ANOVA : (univariate)

Let $X_{\ell 1}, \ldots, X_{\ell n_\ell} \overset{iid}{\sim} N(\mu_\ell, \sigma^2)$

$\ell = 1, 2, \ldots, g$

we want to test

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_g$

we reparameterize

$$\underbrace{\mu_\ell = \mu}_{\text{common / overall effect}} + \underbrace{(\mu_\ell - \mu)}_{\text{treatment effect}} = \mu + \tau_\ell$$

The H_0 can be rewritten as

$H_0 : \tau_1 = \tau_2 = \ldots = \tau_g = 0$

$X_{\ell j} = \mu + \tau_\ell + \varepsilon_{\ell j}$ → random error

notice $\varepsilon_{\ell j} \overset{iid}{\sim} N(0, \sigma^2)$

Similarly observation can be written as

$$x_{\ell j} = \underbrace{\bar{x}}_{\text{estimated overall mean}} + \underbrace{(\bar{x}_\ell - \bar{x})}_{\text{estimated treatment effect}} + \underbrace{(x_{\ell j} - \bar{x}_\ell)}_{\text{residual}}$$

→ Under H_0, the corrected sum of squares

$$\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2$$

Observe that

$$\sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2 = n_\ell (\bar{x}_\ell - \bar{x})^2 + \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2$$

$$\Rightarrow \underbrace{\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})^2}_{\substack{\text{Total (corrected)} \\ \text{sum of squares} \\ (SS_{cor})}} = \underbrace{\sum_{\ell=1}^{g} n_\ell (\bar{x}_\ell - \bar{x})^2}_{\substack{\text{Treatment (between)} \\ \text{sum of squares} \\ (SS_{tr})}} + \underbrace{\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)^2}_{\substack{\text{Residual (within)} \\ \text{sum of squares} \\ (SS_{res})}}$$

Remark: For estimating τ_i's we need a contraint that

$$\sum_{\ell=1}^{g} n_\ell \tau_\ell = 0$$

ANOVA Table

Source	Sum of Squares	Degrees of freedom
Treatment	SS_{tr}	$g-1$
Residual	SS_{res}	$n-g$
Total	SS_{cor}	$n-1$

where,

$$n = \sum_{\ell=1}^{g} n_\ell$$

→ Under H_0, Using Cocharan's Theorem

$$\left.\begin{aligned} SS_{tr} &\sim \chi^2_{g-1} \\ SS_{res} &\sim \chi^2_{n-g} \end{aligned}\right\} \text{indep}$$

$$\Rightarrow \frac{SS_{tr}/(g-1)}{SS_{res}/(n-g)} \sim F_{g-1,\,n-g}$$

Remark: For large n_ℓ $\forall \ell = 1, 2, \dots, g$, the assumption of normality can be dropped, due to CLT.

Multivariate ANOVA: (MANOVA)

$SS_{tr} / g-1$, $SS_{res} / n-g$ → random error

$\underline{\text{Pop}^n 1}$: $X_{11} \quad \dots \quad X_{1n_1} \overset{iid}{\sim} N_p(\mu_1, \Sigma)$

$\vdots$

$\underline{\text{Pop}^n g}$: $X_{g1} \quad \dots \quad X_{gn_g} \overset{iid}{\sim} N_p(\mu_g, \Sigma)$

The samples from different population are indep.

we want to test

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_g$$

against H_A: H_0 not true

Reparameterization:

$$X_{\ell j} = \underset{\substack{\downarrow \\ \text{overall} \\ \text{effect}}}{\mu} + \underset{\substack{\downarrow \\ \text{treatment} \\ \text{effect}}}{(\mu_\ell - \mu)} + \underset{\substack{\downarrow \\ \text{random} \\ \text{error}}}{\varepsilon_{\ell j}}$$

$$= \mu + \tau_\ell + \varepsilon_{\ell j}$$

→ For estimating τ_ℓ's, we need a constraint.

$$\sum_{\ell=1}^{g} n_\ell \tau_\ell = \underline{0}$$

$$\bar{X}_\ell = \frac{1}{n_\ell} \sum_{j=1}^{n_\ell} x_{\ell j} \quad ; \quad \bar{X} = \frac{1}{n} \sum_{\ell=1}^{g} n_\ell \bar{x}_\ell$$

Equivalent decomposition for sample observation:

$$x_{\ell j} = \underbrace{\bar{x}}_{\substack{\text{estimated} \\ \text{overall mean}}} + \underbrace{(\bar{x}_\ell - \bar{x})}_{\substack{\text{estimated} \\ \text{treatment} \\ \text{effect}}} + \underbrace{(x_{\ell j} - \bar{x}_\ell)}_{\substack{\text{residual} \\ \text{effect}}}$$

→ Similar algebra as in ANOVA yields:

$$\underbrace{\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x})(x_{\ell j} - \bar{x})^T}_{\substack{\text{Total corrected SS} \\ \text{and cross products} \\ \text{(SSCP)}}} = \underbrace{\sum_{\ell=1}^{g} n_\ell (\bar{x}_\ell - \bar{x})(\bar{x}_\ell - \bar{x})^T}_{\text{Treatment (between) SSCP}} + \underbrace{\sum_{\ell=1}^{g} \sum_{j=1}^{n_\ell} (x_{\ell j} - \bar{x}_\ell)(x_{\ell j} - \bar{x}_\ell)^T}_{\substack{\text{Residual (within)} \\ \text{SSCP}}}$$

$$\Rightarrow \quad T = B + W$$

Observation:

$$W = (n_1 - 1) S_1 + (n_2 - 1) S_2 + \dots + (n_g - 1) S_g \sim \text{Wishart}$$

The H_0 can be rewritten as

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_g = 0$$

MANOVA Table:

Source	SSCP	degrees of freedom
Treatment	B	$g-1$
Residual	W	$n-g$
Total	T	$n-1$

Wilk's Test:

$$\Lambda^* = \frac{|W|}{|B+W|}$$, known as Wilk's lambda

Distribution : Pg 300 (AMSA) 5th edition (Table 6.3)

Also see Bartlett test on same page

Two-way treatment effect:

$$\mu_{\ell k} = \mu + (\mu_\ell - \mu) + (\mu_k - \mu) + (\mu_{\ell k} - \mu_\ell - \mu_k + \mu)$$

	Age group 1	. . .	Age group m
Treat 1			
⋮			
Treat n			