alpha_i + \sum_{j=1}^{p} p\beta_j + \sum_{k=1}^{p} p\gamma_k = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}$$

$$p\mu + p\alpha_i + \sum_{j=1}^{p}\beta_j + \sum_{k=1}^{p}\gamma_k = \sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}$$, i= 1,2 …p

And

,

$$p\mu + \sum_{i=1}^{p}\alpha_i + p\beta_j + \sum_{k=1}^{p}\gamma_k = \sum_{i=1}^{p}\sum_{k=1}^{p} y_{ijk}$$, j=1.2,…p

$$p\mu + \sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{p}\beta_j + p\gamma_k = \sum_{i=1}^{p}\sum_{j-1}^{p} y_{ijk}$$

,

These normal equations are not linearly independent, .Adding independent constraint, $\sum_{i=1}^{p}\alpha_i = 0$, $\sum_{j=1}^{p}\beta_j = 0$ $\sum_{k=1}^{p}\gamma_k = 0$ and Solving we get the solutions as

$$\hat{\mu} = \bar{y}..$$

$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$ i=1,2..p

$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{..}$ and

$\hat{\gamma}_k = \bar{y}_{..k} - \bar{y}_{..}$ where

$\overline{y_{i..}}$ average of observations under the i[th] treatments, $\overline{y_{.j.}}$ average of observations under the j[th] Blocks. and $\overline{y_{..k}}$ average of observations under the kth row. $\overline{y_{...}}$ represents grand average of all the observations.

4. **Decomposition of the total sum of squares**

Substituting the values of the estimators μ, α_i, β_j in the model we get

$Y_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \epsilon_{ijk}$

Or

$$Y_{ijk} - \bar{y}_{...} = (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..K} - \bar{y}_{...}) + (y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} - \overline{y}_{..k} + 2\overline{y}_{...}),$$

Squaring both sides and summing over all the observations we get

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk} - \overline{y_{...}})^2 = \sum_{i=1}^{p} p(\overline{y}_{i..} - \overline{y}_{...})^2 + \sum_{j=1}^{p} p(\overline{y}_{.j.} - \overline{y}_{...})^2 + \sum_{j=1}^{p} p(\overline{y}_{..k} - \overline{y}_{...})^2 + \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} - \overline{y}_{..k} + 2\overline{y}_{...})^2$$

and all the cross product vanishes.

OR

SST=SSTR+SSB+SSR+SSE where

SST= $\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk}-\overline{y_{...}})^2$ = $\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}^2 - \frac{y_{...}^2}{p^2}$

SSTR = $\sum_{i=1}^{p} p(\overline{y_{i.}}-\overline{y_{..}})^2$ = $\sum_{i=1}^{p}\frac{y_{i...}^2}{p}-\frac{y_{..}^2}{p^2}$

SSB = $\sum_{j=1}^{p} p(\overline{y_{.j}}-\overline{y_{..}})^2$ = $\sum_{i=1}^{k}\frac{y_{.j.}^2}{p}-\frac{y_{...}^2}{p^2}$

SSR = $\sum_{j=1}^{p} p(\overline{y_{.k}}-\overline{y_{..}})^2$ = $\sum_{i=1}^{k}\frac{y_{..k.}^2}{p}-\frac{y_{...}^2}{p^2}$

And SSE = $\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}(y_{ijk}-\overline{y_{i..}}-\overline{y_{.j.}}-\overline{y_{..k}}+2\overline{y_{...}})^2$

Statistical Analysis

Degrees of freedom of various sum of squares:

The TSS is computed with p^2 quantities, which will carry p^2-1 degrees of freedom, SSTR will have p-1 degrees of freedom, SSB will have p-1 degrees of freedom and SSR will have p-1 degrees of freedom and SSE will carry (p^2-1) – (p-1) –(p-1) –(p-1)= (p-1) (p-2) degrees of freedom

We have assumed that the errors ϵ_{ijk} are normally and independently distributed with mean zero and variance σ^2. The observations y_{ijk} are normally and independently distributed with mean μ + αi+ βj+ γ_k and variance σ^2. Thus SST is a sum of squares in normally distributed random variables hence can be shown that SST/σ^2 is distributed as chi-square with p^2-1 degrees of freedom .and that SSTR /σ^2 is chi - square variate with p-1 degrees of freedom, SSB/σ^2 is chi - square variate with p-1 degrees of freedom and SSR/σ is chi - square variate with p-1 degrees of freedom if the null hypothesis true. Further we can show that SSE/σ2 is chi-square (p-1) x (p-2) degrees of freedom . It also implies that SSTR/ σ^2, SSB/ σ^2, SSR/ σ^2 and SSE/σ^2 independently distributed chi-square random variables. Therefore under H_0.

$$F_A=\frac{SSTR/(p-1)}{SSE/(p-1)(p-2)}=\frac{MSSTR}{MSSE}$$

is distributed as F with p-1 and (p-1)(p-2) degrees of freedom.

$$F_B = \frac{SSB/(p-1)}{SSE/(p-1)(p-2)} = \frac{MSSB}{MSSE}$$ is distributed as F with p-1 and (p-1)(p-2) degrees of freedom.

$$F_C = \frac{SSR/(p-1)}{SSE/(p-1)(p-2)} = \frac{MSSR}{MSSE}$$ is distributed as F with p-1 and (p-1)(p-2) degrees of freedom.

Decision Rule:
Reject H_A , H_B , H_C at α level if $F_{A(cal)}$, $F_{B(cal)}$, $F_{C(cal)} > F_{\alpha,,p-1,\ (p-1)(p-2)}$

5. Analysis of Variance Table:

Sources of variation	Degree of freedom	Sum of squares	Mean sum of squares	F-Value
Treatments	p-1	SSTR	MSSTR = SSTR/(p-1)	FA = MSSTR/ MSSE
Blocks	p-1	SSB	MSSB = SSB/(p-1)	FB = MSS**B**/ MSSE
Rows	p-1	SSR	MSSR = SSR/(P-1)	FB = MSS**R**/ MSSE
Error	(p-1) (p-2)	SSE	MSSE = SSE/(p-1) (p-2)	
total	P^2-1	TSS		

Refer to F table and write the conclusion.

6. **Advantages and disadvantages of LSD:**

Advantages

1. Controls more variation than CRD or RBD designs because of 2-way stratification. Results in a smaller mean square for error.
2. Simple analysis of data
3. Analysis is simple even with missing plots.
4. LSD is an incomplete 3 way layout. Its advantages over the 3 way layout is that instead of p^3 experimental units only p^2 units are needed.
5. More than one factor can be investigates simultaneously and with fewer trials than more complicated designs

Disadvantages of LSD

1. The number of treatments, rows and columns must be the same
2. Number of treatments is limited to the number of replicates which in normal situations difficult to arrive at (exceeds 10 generally).

3. If have less than 5 treatments, the df for controlling random variation is relatively large and the df for error is small.
4. The effect of each treatment must be approximately the same across rows and columns.
5. In an field layout, RBD is much easy to manage than LSD as the former can perform a square or rectangular field easily or field of any shape where as for the later approximately a square field is necessary.

8. Example: An experiment was carried out to determine the effect of claying the ground on the field of barley grains, amount of clay used were as follows:

A. No clay
B. Clay at 100 per acre
C. Clay at 200 per acre
D. Clay at 300 per acre.
E. The yields were in plots of 8 meters and are given in the following table

Column→ Row ↓	I	II	III	IV	Row Total ($y_{...k}$)
I	D 29.1	B 18.9	C 29.4	A 5.7	83.1
II	C 16.4	A 10.2	D 21.2	B 19.1	66.9
III	A 5.4	D 38.8	B 24.0	C 37.0	105.2
IV	B 24.9	C 41.7	A 9.5	D 28.9	105.0
Column Total($y_{.j.}$)	75.8	109.6	84.1	90.7	360.2

Solution:

We assume the three way ANOVA model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \qquad (4)$$

Where μ general effect
α_i – i^{th} treatment effect
βj – j^{th} block effect;
γk – k^{th} row effect;
ϵijk- error term

We test the hypothesis

1. H_A: $\alpha_1 = \alpha_2 = \ldots \alpha_p = 0$

And the alternative hypothesis is

H_{A1}: atleast one $\alpha_i \neq \alpha_j$ for all i,j

2. H_B: $\beta_1 = \beta_2 = \ldots \beta_p = 0$

And the alternative hypothesis is

H_{B1}: atleast one$\beta_i \neq \beta_j$ for all i,j

3. H_C: $\gamma_1 = \gamma_2 = \ldots \gamma_p = 0$

And the alternative hypothesis is

H_{C1}: atleast one $\gamma_i \neq \gamma_j$ for all i,j

Step 1:Calculation of correction factor(CF): $\frac{y_{...}^2}{p^2}$ = $360.2^2/16 = 8109.0025$

Step;2: Calculation of Total sum of squares: $\sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} y_{ijk}^2 - CF$ =

$(29.1^2+18.9^2+\ldots+28.9^2)$-CF =1943.0775

Step;3: Calculation of Treatment sum of squares= $\sum_{i=1}^{p} \frac{y_{i..}^2}{p} - \frac{y_{...}^2}{p^2}$ =

$(30.8^2+86.9^2+124.5^2+118.0^2)/4$ –CF = 1372.1225

(Total(A) =30.8, **Total(B) =**86.9, **Total(C) =124.5, Total(D) =118.0)**

Step;4: Calculation of Block sum of squares= $\sum_{j=1}^{p} \frac{y_{.j.}^2}{p} - \frac{y_{...}^2}{p^2}$ =

$(75.8^2+109.6^2+84.1^2+90.7^2)/4$ –CF = 155.2725

Step;5: Calculation of Row sum of squares= $\sum_{k=1}^{p} \frac{y_{..k}^2}{p} - \frac{y_{...}^2}{p^2}$ =

$(83.1^2+66.9^2+105.2^2+105.0^2)/4$ –CF = 259.3125

Step;6: Calculation of Error sum of squares

ESS = SST- SSTR-SSB-SSR = 1943.0775-1372.1225-155.2725-259.3125 = 156.37

Anova Table:

Table value for F_α is $F_{.05, 3,18} = 3.16$

Sources of variation	Degree of freedom	Sum of squares	Mean sum of squares	F-Value
Treatments	p-1=3	1372.1225	457.3742	457.3742/26.0616 =17.55
Blocks	p-1=3	155.2725	51.7575	51.7575/26.0616 =1.98
Rows	p-1=3	259.3125	86.4375	86.4375/26.0616 =3.32
Error	(p-1) (p-2)=6	156.37	26.0616	
total	P^2-1=15	1943.0775		

Table value: $F_{.05, 3,6} = 4.76$

Since $F_{A\ cal}(17.55) > F_{.05, 3,6}$ we reject H_A. Hence we conclude that all treatment means are not equal

Since $F_{B\ cal}(1.98) < F_{.05, 3,6}$ we do not reject H_B. Hence we conclude that all block means are equal

Since $F_{c\ cal}(3.32) < F_{.05, 3,6}$ we do not reject H_C. Hence we conclude that all row means are equal.

Hence we can conclude that the variation due to rows and columns are not significant but the different levels of clay have significant effect on the yield.

9. Conclusion:

In this session we discussed the definition and layout of LSD. We defined the model and hypothesis to be tested under LSD. We derived the least square estimated of parameters of the model. We have discussed how the total variation split into different component parts. We discussed the analysis of LSD . The entire analysis is summarised in ANOVA Table. We discussed the advantages and disadvantages of LSD. WE discussed the entire analysis of LSD with the help of solving an example.