Introduction to Biostatistics (BIOS:4120)
Breheny

Assignment 13
Due: Tuesday, May 5

1. An investigator is exploring whether the expression levels of genes significantly differ between a sample of healthy individuals and a sample of individuals with Type 2 diabetes. He performs a separate t-test comparing the two samples for 5,000 different genes, and uses $\alpha = .05$ as his cutoff. His analysis identifies 411 genes as having different expression levels between the two samples.

   (a) The investigator reasons that because he carried out his t-tests using a type I error rate of 5%, he should expect about 5% of the 411 genes that he discovered to be type I errors. Is this reasoning correct or incorrect? If it is incorrect, what's wrong with it?

   (b) What is the investigator's false discovery rate?

2. To illustrate how multiple comparisons can produce significant associations with no clinical plausibility, Canadian investigators conducted a study of the association between astrological signs and common reasons for hospitalization. They tested 24 such associations.

   (a) How many statistically significant findings (*i.e.*, with $p < 0.05$) would you expect the investigators to discover in their study?

   (b) If we apply the Bonferroni correction, what number should we compare our p-values to in order to maintain a 5% overall probability of making a single type I error?

   (c) The study obtained two "significant" findings: individuals born under Leo had a higher probability of gastrointestinal hemorrhage ($p = 0.0447$), while Sagittarians had a higher probability of humerus fracture ($p = 0.0123$) compared to all other signs combined. Are these findings statistically significant in light of the multiple comparisons that the investigators performed?

3. German researchers carried out a study of two different treatments for heart attacks in a randomized trial involving 421 patients suffering from acute myocardial infarctions. They performed hypothesis tests for 15 different cardiac outcomes.

   (a) In order to keep the overall probability of making a type I error at 5%, what significance level should they test each individual hypothesis at?

   (b) The hypothesis test for the most important outcome, mortality, was $p = .0095$. Is this statistically significant according to the cutoff you defined in part (a)?

   (c) Of the 15 hypotheses, 4 (including the test for mortality mentioned above) were significant at the level $\alpha = .01$. What is the false discovery rate associated with this α level?

   (d) The investigators conclude that there is a statistically significant difference in the mortality rates of the two treatments. Comment on whether this statement is or is not justified in light of the multiple comparisons that they have made. In particular, state whether you agree with their conclusion.

4. In a study published in the *Journal of Gerontological Nursing*, investigators used age and education level to predict the capacity to direct attention (CDA) in elderly women. CDA was measured by how well the women accomplished a variety of tasks in the presence of competing and distracting stimuli.

The residual sum of squares for the most simple model (*i.e.*, all women have equal CDA, everything else is just random variability) was 1061. The residual sum of squares for the investigators' model, which used age and education to predict CDA, was 668. What percent of the total variability did the investigators' model explain?

5. Adriamycin (ADR) is a commonly prescribed drug in cancer chemotherapy. One of its unfortunate side-effects is damage to cardiac muscle. In a study done by researchers in the Toxicology department at the University of Kentucky, mice were randomized to receive either ADR or placebo (a saline solution). This experiment was conducted on two groups of mice: regular ("wild-type") mice and mice with a genetic mutation that makes them more susceptible to oxidative damage ("knock-down" mice). Thus, there are four groups of mice: WT-ADR, WT-Sal, KD-ADR, KD-Sal. The outcome in this experiment was concentration of the metabolite fumarate (a key intermediary in oxidative phosphorylation, a vital biochemical pathway in most multicellular organisms) in heart tissue. The data for this experiment is on the course website.

   (a) Fit a model in which each group is allowed to have its own average fumarate concentration. What percent of the variability does this model explain?

   (b) The model you fit in (a) reduces the unexplained variability; however, it also has more parameters than a simpler model which states that all mice have the same average fumarate levels. How many additional parameters does the more complex model have?

   (c) What is the probability that the reduction in unexplained variability would be as large or larger than what you observed, just by random chance?

   (d) Carry out all two-group t-tests with the Tukey adjustment for multiple comparisons. Which comparisons are significant at the 10% level?

   (e) Provide a one-sentence summary of this study's findings.