Analysis of Variance and Design of Experiments

Experimental Designs and Their Analysis

:::

Lecture 18

Completely Randomized Design

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

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

Completely randomized design (CRD)

The CRD is the simplest design. Suppose there are *v* treatments to be compared.

- **All experimental units are considered the same and no division or grouping among them exist.**
- **In CRD, the *v* treatments are allocated randomly to the whole set of experimental units, without making any effort to group the experimental units in any way for more homogeneity.**
- **Design is entirely flexible in the sense that any number of treatments or replications may be used.**

Completely randomized design (CRD)

- **The number of replications for different treatments need not be equal and may vary from treatment to treatment depending on the knowledge (if any) on the variability of the observations on individual treatments as well as on the accuracy required for the estimate of individual treatment effect.**

Completely randomized design (CRD)

Example:

Suppose there are 4 treatments and 20 experimental units, then

- ✓ **the treatment 1 is replicated, say 3 times and is given to 3 experimental units,**
- ✓ **the treatment 2 is replicated, say 5 times and is given to 5 experimental units,**
- ✓ **the treatment 3 is replicated, say 6 times and is given to 6 experimental units and**
- ✓ **finally, the treatment 4 is replicated [20 - (6 + 5 + 3) =]6 times and is given to the remaining 6 experimental units.**

Completely randomized design (CRD)

- **All the variability among the experimental units goes into experimented error.**
- **CRD is used when the experimental material is homogeneous.**
- **CRD is often inefficient.**
- **CRD is more useful when the experiments are conducted inside the lab.**
- **CRD is well suited for the small number of treatments and for the homogeneous experimental material.**

Completely randomized design (CRD): Layout of CRD

Following steps are needed to design a CRD:

- **Divide the entire experimental material or area into a number of experimental units, say *n*.**
- **Fix the number of replications for different treatments in advance (for given total number of available experimental units).**
- **No local control measure is provided as such except that the error variance can be reduced by choosing a homogeneous set of experimental units.**

Completely randomized design (CRD) : Procedure

Let the v treatments are numbered from 1,2,...,v and

n_i be the number of replications required for i^{th} treatment such that $\sum_{i=1}^{v} n_i = n.$

- **Select n_1 units out of n units randomly and apply treatment 1 to these n_1 units.**

  (Note: This is how the randomization principle is utilized is CRD.)

- **Select n_2 units out of (n - n_1) units randomly and apply treatment 2 to these n_2 units.**

Completely randomized design (CRD) : Procedure

- **Continue with this procedure until all the treatments have been utilized.**

- **Generally, the equal number of treatments are allocated to all the experimental units unless no practical limitation dictates or some treatments are more variable or/and of more interest.**

Completely randomized design (CRD): Analysis

There is only one factor which is affecting the outcome – treatment effect. So the set-up of one-way analysis of variance is to be used.

y_{ij} **: Individual measurement of** j^{th} **experimental units for** i^{th} **treatment** $i = 1,2,...,v$, $j = 1,2,...,n_i$.

y_{ij} **: Independently distributed following** $N(\mu + \alpha_i, \sigma^2)$ **with** $\sum_{i=1}^{v} n_i \alpha_i = 0.$

μ **: overall mean**

α_i **:** i^{th} **treatment effect**

Completely randomized design (CRD): Analysis

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_v = 0$$

***H*₁** **: All $\alpha_i^{'}s$ are not equal.**

The data set is arranged as follows:

Treatments

1	2	...	v
y_{11}	y_{21}	$\cdots$	y_{v1}
y_{12}	y_{22}	$\cdots$	y_{v2}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_{1n_1}	y_{2n_2}	$\cdots$	y_{vn_v}
T_1	T_2	...	T_v

where $T_i = \sum_{j=1}^{n_i} y_{ij}$ **is the treatment total due to *i*th effect,**

$G = \sum_{i=1}^{v} T_i = \sum_{i=1}^{v} \sum_{j=1}^{n_i} y_{ij}$ **is the grand total of all the observations.**

Completely randomized design (CRD): Analysis

In order to derive the test for H_0, we can use either the likelihood ratio test or the principle of least squares.

Since the likelihood ratio test has already been derived earlier, so we choose to demonstrate the use of the least-squares principle.

The linear model is $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, 2, ..., v, \; j = 1, 2, ..., n_i$

where ε_{ij} 's are identically and independently distributed random errors with mean 0 and variance σ^2.

The normality assumption of $\varepsilon's$ is not needed for the estimation of parameters but will be needed for deriving the distribution of various involved statistics and in deriving the test statistics.

Completely randomized design (CRD): Analysis

Let

$$S = \sum_{i=1}^{v}\sum_{j=1}^{n_i}\varepsilon_{ij}^2 = \sum_{i=1}^{v}\sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha_i)^2.$$

Minimizing *S* with respect to μ and α_i, we get the normal equations

$$\frac{\partial S}{\partial \mu} = 0 \Rightarrow -2\sum_{i=1}^{v}\sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha_i) = 0$$

or $$n\mu + \sum_{i=1}^{v} n_i\alpha_i = \sum_{i=1}^{v}\sum_{j=1}^{n_i} y_{ij}$$

$$\frac{\partial S}{\partial \alpha_i} = 0 \Rightarrow -2\sum_{j=1}^{n_i}(y_{ij}-\mu-\alpha_i) = 0, i = 1,2,...,v$$

or $$n_i\mu + n_i\alpha_i = \sum_{j=1}^{n_i} y_{ij}$$

Completely randomized design (CRD): Analysis

Solving them using $\sum_{i=1}^{v} n_i \alpha_i = 0$**, we get**

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

$$\hat{\alpha}_i = \bar{y}_{io} - \bar{y}_{oo}$$

where

$\bar{y}_{io} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ **is the mean of observation receiving the *i*th treatment**

and

$\bar{y}_{oo} = \frac{1}{n}\sum_{i=1}^{v}\sum_{j=1}^{n_i} y_{ij}$ **is the mean of all the observations.**

Completely randomized design (CRD): Analysis

The fitted model is obtained after substituting the estimate $\hat{\mu}$ and $\hat{\alpha}_i$ in the linear model, we get

$$y_{ij} = \bar{y}_{oo} + (\bar{y}_{io} - \bar{y}_{oo}) + (y_{ij} - \bar{y}_{io})$$

or $$(y_{ij} - \bar{y}_{oo}) = (\bar{y}_{io} - \bar{y}_{oo}) + (y_{ij} - \bar{y}_{io}).$$

Squaring both sides and summing over all the observation, we have

$$\sum_{i=1}^{v}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{oo})^2 = \sum_{i=1}^{v} n_i(\bar{y}_{io} - \bar{y}_{oo})^2 + \sum_{i=1}^{v}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{io})^2$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

or (Total sum of squares) = (Sum of squares due to treatment effects) + (Sum of squares due to error)

or $$TSS = SSTr + SSE$$

Completely randomized design (CRD): Analysis

- **Since** $\sum_{i=1}^{v}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{oo})=0,$ **so *TSS* is based on the sum of (*n* -1) squared quantities. Thus *TSS* carries only (*n* - 1) degrees of freedom.**

- **Since** $\sum_{i=1}^{v} n_i(\bar{y}_{io}-\bar{y}_{oo})=0,$ **so *SSTr* is based only on the sum of (*v* -1) squared quantities. Thus *SSTr* carries only (*v* - 1) degrees of freedom.**

- **Since** $\sum_{i=1}^{n_i} n_i(y_{ij}-\bar{y}_{io})=0$ **for all *i* = 1,2,...,*v*, so *SSE* is based on the sum of squaring *n* quantities like** $(y_{ij}-\bar{y}_{io})$ **with *v* constraints** $\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{io})=0,$ **so SSE carries (*n* – *v*) degrees of freedom.**

Completely randomized design (CRD): Analysis

- **Using the Fisher-Cochran theorem,**

  TSS* = *SSTr* + *SSE **with degrees of freedom partitioned as**

  (n – 1) = (v - 1) + (n – v).

Moreover, equality in *TSS* = *SSTr* + *SSE* has to hold exactly.

To ensure that the equality holds exactly, we find one of the sums of squares through subtraction.

Generally, it is recommended to find SSE by subtraction as

***SSE* = *TSS* - *SSTr* .**

Completely randomized design (CRD):Analysis

$$TSS = \sum_{i=1}^{v}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{io})^2$$

$$= \sum_{i=1}^{v}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n}$$

$G = \sum_{i=1}^{v}\sum_{j=1}^{n_i} y_{ij}$ **: Grand total,** $\frac{G^2}{n}$ **: correction factor**

$$SSTr = \sum_{j=1}^{n_i} n_i(\bar{y}_{io} - \bar{y}_{oo})^2$$

$$= \sum_{i=1}^{v}\left(\frac{T_i^2}{n_i}\right) - \frac{G^2}{n}$$

$T_i = \sum_{j=1}^{n_i} y_{ij}$ **Treatment total.**

Completely randomized design (CRD): Analysis

Now under $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_v = 0$ **, the model become**

$$Y_{ij} = \mu + \varepsilon_{ij},$$

and minimizing $S = \sum_{i=1}^{v} \sum_{j=1}^{n_i} \varepsilon_{ij}^2$

with respect to μ **gives** $\frac{\partial S}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \frac{G}{n} = \bar{y}_{oo}.$

The ***SSE*** **under** H_0 **becomes** $SSE = \sum_{i=1}^{v} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{oo})^2$

and thus ***TSS*** **=** ***SSE.*** **This** ***TSS*** **under** H_0 **contains the variation only due to the random error whereas the earlier** ***TSS*** **=** ***SSTr*** **+** ***SSE*** **contains the variation due to treatments and errors both.**

Completely randomized design (CRD): Analysis

This *TSS* under *TSS* = *SSE* contains the variation only due to the random error whereas the earlier *TSS* = *SSTr* + *SSE* contains the variation due to treatments and errors both.

The difference between the two will provides the effect of treatments in terms of sum of squares as

$$SSTr = \sum_{i=1}^{v} n_i (\overline{y}_i - \overline{y}_{oo})^2$$

Completely randomized design (CRD): Distributions and decision rules:

Using the normal distribution property of ε_{ij}'s, we find that *y_{ij}'s* are also normal as they are the linear combination of ε_{ij}'s.

Under *H_0*

$$-\ \frac{SSTr}{\sigma^2} \sim \chi^2(v-1)$$

$$-\ \frac{SSE}{\sigma^2} \sim \chi^2(n-v).$$

***SSTr* and *SSE* are independently distributed.**

Under *H_0*, $\dfrac{MStr}{MSE} \sim F(v-1, n-v)$

Reject *H_0* at α^* level of significance if $F > F_{\alpha^*, v-1, n-v}$.

[Note: We denote the level of significance here by α^* because α has been used for denoting the factor]

Completely randomized design (CRD): ANOVA table

The analysis of variance table for CRD is

Source of variation	Degrees of freedom	Sum of squares	Mean squares	*F* - value
Between treatments	$v-1$	*SSTr*	*MSTr*	$\frac{MSTr}{MSE}$
Within populations	$n-v$	*SSE*	*MSE*	
Total	$n-1$	*TSS*		

Completely randomized design (CRD): Expectations

$$E(SSE) = \sum_{i=1}^{v} \sum_{j=1}^{n_i} E(y_{ij} - y_{io})^2$$

$$= \sum_{i=1}^{v} \sum_{j=1}^{n_i} E(\varepsilon_{ij} - \overline{\varepsilon}_{io})^2$$

$$= \sum_{i=1}^{v} \sum_{j=1}^{n_i} E(\varepsilon_{ij}^2) - \sum_{i=1}^{v} n_i E(\overline{\varepsilon}_{io.}^2)$$

$$= n\sigma^2 - \sum_{i=1}^{v} n_i \frac{\sigma^2}{n_i}$$

$$= (n - v)\sigma^2$$

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

Completely randomized design (CRD): Expectations

$$E(SSTr) = \sum_{i=1}^{v} n_i E(\overline{y}_{io} - \overline{y}_{oo})^2$$

$$= \sum_{i=1}^{v} n_i E(\alpha_i + \overline{\varepsilon}_{io} - \overline{\varepsilon}_{oo})^2$$

$$= \sum_{i=1}^{v} n_i \alpha_i^2 + \left[\sum_{i=1}^{v} n_i E(\overline{\varepsilon}_{io}^2) - nE(\overline{\varepsilon}_{oo}^2) \right]$$

$$= \sum_{i=1}^{v} n_i \alpha_i^2 + \left[\sum_{i=1}^{v} n_i \frac{\sigma^2}{n_i} - n\frac{\sigma^2}{n} \right]$$

$$= \sum_{i=1}^{v} n_i \alpha_i^2 + (v-1)\sigma^2$$

$$E(MSTr) = E\left(\frac{SStr}{v-1}\right) = \frac{1}{v-1}\sum_{i=1}^{v} n_i \alpha_i^2 + \sigma^2.$$

In general $E\left(MSTr\right) \neq \sigma^2$

but under H_0, **all** $\alpha_i = 0$ **and so** $E(MSTr) = \sigma^2$