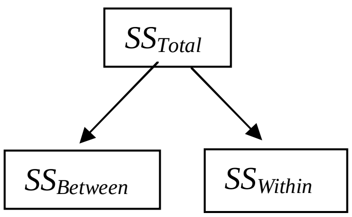
Computational Formulas for ANOVA

One-Way ANOVA

Let a = # of levels of the independent variable = # of groups
N = total # of observations in the experiment
n_1 = # of observations in group 1, etc.

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_a$

ANOVA analyzes sample variances to draw inferences about population means. Sample variances can always be calculated as *SS/df* and these sample variances are called mean squares (*MS*):



$$SS_{Total} = \sum X^2 - \frac{\left(\sum X\right)^2}{N} \qquad df_{Total} = N - 1$$

$$SS_{Between} = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} + \ldots + \frac{\left(\sum X_a\right)^2}{n_a} - \frac{\left(\sum X\right)^2}{N} \qquad df_{Between} = a - 1$$

$$s^2 = \frac{SS}{df} = MS \qquad F = \frac{MS_{Between}}{MS_{Within}}$$

$$SS_{Within} = SS_{Total} - SS_{Between} \qquad df_{Within} = N - a$$

Example.

	X_1 Placebo	X_2 Drug A	X_3 Drug B	
	9	5	2	
	8	4	4	
	8	5	3	
	6	8	1	
	9	3	5	
Sum	40	25	15	80
M	8	5	3	5.333
s	1.224745	1.870829	1.581139	

ANOVA Summary Table

Source	*SS*	*df*	*MS*	*F*	*p*
Between	63.333	2	31.667	12.67	0.0011
Within	30.000	12	2.500		
Total	93.333	14	6.667		

$SS_T = 9^2 + 8^2 + \ldots + 1^2 + 5^2 - (80)^2/15 = \mathbf{93.333}$
$SS_B = (40^2 + 25^2 + 15^2)/5 - 80^2/15 = \mathbf{63.333}$
$SS_W = 93.333 - 63.333 = \mathbf{30.000}$

An alternative computational approach emphasizing the conceptual basis of ANOVA is given below.

$$\hat{\sigma}_T^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N - 1} = MS_{Total}$$

This is the variance of all scores in the experiment = 6.667.

$$\hat{\sigma}_W^2 = \frac{s_1^2 + s_2^2 + \ldots + s_a^2}{a} = MS_{Within}$$

This is the average of the variances within the groups = 2.50.
$(1.22^2 + 1.87^2 + 1.58^2)/3 = 2.50$.

$$\hat{\sigma}_B^2 = n\hat{\sigma}_M^2 = MS_{Between}$$

This is *n* times the variance of the means = 5(6.333) = 31.667.

Multiple Comparisons:

$$LSD = t_{Crit}\sqrt{\frac{2MS_{Error}}{n}}$$

t_{Crit} is the critical value from a *t*-table using the *df* of the error term from the ANOVA table. The error term is always the denominator of the *F*-ratio. Thus, in the above example, the error *df* would be 12. The MS_{Error} would be 2.50; *n* is always the number of observations each mean you're comparing is based on.

Two-Way Factorial ANOVA

2 × 3 factorial design

DV = Years

		Crime (C)		
Attractiveness of Offender (A)		Forgery	Swindle	Burglary
	Attractive	3	6	2
		4	7	4
		5	8	6
	Unattractive	4	3	4
		6	4	6
		8	5	8

Let:

a = # of levels of the independent variable A
c = # of levels of the independent variable C
ac = # of cells in the experiment
N = total # of observations in the experiment
n_1 = # of observations in cell 1, etc.

Table of Totals

	Forgery	Swindle	Burglary	Marginal Totals
Attractive	12	21	12	45
Unattractive	18	12	18	48
Marginal Totals	30	33	30	93

Table of Means

	Forgery	Swindle	Burglary	Marginal Means
Attractive	4	7	4	5
Unattractive	6	4	6	5.33
Marginal Means	5	5.5	5	5.17

$$SS_{Total} = \sum X^2 - \frac{\left(\sum X\right)^2}{N}$$

$$SS_{Between} = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} + \ldots + \frac{\left(\sum X_{ac}\right)^2}{n_{ac}} - \frac{\left(\sum X\right)^2}{N}$$

$$SS_{Within} = SS_{Total} - SS_{Between}$$

$$SS_A = \sum \frac{\left(\sum \textit{ for each row}\right)^2}{n \textit{ for each row}} - \frac{\left(\sum X\right)^2}{N}$$

$$SS_C = \sum \frac{\left(\sum \textit{ for each column}\right)^2}{n \textit{ for each column}} - \frac{\left(\sum X\right)^2}{N}$$

$$SS_{AC} = SS_{Between} - SS_A - SS_C$$

$SS_T = 537 - 93^2/18 =$ **56.50**
$SS_B = 12^2/3 + 21^2/3 + \ldots - 93^2/18 =$ **26.50**
$SS_W = 56.50 - 26.50 =$ **30.00**
$SS_A = 45^2/9 + 48^2/9 - 93^2/18 =$ **0.50**
$SS_C = 30^2/6 + 33^2/6 + 30^2/6 - 93^2/18 =$ **1.00**
$SS_{A\times C} = 26.50 - 0.50 - 1.00 =$ **25.00**

ANOVA Summary Table

Source	*SS*	*df*	*MS*	*F*	*p*
A	0.50	1	0.500	0.20	0.6627
C	1.00	2	0.500	0.20	0.8214
AC	25.00	2	12.500	5.00	0.0263
Within	30.00	12	2.500		
Total	56.50	17	3.324		

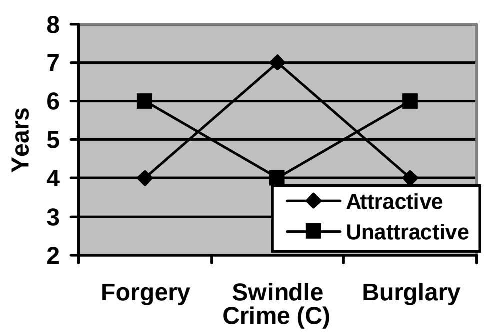
$df_A = a - 1$

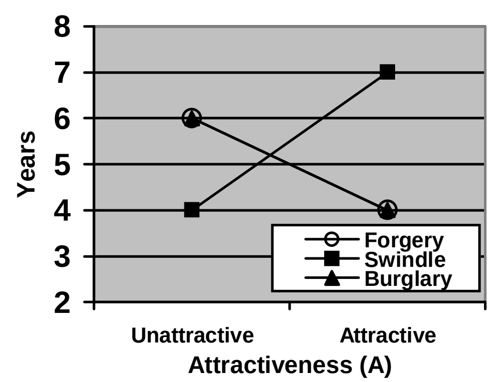
$df_C = c - 1$

$df_{AC} = (a-1)(c-1)$

$df_{Within} = N - ac$

$df_{Total} = N - 1$





Null hypotheses:

A main effect H$_0$: $\mu_A = \mu_U$

C main effect H$_0$: $\mu_F = \mu_S = \mu_B$

AC interaction H$_0$: $(\mu_{AF} - \mu_{UF}) = (\mu_{AS} - \mu_{US}) = (\mu_{AB} - \mu_{UB})$
or equivalently
H$_0$: parallel lines in the cell mean plot