BASIC EXPERIMENT DESIGNS

i) Completely Randomised Design (CRD)
ii) Randomised Block Design (RBD)
iii) Latin Square Design (LSD)

Completely Randomisation Design (CRD).

Consider the problem of determining whether or not different types of tyres exhibit different amounts of tread loss after 20,000km of driving.
A manager wishes to consider 4 tyres that are available and make some decision about which type or brand might show the least amount of tread wear or loss.
The brands to be considered are A, B, C, and D and she wants to try these 4 brands under actual driving conditions. The variable to be measured is the difference in maximum tread thickness on the tyre between the time it is mounted on the wheel of a car and after it has completed 20,000km on a car. The measured variable y_{ij} is this difference in thickness in $\frac{1}{1000}$ cm.
The single factor of interest is brands/tyre types (α_i, i = 1, 2 3, 4)

Since the tyres must be tried on cars and since some measure of error is necessary (replication) more than one tyre of each brand must be used and a set of 4 of each brand would seem quite practical. A car normally uses 4 tyres.
This means that we are going to have 16 experimental units (tyres), 4 each for the 4 brand and if we designate each of the cars as I, II, III, IV, one might put the brand A tyres on car I, brand B on car II and so on with the design as shown in the table below.

Table 1: Design 1 of the tyre-type problem

I	II	III	IV
A	B	C	D
A	B	C	D
A	B	C	D
A	B	C	D

If one looks at design 1, there is a problem since averages for brands are also averages for cars. If the cars travel over different terrains using different drivers, any apparent brand differences are also car differences and therefore this design is called a completely confounded one since we cannot distinguish between brands and cars in this analysis.
A second attempt at the design may be to try the CRD.

Advantages of the CRD

i) Simplicity in the layout of design.

ii) The statistical analysis and interpretation of results are straight forward.
iii) This design does not require the use of equal sample sizes for each treatment level i.e one can use unequal number of observations per treatment.
iv) It allows the maximum number of degrees of freedom for the error sum of squares.
v) It does not require an experimental unit to participate under more than one treatment level therefore the sample size is maximised.

Disadvantages

i) Effects of differences among subjects are controlled by random assignment of subjects to treatment levels. For this to be more effective, subjects should be relatively homogeneous or a large number of subjects should be used (as many replications as possible).
ii) When many treatment levels are included in the experiment, the required sample size may be prohibitive especially from the point of view of costs.
iii) This design does not offer the possibility of evaluating the interaction effect (you can find out the interaction effect if you have more than one factor).

Analysis

To carry out ANOVA for one factor experiment, the total sums of squares (SST) is partitioned or broken down into sums of squares due to the treatments which we donate as SSB and the error sums of squares (SSE)

SST = SSB + SSE

Randomisation and lay out of the CRD with equal number of observations/treatments

	1	**2**	**3 ....**	**a**		
	y_{11}	y_{21}	y_{31} ……	y_{a1}		
	y_{12}	y_{22}	y_{32} ……	y_{a2}		
	y_{13}	y_{23}	y_{33} ……	y_{a3}		
	:	:	:	:		
	:	:	:	:		
	y_{1n}	y_{2n}	y_{3n}	y_{an}		
Totals	$\mathbf{y_{1o}}$	$\mathbf{y_{2o}}$	$\mathbf{y_{3o}}$	$\mathbf{y_{ao}}$	**=>**	$\mathbf{y_{oo}}$ **(grand total)**
Means	$\bar{y}_{1o}$	$\bar{y}_{2o}$	$\bar{y}_{3o}$	$\bar{y}_{ao}$		$\bar{y}_{oo}$ **(grand mean)**

y_{ij} = j^{th} observation for the i^{th} treatment

y_{23} = 3^{rd} observation for the 2^{nd} treatment

The first column represents a random sample of size n for treatment 1.........., a^{th} column represents a random sample of size **n** for treatment **a**

y_{io} and $\bar{y}_{io}$ are the total and mean respectively of the observations in the i^{th} treatment.

y_{oo} and $\bar{y}_{oo}$ are the grand total and mean of all the **N = an** observations

$\bar{y}_{oo} = \frac{y_{oo}}{N}$

i.e. Total variation = Variation between the treatments + Variation due to error within treatments

SST = SSB + SSE

We can compute

$$SST = \Sigma_i \Sigma_j \left(y_{ij} - \bar{y}_{oo}\right)^2 \quad \Longrightarrow \Sigma_i \Sigma_j \left[\, y_{ij}^2 + \bar{y}_{oo}^2 - 2 y_{ij} \bar{y}_{oo}\right]$$

$$= \Sigma_i \Sigma_j \, y_{ij}^2 - \frac{y_{oo}^2}{N} \text{ (prove)}$$

$\frac{y_{oo}^2}{N}$ is called the correction factor.

$\Sigma_i \Sigma_j y_{ij}^2$ is uncorrected total sums of squares.

$SSB = n\sum_i (\bar{y}_{io} - \bar{y}_{oo})^2$ **n** is the number of observations per treatment

$$= \frac{1}{n} \Sigma\, y_{io}^2 - \frac{y_{oo}^2}{N} \text{ (prove)}$$

$\frac{1}{n} \Sigma\, y_{io}^2 \rightarrow$ uncorrected SSB

$\frac{y_{oo}^2}{N} \rightarrow$ correction factor

$$SSE = \Sigma_i \Sigma_j y_{ij}^2 - \frac{1}{n} \Sigma\, y_{io}^2$$

OR

SSE = SST – SSB

ANOVA TABLE

Source of variation	**df**	**ss**	**ms**	**F-ratio**
Between treatments	$a - 1$	SSB	MSB = $\frac{SSB}{(a-1)}$	Fc = $\frac{MSB}{MSE}$
Error	$N - a$	SSE	MSE - $\frac{SSE}{(N-a)}$	–
Total	$\boldsymbol{N - 1}$	**SST**	–	–

Tabulated value

$F_T = F_{0.05, (a-1), (N-a)}$

$(a - 1)$ – degrees of freedom for the factor of interest.

$(N - a)$ – degrees of freedom for the error term.

Rejection criteria

$F_c \geq F_T$, Reject H_o => Treatment effects are not equal to zero/Treatment means are not equal

$F_c < F_T$, Accept $H_{o} \Rightarrow$ Treatment effects are equal to zero/Treatment means are equal

Example 1

The data in the table below gives the number of hours of pain relief provided by 4 different types of headache tablets administered to 24 people. The 24 experimental units were randomly divided into 4 groups and each group was treated with a different brand/type. Do the different drug types give significantly different hours of pain relief?

Brands

	1	**2**	**3**	**4**
	12.2	4.9	8.0	4.6
	9.5	10.6	12.1	6.1
	11.6	7.0	5.7	5.0
	13.0	8.3	8.6	3.8
	10.1	5.5	7.2	8.2
	9.6	11.7	12.4	7.7
$\boldsymbol{y_{io}}$	**66.0**	**48.0**	**54.0**	**36.0**
$\boldsymbol{\bar{y}_{io}}$	**11.0**	**8.0**	**9.0**	**6.0**

y_{oo}=204

Steps

Model

$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$

Where y_{ij} - j^{th} observation under the i^{th} drug type

μ - general mean

α_i - i^{th} treatment effect (i^{th} drug type effect)

ϵ_{ij} - error term

Hypothesis

H_o: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ OR $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$(say)

H_A: Some $\alpha_i \neq 0$ Some $\mu_{io} \neq \mu$

Rejection region

Reject H_o if $F_c \geq F_T$ ($F_{0.05, 3, 20}$) = 3.10

Accept H_o if $F_c < F_{0.05, 3, 20}$ = 3.10

Computations

Correction factor = $\frac{y_{oo}^2}{N}$ = $\frac{204^2}{24}$ = 1734.0

SST = $\Sigma_i \Sigma_j\, y_{ij}^2 - \frac{y_{oo}^2}{N}$

= [$12.2^2 + 9.5^2 + ... + 7.7^2$] − Correction factor

= 1912.7 – 1734.0 = 178.7

SSB = $\frac{1}{n} \Sigma\, y_{io}^2 - \frac{y_{oo}^2}{N}$ n – no of observations per treatment

= $\frac{1}{6}$ [$66.0^2 + 48.0^2 + 54.0^2 + 36.0^2$] – C.F

= 1812.0 – 1734.0 = 78.0

SSE = SST – SSB = 178.7 – 78.0 = 100.7

Source of variation	**Df**	**ss**	**ms**	**F-ratio**
Between treatments	3	78.0	MSB = 26.0	$F_c = \frac{26.0}{5.035}$ = 5.1638
Errors	20	100.7	MSE = 5.035	
Total	**23**	**178.7**		

Conclusion

$F_c = 5.1638 > F_T = 3.10$; Reject H_o implying;

Drug types are significantly different in terms of the hours of pain relief they give.

Note: The sums of squares are NEVER negative.

Lay out of CRD with unequal number of observations/treatment (n_i)

	1	**2**	**3 .... a**		
	y_{11}	y_{21}	y_{31}	y_{a1}	
	y_{12}	y_{22}	y_{32}	y_{a2}	
	:	:	:	:	
	:	:	:	:	
	y_{1n1}	y_{2n2}	y_{3n3}	y_{ana}	
Totals	y_{1o}	y_{2o}	y_{3o}	y_{ao}	y_{oo}
Means	$\bar{y}_{1o}$	$\bar{y}_{2o}$	$\bar{y}_{3o}$	$\bar{y}_{ao}$	$\bar{y}_{oo}$

N = $\sum_{i=1}^{a} ni$

SSB = $\Sigma\left[\frac{y_{io}^2}{ni}\right] - \frac{y_{oo}^2}{N}$ $\left[\sum_{i=1}^{a}\sum_{j=1}^{ni}(\bar{y}_{io} - \bar{y}_{oo})^2\right]$

SST and SSE remain the same as before.

Example 2

Four groups of students were subjected to different teaching techniques and tested at the end of a specified period of time. The table below gives the performance in percentages. Are the teaching techniques significantly different judging from the performance of the students?

Teaching techniques

	1	**2**	**3**	**4**
	65	75	59	94
	87	69	78	89
	73	83	67	80
	79	81	62	88
	81	72	83	
	69	79	76	
		90		
n_i	**6**	**7**	**6**	**4**
y_{io}	**454**	**549**	**425**	**351**
$\bar{y}_{io}$	**75.67**	**78.43**	**70.83**	**87.75**

N = 23 N = $\sum_{i=1}^{a} ni$

y_{oo} = 1779

$\bar{y}_{oo} = \frac{1779}{23}$ = 77.34

Model

y_{ij} = μ + α_i+ $\in_{ij}$

Where y_{ij} - j^{th} observation under the i^{th} teaching technique

μ - grand mean

α_i - i^{th} treatment effect (i^{th} teaching technique effect)

ϵ_{ij} - error term

Hypothesis

$H_o: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ OR $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$(say)

H_o: Some $\alpha_i \neq 0$ Some $\mu_{io} \neq \mu$

Rejection region

Reject H_o if $F_c \geq F_T$ ($F_{0.05, 3, 19}$) = 3.13

Accept H_o if $F_c < F_{0.05, 3, 19}$ = 3.13

Computations

Correction factor = $\frac{y_{oo}^2}{N} = \frac{1779^2}{23}$ = 137,601.8

SST = $\Sigma_i \Sigma_j \, y_{ij}^2 - \frac{y_{oo}^2}{N}$ = [$65^2 + 87^2 + \ldots\ldots\ldots + 88^2$] − Correction factor

= 139511 – 137,601.8 = 1909.2

SSB = $\Sigma \left[\frac{y_{io}^2}{ni}\right] - \frac{y_{oo}^2}{N}$ n – no of observations per treatment

= $\left[\frac{454^2}{6} + \frac{549^2}{7} + \frac{425^2}{6} + \frac{351^2}{4}\right]$ – C.F = 138314.4 – 137601.8 = 712.6

SSE = SST – SSB = 1909.2 – 712.6 = 1196.6

Source of variation	**df**	**Ss**	**ms**	**F-ratio**
Between treatments	3	712.6	MSB = 237.5	$F_c = \frac{237.5}{63.0}$ = 3.769
Errors	19	1196.6	MSE = 62.9 ≅ 63.0	
Total	**22**	**1909.2**		

Conclusion

$F_c = 3.769 > F_T = 3.13$; Reject H_o implying;

The teaching techniques are significantly different judging from the performance of the students.

Estimation of effects

Given the data table e.g example 2 and the ANOVA table, we can make the following inferences/conclusions concerning the population from which the data was obtained.

i) The grand mean $\bar{y}_{oo}$ is an unbiased estimator of μ_{oo} the population mean.

ii) The treatment mean $\bar{y}_{io}$ is an unbiased estimator of the mean of the treatment population. E.g $\bar{y}_{io}$ = 75.67

iii) The difference between any 2 treatment effects is estimated unbiasedly by the difference between the treatment means i.e $\alpha_1 - \alpha_2 = \bar{y}_{1o} - \bar{y}_{2o}$

iv) The factor effect is estimated unbiasedly by the difference between its treatment means and the grand mean i.e $\bar{y}_{io} - \bar{y}_{oo}$ and thus can be positive or negative.

v) The mean square error (MSE) value is an unbiased estimator of the common variance $\sigma^{2.}$ e.g σ^2 = 63 in example 2.

Assumptions of ANOVA

In applying the ANOVA techniques, certain assumptions should be kept in mind and these include;

i) The process is repeatable (can be replicated)

ii) The population distribution being sampled is normal.

iii) The variances of all the levels of a factor are homogeneous.