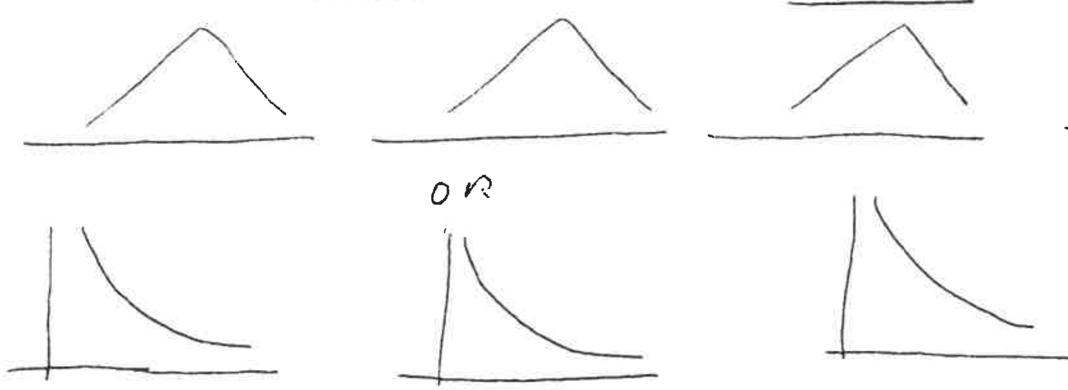
The Kruskal-Wallis test can only be used for a single factor analysis. The ANOVA F-test can be extended to studies with several factors, RCBD, etc.

<u>These tests assume only:</u>

(1) Independent obsns on the factor levels

(2) The factor level populations have approx. the same <u>shape</u> and same <u>spread</u>



shape shape
same spread

(3) the response variable data is at least ordinal (ranked)

<u>ANOVA Test on Ranks</u>

1. Rank <u>all</u> the n_T Y_{ij} from smallest to largest (using average rank for ties)

2. Find $SSTO = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_{..})^2 = \sum_{i=1}^{a} \sum_{j=1}^{n_i} R_{ij}^2 - \frac{R_{..}^2}{n_T}$

where $\boxed{R_{ij} = \text{rank of obsn } Y_{ij}}$

a

$$SSTR = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\bar{R}_{i.} - \bar{R}_{..})^2 = \sum_{i=1}^{a} \frac{R_{i.}^2}{n_i} - \frac{R_{..}^2}{n_T}$$

Carry out Hypothesis Test

$H_0: \mu_1 = \cdots = \mu_a$ vs H_A: not all μ_i are equal

test stat.: $F_R^* = \dfrac{MSTR}{MSE}$ for the RANKS of the Y_{ij}

R.R.: $F_R^* > F_{a-1, n_T-a, 1-\alpha}$

Kruskal-Wallis Test

Test statistic is

$$H = \frac{12}{n_T(n_T+1)} \sum_{i=1}^{a} \frac{R_{i.}^2}{n_i} - 3(n_T+1)$$

R.R.: Reject H_0 if $H > \chi^2_{a-1, 1-\alpha}$

N.B. H has an approx. χ^2 distn with $(a-1)$ d.f. provided the n_i are large enough ($n_i \geq 5$)

BACK TO SOME PROOFS

$$\sum_{j=1}^{n_i} e_{ij} \equiv 0 \quad \text{for each } i = 1, \cdots, a$$

proof: $\sum_{j=1}^{n_i} e_{ij} = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.}) = \sum_{j=1}^{n_i} Y_{ij} - n_i \bar{Y}_{i.}$

$$= \sum_{j=1}^{n_i} Y_{ij} - n_i \left(\frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \right) = \sum_{j=1}^{n_i} Y_{ij} - \sum_{j=1}^{n_i} Y_{ij} \equiv 0$$

$$\underline{E\{MSE\} = \sigma^2}$$

<u>proof</u> Under our model assumptions

$$E\{MSE\} = E\left\{ \frac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2}{n_T - a} \right\}$$

$$= E\left\{ \frac{\sum_{i=1}^{a} (n_i - 1) s_i^2}{n_T - a} \right\} \quad \text{where } s_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2}{n_i - 1}$$

$$= \frac{1}{n_T - a} \sum_{i=1}^{a} (n_i - 1) E\{s_i^2\}$$

= ordinary sample variance for factor level i

$$= \frac{1}{n_T - a} \sum_{i=1}^{a} (n_i - 1) \sigma^2$$

since by assumptions of independence & common variance, each s_i^2 is an unbiased estimator of σ^2

$$= \frac{\sigma^2}{n_T - a} \sum_{i=1}^{a} (n_i - 1)$$

$$= \frac{\sigma^2}{n_T - a} (n_T - a)$$

$$= \sigma^2$$

q.e.d.

$$\underline{E\{MSTR\} = \sigma^2 + \frac{\sum_{i=1}^{a} n_i (\mu_i - \mu_.)^2}{a - 1}, \quad \mu_. = \frac{\sum_{i=1}^{a} n_i \mu_i}{n_T}}$$

Note the following equivalences that will be needed.

1. $E\{\bar{Y}_{i.}\} = \mu_i \qquad Var(\bar{Y}_{i.}) = \dfrac{\sigma^2}{n_i}$

2. $E\{\bar{Y}_{..}\} = E\left\{ \dfrac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} Y_{ij}}{n_T} \right\} = \dfrac{1}{n_T} \sum_{i=1}^{a} \sum_{j=1}^{n_i} E\{Y_{ij}\}$

$$= \frac{1}{n_T} \sum_{i=1}^{a} \sum_{j=1}^{n_i} \mu_i = \frac{1}{n_T} \sum_{i=1}^{a} n_i \mu_i = \mu.$$ as defined above

3. $Var(\bar{Y}_{..}) = Var\left(\dfrac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} Y_{ij}}{n_T} \right)$

$$= \frac{1}{n_T^2} \sum_{i=1}^{a} \sum_{j=1}^{n_i} Var(Y_{ij})$$ since all Y_{ij} are independent r.v.s $\therefore$ cov $= 0$

$$= \frac{1}{n_T^2} \sum_{i=1}^{a} \sum_{j=1}^{n_i} \sigma^2$$

$$= \frac{1}{n_T^2} \sum_{i=1}^{a} n_i \sigma^2 = \frac{\sigma^2}{n_T^2} \sum_{i=1}^{a} n_i = \frac{\sigma^2 n_T}{n_T^2} = \frac{\sigma^2}{n}$$

4. Remember that in general

$$\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$$

$$E\{X^2\} = Var(X) + [E\{X\}]^2$$

<u>proof</u> of $E\{SSTR\} = \sigma^2 + \dfrac{\sum_{i=1}^{a} n_i (\mu_i - \mu_.)^2}{a-1}$

Now $E\{SSTR\} = E\left\{ \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\bar{Y}_{i.} - \bar{Y}_{..})^2 \right\}$

$$= E\left\{\sum_{i=1}^{a} n_i(\bar{Y}_{i.}-\bar{Y}_{..})^2\right\} = E\left\{\sum_{i=1}^{a} n_i\left[\bar{Y}_{i.}^2 - 2\bar{Y}_{i.}\bar{Y}_{..} + \bar{Y}_{..}^2\right]\right\}$$

$$= E\left\{\sum_{i=1}^{a} n_i \bar{Y}_{i.}^2 - 2\bar{Y}_{..}\sum_{i=1}^{a} n_i \bar{Y}_{i.} + \bar{Y}_{..}^2 \sum_{i=1}^{a} n_i\right\}$$

$$= E\left\{\sum_{i=1}^{a} n_i \bar{Y}_{i.}^2 - 2\bar{Y}_{..}Y_{..} + n_T \bar{Y}_{..}^2\right\} \qquad \left[\sum_{i=1}^{a} n_i \bar{Y}_{i.} = Y_{..}\right]$$

$$= E\left\{\sum_{i=1}^{a} n_i \bar{Y}_{i.}^2 - 2\bar{Y}_{..} n_T \bar{Y}_{..} + n_T \bar{Y}_{..}^2\right\}$$

$$= E\left\{\sum_{i=1}^{a} n_i \bar{Y}_{i.}^2 - n_T \bar{Y}_{..}^2\right\} = \sum_{i=1}^{a} n_i E\{\bar{Y}_{i.}^2\} - n_T E\{\bar{Y}_{..}^2\}$$

$$= \left.\begin{array}{l} \displaystyle\sum_{i=1}^{a} n_i\left[\mathrm{Var}(\bar{Y}_{i.}) + (E\{\bar{Y}_{i.}\})^2\right] \\ \qquad - n_T\left[\mathrm{Var}(\bar{Y}_{..}) + (E\{\bar{Y}_{..}\})^2\right] \end{array}\right\} \text{ using (4) above}$$

$$= \sum_{i=1}^{a} n_i\left[\frac{\sigma^2}{n_i} + \mu_i^2\right] - n_T\left[\underbrace{\frac{\sigma^2}{n_T}}_{\text{from (3)}} + \underbrace{\mu_.^2}_{\text{from (2)}}\right]$$

$$= a\sigma^2 + \sum_{i=1}^{a} n_i \mu_i^2 - \sigma^2 - n_T \mu_.^2$$

$$= (a-1)\sigma^2 + \sum_{i=1}^{a} n_i \mu_i^2 - n_T \mu_.^2$$

$$= (a-1)\sigma^2 + \sum_{i=1}^{a} n_i(\mu_i - \mu_.)^2$$

$$\therefore E\{MSTR\} = E\left\{\frac{SSTR}{a-1}\right\} = \sigma^2 + \frac{\sum_{i=1}^{a} (\mu_i - \mu_.)^2}{a-1}$$

q.e.d.

Comparing Factor Level Means

Once it has been determined that the treatment means do differ, you generally want to look at how they differ, That is, you want to make comparisons between them.

For example:

You might want to see whether the mean yield of a chemical increases with the amount of catalyst; or compare several drugs to a standard drug; oar compare high fibre diets to low fibre diets; etc

Thus you will want to do hypothesis tests and find confidence intervals for treatment means and comparisons of treatment means.

Contrasts

Definition

A linear combination of the treatment means $L = \sum_{i=1}^{a} c_i \mu_i$, where the c_i are constants, is a contrast

if $\sum_{i=1}^{a} c_i = 0$

Eg 1
Suppose an experiment was run to test the lifetimes of : alkaline brand name batteries (1), alkaline store brand (2), heavy duty name brand (3), and heavy duty store brand (4). The null hypothesis that the mean lifetimes were equal for all 4 brands was rejected. It is now desired to compare the average lifetimes of alkaline vs heavy duty batteries. The contrast

$L = \mu_1 + \mu_2 - (\mu_3 + \mu_4)$ will do this.

NOTE that $L_2 = (\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4)/2$ and $L_3 = 4(\mu_1 + \mu_2) - 4(\mu_3 + \mu_4)$

are both equivalent contrasts to L for this comparison.

- choose $L = \mu_1 + \mu_2 - (\mu_3 + \mu_4)$ since it is the simplest

Eg 2
Suppose that you want to compare the effect of each of 3 new drugs (1, 2, 3) with a standard drug (4). Then we want the 3 contrasts

$L_1 = \mu_1 - \mu_4$ $L_2 = \mu_2 - \mu_4$ $L_3 = \mu_3 - \mu_4$ That is 3 pairwise comparisons.

NOTE:

In eg 1 above there was interest only in a **single** comparison. In eg 2 there were 3 comparisons of interest.

There is a difference in the methods that should be used for carrying out several comparisons and those for a single comparison. Will first give the C.I.'s and hypothesis tests about a single mean or a single contrast and then look at Multiple Comparison methods.

Confidence Intervals and Hypothesis Tests for SINGLE Factor Level Means and SINGLE Contrasts

For a Single Treatment Mean (model: $Y_{ij} = \mu_i + \varepsilon_{ij}$, $\varepsilon_{ij} \sim NID(0, \sigma^2)$)

$\bar{Y}_{i.} \sim NID\left(\mu_i, \frac{\sigma_i^2}{n_i}\right)$

and $\dfrac{\bar{Y}_{i.} - \mu_i}{\sqrt{\frac{MSE}{n_i}}}$ has a T-distn. with $(n-a)$ d.f.

$\therefore$ a $100(1-\alpha)\%$ C.I. for μ_i is

$\left(\bar{Y}_{i.} \pm t_{n-a,\,\alpha/2} \sqrt{MSE/n_i}\right)$

Hypothesis Tests

$H_0: \mu_i = \mu_0$ $\quad$ μ_0 some specified value

H_A: $\mu \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$

test stat.: $T^* = \dfrac{\bar{Y}_{i.} - \mu_0}{\sqrt{MSE/n_i}}$

N.B. The above are robust to departures from normality (by C.L.T.) but NOT to unequal variances (in which case MSE/n_i does not estimate the unequal σ_i^2).

<u>Single Contrast (or L.C.)</u> $L = \sum_{i=1}^{a} c_i \mu_i$ — could be any linear combination

$\hat{L} = \sum_{i=1}^{a} c_i \bar{Y}_{i.}$ is an unbiased estimator of L with $\mathrm{Var}\{\hat{L}\} = \sum_{i=1}^{a} c_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum_{i=1}^{a} \frac{c_i^2}{n_i}$

so $\dfrac{\hat{L} - L}{\sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i}}$ has a T-distn with $(n-a)$ d.f.,

giving

a $100(1-\alpha)\%$ C.I. for a <u>single</u> L is

$$\left(\hat{L} \pm t_{n-a,\alpha/2} \sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i}\right)$$

<u>N.B.</u> If comparing 2 means μ_i & μ_k this reduces to $\left(\bar{Y}_{i.} - \bar{Y}_{k.} \pm t_{n-a,\alpha/2} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}\right)$

<u>Hypothesis Tests</u>

$H_0: L = L_0$ (L_0 a particular value - usually 0)

H_A: $L \neq L_0$, $L > L_0$, $L < L_0$

<u>test stat</u>: $T^* = \dfrac{\hat{L} - L_0}{\sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i}}$

& if $L = \mu_i - \mu_k$ then

$H_0: \mu_i - \mu_k = D_0$

H_A: $\mu_i - \mu_k \neq D_0$, $\mu_i - \mu_k > D_0$, $\mu_i - \mu_k < D_0$

<u>test stat</u>: $T^* = \dfrac{\bar{Y}_{i.} - \bar{Y}_{k.} - D_0}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}}$

Multiple Comparison Methods

- Usually want to carry out several comparisons involving the means

- Also often want to look at the data produced by the experiment (i.e. results) and choose interesting comparisons suggested by those results. This is called DATA SNOOPING.

Note:

When you "data snoop" for interesting comparisons you are implicitly looking at all possible comparisons of that type.

Eg

Suppose in class eg 4, you look at the $\overline{Y_{i.}}$ and decide to compare μ_E to μ_D since $\overline{Y_{E.}}$ is the largest and $\overline{Y_{D.}}$ is the smallest. BUT this is **not** a single comparison, since in order to pick it, **all** the $\overline{Y_{i.}}$ had to be looked at. You implicitly looked at all $C^5_2 = 10$ possible comparisons.

Thus want comparison methods which control the

OVERALL (or FAMILY) significance level (error rate)

Per Comparison Error Rate (significance level) = P(reject H_0 | H_0 true)

Overall or Family Error Rate = P(reject at least 1 H_0 | all H_0 true)

Summary of Multiple Comparison Methods

1. Bonferroni for PREPLANNED comparisons:

- Applies to any g preplanned contrasts or means or linear combinations
- Can NOT be used for DATA SNOOPING
- If # comparisons g is small Bonferroni method gives shorter C.I.'s than other methods
- Can be used for any design

2. Scheffe (Automatically covers All Possible Comparisons)

- Applies to any g contrasts (but not linear combinations that are NOT contrasts)
- Gives shorter C.I.'s than Bonferroni if g is large and larger C.I.'s if g is small (but remember that Bonferroni can NOT be used unless the comparisons were decided on in advance of looking at the data (i.e. preplanned)
- Can be used for any design
- Allows for data snooping since Scheffe method automatically considers all possible contrasts
- Should NOT, however, be used if one is interested ONLY in pairwise comparisons. In this case TUKEY is better.

3. Tukey (Automatically covers All Possible Pairwise Comparisons)

- Always best for all pairwise comparisons
- Can be used for CRD, RCBD, BIBD
- Allows data snooping since automatically considers all possible pairs
- If g small and comparisons are PREPLANNED, then Bonferroni should use as it gives shorter C.I.'s

4. Dunnett for Treatments vs a Control

- Best for all treatment vs control contrasts $\mu_i - \mu_{control}$
- Can be used for CRD, RCBD, BIBD
- By definition a preplanned method

Bonferroni Multiple Comparisons

These are suitable for any group of g comparisons (C.I.'s or hyp. tests) that are preplanned. They use g T-tests (or T-distn C.I.'s) where the per comparison significance level for each comparison is

α / g with α being the desired family significance level.

This ensures that the overall or family significance level $\leq \alpha$ since

$P(\text{reject at least 1 of the } g\ H_0 \mid \text{all } g\ H_0 \text{ true})$

$= P(A_1 \cup A_2 \cup \cdots \cup A_g)$ where A_k = event reject kth H_0 when it is true

$= 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_g)$

$\leq 1 - [1 - P(A_1) - P(A_2) - \cdots - P(A_g)]$ - by the Bonferroni inequality

$= P(A_1) + P(A_2) + \cdots + P(A_g)$

$= \frac{\alpha}{g} + \frac{\alpha}{g} + \cdots + \frac{\alpha}{g} = \frac{g\alpha}{g} = \alpha$

$\therefore$ family significance level $\leq \alpha$

Conversely, for a set of g C.I.'s the

family confidence level $> 1-\alpha$

Thus use:

For a set of g Linear Combinations $L_1, \ldots, L_g$

$$L_k = \sum_{i=1}^{a} c_i \mu_i$$

$100(1-\alpha)\%$ Confidence Intervals, $k = 1, 2, \ldots, g$

$$\left(\hat{L}_k \pm t_{n_T-a, \frac{\alpha}{2g}} \sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i} \right)$$

For g Hypothesis Tests (sign. level $\leq \alpha$)

$H_0: L_k = L_{k0}$

test stat: $T^* = \dfrac{\hat{L}_k - L_{k0}}{\sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i}}$

$H_A: L_k \neq L_{k0}$ R.R. $|t^*| > t_{n_T-a, \frac{\alpha}{2g}}$

$H_A: L_k > L_{k0}$ R.R.: $t^* > t_{n_T-a, \alpha/g}$

$H_A: L_k < L_{k0}$ R.R.: $t^* < -t_{n_T-a, \alpha/g}$

For g pairwise comparisons: $\mu_i - \mu_k$

the above reduce to

$$\left(\bar{Y}_{i.} - \bar{Y}_{k.} \pm t_{n_T-a, \alpha/2g} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_k}\right)} \right)$$

and $H_0: \mu_i - \mu_k = D_0$ test stat: $T^* = \dfrac{\bar{Y}_{i.} - \bar{Y}_{k.} - D_0}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}}$

N.B Each of the $A_1, A_2, \ldots, A_g$ could have different probabilities so in general

$$P(A_1 \cup A_2 \cup \cdots \cup A_g) \leq P(A_1) + \cdots + P(A_g) = \alpha_1 + \alpha_2 + \cdots + \alpha_g$$

which will be $\leq \alpha$ if $\sum_{i=1}^{g} \alpha_i = \alpha$

Scheffe Method for All Possible Contrasts

$$L = \sum_{i=1}^{a} c_i \mu_i \quad \text{where} \quad \sum_{i=1}^{a} c_i = 0$$

Note that while Bonferroni applies to any linear combination, the Scheffe method only covers contrasts.

- automatically covers all possible contrasts with family error rate $\leq \alpha$
  $\therefore$ can be used for data snooping
- since never actually do all possible contrasts (except if data snooped), in practice error rate $< \alpha$

Simultaneous $100(1-\alpha)\%$ C.I.'s for ALL Possible L_k

$$\left(\hat{L}_k \pm \sqrt{(a-1) F_{a-1, n_T-a, 1-\alpha}} \sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i} \right)$$

For testing all possible contrasts with family error rate < α

$H_0 : L_k = 0$
$H_A : L_k \neq 0$ } for all possible contrasts $L_k = \sum_{i=1}^{a} c_i \mu_i$

Reject H_0 if

$$|\hat{L}_k| > \sqrt{(a-1) F_{a-1, n_T-a, \alpha}} \sqrt{MSE \sum_{i=1}^{a} c_i^2 / n_i}$$

Tukey Method for PAIRWISE Comparisons: $\mu_i - \mu_k$

- can be used for data snooping since automatically covers all possible pairs
- always best method when interested in all pairwise comparisons (it gives the narrowest C.I.'s in this case)
- based on the Studentized Range Distribution
- if only want some paired comparisons that are PREPLANNED, then use either Tukey or Bonferroni, whichever gives the shorter C.I.s.