One-way analysis of variance
(one between-subjects factor)

ANOVA summary table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic*
Between groups	SS_B	$k-1$	MS_B	$F=\frac{MS_B}{MS_W}$
Within groups	SS_W	$n-k$	MS_W	
Total	SS_T	$n-1$		

Sums of squares

$$SS_T = \sum_i \sum_j x_{ij}^2 - \frac{T^2}{n}$$

$$SS_B = \sum_i \frac{T_i^2}{n_i} - \frac{T^2}{n}$$

$$SS_W = \sum_i \sum_j \left(x_{ij} - \bar{x}_i\right)^2$$

$$\left[SS_W = SS_T - SS_B\right]$$

What the letters stand for

k = number of levels of the factor

n = total sample size $= \sum_i n_i$

n_i = size of ith sample (i.e. for ith level of factor)

x_{ij} = jth member of ith sample

$T = \sum_i \sum_j x_{ij}$ = overall total

$T_i = \sum_j x_{ij}$ = total of ith sample

$\bar{x}_i$ = mean of ith sample

Worked example model: $x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim$ independent $\mathrm{N}\left(0, \sigma^2\right)$

(NB the use of computers allows much larger samples to be worked with easily)

To see whether the mean height of women varies with ethnic background, a random sample of adult women have their heights measured, with the following results.

Ethnic background	**Height (cm)**						T_i	n_i	$\bar{x}_i$
White British	161.7	154.4	165.8	173.6	173.0	177.0	1005.5	6	167.583
Black British	154.7	151.8	167.9	161.0	155.7		791.1	5	158.22
Asian British	162.5	154.1	137.5	169.3			623.4	4	155.85
						TOTAL	2420.0	15	

The population variance of the heights for each group is assumed to be the same. The populations are assumed to be Normally distributed.

Null hypothesis: The population mean height for each group is the same.
Alternative hypothesis: At least one population mean height differs from the others.

$k=3$, $n=15$, $T=2420.0$

ANOVA table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic*
Between groups	403.107	2	201.553 5	2.215
Within groups	1092.106	12	91.008 9	
Total	1495.213	14		

The test statistic has an F distribution with parameters (2, 12) and is not significant at the 5% level (The upper 5% point is 3.89). There is insufficient evidence of a difference in (population) mean heights between age groups.

† Mean squares
Each mean square is calculated by dividing the sum of squares by the degrees of freedom.

***Distributions of test statistics**
All test statistics in the table above have an F distribution with parameters "degrees of freedom of numerator" and "degrees of freedom of denominator".

Two-way analysis of variance
(no interaction)

ANOVA summary table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic*
Factor A	SS_{A}	$a-1$	MS_{A}	$F=\frac{MS_{\text{A}}}{MS_{\text{R}}}$
Factor B	SS_{B}	$b-1$	MS_{B}	$F=\frac{MS_{\text{B}}}{MS_{\text{R}}}$
Residual	SS_{R}	$n-a-b+1$ (by subtraction)	MS_{R}	
Total	SS_{T}	$n-1$		

Sums of squares

$$SS_{\text{T}}=\sum_i\sum_j\sum_k x_{ijk}{}^2-\frac{T^2}{n}$$

$$SS_{\text{A}}=\sum_i\frac{T_i^2}{rb}-\frac{T^2}{n}$$

$$SS_{\text{B}}=\sum_j\frac{T_j^2}{ra}-\frac{T^2}{n}$$

$$SS_{\text{R}}=SS_{\text{T}}-(SS_{\text{A}}+SS_{\text{B}})$$

What the letters stand for

a = number of levels of factor A
b = number of levels of factor B
r = number of data values for each combination of levels of A and B
$n=rab=$ total sample size
x_{ijk} = kth member of sample at level i of factor A & level j of factor B
$T=\sum_i\sum_j\sum_k x_{ijk}=$ overall total
$T_i=$ total of data at level i of factor A
$T_j=$ total of data at level j of factor B

Worked example model: $x_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij}$, where $\varepsilon_{ij}\sim$ independent $\text{N}\left(0,\sigma^2\right)$

(NB the use of computers allows much larger samples to be worked with easily)

Two laboratories test the calorific content of four brands of digestive biscuit. Three of each variety of biscuit are tested by each laboratory with the following results. The number of calories per biscuit is shown.

	Lab 1			**Lab 2**			T_i
Biscuit I	73.6	72.2	74.3	70.9	73.7	75.5	440.2
Biscuit II	69.3	67.6	70.1	70.6	69.6	69.9	417.1
Biscuit III	75.8	75.7	76.4	72.5	73.0	71.6	445.0
Biscuit IV	71.1	66.7	69.8	70.5	71.2	70.4	419.7
T_j			862.6			859.4	T = 1722

The population variance of the number of calories per biscuit is assumed to be the same for each combination of laboratory and biscuit brand. It is assumed that there is no interaction between the factors (see page 5 for "with interaction"). The populations are assumed to be Normally distributed.

Null hypotheses: (a) The population mean number of calories per biscuit is the same for each brand.
(b) The population mean number of calories per biscuit is the same for each of the laboratories.
Alternative hypotheses: (a) At least one population mean calorie count differs from the other brands.
(b) At least one population mean calorie count differs from the other laboratories.

Factor A is biscuit brand. Factor B is laboratory. $a=4$, $b=2$, $r=3$, $n=24$

ANOVA table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic *
Factor A	100.09	3	33.363	12.03
Factor B	0.4267	1	0.4267	0.154
Residual	52.703	19	2.774	
Total	153.22	23		

The critical value for $F_{3,19}$ at the 5% level is 3.13. 12.03 is bigger than this so there is evidence that not all brands of biscuit have the same population mean calorie count.

The critical value for $F_{1,19}$ at the 5% level is 4.38. 0.137 is less than this so there is no evidence of a difference between laboratories in respect of the population mean calorie count.

† Mean squares
Each mean square is calculated by dividing the sum of squares by the degrees of freedom.

***Distributions of test statistics**
All test statistics in the table above have an F distribution with parameters "degrees of freedom of numerator" and "degrees of freedom of denominator".

Analysis of variance for randomised blocks

ANOVA summary table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic *
Factor A (treatment)	SS_A	$a-1$	MS_A	$F=\frac{MS_A}{MS_R}$
Factor B (blocks)	SS_B	$b-1$	MS_B	$F=\frac{MS_B}{MS_R}$
Residual	SS_R	$n-a-b+1$ (by subtraction)	MS_R	
Total	SS_T	$n-1$		

Sums of squares

$$SS_T=\sum_i\sum_j\sum_k {x_{ijk}}^2-\frac{T^2}{n}$$

$$SS_A=\sum_i\frac{T_i^2}{rb}-\frac{T^2}{n}$$

$$SS_B=\sum_j\frac{T_j^2}{ra}-\frac{T^2}{n}$$

$$SS_R=SS_T-(SS_A+SS_B)$$

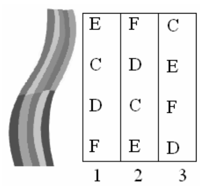
What the letters stand for

a = number of levels of factor A
b = number of levels of factor B
r = number of data values for each combination of levels of A and B
$n=rab=$ total sample size
x_{ijk} = kth member of sample at level i of factor A & level j of fac
$T=\sum_i\sum_j\sum_k x_{ijk}=$ overall total
$T_i=$ total of data at level i of factor A
$T_j=$ total of data at level j of factor B

Worked example model: $x_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij}$, where $\varepsilon_{ij}\sim$ independent $N(0,\sigma^2)$

(NB the use of computers allows much larger samples to be worked with easily)
Four varieties of garden pea (C, D, E and F) are planted in a randomised block design in strips in a field which has a stream flowing down one side. The treatment factor is pea variety. The blocking (nuisance) factor is distance from the stream. The mean yield per plant (grammes) is shown in the table below.

	Strip 1	**Strip 2**	**Strip 3**	T_i
Pea C	294	274	305	873
Pea D	324	335	256	915
Pea E	322	278	286	886
Pea F	263	280	285	828
T_j	1203	1167	1132	T =3502



The population variance of the mean yield is assumed to be the same for each combination of variety and strip. It is assumed that there is no interaction between the factors. The populations are assumed to be Normally distributed. $a=4$, $b=3$, $r=1$, $n=12$.

Null hypotheses: (a) The population mean yield is the same for each pea variety.
(b) The population mean yield is the same for each strip.
Alternative hypotheses: (a) At least one population mean yield for a variety differs from the other varieties.
(b) At least one population mean yield for a strip differs from the other strips.

ANOVA table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic*
Variety	1311	3	437	0.526
Strips	630.17	2	315.08	0.3796
Residual	4980.5	6	830.08	
Total	6831.7	11		

The critical value for $F_{3,11}$ at the 5% level is 3.59. 0.526 is less than this so there is no evidence of a difference in population mean yield between pea varieties.

The critical value for $F_{2,11}$ at the 5% level is 3.98. 0.3796 is less than this so there is no evidence of a difference between strips in respect of population mean yield.

† Mean squares
Each mean square is calculated by dividing the sum of squares by the degrees of freedom.

***Distributions of test statistics**
All test statistics in the table above have an F distribution with parameters "degrees of freedom of numerator" and "degrees of freedom of denominator".

Analysis of variance for Latin square

ANOVA summary table

Source	Sum of squares	Degrees of freedom	Mean square †	Test statistic *
Rows	SS_{rows}	$n-1$	MS_{rows}	$F=\frac{MS_{\text{rows}}}{MS_{\text{R}}}$
Columns	SS_{cols}	$n-1$	MS_{cols}	$F=\frac{MS_{\text{cols}}}{MS_{\text{R}}}$
Treatments	SS_{treats}	$n-1$	MS_{treats}	$F=\frac{MS_{\text{treats}}}{MS_{\text{R}}}$
Residual	SS_{R}	$(n-1)(n-2)$ (by subtraction)	MS_{R}	
Total	SS_{T}	n^2-1		

Sums of squares

$$SS_{\text{T}} = \sum_i \sum_j x_{ij}{}^2 - \frac{T^2}{n^2}$$

$$SS_{\text{rows}} = \sum_r \frac{T_r{}^2}{n} - \frac{T^2}{n^2}$$

$$SS_{\text{cols}} = \sum_c \frac{T_c{}^2}{n} - \frac{T^2}{n^2}$$

$$SS_{\text{treats}} = \sum_k \frac{T_k{}^2}{n} - \frac{T^2}{n^2}$$

$$SS_{\text{R}} = SS_{\text{T}} - (SS_{\text{rows}} + SS_{\text{cols}} + SS_{\text{treats}})$$

What the letters stand for

n = number of rows (or columns)

x_{ij} = observation in row i and column j

$T = \sum_i \sum_j x_{ij}$ = overall total

T_r = total of data in row r

T_c = total of data in column c

T_k = total of data at one level of the treatment

Worked example model: $x_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ij(k)}$, where $\varepsilon_{ij(k)} \sim$ independent $\text{N}\left(0, \sigma^2\right)$

An experiment to test whether four varieties of potato give the same mean yield is carried out in a square field subdivided into a square grid of plots of equal size. It is thought that the field might have natural "fertility gradients" both across and down it. To allow for this possibility, a Latin square design is used. The table shows the layout of the field and the yield in kg per plot. The rows and columns represent the possible fertility gradients (these are nuisance factors). The letters A, B, C, D represent the varieties of potato (the factor of interest). n = 4.

					Row Total
	A 19.8	B 21.2	D 22.0	C 18.6	81.6
	D 21.3	A 21.0	C 18.8	B 18.7	79.8
	B 20.8	C 18.3	A 20.7	D 17.1	76.9
	C 20.4	D 16.1	B 18.4	A 14.3	69.2
Col Total	82.3	76.6	79.9	68.7	307.5

Variety	**Total**
A	75.8
B	79.1
C	76.1
D	76.5
Total	307.5

Null hypotheses: (a) The population mean yield is the same for all varieties.
(b) The population mean yield is the same in each row (i.e. no fertility gradient in this direction).
(c) The population mean yield is the same in each column (i.e. no fertility gradient in this direction)
Alternative hypotheses: Each null hypothesis has a corresponding alternative hypothesis that at least one population mean yield differs from the others

ANOVA table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic*
Rows	22.447	3	7.482	2.839
Columns	26.372	3	8.791	3.335
Treatments (varieties)	1.712	3	0.571	0.217
Residual	15.814	6	2.636	
Total	66.344	15		

The critical value for $F_{3,6}$ at the 5% level is 4.76.
2.839 and 3.335 are each smaller than 4.76 so there is no evidence of a significant difference in mean yields between rows or columns

0.217 is also smaller than 4.76 so there is no evidence of difference in mean yield between varieties.

NOTE The Latin square is also useful for situations where there are two "real-interest" factors and one "nuisance" or three "real-interest" factors, provided all factors have the same number of levels and there is no interaction. Simply use the rows and/or columns to represent the additional real-interest factor(s).

† Mean squares
Each mean square is calculated by dividing the sum of squares by the degrees of freedom.

***Distributions of test statistics**
All test statistics in the table above have an F distribution with parameters "degrees of freedom of numerator" and "degrees of freedom of denominator".

Two-way analysis of variance (with interaction)
(two between-subjects factors)

ANOVA summary table

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic *
Factor A	SS_{A}	$a-1$	MS_{A}	$F=\frac{MS_{\text{A}}}{MS_{\text{R}}}$
Factor B	SS_{B}	$b-1$	MS_{B}	$F=\frac{MS_{\text{B}}}{MS_{\text{R}}}$
A×B interaction	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$F=\frac{MS_{\text{AB}}}{MS_{\text{R}}}$
Residual	SS_{R}	$n-ab$ (by subtraction)	MS_{R}	
Total	SS_{T}	$n-1$		

Sums of squares

$$SS_{\text{T}}=\sum_i\sum_j\sum_k x_{ijk}{}^2-\frac{T^2}{n}$$

$$SS_{\text{A}}=\sum_i\frac{T_i^2}{n_i}-\frac{T^2}{n}$$

$$SS_{\text{B}}=\sum_j\frac{T_j{}^2}{n_j}-\frac{T^2}{n}$$

$$SS_{\text{G}}=\sum_g\frac{T_g{}^2}{n_g}-\frac{T^2}{n}$$

$$SS_{\text{AB}}=SS_{\text{G}}-(SS_{\text{A}}+SS_{\text{B}})$$

$$SS_{\text{R}}=SS_{\text{T}}-(SS_{\text{A}}+SS_{\text{B}}+SS_{\text{AB}})$$

What the letters stand for

a = number of levels of factor A
b = number of levels of factor B
n = total sample size
x_{ijk} = kth member of sample at level i of factor A & level j of factor B
$T=\sum_i\sum_j\sum_k x_{ijk}$ = overall total
T_i = total of data at level i of factor A
n_i = number of data items at level i of factor A
T_j = total of data at level j of factor B
n_j = number of data items at level j of factor B
T_g = total of all the data at a particular level of A and of B (there are ab such groups)
n_g = number of data values at a particular level of A and of B

Worked example model: $x_{ijk}=\mu+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijk}$, where $\varepsilon_{ijk}\sim$ independent $\text{N}\left(0,\sigma^2\right)$

(NB the use of computers allows much larger samples to be worked with easily)

In an experiment to test whether students achieve similar marks on computer based tests to those they achieve on paper based test, a random sample of students from three classes sit either a paper based or a computer based test (the questions on both tests are the same). The marks they achieve are shown:

		Class			**Test mode totals**
		Class 1	Class 2	Class 3	
Test mode	Computer	33 38 61 69 Total 201	62 57 63 41 Total 223	62 44 39 45 Total 190	**614**
	Paper	53 47 36 47 Total 183	37 42 45 52 Total 176	30 42 27 35 Total 134	**493**
Class totals		**384**	**399**	**324**	**1107**

The variance for each section of the population is assumed to be the same. The population is assumed normal.

Factor A is "test mode"; this is the factor of interest. Factor B is "class"; this is a nuisance factor.
$a=2$, $b=3$, $n=24$, $n_i=12$ (all i), $n_j=8$ (all j), $n_g=4$ (all g).

† Mean squares
Each mean square is calculated by dividing the sum of squares by the degrees of freedom.

***Distributions of test statistics**
All test statistics in the table above have an F distribution with parameters "degrees of freedom of numerator" and "degrees of freedom of denominator".

See next page for ANOVA table and conclusion.

† Mean squares
Each mean square is calculated by dividing the sum of squares by the degrees of freedom.

***Distributions of test statistics**
All test statistics in the table above have an F distribution with parameters "degrees of freedom of numerator" and "degrees of freedom of denominator".

ANOVA table for two-way analysis of variance (two between-subjects factors)

Null hypotheses: (a) The population mean test mark is the same for either mode of test (paper or computer).
(b) The population mean test mark is the same for each of the classes
(c) There is no interaction between test mode and class in respect of the population mean test mark.
Alternative hypothesis: (a) The population mean test marks differ for the test modes.
(b) At least one population mean test mark differs from the other classes.
(c) There is some interaction between test mode and class in respect of the population mean test mark.

Source	Sum of squares	Degrees of freedom	Mean square†	Test statistic*	Distribution of test statistic	Critical value at 5% level
Test mode	610.04	1	610.04	5.706	$F_{1,18}$	4.41
Class	393.75	2	196.875	1.842	$F_{2,18}$	3.55
A×B interaction	98.585	2	49.2925	0.461	$F_{2,18}$	3.55
Residual	1924.25	18	106.903			
Total	3026.625	23				

$SS_G = 1102.375$

5.706 > 4.41 so there is evidence that there is a difference in the mean marks for different modes of test.

1.842 < 3.55 so there is no evidence of difference in mean marks between the classes.

0.461 < 3.55 so there is no evidence of interaction between class and mode of test.

† Mean squares
Each mean square is calculated by dividing the sum of squares by the degrees of freedom.

***Distributions of test statistics**
All test statistics in the table above have an F distribution with parameters "degrees of freedom of numerator" and "degrees of freedom of denominator".

Kruskal-Wallis one-way analysis of variance

Calculation of test statistic	What the letters stand for	Distribution of test statistic
Rank all data values, sum the ranks for each group and square these values. Test statistic, H, is: $H=\frac{12}{N(N+1)}\left(\sum\frac{R_j^2}{N_j}\right)-3(N+1)$	$N_j =$ size of sample j, $N=\sum N_j$, $R_j =$ sum of ranks for sample j K = number of groups	χ^2_{K-1}

Worked example
(NB the use of computers allows much larger samples to be worked with easily)

An investigation compares customer waiting times in three branches of a bank. The queuing time of a random sample of customers in each branch is measured with the following results (in minutes).

Branch	Waiting times (mins)					
A	0	0	2.9	8.2	0	6.0
B	7.3	0.9	0.7	2.4	4.0	
C	8.4	4.4	0.8	7.0	0.3	

Null hypothesis: The mean waiting times for all three branches are the same.
Alternative hypothesis: The mean waiting times for all three branches are not the same.

Ranks in bold

Branch	Waiting times (mins)						Sum of ranks (R_j)	Size of sample (N_j)
A	0 **2**	0 **2**	2.9 **9**	8.2 **15**	0 **2**	6.0 **12**	42	6
B	7.3 **14**	0.9 **7**	0.7 **5**	2.4 **8**	4.0 **10**		44	5
C	8.4 **16**	4.4 **11**	0.8 **6**	7.0 **13**	0.3 **4**		50	5

$N=16, \quad K=3$

$$H=\frac{12}{N(N+1)}\left(\sum\frac{R_j^2}{N_j}\right)-3(N+1)$$

$$H=\frac{12}{16\times 17}\left(\frac{42^2}{6}+\frac{44^2}{5}+\frac{50^2}{5}\right)-3\times 17$$

$$H\approx 1.1118$$

The 5% critical value for χ^2_2 is 5.991.

1.1118 < 5.991 so there is insufficient evidence, at the 5% level, that there is any difference in mean waiting times between the three branches.

Friedman's two-way analysis of variance by rank

Calculation of test statistic	What the letters stand for	Distribution of test statistic
Put the data in a table. To test whether there is a significant difference between the sets of data in the columns, rank each row and sum the ranks for each column. Test statistic, *M*, is: $M = \frac{12}{NK(K+1)}\sum_{i=1}^{K} R_i^2 - 3N(K+1)$	N = size of each sample (number of data items in each column), K = number of samples (number of data items in each row), R_i = sum of ranks for sample i (sum of ranks of column i)	χ^2_{K-1}

Worked example

Five judges give a mark out of 15 to each of three paintings. Is there a significant difference between the judges?

Painting \ Judge	A	B	C	D	E
sunset 1	7	5	8	6	9
sunset 2	9	8	10	8	11
sunrise	13	13	14	13	13

Null hypothesis: There is no significant difference between the judges.
Alternative hypothesis: There is a significant difference between the judges.

The factor of interest is whether the judges agree. The "nuisance" factor is "painting".

Ranking the rows

Painting \ Judge	A	B	C	D	E
sunset 1	7 **3**	5 **1**	8 **4**	6 **2**	9 **5**
sunset 2	9 **3**	8 **1.5**	10 **4**	8 **1.5**	11 **5**
sunrise	13 **2.5**	13 **2.5**	14 **5**	13 **2.5**	13 **2.5**
rank sum	**8.5**	**5**	**13**	**6**	**12.5**

$N = 3, \quad K = 5$

$$M = \frac{12}{NK(K+1)}\sum_{i=1}^{K} R_i^2 - 3N(K+1)$$

$$M = \frac{12}{3\times5\times6}\left(8.5^2 + 5^2 + 13^2 + 6^2 + 12.5^2\right) - 3\times3\times6$$

$$M = 7.1\dot{3}$$

The critical value at the 5% level for χ^2_4 is 9.488.

$7.1\dot{3} < 9.488$ so there is insufficient evidence of a significant difference in the marks given by the five judges.