<u>ANalysis Of VAriance</u> (ANOVA)

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$$
$$H_a: \text{not all means are equal}$$

$$F_s = \frac{\text{variation } \textit{between} \text{ samples}}{\text{variation } \textit{within} \text{ samples (pooled)}}$$

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_a: \text{not all means are equal}$$

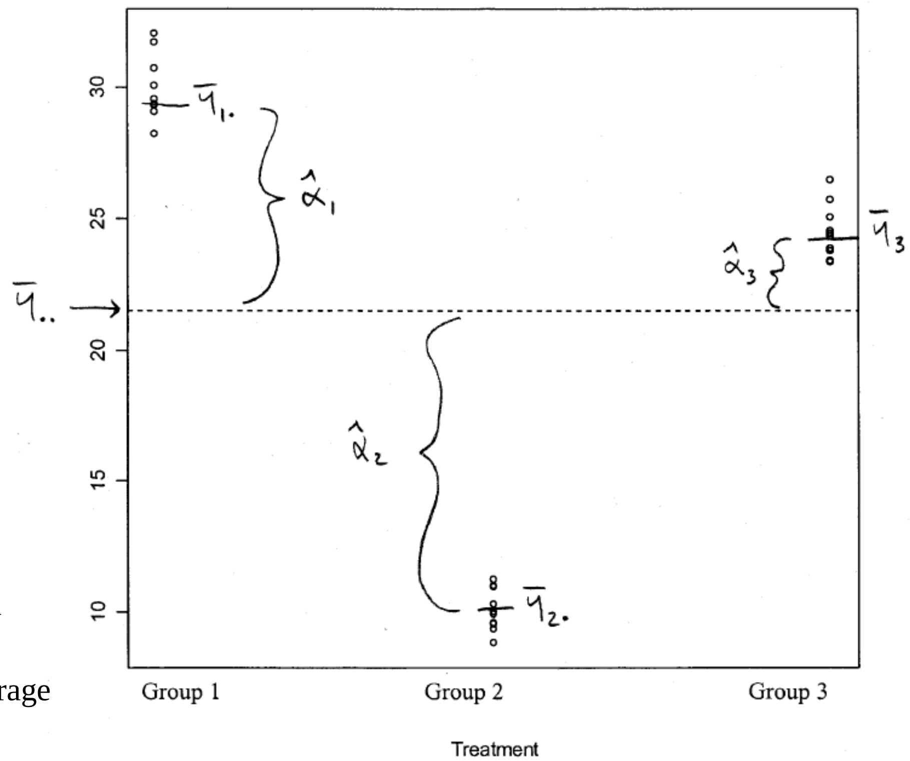
$$\left.\begin{array}{l} H_0: \mu_1 = \mu_2 \text{ at the } \alpha\text{=.05 } l.o.s. \\ H_0: \mu_1 = \mu_3 \text{ at the } \alpha\text{=.05 } l.o.s. \\ H_0: \mu_2 = \mu_3 \text{ at the } \alpha\text{=.05 } l.o.s. \end{array}\right\} \Rightarrow \begin{array}{ll} \alpha\text{=P(type I error) for each test} & \left(=\alpha_{\text{Comparison-Wise}}\right) \\ \text{P(at least one type I error)} >> \alpha & \left(=\alpha_{\text{Experiment-Wise}}\right) \end{array}$$

One-way layout

Treatments: $A_1, A_2, \ldots, A_k$

Y_{ij} : j^{th} observation of the i^{th} treatment

A_1	A_2	…	A_k	
Y_{11}	Y_{21}	…	Y_{k1}	
…	...	...	...	
Y_{1r}	Y_{2r}	…	Y_{kr}	
$Y_{1\cdot}$	$Y_{2\cdot}$	...	$Y_{k\cdot}$	Column Sum
$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$	...	$\bar{y}_{k\cdot}$	Column Average



'Dot-bar' Notation: $Y_{2\bullet} = \sum_{j=1}^{r} Y_{2j}$ $\qquad \bar{Y}_{2\bullet} = \dfrac{\sum_{j=1}^{r} Y_{2j}}{r}$

---

k treatments $\quad i = 1,2, \ldots, k$

r replicates $\quad j = 1, 2, \ldots, r$

Cell Means Model

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

Treatment Effects Model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

μ : constant

α_i : constant for the i^{th} group

$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$

$$\hat{\mu} = \bar{Y}_{\bullet\bullet}$$

$$\hat{\alpha}_i = \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$$

$$\hat{\varepsilon}_{ij} = Y_{ij} - \bar{Y}_{i\bullet}$$

Least Squares Estimates (LSE)

$$SSE = S = \sum_{i=1}^{k}\sum_{j=1}^{r}\varepsilon_{ij}^2 = \sum_{i=1}^{k}\sum_{j=1}^{r}\left(Y_{ij} - \mu - \alpha_i\right)^2$$

$$Y_{ij} - \hat{\mu} = \hat{\alpha}_i + \hat{\varepsilon}_{ij}$$

$$Y_{ij} - \bar{Y}_{\bullet\bullet} = (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}) + (Y_{ij} - \bar{Y}_{i\bullet})$$

$$\sum_i \sum_j \left[Y_{ij} - \bar{Y}_{\bullet\bullet} \right]^2 = \sum_i \sum_j \left[(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}) + (Y_{ij} - \bar{Y}_{i\bullet}) \right]^2$$

$$= \sum_i \sum_j (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + 2\sum_i \sum_j (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})(Y_{ij} - \bar{Y}_{i\bullet}) + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\bullet})^2$$

$$= \sum_i \sum_j (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\bullet})^2$$

$$SS(Total) = SS(TRT) + SS(Error)$$

$$SSY = SST + SSE$$

	Source	d.f.	SS	MS	E(MS)
(Between)	Treatment (T)	*k-1*	$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet} \right)^2$	*SST / (k-1)*	$\sigma_e^2 + \frac{r\sum_{i=1}^{k} \alpha_i^2}{k-1}$
(Within)	Error (E)	*N-k*	$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i\bullet} \right)^2$	*SSE / (N-k)*	σ_e^2
	Total	*N-1*	$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{\bullet\bullet} \right)^2$		

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

H_a : not all means are equal

$$\frac{MST}{MSE} = F_s \sim F_{t-1,N-k} \ (central) \text{ under } \mathrm{H}_0$$

$$\sim F_{k-1,N-k}^{NC(\lambda)} \ (Non-central) \text{ under } \mathrm{H}_A \qquad \lambda = r\sum_{i=1}^{k} \alpha_i^2 / \sigma_e^2$$

$$\text{Power} = P\left(F^{NC(\lambda)} > F_{\alpha}^{Central} \mid H_A \right)$$

Contrasts

$L = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k \qquad \sum_{i=1}^{k} c_i = 0$ A linear combination of *population* means

$\hat{L} = c_1\bar{Y}_{1\bullet} + c_2\bar{Y}_{2\bullet} + \cdots + c_k\bar{Y}_{k\bullet} \qquad \sum_{i=1}^{k} c_i = 0$ A linear combination of *sample* means

Vector/Matrix Notation:

$$\underset{\sim}{D} = \begin{pmatrix} d_1 \\ \vdots \\ d_k \end{pmatrix} \quad \underset{\sim}{\bar{Y}} = \begin{pmatrix} \bar{Y}_{1\bullet} \\ \vdots \\ \bar{Y}_{tk\bullet} \end{pmatrix} \quad \rightarrow \quad \hat{L} = (d_1 \cdots d_k) \begin{pmatrix} \bar{Y}_{1\bullet} \\ \vdots \\ \bar{Y}_{k\bullet} \end{pmatrix} = \sum_{i=1}^{k} d_i \bar{Y}_{i\bullet} = \underset{\sim}{D} \bullet \underset{\sim}{\bar{Y}} = \underset{\sim}{D}^T \underset{\sim}{\bar{Y}}$$

Example: 4 contrasts

$$L_1 = \mu_1 - \mu_2 \qquad L_3 = \mu_2 - \mu_3$$

$$L_2 = \mu_1 - \mu_3 \qquad L_4 = \frac{\mu_1 + \mu_2}{2} - \mu_3$$

The above set of contrasts are not mutually independent

e.g., $L_2 - \frac{1}{2}L_1 = (\mu_1 - \mu_3) - \frac{1}{2}(\mu_1 - \mu_2) = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = L_4$

A set of contrasts are linearly independent if they are orthogonal.

- Consider two vectors : $\underset{\sim}{A}$ & $\underset{\sim}{B}$

$$\underset{\sim}{A} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \underset{\sim}{B} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

$$\underset{\sim}{A} \bullet \underset{\sim}{B} = \underset{\sim}{A}^T \underset{\sim}{B} = a_1b_1 + a_2b_2 + a_3b_3 = \|\underset{\sim}{A}\|\,\|\underset{\sim}{B}\| \cos\theta \qquad \theta = 90^o \rightarrow \cos\theta = 0$$

- Two contrasts $\left\{ \begin{matrix} C = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k \\ D = d_1\mu_1 + d_2\mu_2 + \cdots + d_k\mu_k \end{matrix} \right\}$ are orthogonal if $\sum_{i=1}^{k} c_i d_i = 0$

- Which of $L_1, \ldots, L_4$ above are orthogonal?

$$L = \sum_{i=1}^{k} c_i \mu_i$$

<u>Bonferroni's Method</u> $\hat{L} = \sum_{i=1}^{k} c_i \bar{Y}_{i\bullet}$

- m <u>planned</u> contrasts
- $t_s = \dfrac{\hat{L}}{SE(\hat{L})} = \dfrac{\sum_{i=1}^{k} c_i \bar{Y}_{i\bullet}}{\sqrt{MSE \sum_{i=1}^{k} c_i^2 / r_i}}$ Reject H$_0$ if $t_s > t_{\frac{\alpha/2}{m}, N-k}$
- CI: $\hat{L} \pm t_{\frac{\alpha/2}{m}, N-k} \sqrt{MSE \sum_{i=1}^{k} c_i^2 / r_i}$

<u>Tukey(-Kramer) HSD Method</u> $\hat{L} = \bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}$

- Originally for <u>all pairwise contrasts</u> with equal sample sizes (r_i=r)

  Based on Studentized Range Distribution $Q = \dfrac{\max(\bar{Y}_{i\bullet}) - \min(\bar{Y}_{i\bullet})}{\sqrt{MSE / r}}$
- $W = \dfrac{|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}|}{\left(1/\sqrt{2}\right)\sqrt{MSE(1/r_i + 1/r_j)}}$ Reject H$_0$ if $W > q_\alpha(k, df_{Error})$
- CI: $\hat{L} \pm q_\alpha(k, df_{Error}) \left(1/\sqrt{2}\right) \sqrt{MSE(1/r_i + 1/r_j)}$

<u>Scheffé's Method</u> $\hat{L} = \sum_{i=1}^{k} c_i \bar{Y}_{i\bullet}$

- <u>all</u> possible contrasts
- CI: $\hat{L} \pm \sqrt{(k-1) F_{\alpha, k-1, N-k}} \sqrt{MSE \sum_{i=1}^{k} c_i^2 / r_i}$

<u>Example:</u>

```
        low   med  high
N        47   95     58
MEAN 50.62 51.93 55.91
SD    9.49  9.11  9.44

1) Compute a 95% CI for the average of (low & med) vs. high using Scheffe's
   method.
2) Verify that the MSE is 86.4 using the summary data in the table.
```