Analysis of Variance and Design of Experiments

Experimental Designs and Their Analysis

:::

Lecture 21

Analysis in Latin Square Design and Missing Plot Technique

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Slides can be downloaded from http://home.iitk.ac.in/~shalab/sp

Latin Square:

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

For example, to write a Latin square of order 4,

- **choose four Latin letters as symbols – A, B, C and D.**
- **Write them in a way such that each of the letters out of A, B, C and D occurs once and only once in each row and each column.**

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Analysis of LSD (one observation per cell):

In designing a LSD of order *p*,

- **choose one Latin square at random from the set of all possible Latin squares of order *p*.**
- **Select a standard Latin square from the set of all standard Latin squares with equal probability.**
- **Randomize all the rows and columns as follows:**
  - ❑ **Choose a random number, less than *p*, say n_1 and then 2nd row is the n_1^{th} row.**
  - ❑ **Choose another random number less than *p*, say n_2 and then 3rd row is the n_2^{th} row and so on.**
  - ❑ **Then do the same for column.**

Analysis of LSD (one observation per cell):

For Latin squares of order less than 5, fix first row and then randomize rows and then randomize columns.

In Latin squares of order 5 or more, need not to fix even the first row. Just randomize all rows and columns.

Analysis of LSD (one observation per cell): Example

Suppose following Latin square is chosen

A	B	C	D	E
B	C	D	E	A
D	E	A	B	C
E	A	B	C	D
C	D	E	A	B

Now randomize rows, e.g., 3rd row becomes 5th row and 5th row becomes 3rd row. The Latin square becomes

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
E	A	B	C	D
D	E	A	B	C

Analysis of LSD (one observation per cell): Example

Now randomize columns, say 5th column becomes 1st column, 1st column becomes 4th column and 4th column becomes 5th column

E	B	C	A	D
A	C	D	B	E
D	A	B	E	C
C	E	A	D	B
B	D	E	C	A

Now use this Latin square for the assignment of treatments.

Analysis of LSD (one observation per cell):

y_{ijk} : **Observation on *k*th treatment in *i*th row and *j*th block,**

$i = 1, 2, \ldots, v,\ \ j = 1, 2, \ldots, v,\ \ k = 1, 2, \ldots, v.$

Triplets (*i, j, k*) take on only the v^2 values indicated by the chosen particular Latin square selected for the experiment.

y_{ijk}**'s are independently distributed as** $N(\mu + \alpha_i + \beta_j + \tau_k, \sigma^2)$.

Analysis of LSD (one observation per cell):

Linear model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk}, i = 1, 2, \ldots, v; j = 1, 2, \ldots, v; k = 1, 2, \ldots, v$$

where ε_{ijk} **are random errors which are identically and independently distributed following** $N(0, \sigma^2)$ **with**

$$\sum_{i=1}^{v} \alpha_i = 0, \quad \sum_{j=1}^{v} \beta_j = 0, \quad \sum_{k=1}^{v} \tau_k = 0,$$

α_i : **main effect of rows**

β_j : **main effect of columns**

τ_k : **main effect of treatments.**

Analysis of LSD (one observation per cell):

The null hypothesis under consideration are

$$H_{0R} : \alpha_1 = \alpha_2 = ... = \alpha_v = 0$$

$$H_{0C} : \beta_1 = \beta_2 = ... = \beta_v = 0$$

$$H_{0T} : \tau_1 = \tau_2 = ... = \tau_v = 0.$$

Analysis of LSD (one observation per cell):

The analysis of variance can be developed on the same lines as earlier.

Minimizing $S = \sum_{i=1}^{v} \sum_{j=1}^{v} \sum_{k=1}^{v} \varepsilon_{ijk}^2$ **with respect to** μ, α_i, β_j **and** τ_k **given the least squares estimate as**

$$\hat{\mu} = \bar{y}_{ooo}$$

$$\hat{\alpha}_i = \bar{y}_{ioo} - \bar{y}_{ooo} \qquad i = 1, 2, ..., v$$

$$\hat{\beta}_j = \bar{y}_{ojo} - \bar{y}_{ooo} \qquad j = 1, 2, ..., v$$

$$\hat{\tau}_k = \bar{y}_{ook} - \bar{y}_{ooo} \qquad k = 1, 2, ..., v.$$

Analysis of LSD (one observation per cell):

Using the fitted model based on these estimators, the total sum of squares can be partitioned into the mutually orthogonal sum of squares *SSR, SSC, SSTr* and *SSE* as

$$TSS = SSR + SSC + SSTr + SSE$$

where

***TSS*: Total sum of squares** $\sum_{i=1}^{v}\sum_{j=1}^{v}\sum_{k=1}^{v}(y_{ijk}-\bar{y}_{ooo})^2 = \sum_{i=1}^{v}\sum_{j=1}^{v}\sum_{k=1}^{v}y_{ijk}^2 - \frac{G^2}{v^2}$

***SSR*: Sum of squares due to rows =** $v\sum_{i=1}^{v}(\bar{y}_{ioo}-\bar{y}_{ooo})^2 = \frac{\sum_{i=1}^{v}R_i^2}{v} - \frac{G^2}{v^2};$

where $R_i = \sum_{j=1}^{v}\sum_{k=1}^{v}y_{ijk}.$

Analysis of LSD (one observation per cell):

***SSC*: Sum of squares due to column** $= v\sum_{j=1}^{v}(\bar{y}_{ojo} - \bar{y}_{ooo})^2 = \frac{\sum_{j=1}^{v} C_j^2}{v} - \frac{G^2}{v^2};$

where $C_j = \sum_{i=1}^{v}\sum_{k=1}^{v} y_{ijk}.$

***SSTr* : Sum of squares due to treatment** $= v\sum_{k=1}^{v}(\bar{y}_{ook} - \bar{y}_{ooo})^2 = \frac{\sum_{k=1}^{v} T_k^2}{v} - \frac{G^2}{v^2};$

where $T_k = \sum_{i=1}^{v}\sum_{j=1}^{v} y_{ijk}.$

Degrees of freedom carried by *SSR, SSC* and *SSTr* are (v - 1) each.

Degrees of freedom carried by *TSS* are $v^2 - 1$.

Degrees of freedom carried by *SSE* are (v - 1)(v - 2).

Analysis of LSD (one observation per cell):

Thus

- **under** H_{0R}, $F_R = \dfrac{MSR}{MSE} \sim F((v-1),(v-1)(v-2))$
- **under** H_{0C}, $F_C = \dfrac{MSC}{MSE} \sim F((v-1),(v-1)(v-2))$
- **under** H_{0T}, $F_T = \dfrac{MSTr}{MSE} \sim F((v-1),(v-1)(v-2)).$

Decision rules:

Reject H_{0R} **at level** α **if** $F_R > F_{1-\alpha;(v-1),(v-1)(v-2)}$

Reject H_{0C} **at level** α **if** $F_C > F_{1-\alpha;(v-1),(v-1)(v-2)}$

Reject H_{0T} **at level** α **if** $F_T > F_{1-\alpha;(v-1),(v-1)(v-2)}.$

If any null hypothesis is rejected, then use multiple comparison test.

Analysis of LSD (one observation per cell): ANOVA Table

The analysis of variance table is as follows

Source of variation	Degrees of freedom	Sum of squares	Mean squares	F - value
Rows	$v - 1$	SSR	MSR	F_R
Columns	$v - 1$	SSC	MSC	F_C
Treatments	$v - 1$	$SSTr$	$MSTr$	F_T
Error	$(v - 1)(v - 2)$	SSE	MSE	
Total	$v^2 - 1$	TSS		

Analysis of LSD (one observation per cell):

The expectations of mean squares are obtained as

$$E(MSR) = E\left(\frac{SSR}{v-1}\right) = \sigma^2 + \frac{v}{v-1}\sum_{i=1}^{v}\alpha_i^2$$

$$E(MSC) = E\left(\frac{SSC}{v-1}\right) = \sigma^2 + \frac{v}{v-1}\sum_{j=1}^{v}\beta_j^2$$

$$E(MSTr) = E\left(\frac{SSTr}{v-1}\right) = \sigma^2 + \frac{v}{v-1}\sum_{k=1}^{v}\tau_k^2$$

$$E(MSE) = E\left(\frac{SSE}{(v-1)(v-2)}\right) = \sigma^2.$$

Missing plot techniques:

It happens many time in conducting the experiments that some observation are missed.

This may happen due to several reasons.

For example, in a clinical trial, suppose the readings of blood pressure are to be recorded after 3 days of giving the medicine to the patients. Suppose the medicine is given to 20 patients and one of the patients doesn't turn up for providing the reading.

Similarly, in an agricultural experiment, the seeds are sown and yields are to be recorded after few months. Suppose some cattle destroy the crop of any plot or the crop of any plot is destroyed due to storm, insects etc.

Missing plot techniques:

In such cases, one option is to

- somehow estimate the missing value on the basis of available data,
- replace it back in the data and make the data set complete.

Now conduct the statistical analysis on the basis of completed data set as if no value was missing by making necessary adjustments in the statistical tools to be applied.

Such an area comes under the purview of "missing data models" and a lot of development has taken place.

Missing plot techniques:

We discuss here the classical missing plot technique proposed by Yates which involve the following steps:

- **Estimate the missing observations by the values which makes the error sum of squares to be minimum.**
- **Substitute the unknown values by the missing observations.**
- **Express the error sum of squares as a function of these unknown values.**
- **Minimize the error sum of squares using the principle of maxima/minima, i.e., differentiating it with respect to the missing value and put it to zero and form a linear equation.**

Missing plot techniques:

- **Form as much linear equation as the number of unknown values (i.e., differentiate the error sum of squares with respect to each unknown value).**
- **Solve all the linear equations simultaneously and solutions will provide the missing values.**
- **Impute the missing values with the estimated values and complete the data.**
- **Apply analysis of variance tools.**
- **The error sum of squares thus obtained is corrected but the treatment sum of squares is not corrected.**

Missing plot techniques:

- **The number of degrees of freedom associated with the total sum of squares is subtracted by the number of missing values and adjusted in the error sum of squares.**
- **No change in the degrees of freedom of sum of squares due to treatment is needed.**

Missing observations in RBD: One missing observation

Suppose one observation in $(i, j)^{th}$ cell is missing and let this be x.

The arrangement of observations in RBD then will be as follows:

	Treatments (Factor B)							Block totals
		1	2		j		v	
Blocks (Factor A)	1	y_{11}	y_{12}	…	y_{1j}	…	y_{1v}	B_1
	2	y_{21}	y_{22}	…	y_{2j}	…	y_{2v}	B_2
	. . .	. . .	. . .		. . .		. . .	. . .
	i	y_{i1}	y_{i2}	…	$y_{ij} = x$	…	y_{iv}	$B_i = y'_{io} + x$
	. . .		. . .		. . .		. . .	. . .
	b	y_{b1}	y_{b2}	…	y_{bj}	…	y_{bv}	B_b
Treatment totals		T_1	T_2	…	$T_j = y'_{oj} + x$	…	T_v	Grand total $G = y'_{oo} + x$

where

y'_{oo} : total of known observations

y'_{io} : total of known observations in i^{th} block.

y'_{oj} : total of known observations in j^{th} treatment.

Missing observations in RBD: One missing observation

Correction factor $(CF) = \dfrac{(G')^2}{n} = \dfrac{(y'_{oo} + x)^2}{bv}$

$$TSS = \sum_{i=1}^{b} \sum_{j=1}^{v} y_{ij}^2 - CF$$

$$= (x^2 + \text{ terms which are constant with respect to } x) - CF$$

$$SSBl = \frac{1}{v}[(y'_{io} + x)^2 + \text{ terms which are constant with respect to } x] - CF$$

$$SSTr = \frac{1}{b}[(y'_{oj} + x)^2 + \text{ terms which are constant with respect to } x] - CF$$

$$SSE = TSS - SSBl - SSTr$$

$$= x^2 - \frac{1}{v}(y'_{io} + x)^2 - \frac{1}{b}(y'_{oj} + x)^2 + \frac{(y'_{oo} + x)^2}{bv}$$

$$+ \text{ (terms which are constant with respect to } x) - CF.$$

Missing observations in RBD: One missing observation

Find *x* such that *SSE* is minimum

$$\frac{\partial(SSE)}{\partial x} = 0 \Rightarrow 2x - \frac{2(y_{io}^{'} + x)}{v} - \frac{2(y_{oj}^{'} + x)}{b} + \frac{2(y_{oo}^{'} + x)}{bv} = 0$$

or $$x = \frac{by_{io}^{'} + vy_{oj}^{'} - y_{oo}^{'}}{(b-1)(v-1)}$$

The second-order derivative condition for *x* to provide minimum *SSE* can be easily verified.

Missing observations in RBD: Two missing observations

If there are two missing observation, then let they be *x* and *y*.

- **Let the corresponding row sums (block totals) are** $(R_1 + x)$ **and** $(R_2 + y).$

- **Column sums (treatment totals) are** $(C_1 + x)$ **and** $(C_2 + y).$

- **Total of known observations is *S*.**

Then

$$SSE = x^2 + y^2 - \frac{1}{v}[(R_1 + x)^2 + (R_2 + y)^2] - \frac{1}{b}[(C_1 + x)^2 + (C_2 + y)^2]$$

$$+ \frac{1}{bv}(S + x + y)^2 + \text{ terms independent of } x \text{ and } y.$$

Missing observations in RBD: Two missing observations

Now differentiate *SSE* with respect to *x* and *y*, as

$$\frac{\partial(SSE)}{\partial x} = 0 \Rightarrow x - \frac{R_1 + x}{v} - \frac{C_1 + x}{b} + \frac{S + x + y}{bv} = 0$$

$$\frac{\partial(SSE)}{\partial y} = 0 \Rightarrow y - \frac{R_2 + y}{v} - \frac{C_2 + y}{b} + \frac{S + x + y}{bv} = 0.$$

Thus solving the following two linear equations in *x* and *y*, we obtain the estimated missing values

$$(b-1)(v-1)x = bR_1 + vC_1 - S - y$$

$$(b-1)(v-1)y = bR_2 + vC_2 - S - x.$$

Adjustments to be done in analysis of variance:

i. **Obtain the within block sum of squares from incomplete data.**
ii. **Subtract correct error sum of squares from (i). This gives the correct treatment sum of squares.**
iii. **Reduce the degrees of freedom of error sum of squares by the number of missing observations.**
iv. **No adjustments in other sums of squares are required.**

Missing observations in LSD:

Let

- **x be the missing observation in (i, j, k)th cell, i.e.**

$$y_{ijk}, i = 1, 2, .., v, \ j = 1, 2, .., v, k = 1, 2, .., v.$$

- ***R*: Total of known observations in i^{th} row**
- ***C*: Total of known observations in j^{th} column**
- ***T*: Total of known observation receiving the k^{th} treatment.**
- ***S*: Total of known observations**

Missing observations in LSD:

Now

Correction factor $(CF) = \dfrac{(S+x)^2}{v^2}$

Total sum of squares $(TSS) = x^2$ **+ term which are constant with respect to** x **-** CF

Row sum of squares $(SSR) = \dfrac{(R+x)^2}{v}$ **+ term which are constant with respect to** x **-** CF

Column sum of squares $(SSC) = \dfrac{(C+x)^2}{v}$ **+ term which are constant with respect to** x **-** CF

Treatment sum of squares$(SSTr) = \dfrac{(T+x)^2}{v}$ **+ term which are constant with respect to** x **-** CF

Sum of squares due to error $(SSE) \;\; = TSS \;-\; SSR \;-\; SSC \;-\; SSTr$

$$= x^2 - \frac{1}{v}\left[(R+x)^2 + (C+x)^2 + (T+x)^2\right] + \frac{2(S+x)^2}{v^2}$$

Missing observations in LSD:

Choose *x* such that *SSE* is minimum. So

$$\frac{d(SSE)}{dx} = 0$$

$$\Rightarrow 2x - \frac{2}{v}\left(R + C + T + 3x\right) + \frac{4(S + x)}{v^2} = 0$$

or $x = \dfrac{V(R + C + T) - 2S}{(v-1)(v-2)}$

Adjustment to be done in analysis of variance:

Do all the steps as in the case of RBD.

To get the correct treatment sum of squares, proceed as follows:

- **Ignore the treatment classification and consider only row and column classification.**
- **Substitute the estimated values at the place of missing observation.**
- **Obtain the error sum of squares from complete data, say SSE_1 .**
- **Let SSE_2 be the error sum of squares based on LSD obtained earlier.**
- **Find the corrected treatment sum of squares = SSE_2 - SSE_1**
- **Reduce of degrees of freedom of error sum of squares by the number of missing values.**