Class 36: Topic 22: Comparing Two Means

**Held:** Monday, 28 April 2008

**Summary:** We consider tests that can be used to determine what samples of two populations reveal about the differences of means between those two populations.

**Notes:**
- Due Wednesday: 22-5, 22-6, 22-9, 22-14, 22-27.
- Handouts: Applet: Test of Significance Calculator; R Notes for Topic 22.

**Overview:**
- Comparing two means.
- A complication: Degrees of freedom.
- Using R.

**Comparing Two Means**
- Much like comparing two proportions.
- We typically call this test a two-sample $t$-test.
- Null hypothesis is typically “the two means are the same”.
- Three possible alternative hypotheses (which affect how we assign $p$-values).
  - The first is greater than the second
  - The first is less than the second
  - The two differ
- We use a different formula for the standard error: $\sqrt{s_1^2/n_1+s_2^2/n_2}$.
- We look up the $p$-value in the $t$-table, rather than than the standard normal probabilities table.
  - Remember, the standard error for means seems to vary more from the standard deviation than does the standard error for proportions.
  - Interestingly, the $t$-distribution is not a precise representation of this sampling distribution. It is, however, close enough.
- We use $t^*$ rather than $z^*$ for computing confidence intervals.

**Degrees of Freedom**
- Our book notes that the authors seem to compute degrees of freedom differently than do many statistics packages.
  - This is an effect of the use of the $t$-table to approximate the distribution.
- When you’re doing two-sample $t$-tests by hand, use the smaller of $n_1-1$ and $n_2-1$. 
• Most computers use a different formula.
  \[
  df = \frac{(s1^2/n1 + s2^2/n2)^2/((s1^2/n1)^2/(n1-1) + (s2^2/n2)^2/(n2-1))}
  \]
• No, I don’t know where that comes from.

Using R

• If you have the original samples, you can use a two-sample \( t \)-test in R with \texttt{t.test}.

  \texttt{t.test(Sample1,Sample2)}

• You can also use \texttt{t.test} to compute a confidence interval.

  \texttt{t.test(Sample1,Sample2,conf.level=.##)}

• If you simply know the test statistic and degrees of freedom, you can compute the \( p \)-value with \texttt{pt}.
• You can also compute \( t^* \) values using \texttt{qt}, but, once again, you need to supply the area of the curve to the left of that value. For a 95\% confidence interval and 5 degrees of freedom, we would write

  \texttt{qt(.975,df=5)}