Multiple Comparison Method Overview (summarized from Kirk, Chapter 4)

Tukey described two kinds of Type I error rates for a family of contrasts (tests that are related in content and use).

\[
\text{per contrast error} = \frac{\text{# contrasts falsely declared sig}}{\text{number of contrasts}} \times \text{prob any one contrast is a Type I error}
\]

\[
\text{family-wise error} = \frac{\text{# families with } \geq 1 \text{ contrast falsely declared sig}}{\text{number of families}}
\]

As family size goes up, the family-wise and per contrast error rates diverge.

- If there are two concepts of error, then there should be corresponding concepts of power.
  (a) Overall power: probability of rejecting a false complete null hypothesis
     - A complete null hypothesis says all means are equal.
  (b) P-subset power:  
     per pair power: \( H_0: \mu_j = \mu_{j'} \)
     per triplet power: \( H_0: \mu_j = \mu_k = \mu_l \)
     These are usually average probabilities (expresses power only for a test at the average).
     (Ramsey, 1978)
     Any-pair power: \( p \) of detecting at least 1 true difference among all pairs of means (focuses on largest mean difference).
     All pairs power: \( p \) of detecting all true differences among all pairs of means (focuses on smallest mean difference).
  (c) Which type of power is most important in your study?

Categories of multiple comparison procedures

1. Single-step procedures, or “simultaneous” contrast procedures
   a. Scheffe method
      - Can make confidence intervals
      - Allows all possible contrasts
      - Has poor power
      - Controls family-wise \( \alpha \)
   b. Tukey’s test
      - Can make confidence intervals
      - All possible pairwise combinations
      - Controls \( \alpha \) family-wise
      - Has better power than Scheffe for pairwise tests
      - Need equal n’s
      - Use Tukey-Kramer if n’s unequal
Handout #10.5, p. 2

\[ q_{TK} = \frac{\hat{\psi}}{\sqrt{MS_{error}\{[(1/n_1) + (1/n_2)]/2\}}} \]

c. Dunnett’s test for pairwise contrasts with a single control group mean. Use when:
   · sample sizes are equal
   · you want to compare to a control group
     controls \( \alpha_{FW} \) for the family, not beyond the family.
   · It is structured like a t-test.
   · A modified procedure is available when unequal n’s

d. Dunn or Bonferroni procedure
   Controls \( \alpha_{FW} \) by testing each member of a family of contrasts at \( p = \alpha/\# \text{ contrasts} \)
   \( \alpha_{FW} \) is approximately equal to \( p \).
   · Power depends on number of contrasts to be done

e. Dunn-Sidak
   · Slightly more powerful version of Dunn.
   · Controls \( \alpha_{FW} \), but you divide by something that
     is slightly less than number of contrasts (by using special table).

Dunn and Dunn-Sidak are: Robust with respect to Type I error under violations of normality and homogeneity of variance

2. Step-down procedures. Can’t construct confidence intervals.
   First you test a more global \( H_0 \), then a less global one.

a. Fisher’s LSD or protected t-test
   · Only do pairwise t-tests if omnibus \( H_0 \) is rejected.
   · Protects \( \alpha_{FW} \) for 3 groups
   · Does not protect \( \alpha_{FW} \) for > 3 groups.
b. Newman-Keuls
   - Order the means from smallest to largest.
   - Test smallest vs. largest (separated by p steps)
   - If significant, then the two most extreme Hs embedded are tested.
     \[ Y_5 - Y_1; \text{ then } Y_4 - Y_1 \text{ and } Y_5 - Y_2 \]

\[
\begin{array}{cccccc}
Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \\
\hline
\end{array}
\]

\[ * \]
\[ p \text{ steps} \]

- When number of steps > 3, \( \alpha_{FW} \) is NOT controlled
- Is relatively powerful because of this!

The step-down idea is important, and is used in newer step-down procedures that do control \( \alpha_{FW} \) when the number of groups is > 3.

c. Fisher-Hayter test
   - Step 1: test omnibus \( H_0 \) at \( \alpha \) using F
   - Step 2: Use Tukey-Kramer’s method, but use a-1 means in the table instead of a.
     - Has good power, better than Tukey and Dunn-Sidak for all possible pairwise comparisons.
     - Is easy to use.
     - Controls \( \alpha_{FW} \)
     - Allows only pairwise contrasts
     - Power almost as good as best newer methods

-- What if you want to mix planned and post-hoc tests?
   - Kirk (1995) suggests if you have one or two planned contrasts, do them.
     - Subtract their SS from SS_A.
     - Conduct omnibus test on residual with df = df_A - # planned contrasts
     - If significant, go ahead and use Fisher-Hayter or whatever with reduced df.