One-Way Analysis of Variance (ANOVA)
"Completely Randomized Single-Factor Experiment"

Warning: Differences in notation can be encountered in different reference sources

Comparison of m (= a) Treatment Means

We consider m populations, each representing one level of treatment, with observations as shown below:

Observations

Treatments	1	2	⋯		⋯	n	
1	Y_{11}	Y_{12}	⋯	Y_{1j}	⋯	Y_{1n}	$\mu_{1.}$
2	Y_{21}	Y_{22}	⋯	Y_{2j}	⋯	Y_{2n}	$\mu_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮
i	Y_{i1}	Y_{i2}		Y_{ij}		Y_{in}	$\mu_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮
m	Y_{m1}	Y_{m2}	⋯	Y_{mj}	⋯	Y_{mn}	$\mu_{m.}$

$\mu_{i.}$ = mean of treatment i population

One-Way Analysis of Variance (ANOVA)

In the ANOVA model,

τ_i, the treatment effect,

can be indicated by $\mu_{i.} - \mu$,

so the model is

either $\quad Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

or $\quad Y_{ij} = \mu + (\mu_{i.} - \mu) + (Y_{ij} - \mu_{i.})$

Expressed another way,

$$(Y_{ij} - \mu) = (\mu_{i.} - \mu) + (Y_{ij} - \mu_{i.})$$

One-Way Analysis of Variance (ANOVA)

Now select random samples. If n observations are taken for each treatment, we have:

Observations

Treatments	1	$\cdots$	j	$\cdots$	n	Totals	Means
1	y_{11}	$\cdots$	y_{1j}	$\cdots$	y_{1n}	$T_{1.}$	$\bar{y}_{1.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
i	y_{i1}	$\cdots$	y_{ij}	$\cdots$	y_{in}	$T_{i.}$	$\bar{y}_{i.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
m	y_{m1}	$\cdots$	y_{mj}	$\cdots$	y_{mn}	$T_{m.}$	$\bar{y}_{m.}$
						$T_{..}$	$\bar{y}_{..}$

(definitions on next page)

One-Way Analysis of Variance (ANOVA)

where

$T_{i.}$ = total of observations taken under treatment i

$\bar{y}_{i.}$ = observed mean for treatment i

$T_{..}$ = sum of all observation values

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij} = \sum_{i=1}^{m} T_{i.}$$

N = total number of observations in experiment $= mn$

$\bar{y}_{..}$ = grand mean $= \frac{1}{N} \sum_{i=1}^{m} n \, \bar{y}_{i.}$ or $\frac{1}{m} \sum_{i=1}^{m} \bar{y}_{i.}$

Note that it is permissible for the number of observations for each treatment to differ, yielding n_i, $i = 1, \ldots, m$ sample sizes.

Then $N = \sum_{i=1}^{m} n_i$ and $\bar{y}_{..} = \frac{1}{N} \sum_{i=1}^{m} n_i \bar{y}_{i.}$

One-Way Analysis of Variance (ANOVA)

Substituting sample statistics for population parameters in our model (p 362-2),

$$(y_{ij} - \bar{y}_{..}) = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

This states that the deviation of any observation from the grand mean can be broken into two parts:

(1) the deviation of the treatment mean from the grand mean; i.e., $(\bar{y}_{i.} - \bar{y}_{..})$

<u>plus</u> (2) the deviation of an observation from its own treatment mean; i.e., $(y_{ij} - y_{i.})$

Squaring both sides, summing over i and j, and simplifying, we obtain:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

This result is referred to as the

Fundamental Equation of Analysis of Variance

One-Way Analysis of Variance (ANOVA)

If two independent unbiased estimates of the same variance are compared, their ratio can be shown to be F-distributed, with $(m-1), (N-m)$ degrees of freedom

Thus if

$$H_0: \mu_{1.} = \mu_{2.} = \cdots = \mu_{i.} = \cdots = \mu_{m.} \quad \text{is } \underline{\text{true}}$$

the test of hypothesis may be made using the critical region appropriate to

$$F_{(m-1),(N-m)} = \frac{\sum_{i=1}^{m} n_i (\bar{y}_{i.} - \bar{y}_{..})^2 / (m-1)}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 / (N-m)}$$

Note: H_1: At least one mean is significantly different from the others.

One-Way Analysis of Variance (ANOVA)

From the perspective of computation:

(1) Determine what is often referred to as the "correction factor" or "correction term:"

$$y_{..}^2/N = T_{..}^2/N$$

i.e., square the grand total of all observation values and divide the result by the total number of observations.

(2) Determine the Total Sum of Squares, SS_T:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}^2 - T_{..}^2/N$$

i.e., square each observation, sum the squared observations, and subtract the correction term.

(3) Determine the sum of squares <u>between</u> the treatments, or Treatment Sum of Squares, SS_τ:

$$\sum_{i=1}^{m} \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N} \qquad \text{or} \qquad \sum_{i=1}^{m} \frac{y_{i.}^2}{n_i} - \frac{T_{..}^2}{N}$$

i.e., sum the observation values for each treatment, square the sum, divide by the number of observations for the treatment, and subtract the correction factor.

(continued)

One-Way Analysis of Variance (ANOVA)

(4) Determine the sum of squares <u>within</u> the treatments, or Error Sum of Squares, SS_E:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{m}\frac{T_{i.}^2}{n_i} = SS_T - SS_\tau$$

i.e., subtract the Treatment Sum of Squares from the Total Sum of Squares

(5) Specify the degrees of freedom:

$dof_\tau = m-1$ (treatments)
$dof_E = N-m$ (errors)
$dof_T = N-1$ (total)

(6) Determine the Mean Squares:

$MS_\tau = SS_\tau / dof_\tau$
$MS = SS_E / dof_E$

(note there is <u>no</u> Mean Square for the Total Sum of Squares, SS_T)

(7) Determine the calculated test statistic:

$$F_{(m-1),(N-m)} = MS_\tau / MS_E$$

One-Way Analysis of Variance (ANOVA)

ANOVA computation results are reported in a standard format. For the one-way ANOVA:

Source	Sum of Squares	dof	Mean Squares	$\mathcal{F}$
Treatment	SS_τ	$m-1$	MS_τ	$\mathcal{F}_{test}$
Error	SS_E	$N-m$	MS_E	
Total	SS_T	$N-1$		

In the context of hypothesis testing:

$H_0: \mu_{1.} = \cdots = \mu_{i.} = \cdots \mu_{m.}$
H_1: at least one mean is not equal

For given α, identify theoretical $\mathcal{F}_{(m-1),(N-m)}$

Determine the calculated test statistic, $\mathcal{F}_{test}$
(see ANOVA table)

If $\mathcal{F}_{test} > \mathcal{F}_{theoretical}$, Reject H_0
Otherwise, Do Not Reject H_0

Example: One-Way ANOVA
"Completely Randomized Single-Factor Experiment"

A sample of values of the tensile strength of a titanium alloy sheet metal was obtained by measuring four sheets, each sheet at three different points (corner, middle, edge). The problem is to determine whether the tensile strength is the same at all points of the sheet.

Measurement Location	Sample Values			
corner	137	142	128	137
middle	140	139	117	137
edge	142	140	133	141

Three treatments $\Rightarrow m = 3$, so $i = 1, \cdots, 3$
Four sample values per treatment $\Rightarrow j = 1, \cdots, 4$

(continued)

Location	Observations				$T_{i.}$	$\sum_{j=1}^{n} y_{ij}^2$
corner	137	142	128	137	544	74086
middle	140	139	117	137	533	71399
edge	142	140	133	141	556	77334
				$T_{..} \rightarrow$	1633	222799 $\leftarrow \sum_i \sum_j y_{ij}^2$

$\sum_{i=1}^{m} \frac{T_{i.}^2}{n_i} = 222290.25$ (note in this case $n_1 = n_2 = n_3$)

$T_{..}^2/N = 222224.08$ (correction term)

$$SS_T = \sum_i \sum_j y_{ij}^2 - \frac{T_{..}^2}{N} = 222799 - 222224.08 = 574.9$$

$$SS_\tau = \sum_i \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N} = 222290.25 - 222224.08 = 66.2$$

$$SS_E = 574.9 - 66.2 = 508.7$$

(note need for precision)

Source	dof	SS	MS	$\mathcal{F}$
Location	2	66.2	33.1	0.59
Error	9	508.7	56.5	
Total	11	574.9		

$H_0: \mu_1 = \mu_2 = \mu_3$
H_1: not all μ_i are equivalent
Assume $\alpha = 0.5$, so that $\mathcal{F}_{2,9} = 4.26$ for critical value

Since $\mathcal{F}_{test}$ $(= 0.59) < \mathcal{F}_{crit}$ $(= 4.26)$
Do <u>not</u> Reject H_0
$\Rightarrow$ tensile strength is the same at all points on the sheet

Two-Way Analysis of Variance (ANOVA)
aka
Randomized Block Design

In a Randomized Block Design, there are two identified factors considered, in addition to the random error.

A matrix is created, and one observation is made per cell.

- A Randomized Complete Block Design is used when it is possible to apply all treatments in every block (e.g, a design involving the testing of 4 different tires on 4 different vehicles)

- A Randomized Incomplete Block Design is used when it is not possible to apply all treatments in every block (e.g., a design involving the testing of 6 different tires on 4 different vehicles)

- Because there is only one observation per cell in both designs, the loss of an observation must be addressed. The usual procedure is to replace a missing value with one which minimizes the sum of the squares of the errors, and to reduce the degrees of freedom for the error term accordingly.

We shall consider only the Randomized Complete Block Design.

Two-Way Analysis of Variance
Randomized Complete Block Design

The model is: $Y_{ij} = \mu + A_i + B_j + \varepsilon_{ij}$

where

Y_{ij} = the single observation on the i^{th} treatment of factor A and the j^{th} block of factor B

μ = the common effect for the entire experiment

A_i = the effect of the i^{th} treatment of factor A

B_j = the effect of the j^{th} block of factor B

ε_{ij} = the random error present in Y_{ij}

Note that ε_{ij} is actually comprised of ε_{ij} plus an interaction effect, AB. However, in this model, the interaction effect is hopelessly confounded with the error.

Two-Way Analysis of Variance
Randomized Complete Block Design

Sample Layout:

		Factor B; Blocks					
	B_1	B_2	$\cdots$	B_j	$\cdots$	B_l	
Factor A; Treatments A_1	y_{11}	y_{12}	$\cdots$	y_{1j}	$\cdots$	y_{1l}	$T_{1.}$
A_2	y_{21}	y_{22}	$\cdots$	y_{2j}	$\cdots$	y_{2l}	$T_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	y_{i1}	y_{i2}	$\cdots$	y_{ij}	$\cdots$	y_{il}	$T_{i.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_m	y_{m1}	y_{m2}	$\cdots$	y_{mj}	$\cdots$	y_{ml}	$T_{m.}$
	$T_{.1}$	$T_{.2}$	$\cdots$	$T_{.j}$	$\cdots$	$T_{.l}$	$T_{..}$

$$T_{..} = \sum_{j=1}^{l} T_{.j} = \sum_{i=1}^{m} T_{i.}$$

Note: Treatments are randomly assigned to blocks

e.g., Let treatments be 4 brands of tire
Let block be 4 wheels on vehicle

then tire brand is randomly assigned to wheel location on vehicle

Two-Way Analysis of Variance
Randomized Complete Block Design

Computations:

Degrees of freedom: for A: $(m-1)$
for B: $(l-1)$
for Error: $(m-1)(l-1)$
for Total: $(ml-1)$

Sums of squares: correction term $= T_{..}^2/ml$

(Total) $$SS_T = \sum_{i=1}^{m}\sum_{j=1}^{l} y_{ij}^2 - T_{..}^2/ml$$

(Factor A) $$SS_A = \frac{1}{l}\sum_{i=1}^{m} T_{i.}^2 - T_{..}^2/ml$$

(Factor B) $$SS_B = \frac{1}{m}\sum_{j=1}^{l} T_{.j}^2 - T_{..}^2/ml$$

(Error) $$SS_E = SS_T - (SS_A + SS_B)$$

Mean squares: $MS_A = SS_A/dof_A$
$MS_B = SS_B/dof_B$
$MS_E = SS_E/dof_E$

Calculated test statistics:

For Factor A (Treatments) $F_{dof_A, dof_E} = MS_A/MS_E$

For Factor B (Blocks) $F_{dof_B, dof_E} = MS_B/MS_E$

Two-Way Analysis of Variance
Randomized Complete Block Design

Summary:

Source	Sums of Squares	dof	Mean Squares	F
Factor A, Treatments	$\frac{1}{l}\sum_i T_{i\cdot}^2 - T_{\cdot\cdot}^2/ml$	$m-1$	SS_A/dof_A	dof_A, dof_E MS_A/MS_E
Factor B, Blocks	$\frac{1}{m}\sum_j T_{\cdot j}^2 - T_{\cdot\cdot}^2/ml$	$l-1$	SS_B/dof_B	dof_B, dof_E MS_B/MS_E
Error, ε_{ij}	$SS_T-(SS_A+SS_B)$	$(m-1)(l-1)$	SS_E/dof_E	—
Total	$\sum_i\sum_j y_{ij}^2 - T_{\cdot\cdot}^2/ml$	$ml-1$	—	—