Statistics 514: Design of Experiments

Topic 7b Summary

Testing Multiple Contrasts

- F-test tests all possible contrasts (reduced power for any one contrast).
- Overall error rate – Pr(any Type I error)
  - Goes to 1 as number of tests increases (if nothing done)
  - Usually per-comparison error rates α' should be smaller
- Independent tests (orthogonal contrasts)

$$\alpha_{contrast} = 1 - (1 - \alpha_{overall})^{1/m}$$

- *Scheffé's Method*
  - For unplanned, general comparisons
  - Substitutes $a - 1$ for 1 degree of freedom
  - Nothing significant if F-test is not significant.
- Comparisons of means
  - Check difference of means against *Critical Difference*
- *Fisher's LSD*
  - Compare t with no adjustment. (Note MS_E in denominator.)
- *Tukey's HSD*
  - Controls experimentwise error over all comparisons of means (not general contrasts).
  - Rectangular confidence regions
- *Bonferroni*
  - Use $\alpha_{adj} = \alpha_{overall}/m$
  - For planned comparisons
  - Conservative (not powerful)
- Others
  - Duncan's Multiple Range Test
  - Newman Keuls
  - Dunnett
  - Comparison with Best

Confidence Intervals

- If overall rate is α, then set each confidence interval to be at $100(1 - \alpha_{adj})$ level (or adjust t if Scheffé or Tukey)
- Get larger confidence intervals
- Thus, the probability that at least one confidence interval does not contain true value (in our case, the population mean) is approximately α.
- Interpretation important

Issues with Multiple Comparisons

- *Error Rates*
  - Comparisonwise
  - Experimentwise
  - False Discovery Rate
  - Others
- *Context*
  - Which types of comparisons?
  - Consistency with F-ratio?
  - Only want to use p-values?
  - Tests independent?
- Presentation (criteria beyond error rates)
  - Just p-values
  - Confidence intervals
- Definition of power

Controversies

- P-values are hypothetical, unintuitive, approximate, incomplete ...
- Multiple testing adjustments are arbitrarily applied, antithetical to discovery, approximate, open to manipulation ...
- but data snooping can be more disastrous
- Ultimately, results depend on intention of user.