

Analysis of variance for Completely Randomised Design

Updated: March 2025

Analysis of Variance - Completely Randomised Designs

Analysis of research data often involves the comparison of observed outcomes from two or more groups. For example, suppose you have an experiment that compares a control group against two or more experimental groups. Do you use t-tests for this analysis, or some other technique?

For experiments with outcomes whose values are continuous measures and considered to be statistically normal, it makes sense to compare group means.

In the case where there are only two means to compare, it is normal to use a t-test for independent groups to determine if there are statistically significant differences.

When an experiment consists of more than two groups, the analysis becomes a bit more complicated. Suppose you are comparing the means of three groups and you're interested in knowing which means are significantly different from the other means.

Consider the following situation.

Three types of medium sized cars assembled in New Zealand have been test driven by a motoring magazine and compared on a variety of criteria. In the area of fuel efficiency performance, five cars of each brand were each test driven 1000 km; the km per litre data are obtained as follows:

	Kilometres per litre					Total	Mean
Brand A	7.6	8.4	8.0	7.6	8.4	40.0	8.0
Brand B	7.8	8.0	9.1	8.5	9.6	43	8.6
Brand C	9.6	10.4	9.2	9.7	10.6	49.5	9.9

The task is to determine if there is a difference in the mean fuel consumption for the three makes of car.

It may seem logical to perform t-tests for all pairs of means – (Brand A vs. Brand B), (Brand A vs. Brand C) and (Brand B vs. Brand C). In other words, perform t-tests on all possible comparisons. However, there is a fundamental problem with this technique. The p-value associated with each t-test is determined as if only one t-test is performed per experiment. If three t-tests are performed in a single experiment, then the p-values for these tests are no longer accurate. The more comparisons you make the more likely it is that you will say there is a significant difference between two means when there actually isn't.

The solution to this multiple comparisons' problem is NOT to perform all possible t-tests. Instead, a two-step procedure is used - an **Analysis of Variance (ANOVA)** followed by a **Multiple Comparison Test**.

- The first part answers the question, "Is there at least one mean that is significantly different from one other mean?" If the p-value for the ANOVA results is less than your chosen significance level (usually 0.05), you have evidence that at least one mean is different. If the p-value is not significant your analysis is over and you conclude that there is no difference between any pair of means.
- If the ANOVA's p-value is significant, proceed to the second stage of the analysis to answer the question, "Which means are significantly different from which other means?" This is the multiple comparison stage.

One-way Analysis of Variance of CRD

Assumptions that are made:

- Each sample consists of units that are randomly selected from each group with the sample from one group being unrelated to (i.e. independent of) the sample from the other groups.
- The units sampled from each group are independent of each other and normally distributed
- The variances of the populations are assumed to be equal.
- Sample sizes do not have to be the same, although large differences in sample size may affect the outcome.

State hypotheses

$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ i.e all means are equal; NO differences between the groups.

$H_A : \mu_1 \neq \mu_2 \neq \dots \neq \mu_k$ i.e. the means of at least two groups are not equal.

Construct an ANOVA table:

SOURCE	df	SS	MS	F	p
Treatment *					
Error					
Total					

Note that the table is similar to the Linear Regression table, with a row for 'Treatment' instead of 'Regression'.

* Treatment can also be referred to as 'Group'. This refers to the various treatments that are to be compared. k is generally used to represent the number of treatments involved.

Fill

Calculate Degrees of Freedom (df) for each source of variation

- $df_{\text{treatment}} = (k-1)$; $df_{\text{total}} = (n-1)$; $df_{\text{error}} = df_{\text{total}} - df_{\text{treatment}} = (n-k)$

in

Calculate the Sums of Squares:

- Identify which columns or rows represent the treatment values and calculate these totals.
- Calculate the Correction Factor
- Calculate the Total Sum of Squares
- Calculate the Group (Treatment) Sum of Squares
- Calculate the Error Sum of Squares as $SS_{\text{error}} = SS_{\text{total}} - SS_{\text{treatments}}$

Formulae
over
page



results

Complete the remaining calculations:

- Calculate the Mean Square for Groups (Treatments):

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{k-1}$$

- Calculate the Mean Square for Error: $MS_{\text{error}} = \frac{SS_{\text{Error}}}{n-k}$
- Calculate the F Value
- Find the p-value

are

completed

ANOVA FORMULAE FOR A CRD (ONE WAY ANALYSIS)

n = total number of data values; k = number of treatments (or groups) to be compared.

Correction Factor:	$CF = \frac{(\sum \text{all entries})^2}{\text{number of entries}} = \frac{(\sum y)^2}{n}$	<ul style="list-style-type: none"> Add all entries Square the result Divide by n
Total Sum of Squares:	$SS_{Total} = \sum (\text{each entry})^2 - CF$ $= \sum y^2 - \frac{(\sum y)^2}{n}$	<ul style="list-style-type: none"> Square each entry Add all results Subtract CF
Sum of Squares for Treatments (also SS between treatments or between groups):	$= SS_{Treatments} = SS_{Between Groups}$ $= \sum \left(\frac{\text{group total}^2}{\text{number in group}} \right) - CF$ $= \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \dots + \frac{T_k^2}{n_k} - \frac{(\sum y)^2}{n}$ <p>This <i>Sum of Squares</i> calculation reflects the differences between the treatments.</p>	<ul style="list-style-type: none"> Square each Treatment total Divide by no. of groups in total Add results Subtract CF
Sum of Squares for Error:	$SS_{Error} = SS_{Total} - SS_{Between Groups} *$ <p>SS within treatments ie differences due to random sampling error</p>	
Mean Square for Treatments:	$MS_{Treatments} = \frac{SS_{treatment}}{k - 1}$	
Mean Square for Error:	$MS_{error} = \frac{SS_{error}}{n - k}$	Sometimes referred to as EMS (Error Mean Square)
Test statistic F:	$\frac{MS_{Treatments}}{MS_{error}}$	
Standard Error:	$s = \sqrt{MSE}$ $SEM = \sqrt{\frac{MSE}{n}} \text{ (one sample)}$ $SED = \sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \text{ (two samples)}$	<p>Note $MSE = MS_{error}$</p> <p>$SEM = S.E.(\text{mean})$</p> <p>$SED = SE(\text{diff. btwn means})$</p> <p>$n_1, n_2$ are the number of data values in group1, group2</p>
<p>* SS_{error} can also be calculated by $(n_1 - 1)s_1^2 + \dots + (n_k - 1)s_k^2$</p>		

Significance level and the p-value

The **significance level** is denoted by α – it's the probability of *rejecting* the null hypothesis when you *shouldn't*. A value of $\alpha = 0.05$ is generally used, meaning that you are prepared to make the wrong decision just 5% of the time.

Think of the **p-value** as the probability of getting the calculated F test value (or an even larger value) *when the null hypothesis is actually true*.

When the p-value is smaller than the significance level we reject H_0 .

Returning to the Car Brand Problem:

	Score					Total
Brand A	7.6	8.4	8.0	7.6	8.4	40.0
Brand B	7.8	8.0	9.1	8.5	9.6	43
Brand C	9.6	10.4	9.2	9.7	10.6	49.5

$$H_0 : \mu_{Brand A} = \mu_{Brand B} = \mu_{Brand C}$$

$$H_A : \mu_{Brand A} \neq \mu_{Brand B} \neq \mu_{Brand C} \text{ i.e. the means of at least two groups are not equal.}$$

Enter all values in calculator. (You can leave it in LR mode.) $n = 15$, $k = 3$

(i) $df_{treatment} = 2$; $df_{total} = 14$; $df_{error} = df_{total} - df_{treatment} = 12$

(ii) Calculate the correction factor

$$CF = \frac{(\sum \text{all entries})^2}{\text{number of entries}} = \frac{(\sum y)^2}{n} = \frac{132.5^2}{15} = 1170.42.$$

You can use your Alpha or Shift key to process these first two.

(iii) Calculate the Total Sum of Squares

$$SS_{Total} = \sum (\text{each entry})^2 - CF = \sum y^2 - \frac{(\sum y)^2}{n}$$
$$= 1184.11 - \frac{132.5^2}{15} = 13.69$$

(iv) Calculate the Group (Treatment) Sum of Squares

You now have to use the treatment totals.

$$SS_{Between Groups} = \sum \left(\frac{\text{group total}^2}{\text{number in group}} \right) - CF = \frac{40^2}{5} + \frac{43^2}{5} + \frac{49.5^2}{5} - \frac{132.5^2}{15} = 9.43$$

(v) Calculate the Error Sum of Squares

$$SS_{error} = SS_{total} - SS_{treatment} = 13.69 - 9.43 = 4.26$$

(vi) Calculate

$$MS_{treatment} = \frac{SS_{treatment}}{k - 1} \quad \text{and} \quad MS_{error} = \frac{SS_{error}}{n - k}$$
$$= \frac{9.43}{2} = 4.715 \quad \quad \quad = \frac{4.26}{12} = 0.355$$

(vii) **Calculate F**

$$F = \frac{MS_{treatment}}{MS_{error}} = \frac{4.715}{0.355} = 13.28$$

(viii) **Find the p value and compare with α , the significance level**

Numerator $df = 2$; denominator $df = 12$

df denominator	df numerator			
	α	1	2	...
1	...			
...	...			
12	0100		2.81	
	0050		3.89	
	0.025		5.10	
	0.010		6.93	
	0.001		12.97	

Since the calculated value for F (13.28) is greater than 12.97, the level of significance, p, is less than 0.001. This can be written as simply 0.001 or $p < 0.001$.

Note that because computers are able to store information to a greater degree of accuracy, the p-value often appears as zero (through rounding), or as a specific value. In answering questions, you are expected to use tables to look up critical values to compare with your test result. You can further demonstrate your understanding of how significant your conclusion is by stating that your result is: significant ($p < 0.05$), very significant ($0.005 < p < 0.01$) or highly significant ($p < 0.001$).

(P-values do not simply provide you with a “Yes” or “No” answer. They provide a sense of the strength of the evidence for the null hypothesis. The lower the p-value, the weaker the support for H_0 i.e., the stronger the evidence against H_0 . Once you know how to read p-values, you can more critically interpret journal articles, and decide for yourself if you agree with the conclusions of the author.)

(vii) **Complete the ANOVA table** with group $df = (k-1)$ and error $df = (n-k)$ respectively:

SOURCE	df	SS	MS	F	p
Treatment	2	9.43	4.715	13.28	<0.001
Error	12	4.26	0.355		
Total	14	13.69			

Since this result is highly significant, we can reject H_0 in favour of H_A . That is, there is strong evidence that the mean km/l is different in at least one of the three makes of car.

The next step would be to do a multiple comparison test. This will be dealt with later.

Test questions (multi-choice)**Answers p.13/14**

1. The dry shear strength of birch plywood bonded with different resin glues was studied with a completely randomised designed experiment.

Here are the data:

	Glue A	Glue C	Glue F
	102	100	220
	58	102	243
	45	80	189
	79	119	176
	68		176
	63		
	117		
Total	532	401	1004

Here is a partial analysis of variance table:

Analysis of Variance

SOURCE	df	SS	MS	F	p
Treatment	*	****	****	37.99	0.000
Error	**	8168	628		
Total	**	****			

- i. For the shear strength of birch plywood data, what are the error degrees of freedom?
- A. 14 B. 12 C. 13 D. 2 E. 8
- ii. For the shear strength of birch plywood data, what is the standard error of the difference between the mean of Glue C and the mean of Glue F?
- A. 3675.6 B. 16.81 C. 282.6 D. 6.265 E. 60.63
- iii. What is the sum of squares for Glue?
- A. 8168 B. 3182.5 C. 100592 D. 47737 E. 3456.2
2. Pott (1992) looked at the effect of feeding dietary molybdenum (Mo) on the Mo concentration in sheep kidneys. 20 sheep were randomly allocated to one of the four treatment groups.

Group	Treatment	Mo concentration (ppm)					
1	0	2.8	3.2	3.6	3.4	4.1	
2	15	4.6	2.7	4.4	3.3	5.1	
3	30	4.5	8.6	5.0	4.9	3.7	
4	45	7.8	5.9	9.2	9.7	15.7	

What is the Treatment sum of squares?

- A. 192458 B. 10.129 C. 3.183 D. 118.998 E. 39.67

3. A pot trial examining tree grafting was set out as a completely randomised design experiment with six treatments. The treatments had different numbers of replicates as follows:

Treatment A I	3
Treatment S38	3
Treatment G_9	5
Treatment A2	6
Treatment GH	4
Control	8

What are the error DF?

- A. 6 B. 28 C. 23 D. 5 E. 29

Long Answer Questions

1. Independent random samples were selected from three populations with results as shown in the table.

Sample 1	Sample 2	Sample 3
2.1	4.4	1.1
3.3	2.6	.2
0.2	3.0	2.0
	1.9	

Calculate an ANOVA table, and use the results to test the hypothesis that there is no difference in the means against the alternative hypothesis that at least one population mean is different from the other two.

2. A partially completed ANOVA table for a completely randomised design is shown:

SOURCE	Df	SS	MS	<i>F</i>	<i>p</i>
Treatments	4	24.7	—	—	
Error	—	—	—		
Total	34	62.4			

- Complete the ANOVA table.
 - How many treatments are involved in the experiment?
 - Do the data provide enough evidence to indicate a difference between the population means?
3. Nine Pigs were used in a feeding trial where different levels of vitamin B₁₂ were added to the diet (11, 22 and 44 kg⁻¹ of feed). The daily average weight gain (gm day⁻¹) were as follows.

B ₁₂ mg kg ⁻¹ feed	Weight gain (gm day ⁻¹)		
Treat A (11)	689	708	699
Treat B (22)	739	712	699
Treat C (44)	653	689	739

Calculate the Between Group sum of squares

4. In a trial studying the mortality of typhoid strains, 224 mice were inoculated with three strains of typhoid and the number of days the mice survived recorded. The analysis of the variance table is:

Source	DF	SS	MS	F	P
Strains	2	359.79	179.90	31.10	0.000
Error	221	1278.42	5.78		
Total	223	1638.21			

Strain	Number	Mean
9D	31	4.032
11C	60	7.367
DSC1	133	7.979

What is the standard error of the difference between the means for strain 9D and 11C?

5. The standard error of the difference between the mean for 11C and DSC1 is 0.374. What is the value of the t statistic to test if the means are significantly different?
6. Twenty people, patients with high systolic blood pressure, were randomly allocated to one of five treatment protocols: a control using a standard drug, and four others with the standard and various other components. The ANOVA is as follows:

Source	DF	SS	MS	F	P
Drug	4	388.70	97.18	15.10	0.000
Error	15	96.50	6.43		
Total	19	485.20			

Drug	N	Mean
S	4	25.00
S+AL	4	15.75
S+AH	4	11.50
S+BL	4	19.00
S+BH	4	17.75

The t statistic to test for the difference between the mean of 'S' (25.00) and the mean of 'S+BL' (19.00) is 3.35. Determine if the means are significantly different, **and** at what level. Justify your answer.

Answers Completely Randomised Design:
Multi choice

1. i) C 13

$$df_{treatment} = (3-1) = 2; df_{total} = (16-1) = 15; df_{error} = (16-3) = 13$$

$$sed = \sqrt{MS_{error} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{(628 \left(\frac{1}{4} + \frac{1}{5} \right))} = 16.81$$

ii) B 16.81

iii) D 47737

$$SS_{Glue} = \sum \left(\frac{glue\ total^2}{glue\ n} \right) - \frac{grand\ total^2}{total\ n}$$

$$= \left(\frac{532^2}{7} + \frac{401^2}{4} + \frac{1004^2}{5} \right) - \frac{1937^2}{16} = 47737.3875$$

2. D 118.998 Enter data into calculator in STAT mode to get totals.

Use formula:

$$SS_{groups} = \sum \left(\frac{group\ total^2}{number\ in\ group} \right) - CF$$

$$= \frac{17.1^2 + 20.1^2 + 26.7^2 + 48.3^2}{5} - \frac{112.2^2}{20} = 118.998$$

3. $n = 29, k = 6, \Rightarrow df_{error} = 23$

Long answers

1 Calculations: $n = 10, k = 3 \Rightarrow df = 2, 7, 9$ respectively

$$CF = \frac{(\sum all\ entries)^2}{number\ of\ entries} = \frac{(\sum y)^2}{n}$$

$$= \frac{(20.8)^2}{10} = 43.264$$

$$SS_{treatment} \stackrel{10}{=} SS_{BG} = SS_{Between\ Groups}$$

$$= \sum \left(\frac{group\ total^2}{number\ in\ group} \right) - CF$$

$$= \frac{(5.6)^2}{3} + \frac{(11.9)^2}{4} + \frac{(3.3)^2}{3} - \frac{20.8^2}{10}$$

$$= 49.486 - 43.264 = 6.222$$

$$SS_{Total} = \sum (each\ entry)^2 - CF$$

$$= 2.1^2 + 3.3^2 + 0.2^2 + \dots + 0.2^2 + 2.0^2 - \frac{20.8^2}{10}$$

$$= 59.32 - 43.264 = 16.056$$

$$SS_{Error} = SS_{Total} - SS_{Between\ Groups}$$

$$= 16.056 - 6.222 = 9.834$$

$$MS_{treatment} = \frac{6.222}{2} = 3.111 \quad MS_{error} = \frac{9.834}{7} = 1.405 \quad F = \frac{3.111}{1.405} = 2.21$$

ANOVA:

SOURCE	df	SS	MS	F	p
Treatments	2	6.222	3.111	2.21	> 0.1
Error	7	9.834	1.405		
Total	9	16.056			

To determine p-value : (df = 2,7), so from tables,

$$F_{0.1} = 3.26$$

$$F_{0.05} = 4.74$$

$$F_{0.01} = 9.55$$

Since calculated F value ($F_{\text{calc}} = 2.21$) is less than F_{table} for $\alpha = 0.1$, we write $p > 0.1$
 \Rightarrow there is not enough evidence to reject H_0 .

That is, assuming no difference between the treatments, (i.e., H_0 is True), there is a greater than 10% probability of getting these results by chance. This is not small enough to be able to reject H_0 . (Using a significance level of 0.05 we would reject H_0 only if the p-value was smaller than 0.05).

2. a)

SOURCE	df	SS	MS	F	p
Treatments	4	24.7	6.175	4.91	$0.001 < p < 0.01$
Error	30	37.7	1.257		
Total	34	62.4			

b) $df_{\text{treatment}} = 4$, $k = 5$ (number of treatments = 5).

c) Calculated $F = 4.91$, $F_{\text{table}} = 4.02$ for $\alpha = 0.01$; $F_{\text{table}} = 6.12$ for $\alpha = 0.001$
Hence, we can reject H_0 , that there is no difference among the population means at $\alpha = 0.01$, but not at $\alpha = 0.001$ Hence $0.001 < p < 0.01$

3 Calculate totals for the treatments:

B ₁₂ mg kg ⁻¹ feed	Weight gain (gm day ⁻¹)			Totals
Treat A (11)	689	708	699	2096
Treat B (22)	739	712	699	2150
Treat C (44)	653	689	739	2081
Total				6327

Calculate the sum of squares between the groups – that is between treatments.

$$SS_{\text{treatment}} = \frac{2096^2}{3} + \frac{2150^2}{3} + \frac{2081^2}{3} - \frac{6327^2}{9} = 878$$

Since the number of pigs in each group is the same, this could also be done:

$$SS_{\text{treatment}} = \frac{(2096^2 + 2150^2 + 2081^2)}{3} - \frac{6327^2}{9} = 878$$

$$4. \quad sed = \sqrt{\left(EMS \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)} = \sqrt{\left(5.78 \times (31^{-1} + 60^{-1})\right)} = 0.532$$

$$5. \quad t = \frac{\bar{y}_1 - \bar{y}_2}{sed} = \frac{(7.367 - 7.979)}{0.374} = -1.64$$

6. The t table for 15 DF lists 2.947 under the 0.005 column. The 3.35 is larger than this but not as large as the 3.733 under the 0.001 column. This is a two sided test, so double the tail value and conclude that the means are significantly different at the 0.01 level. (i.e., $p < 0.01$)