2/9/2010 36-402/608 ADA-II H. Seltman

Handout #9: Multiple Comparisons

1. Essentially equivalent concepts: multiple comparisons, data snooping, post hoc comparisons, unplanned comparisons

2. Examples:

   (a) Ideal situation: Planned orthogonal contrasts are chosen before looking at the data, and examined only when the overall ANOVA null hypothesis is rejected. Result: When studying useless factors, we get a 5% false positive rate. When studying factors that do affect the outcome, the true positive rate is power%, and the false negative rate is 100%-power%.

   (b) Gas chromatography study: Compare 30 chocolate lovers to 30 non-chocolate lovers by measuring the concentration of 100 substances in their urine. Test each at alpha=0.05 with a t-test. Result: If there are really no differences we expect 0 significant p-values 0.5% of the time, 1=3.1%, 2=8.1%, 3=14.0%, 4=17.7%, 5=18.0%, 6=15.0%, 7=10.6%, 8=6.5%, $\geq$ 9=6.3%.

3. If more that the planned orthogonal contrast are checked, the rate of false positive p-values rises. The solution to maintaining type-one error is to raise the cutoff value of the statistic or lower the alpha value to prevent excess false positives. This is necessarily at the expense of power.

4. The appropriate procedure depends on the type of snooping. For ANOVA, with I levels, usually $I-1$ *planned* orthogonal contrasts are tested without a multiple comparison "penalty", then additional test are performed with a penalty.

   (a) If the only tests are $I-1$ levels compared to a control level use Dunnett's test. (These are non-orthogonal.)

   (b) If all pairs (or the most interesting pair) are to be tested, use Tukey's test.

   (c) If a fixed number of tests (m) are to be made, use Bonferroni's correction: $\alpha' = \alpha/m$. E.g., the gas chromatograph study.

   (d) If *any* interesting contrast will be tested, use Scheffe's method where the cutoff F value is $I-1$ times the cutoff for an F test with $I-1$ numerator df and $N-I$ denominator df.

   (e) For massive testing, e.g., genetic array or paired voxels across brain images, use False Discovery Rate (more Thursday).

5. Breakout and Discussion