Fisher's least significant difference (LSD)

Procedure:

1. Perform overall test of H_0: $\mu_1 = \mu_2 = \cdots = \mu_t$ vs. H_a: $\mu_1 \neq \mu_2 \neq \cdots \neq \mu_t$

2. If outcome is "do not reject H_0, then *stop*. Otherwise continue to #3.

3. Perform desired hypothesis tests of (preplanned) paired comparisons using contrast tests:

H_0: $\mu_i = \mu_j$
H_a: $\mu_i \neq \mu_j$ (two sided)

The LSD aspect comes from the rejection region for the test statistic

$$t = \frac{\widehat{l}_{ij} - 0}{\sqrt{\widehat{V}\left(\widehat{l}_{ij}\right)}} = \frac{\overline{y}_{i\cdot} - \overline{y}_{j\cdot}}{\sqrt{s_W^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

One rejects H_0 if $|t| \geq t_{\alpha/2}$ (with $n_T - t$ df). This is the same as: reject H_0 if

$$\left|\overline{y}_{i\cdot} - \overline{y}_{j\cdot}\right| \geq t_{\alpha/2}\sqrt{s_W^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

When the sample sizes are equal (each $n_i = n$), the quantity on the right hand side becomes $t_{\alpha/2}\sqrt{s_{\mathrm{W}}^2(2/n)}$, which is the original meaning of the "least significant difference".

In SAS:

```
   ⋮
PROC GLM;
CLASS TRT;
MODEL Y=TRT;
MEANS TRT / LSD;  (can use T instead of LSD
                            for same results)
```

"Story" is: the overall AOV test protects the EER for the mean comparisons.

- story based on simulations from early '70s
- it is not really true for many means
- still recommended statistical practice, when EER is not taken too seriously
- confidence intervals instead of hypothesis tests are calculated with the usual methods for contrasts:

$$\overline{y}_{i\cdot} - \overline{y}_{j\cdot} \pm t_{\alpha/2}\sqrt{s_{\mathrm{W}}^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Tukey's procedure

Tukey's procedure is based on the **studentized range distribution**

X_1, X_2, ..., X_n a normal random sample; S is an *independent* estimate of σ, then

$$\frac{X_{\text{max}} - X_{\text{min}}}{S} \sim \text{studentized range}$$

The procedure assume equal sample sizes ($= n$) from each population. Then, under the hypothesis that the group means are the same ($= \mu$), the $\overline{Y}_{i\cdot}$ are a random sample from a normal$(\mu, \sigma^2/n)$ distribution, and (a property of the normal) the sample means are independent of S_{W}^2 . Thus, under the hypothesis that the means are the same,

$$\frac{\overline{Y}_{i\cdot}(\text{max}) - \overline{Y}_{j\cdot}(\text{min})}{\sqrt{(S_{\text{W}}^2/n)}}$$

has a studentized range distribution with $(n_{\text{T}} - t)$ df.

This statistic is based on the the *maximum* difference between the means. EER protection is accomplished by using the critical value for this statistic for *all* the mean comparisons.

Hypotheses

$$H_0\text{: } \mu_i - \mu_j = 0$$
$$H_a\text{: } \mu_i - \mu_j \neq 0$$

Test statistic

$$q = \frac{|\overline{y}_{i\cdot} - \overline{y}_{j\cdot}|}{\sqrt{(s_W^2/n)}}$$

Rejection region

reject H_0 if $q \geq q_\alpha$

where q_α is the percentile from the studentized range distribution corresponding to t treatment means and $n_T - t$ df. Selected percentiles appear in the text, Table 10, p. 1115-1116.

Text writes rejection region as:

$$\text{reject } H_0 \text{ if } \left|\overline{y}_{i\cdot} - \overline{y}_{j\cdot}\right| \geq q_\alpha \sqrt{(s_W^2/n)} \ (= W)$$

Simultaneous CIs (family-wide confidence is $100(1-\alpha)\%$) for the differences between means:

$$\overline{y}_{i\cdot} - \overline{y}_{j\cdot} \pm q_\alpha \sqrt{(s_W^2/n)}$$

Modification of Tukey procedure for unequal sample sizes

Tukey-Kramer procedure uses $q_\alpha \sqrt{\frac{1}{2} s_{\mathrm{W}}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$ instead of $q_\alpha \sqrt{(s_{\mathrm{W}}^2/n)}$ in all the formulas. The procedure has not been *proven* to protect the EER, but it has performed well in simulations.

In SAS, the Tukey (or Tukey-Kramer, when sample sizes are unequal) procedure is invoked by

```
    ⋮
PROC GLM;
CLASS TRT;
MODEL Y=TRT;
MEANS TRT / TUKEY;
```

Tukey indeed protects EER, and Tukey-Kramer seems to do so. Comments:

- The tests and CIs are conservative
- Not necessary to protect with overall AOV

Student-Newman-Keuls (SNK) procedure

The SNK procedure is like Tukey's, except that the SNK throws out all the means that are not in between $\bar{Y}_{i\cdot}$ and $\bar{Y}_{j\cdot}$. In SNK, $\bar{Y}_{i\cdot}$ and $\bar{Y}_{j\cdot}$ are the largest and smallest from a normal sample of size r, where $r - 2$ is the number of $\bar{Y}_{k\cdot}$'s that are in between $\bar{Y}_{i\cdot}$ and $\bar{Y}_{j\cdot}$ To compare μ_i with μ_j, SNK uses all the Tukey formulas, except with the studentized range percentile q_α corresponding to r treatment means.

Remarks:

- SNK does not protect EER
- SNK is more conservative than Fisher LSD

SNK is invoked with: MEANS TRT / SNK;

Dunnett's procedure

The Dunnett procedure is specifically for comparing treatments to a particular control or standard. Mean μ_{c} is compared to μ_1, μ_2, ..., μ_{t-1}.

Uses the “many-one t distribution”

Protects EER

Invoked with: MEANS TRT / DUNNETT('A'); Here the character A is how the control observations are coded in the variable TRT. Using DUNNETTU('$\cdots$') gives a one-tailed test for whether each treatment mean is larger than the control, and DUNNETTL('$\cdots$') gives a one-sided test for whether eact treatment is smaller than the control.

Scheffé

The Scheffé approach applies to *all possible* contrasts (and linear hypotheses where the a_i's do not necessarily add to 0)

The procedure is to reject the null hypothesis that $\mu_i - \mu_j = 0$ if

$$\left|\overline{y}_{i\cdot} - \overline{y}_{j\cdot}\right| \geq \sqrt{s_{\mathrm{W}}^{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)(t-1)f_\alpha}$$

where the F distribution percentile f_α has $t-1$ and $n_{\mathrm{T}} - t$ degrees of freedom.

The particular F distribution is the one that applies to the "worst case" contrast or linear hypothesis (values of a_i's for which the critical value is the largest) & thus all other linear hypotheses are obtained for free.

Remarks:

- Scheffé is very conservative for small numbers of comparisons (worse than Bonferroni)
- Scheffé is better than Bonferroni when the number of comparisons is large relative to the number of means

MEANS TRT / SCHEFFE;   (performs Scheffé tests for all differences of means)