Multiplicities in Comparing the Effectiveness of Online and Live Lectures

Performing multiple statistical comparisons can result in an unacceptably high probability of a false conclusion. Indeed Julie Shaffer at the 2010 meeting of the AAAS argued that one of the major reasons for the failure of apparent scientific findings to replicate is the failure to allow for the effects of such multiplicities. A quantitatively literate person should be aware of the problem of multiplicities and be able to spot failures of researchers to take them into account. Unfortunately, the media often does not present enough information about research findings to allow the reader to assess the multiplicities.

The problem of multiplicities is often encountered when a researcher performs a number of subgroup analyses. A good example is found in a study comparing Online and Live lectures (http://www.nber.org/papers/w16089) reported in the New York Times (http://bits.blogs.nytimes.com/2010/09/08/second-thoughts-on-online-education/). According to the New York Times:

> "Certain groups did notably worse online. Hispanic students online fell nearly a full grade lower than Hispanic students that took the course in class ..."

Consider the following problems with the way the New York Times presented the results.

First, the article failed to report that, overall, there was no evidence that live lectures are better than online lectures. The difference between online and live lectures was found only after testing the effect separately for a variety of subgroups.

Second, the article did not report the number of subgroup analyses that were or could have been performed. The authors gathered data on and could potentially have grouped students by University GPA, SAT, ACT, High School GPA, Gender, Ethnicity (Black, Asian, White, Hispanic), and/or Mother's education (5 levels). Failure to present the reader with information about these multiplicities prevents the reader from being able to properly assess the difference between live and online lectures for the Hispanic subgroup. The more potential subgroup analyses, the more cautious one should be in interpreting any given subgroup analysis.

Finally, although the New York Times article reports a respectively large total sample size of 312, it fails to report the sample sizes for the subgroup analyses. The reader has no way of knowing that the major finding reported in the article concerning Hispanic students was based on a comparison of only 8 Hispanic students who viewed online lectures with 25 who attended live lectures.

This New York Times article exemplifies why the quantitatively literate reader has to be skeptical of media reports that rely on subgroup analyses. All too often the media uncritically presents a researcher's conclusions without providing sufficient information for the reader to evaluate them. In short, the quantitatively literate reader must be constantly on the lookout for multiplicities and how they could affect the interpretation of the data.