Some Loose Ends from Ch 1-9

We are (essentially) finished the required material from the text. The program from now until April 5 is:

Wed. March 28 - Loose Ends from Ch 8 and 9.
Fri.March 30 - Review for Quiz Monday (and Final Exam)
Mon. April 2 - Quiz
Wed. April 4 – Feedback from Quiz, tips for Final Exam

Loose Ends:

1. Small sample tests for proportions (pp342-343): It is possible to use the Binomial distribution to produce testing probabilities for samples too small to use the CLT approach. But I have skipped over this since it is rare to need it – there is so little information about p in a sample of size 10 or 20 that to estimate it with this sample makes no sense. Suppose p = 0.5 and consider a random sample of size n=16. Then $\hat{p}$ has a mean of 0.5 and an SD of $(0.5*(1-0.5)/16)^{1/2}$ = .125, so a 2 SD interval for $\hat{p}$ would be (.25, .75). A 95% CI for p would be about this same width – don't you agree it is not precise enough to be useful? By the way, do you see why (.25,.75) is not a 95% CI here? There are two reasons. One is that the 95% CI for p is centered at $\hat{p}$, not p – if we knew p we would not need a CI to estimate it! The other reason is that since the n=16 is too small for the CLT to give a good approx to Normality, and the interval of this width based on a sample of size n=16 might not have a 95% probability of including the population value p.

We are not saying the Binomial model is useless! It tells you want kinds a sequences are likely – for example, how likely is it that drug A gives better results than drug B in five of the first six patients is a clinical study, if drug A is not actually more effective than drug B in a larger group of similar patients? (Answer: about 6/128 or 12/128 depending the context of the question).

2. Statistical vs Practical Significance (p 353-354): In statistics significance is a jargon word with a special meaning – unfortunately it does not mean "important". A "statistically significant result" is one in which the data provide evidence to reject the null hypothesis. The idea (admittedly a weak rationale) is that we tend to learn something new (and "significant") when we find out that our null hypothesis is not true.

An even more difficult aspect of the theory of hypothesis testing is the fact that the conclusion depends on the sample size even fixing the P-value. Suppose we have a huge sample n=1000 and reject the null hypothesis. Since large samples provide very precise information about parameters, we might find that even if the hypothesis is only slightly wrong, and negligibly so, we reject it!. One might argue that all null hypotheses are wrong and we just need a large enough sample to prove it. So we need to be wary of rejecting null hypotheses when we use very large samples. .... A related problem arises when we accept the null hypothesis for a very small sample. In this case we need to b wary of accepting a null hypothesis with a small sample since the hypothesis could be seriously wrong but our sample not large enough to detect that situation.

Add to this the arbitrary choice of the type one error (α) and we can see that there is a lot of judgment required in interpreting hypothesis tests. This simple approach to hypothesis testing is better thought of as "credibility" testing than a rigorous method of decision making.

3. Pooled Estimates of SD in two-sample tests (pp 376-377) I reviewed in class why the Var of a difference was the sum of the Vars of the things differenced. When it is reasonable to assume the two variance estimates are estimating the same variance (that is, when the two populations can be assumed to have the same variance), then we may take advantage of this bit of knowledge by pooling the two estimates to form one "better" estimate. This is described on pp 376-377. But we almost never know that two population variances are equal in a situation where we are interested in the difference of means, and so this is a dangerous assumption, Moreover, the actual advantage is slight except in very small samples. But in very small samples the data gives no hints to confirm the equal variance assumption! Better not to use the pooled variance.

4. Analysis of paired data:

Suppose you have reaction time measurements for twenty class members before and after drinking three cups of strong coffee! An interesting question is whether the reaction time increases or decreases. Obviously the first thing to do with the data is to compute the twenty differences. Then we would test if the mean difference (underlying the sample data) was zero or different from zero. This is just a one-sample procedure (Ch 8, not Ch9). So paired data is usually better analyzed as one sample of difference rather than two independent samples .

5. Differences between two population proportions (pp 391-397): There is really nothing new here – just follow the procedure for the difference of two population means. Of course, you use the short-cut formula when estimating the two sample SDs.