

Analysis of Variance (ANOVA) – Prof. Richard B. Goldstein

Assumptions: k populations are *independent* and *normally distributed* with means $\mu_1, \mu_2, \dots, \mu_k$ and *common variance* σ^2

Hypothesis:
 $H_0: \mu_1 = \mu_2 = \dots = \mu_k$
 $H_1: \text{at least two of the means are not equal}$

Notation:
 y_{ij} = j^{th} observation of i^{th} treatment or group
 $Y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$ the sum of observations of the i^{th} treatment
 $\bar{y}_{i\cdot} = \frac{Y_{i\cdot}}{n_i}$ the mean of the i^{th} group
 $Y_{\cdot\cdot} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}$ the total of all observations
 $\bar{y}_{\cdot\cdot} = \frac{Y_{\cdot\cdot}}{N}$ the mean of all $N = n_1 + n_2 + \dots + n_k$ observations

Model:
 $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ where $\mu = \frac{\sum_{i=1}^k \mu_i}{k}$, α_i is called the *effect* of the i^{th} treatment
 and $\mu_i = \mu + \alpha_i$

Sum of Sqs:
 $\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^k n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2$

$$SST = SSA + SSE \text{ (also } SS_{TOT} = SS_{BET} + SS_W)$$

$$SST = SS_{TOT} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{i=1}^k Y_{i\cdot}\right)^2}{N}$$

$$SSA = SS_{BET} = \sum_{i=1}^k \frac{Y_{i\cdot}^2}{n_i} - \frac{\left(\sum_{i=1}^k Y_{i\cdot}\right)^2}{N} \quad s_1^2 = MSA = \frac{SSA}{k-1} \text{ or } MS_{BET} = \frac{SS_{BET}}{k-1}$$

$$SSE = SS_W = \sum_{i=1}^k \left(\sum_{j=1}^{n_i} y_{ij}^2 - \frac{Y_{i\cdot}^2}{n_i} \right) \quad s_2^2 = MSE = \frac{SSE}{N-k} \text{ or } MS_W = \frac{SS_W}{N-k}$$

$$F = \frac{MSA}{MSE} = \frac{MS_{BET}}{MS_W} \text{ has } k-1 \text{ d.f. in numerator and } N-k \text{ d.f. in denominator}$$

$$s_1^2 \text{ overestimates } \sigma^2 \text{ when } H_1 \text{ is true : } E\left(\frac{SSA}{k-1}\right) = \sigma^2 + \frac{\sum_{i=1}^k n_i \alpha_i^2}{k-1}$$

Single Degree of Freedom Comparisons / Contrasts

$$H_0: \sum_{i=1}^k c_i \mu_i = 0 \quad \text{where} \quad \sum_{i=1}^k c_i = 0$$

$$H_1: \sum_{i=1}^k c_i \mu_i \neq 0$$

$$\text{Test Statistic } f = \frac{\left(\sum_{i=1}^k c_i \bar{y}_{i\cdot} \right)^2}{s^2 \sum_{i=1}^k \left(\frac{c_i^2}{n_i} \right)} = \frac{\left[\sum_{i=1}^k \left(\frac{c_i Y_{i\cdot}}{n_i} \right) \right]^2}{s^2 \sum_{i=1}^k \left(\frac{c_i^2}{n_i} \right)} \quad \text{where } s^2 = \text{MSE with } 1 \text{ and } N - k \text{ d.f.}$$

There can be multiple comparisons if they are orthogonal $\sum_{i=1}^k c_i b_i = 0$

Tukey's procedure - allows for all paired comparisons - there are $k(k-1)/2$ of these

Means i and j have a significant difference if $(\bar{y}_{i\cdot} - \bar{y}_{j\cdot})$ exceeds $q[\alpha, k, v] \sqrt{\frac{s^2}{n}}$

where $n = n_i \forall i$ and q is based upon the *studentized* range distribution (table A.22) or from <http://cse.niaes.affrc.go.jp/miwa/probcalc/s-range/>

Bartlett's Test for equality of variances

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

H_1 : the variances are not all equal

$$\text{Pooled estimate: } s_p^2 = \frac{\sum_{i=1}^k (n_i - 1) s_i^2}{N - k}$$

$$b = \frac{\left[(s_1^2)^{n_1-1} (s_2^2)^{n_2-1} \dots (s_k^2)^{n_k-1} \right]^{1/(N-k)}}{s_p^2}$$

[1] If $n_1 = n_2 = \dots = n_k = n$ we reject H_0 if $b < b_k(\alpha; n)$ Table A.10

[2] If sample sizes are unequal we reject H_0 if $b < b_k(\alpha; n_1, n_2, \dots, n_k)$

where $b_k(\alpha; n_1, n_2, \dots, n_k) \approx \frac{n_1 b_k(\alpha; n_1) + n_2 b_k(\alpha; n_2) + \dots + n_k b_k(\alpha; n_k)}{N}$

Websites:

about the test: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda357.htm>

Javascript: <http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/BartletTest.htm>

Randomized Complete Block Designs (RCB)

Treatment	Block:	1	2	...	j	...	b	Total	Mean
1		y ₁₁	y ₁₂	...	y _{1j}	...	y _{1b}	T _{1.}	$\bar{y}_{1.}$
2		y ₂₁	y ₂₂	...	y _{2j}	...	y _{2b}	T _{2.}	$\bar{y}_{2.}$
⋮		⋮	⋮		⋮		⋮	⋮	⋮
i		y _{i1}	y _{i2}	...	y _{ij}	...	y _{ib}	T _{i.}	$\bar{y}_{i.}$
⋮		⋮	⋮		⋮		⋮	⋮	⋮
k		y _{k1}	y _{k2}	...	y _{kj}	...	y _{kb}	T _{k.}	$\bar{y}_{k.}$
Total		T _{.1}	T _{.2}	...	T _{.j}	...	T _{.b}	T _{..}	
Mean		$\bar{y}_{.1}$	$\bar{y}_{.2}$...	$\bar{y}_{.j}$...	$\bar{y}_{.b}$		$\bar{y}_{..}$

Sum of Sqs:
$$\sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = b \sum_{i=1}^k (\bar{y}_{i.} - \bar{y}_{..})^2 + k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

SST = SSA + SSB + SSE

Source of variation	Sum of squares	Degrees of freedom	Mean Square	Computed f
Treatments	SSA	k - 1	$s_1^2 = \frac{SSA}{k-1}$	$f_1 = \frac{s_1^2}{s^2}$
Blocks	SSB	b - 1	$s_2^2 = \frac{SSB}{b-1}$	
Error	SSE	(k - 1)(b - 1)	$s^2 = \frac{SSE}{(b-1)(k-1)}$	
Total	SST	bk - 1		

H₀: $\mu_1 = \mu_2 = \dots = \mu_k = \mu$
H₁: at least two of the means are not equal