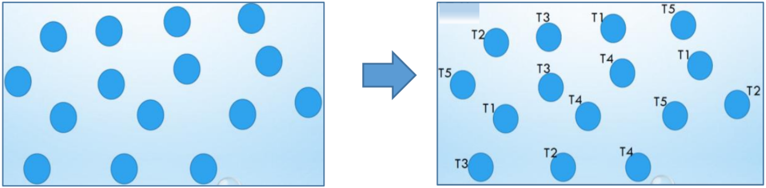
Completely Randomized Design

Norberto E. Milla, Jr.

2025-08-13

Introduction

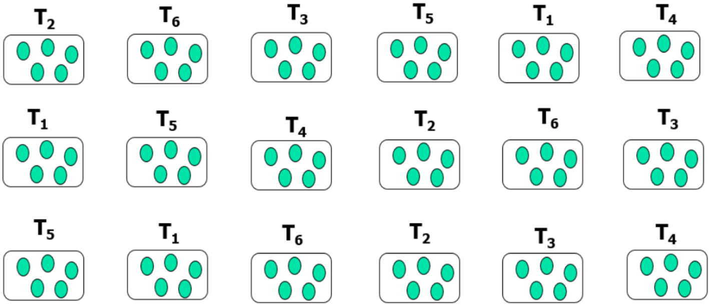
- Simplest experimental design
- Frequently used to compare treatments when environmental conditions are fairly uniform (*homogeneous experimental units*)
- Each treatment is applied at *random* to several experimental units



Introduction

- Advantages:
  - design is extremely easy to construct
  - design is easy to analyze even if the sample sizes might not be the same for each treatment
  - design can be used for any number of treatments
- Disadvantages:
  - it is best suited for situations in which there are relatively few treatments
  - experimental units to which treatments are applied must be as homogeneous as possible
  - if the experimental units are not homogeneous, experimental error would be large and difficult to detect treatment effect
  - suitable only for laboratory and greenhouse experiments where the environment and other conditions can be controlled

Experimental layout



Data Format

Treatment	Replication (Observation)						Treatment Total $y_{i.}$	Treatment Mean $\bar{y}_{i.}$
	1	2	…	j	…	r_i		
1	y_{11}	y_{12}	…	y_{1j}	…	y_{1n_1}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2n_2}	$y_{2.}$	$\bar{y}_{2.}$
.	.	.		.			.	.
.	.	.	…	.	…		.	.
.	.	.		.			.	
i	y_{i1}	y_{i2}		y_{ij}		y_{in_i}	$y_{i.}$	$\bar{y}_{i.}$
.	.	.		.		.	.	.
.	.	.	…	.	…	.	.	
.	.	.		.		.	.	.
p	y_{p1}	y_{p2}		y_{pj}		y_{pn_p}	$y_{p.}$	$\bar{y}_{p.}$
Total							$y_{..}$	$\bar{y}_{..}$

Linear models

Means model

$$Y_{ij} = \mu_i + \epsilon_{ij},\ i = 1, 2, \cdots, t;\ j = 1, 2, \cdots, r_t$$

where:

- Y_{ij} is the response of the j^{th} unit given treatment i
- μ_i is the mean response of units given treatment i
- ϵ_{ij} is the random error component associated with each Y_{ij}, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$

Linear models

Effects model

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, i = 1, 2, \ldots, t; j = 1, 2, \ldots, r_i; \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

where:

- Y_{ij} is the response of the j^{th} unit given treatment i
- μ is the mean response of all units
- τ_i is the effect of treatment i, $\tau_i = \mu_i - \mu$
- ϵ_{ij} is the random error component associated with each Y_{ij}, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$

Estimation of parameters

- the random errors ϵ_{ij} are estimated by

$$e_{ij} = Y_{ij} - \bar{y}_{i.}$$

- minimize sum of squared errors e_{ij} (**Method of least squares**)

$$\sum_{i=1}^{t}\sum_{j=1}^{r_t}\epsilon_{ij}^2 = \sum_{i=1}^{t}\sum_{j=1}^{r_t}(Y_{ij} - \mu - \tau_i)^2$$

- apply partial differentiation to get the estimators that would minimize the above expression

Analysis of fixed effects model

Recall

- The t treatments are deliberately chosen by the researcher or are the only available treatments
- Conclusions apply only to the treatments considered
- Treatment effects τ_i are considered fixed
- Goal: test of equality of treatment means or treatment effects
- Hypotheses:
  - $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$ versus $H_1 : \mu_i \neq \mu_j$, for at least one pair $i \neq j$
  - $H_0 : \tau_i = 0, \forall i$ versus $H_1 : \tau_i \neq 0, \exists i$

Analysis of fixed effects model

Analysis of Variance (ANOVA)

- A method of partitioning the total variance in the response into different components which can be attributed to different sources
  - Systematic variation- effect of manipulated factors
  - Random variation- experimental error
- Method of comparing the effect of treatments on the response variable
- Method of comparing the means of three or more treatments
- Partitioning the total variance observed in the response:

$$\underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_{..})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{t}(\bar{y}_{i.}-\bar{y}_{..})^2}_{SSTr} + \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_{i.})^2}_{SSE}$$

Analysis of variance (ANOVA)

- SSTr is also referred to as **SSB** for sum of squares treatment and SSE as **SSW** for sum of square for experimental error
- When each of the sums of squares is divided by its corresponding degree of freedom we obtain the Mean Square (*MS*) which is a variance

$$S_T^2 = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r_i}(y_{ij} - \bar{y}_{..})^2}{n-1}$$

$$MSTr = S_B^2 = \frac{\sum_{i=1}^{t}(\bar{y}_{i.} - \bar{y}_{..})^2}{t-1}$$

$$MSE = S_W^2 = \frac{\sum_{j=1}^{r_i}(y_{ij} - \bar{y}_{i.})^2}{n-t}$$

- hence, the name Analysis of Variance

Analysis of variance (ANOVA): Test statistic

Recall

- *Theorem*: Let $Y_1, Y_2, \cdots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{\sigma^2}$$

  has a χ^2 distribution with $n-1$ degrees of freedom

- *Theorem*: Suppose U and V are independent chi-square random variables with m and n degrees of freedom, respectively. Then

$$\frac{U/m}{V/n}$$

  has an F distribution with m and n degrees of freedom

Analysis of variance (ANOVA): Test statistic

- $MSTr = \frac{\sum_{i=1}^{t}(\bar{y}_{i.}-\bar{y}_{..})^2}{t-1}$ has a χ^2 distribution with $t-1$ degrees of freedom
- $MSE = \frac{\sum_{j=1}^{r_i}(y_{ij}-\bar{y}_{i.})^2}{n-t}$ has a χ^2 distribution with $n-t$ degrees of freedom
- Hence,

$$F = \frac{MSTr}{MSE}$$

has an F distribution with $t-1$ and $n-t$ degrees of freedom

The ANOVA table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Between Treatments	SSTr	t-1	MSTr	MSTr/MSE
Error (Within Treatments)	SSE	n-t	MSE	
Total	SST	n-1		

An example

The data in Table 1.1 came from an experiment that was conducted to determine how six different kinds of work tasks affect a worker's pulse rate. In this experiment, 78 male workers were assigned at random to six different groups so that there were 13 workers in each group. Each group of workers was trained to perform their assigned task. On a selected day after training, the pulse rates of the workers were measured after they had performed their assigned tasks for 1hr. Unfortunately, some individuals withdrew from the experiment during the training process so that some groups contained fewer than 13 individuals. The recorded data represent the number of heart pulsations in 20 seconds where there are $n = 68$ observations.

An example

TABLE 1.1

Pulsation Data and Summary Information for Six Tasks

	Task					
	1	2	3	4	5	6
	27	29	34	34	28	28
	31	28	36	34	28	26
	26	37	34	43	26	29
	32	24	41	44	35	25
	39	35	30	40	31	35
	37	40	44	47	30	34
	38	40	44	34	34	37
	39	31	32	31	34	28
	30	30	32	45	26	21
	28	25	31	28	20	28
	27	29			41	26
	27	25			21	
	34					
$y_{i.}$	415	373	358	380	354	317
n_i	13	12	10	10	12	11

An example

Computing formulas for the sums of squares

$$SST = \sum_{i=1}^{t}\sum_{j=1}^{r_t} Y_{ij}^2 - \frac{{y_{..}}^2}{n}$$

$$SSTr = \sum_{i=1}^{t} \frac{y_{i.}^2}{r_i} - \frac{{y_{..}}^2}{n}$$

$$SSE = SST - SSTr$$

Totals:

$y_{1.} = 415$, $y_{2.} = 373$, $y_{3.} = 358$, $y_{4.} = 380$, $y_{5.} = 354$, $y_{6.} = 317$, $y_{..} = 2197$, $n = 68$

An example

$$
\begin{aligned}
SST &= \sum_{i=1}^{t}\sum_{j=1}^{r_t} Y_{ij}^2 - \frac{{y_{..}}^2}{n} \\
&= [27^2 + 31^2 + \cdots + 26^2] - \frac{2197^2}{68} \\
&= 73593 - 70982.49 \\
&\approx 2610.51
\end{aligned}
$$

$$
\begin{aligned}
SSTr &= \sum_{i=1}^{t} \frac{y_{i.}^2}{r_i} - \frac{{y_{..}}^2}{n} \\
&= [\frac{415^2}{13} + \frac{373^2}{12} + \cdots + \frac{317^2}{11}] - \frac{2197^2}{68} \\
&= 71676.92 - 70982.49 \\
&\approx 694.43
\end{aligned}
$$

$$
SSE = SST - SSTr = 2610.51 - 694.43 \approx 1916.08
$$

An example

Source of Variation	Sum of Squares	df	Mean Square	F
Between Treatments	694.43	5	138.89	4.49
Within Treatments	1916.08	62	30.90	
Total	**2610.51**	**67**		

- The p-value associated with $F = 4.49$ is obtained using the R command $pf(4.49, 5, 62, lower.tail = F)$ which gives 0.0014807

An example

Using base R:

```r
pulse <- read.csv("Pulse.csv")
mod1 <- with(pulse, aov(Pulse ~ factor(Task)))
summary(mod1)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## factor(Task)  5  694.4   138.9   4.494 0.00147 **
## Residuals    62 1916.1    30.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using ExpDes package

```r
library(ExpDes)
with(pulse, crd(Task, Pulse, quali=T))
```

```
------------------------------------------------
Analysis of Variance Table
------------------------------------------------
           DF      SS      MS     Fc     Pr>Fc
Treatament  5  694.44 138.888 4.4941 0.0014709
Residuals  62 1916.08  30.904
Total      67 2610.51
------------------------------------------------
CV = 17.21 %
```

An example: test of hypothesis

- H_0 : There is no significant difference in the mean pulse rates of workers assigned to different work tasks $[H_0 : \mu_1 = \mu_2 = \cdots = \mu_6]$.
- H_1 : There is significant difference in the mean pulse rates of workers assigned to different work tasks
  $[H_1 : \mu_i \neq \mu_j,$ for at least one pair $i \neq j]$.
- $\alpha = 0.05$
- Test statistic: $F = 4.94$, $p = 0.0014807$
- Decision on H_0: Reject.
- Conclusion: At the 5% level of significance, there is sufficient evidence indicating that a significant *difference* in the mean pulse rates of workers assigned to different tasks.

Measures of precision

Coefficient of variation

- overall index of the reliability (precision) of the experiment
- it is computed as

$$CV = \frac{\sqrt{MSE}}{\overline{Y}_{..}}$$

- the experiment is very precise if $CV \leq 10\%$; with acceptable precision if $10\% < CV \leq 20\%$; not precise if $CV > 20\%$

Measures of precision

Standard error of a treatment mean

- measures the average error in estimating the true treatment mean
- it is computed as

$$se(\bar{y_{i.}}) = \sqrt{\frac{MSE}{r_i}}$$

Standard error of the difference in any two treatment means

- measures the average error in estimating the difference in the true means of treatments *i* and *j*
- it is computed as

$$se(\overline{y}_{i.} - \overline{y}_{j.}) = \sqrt{MSE\left(\frac{1}{r_i} + \frac{1}{r_j}\right)}$$

Analysis of random effects model

Recall

- The treatments is a random sample from a larger population of treatments
- **Goal**: estimate and (test, if any) the variability among treatments in the population
- **Linear model**:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, i = 1, 2, \ldots, t; j = 1, 2, \ldots, r_i; \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

  - τ_i are random variables, $\tau_i \sim N(0, \sigma_\tau^2)$
  - τ_i and ϵ_{ij} are independent, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$
- $V(Y_{ij}) = \sigma_\tau^2 + \sigma_\epsilon^2$

Analysis of random effects model

- the variances σ_τ^2 and σ_ϵ^2 are called **variance components**
- Hence, the random effects model is also known as variance components model
- Goal: Estimation and testing of the variance components
- Hypothesis:$H_0 : \sigma_\tau^2 = 0$ versus $H_1 : \sigma_\tau^2 > 0$
- The partitioning of the total sum of squares still holds, that is, $SST = SSTr + SSE$
- Test statistic is the same as that of the fixed effects model

Analysis of random effects model

- Estimation using Method of Moments (also known as ANOVA Method)

$$\hat{\sigma}^2_\epsilon = MSE$$
$$\hat{\sigma}^2_\tau = \frac{MSTr - MSE}{n_0}$$

  - where:

$$n_0 = \frac{1}{t-1}\left[n - \frac{\sum_i^t r_i^2}{n}\right]$$

Random effects model: an example

A textile company weaves a fabric on a large number of looms. They would like the looms to be homogeneous so that they obtain a fabric of uniform strength. The process engineer suspects that, in addition to the usual variation in strength within samples of fabric from the same loom, there may also be significant variations in strength between looms. To investigate this, he selects four looms at random and makes four strength determinations on the fabric manufactured on each loom. The data and the corresponding ANOVA Table are given on the following slide. Verify

Random effects model: an example

	Observations			
Looms	1	2	3	4
1	98	97	99	96
2	91	90	93	92
3	96	95	97	95
4	95	96	99	98

```r
library(ExpDes)
looms <- read.csv("Looms.csv")
with(looms, crd(looms, ts, quali=T,unfold='0'))
```

```
## ---------------------------------------------------------------------------
## Analysis of Variance Table
## ---------------------------------------------------------------------------
##             DF      SS      MS     Fc      Pr>Fc
## Treatament   3  89.188 29.7292 15.681 0.00018779
## Residuals   12  22.750  1.8958
## Total       15 111.938
## ---------------------------------------------------------------------------
## CV = 1.44 %
##
```

Random effects model: an example

Estimates of the variance components:

$$
\begin{aligned}
n_0 &= \frac{1}{t-1}\left[n - \frac{\sum_{i=1}^{t} r_i^2}{n}\right] \\
&= \frac{1}{4-1}\left[16 - \frac{\sum_{i=1}^{4} 4^2}{16}\right] \\
&= 4
\end{aligned}
$$

$$
\begin{aligned}
\hat{\sigma}_\tau^2 &= \frac{MSTr - MSE}{n_0} \\
&= \frac{29.7292 - 1.8958}{4} \\
&\approx 6.96
\end{aligned}
$$

$$
\hat{\sigma}_\epsilon^2 = MSE = 1.8958
$$

Random effects model: an example

Test of hypothesis:

- $H_0 : \sigma^2_\epsilon = 0$
- $H_1 : \sigma^2_\epsilon > 0$
- $\alpha = 0.05$
- Test statistic: $F = 15.681$, $p = 0.00018779$
- Decision on H_0: Reject.
- Conclusion: At $\alpha = 5\%$, there exists significant *variation* in the strength of fabric among looms.

CRD with subsampling

- experimental unit is the unit of research material to which a treatment is applied
- experimental unit may consist of two or more measurement units
- For example, feed ration (treatment) is randomly assigned to each of 5 pens (experimental units) and each pen contains 4 animals (measurement units)
    - variation among pens given the same feed ration plus variation among animals within each pen

CRD with subsampling

Linear Model:

$$Y_{ijk} = \mu + \tau_i + \epsilon_{ij} + \delta_{ijk}, i = 1, 2, \cdots, t;\ j = 1, 2, \cdots, r_i;\ k = 1, 2, \cdots, n_{ij}$$

where:

- Y_{ijk} = response of the k^{th} (sub)sample of the j^{th} experimental unit given treatment i
- μ = overall mean response
- τ_i = effect of treatment i
- ϵ_{ij} = random error associated with the j^{th} experimental unit given treatment i
- δ_{ijk} = random error associated with the k^{th} (sub)sample of the j^{th} experimental unit given treatment i

CRD with subsampling

Some Remarks:

- **Fixed model**: τ_i are fixed parameters with $\sum_{i=1}^{t} \tau_i = 0$
- **Random model**: τ_i are random parameters with $\tau_i \sim N(0, \sigma_\tau^2)$
- $\tau_i, \epsilon_{ij}, \delta_{ijk}$ are independent with $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ and $\delta_{ijk} \sim N(0, \sigma_\delta^2)$

Partitioning of the sums of squares:

$$\sum_{i=1}^{t}\sum_{j=1}^{r_i}\sum_{k=1}^{n_{ij}}(y_{ijk}-\bar{y}_{...})^2 = \sum_{i=1}^{t}\sum_{j=1}^{r_i} r_i(\bar{y}_{i..}-\bar{y}_{...})^2+\sum_{i=1}^{t}\sum_{j=1}^{r_i} n_{ij}(\bar{y}_{ij.}-\bar{y}_{i..})^2+\sum_{i=1}^{t}\sum_{j=1}^{r_i}\sum_{k=1}^{n_{ij}}(y_{ijk}-\bar{y}_{ij.})^2$$

CRD with subsampling

ANOVA Table

Source of Variation	SS	DF	MS	F
Between Treatments	SSTr	t-1	MSTr	
Experimental Error (Within Treatments)	SSE	$\sum_{i}^{t} r_i - t$	MSE	
Sampling error (Between sampling units)	SS(SE)	$\sum_{i}^{t}\sum_{j}^{r_i} n_{ij} - \sum_{i}^{t} r_i$	MS(SE)	
Total	SST	n-1		

CRD with subsampling

Expected Mean Squares

- **Fixed model** (balance case)
  - Between treatments: $\sigma_\delta^2 + n\sigma_\epsilon^2 + \frac{rs\sum_{i=1}^{t}\tau_i^2}{t-1}$
  - Within treatments: $\sigma_\delta^2 + n\sigma_\epsilon^2$
  - Between (sub)sampling units: σ_δ^2
- **Random model** (balance case)
  - Between treatments: $\sigma_\delta^2 + n\sigma_\epsilon^2 + nr\sigma_\tau^2$
  - Within treatments: $\sigma_\delta^2 + n\sigma_\epsilon^2$
  - Between (sub)sampling units: σ_δ^2

CRD with subsampling

Sequential hypothesis testing

A. Test on the variability of the experimental units

- $H_0 : \sigma_\epsilon^2 = 0$ versus $H_1 : \sigma_\epsilon^2 > 0$
- Test statistic: $F = \frac{MSE}{MS(SE)}$

B. Test on the differences among treatment means

- $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$ versus
  $H_1 : \mu_i \neq \mu_j$, for at least one pair $i \neq j$
- Test statistic (CASE 1, $H_0 : \sigma_\epsilon^2 = 0$ is rejected): $F = \frac{MSTr}{MSE}$
- Test statistic (CASE 2, $H_0 : \sigma_\epsilon^2 = 0$ is not rejected): $F = \frac{MSTr}{MSE_{\text{pooled}}}$

$$MSE_{\text{pooled}} = \frac{SSE + SS(SE)}{df_{SSE} + df_{SS(SE)}}$$

CRD with subsampling: an example

The following data come from an experiment that was conducted in a completely randomized design with sub-sampling: there were 4 treatments, and 4 plots for each treatment. Within each plot 3 measurements (subsamples) were taken. The treatments are different forms of irrigation; the subsamples correspond to 3 small amounts of soil that are measured in each plot. The measurement Y corresponds to a measure of soil moisture. The data are tabulated below. For each treatment and plot combination, the 3 values represent the 3 subsamples.

Treatment	Plot 1			Plot 2			Plot 3			Plot 4		
1	12.6	11.9	12.3	13.0	12.4	12.4	11.3	11.9	10.9	12.5	11.8	11.9
2	12.4	12.1	12.6	11.9	11.6	12.2	14.2	13.3	13.8	12.9	13.7	13.1
3	12.2	11.5	12.0	11.4	11.8	11.0	9.80	10.0	10.4	10.7	11.2	11.2
4	12.9	12.2	12.8	14.2	13.9	13.7	12.5	12.9	12.8	13.3	13.6	12.8

CRD with subsampling: an example

Computing formulas:

$$\text{SST} = \sum_{i=1}^{t}\sum_{j=1}^{r_i}\sum_{k=1}^{n_{ij}} y_{ijk}^2 - \frac{y_{\dots}^2}{n}$$

$$\text{SSTr} = \sum_{i=1}^{t} \frac{y_{i..}^2}{r_i} - \frac{y_{\dots}^2}{n}$$

$$\text{SSE} = \sum_{i=1}^{t}\sum_{j=1}^{n_{ij}} \frac{y_{ij}^2}{n_{ij}} - \sum_{i=1}^{t} \frac{y_{i..}^2}{r_i}$$

$$\text{SS(SE)} = SST - SSTr - SSE$$

CRD with subsampling: an example

Using R:

```
subsample <- read.csv("Subsample.csv")
mod2=aov(y~trt+trt/plot,data=subsample)
summary(mod2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## trt          3  29.41   9.802   76.51 1.10e-14 ***
## trt:plot    12  17.39   1.449   11.31 2.34e-08 ***
## Residuals   32   4.10   0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

CRD with subsampling: an example

Tasks:

1. Verify the entries in the ANOVA table using the computing formulas.
2. Test the necessary hypotheses at 5% level of significance.
3. Estimate the variance components σ_ϵ^2 and σ_δ^2.