STAT 8801 Case Study 9: Cheating

Group 4: Wen, Chris H, Xin, Qi, Joel, Chris D

Is the analysis done by the testing company appropriate?
We think the analysis is appropriate. The analysis was initiated as a request by the court, and there is no indication that this testing company has any vested interest with either party (the school and the student), which is key to the integrity of the results.

What evidence does it provide for or against the student?
An "agreement analysis" is conducted for each pair of students. It first records the number of items both students got wrong and the number of items for which both students chose identical wrong alternatives. Next, it calculates the observed percentage of joint incorrect answers, which is then compared with the expected percentage. Based on this we obtain the z value along with the probability of observing that particular or more extreme z values.
The conclusion is that the probability of observing a z value greater than 3.50 (from the suspected two students) is extremely unlikely, lending support to the school's charge.

Are there issues or problems with this analysis?
This analysis relies on some unspecified/untested assumptions, statistically and substantially.

- The choice of the comparison group is questionable. Examinees from the same school choose identical incorrect answers for various reasons, cheating and being exposed to the same learning environment being two of them. By adding examinees from other schools, this analysis artificially reduces the correlation between observations, leading to a left-shifted empirical distribution and a p-value that might be too small.

- This analysis assumes that, for a given item, the chance of selecting each of the three incorrect responses is equally split. This may not be true given the potential existence of distractors, the incorrect responses designed to be deceptively attractive. Therefore, two independent examinees are still very likely to choose the same incorrect response.

- It is not clear how the α-level is selected. Despite of the convention of using .05 or .01, in this particular case a decision needs to be made about a student's future. We might want to be more conservative, specifying *a priori* the α-level that all parties agree upon.

- The leap from "statistical significance" to the "cheating" conclusion is unfounded. Alternative explanations are equally valid for the significant finding. The two students may prepare for the exam together; or they may adopt the same test-taking strategy (my high school teacher told me to select C whenever I needed to guess an answer.)

Are there other analyses that might provide additional statistical evidence for or against X, or other issues should your client bring up or be ready to respond to?

- We may conduct disagreement analysis, that is, the number of items that one answered correctly and the other answered incorrectly, or the number of items that were answered incorrectly in different ways. It may well reveal different patterns.

- We may do an item-by-item analysis to distinguish the distractors and take this information into account. For example, among the identical incorrect answers of the two suspected students how many are distractors.

- One point worth emphasizing is that these analyses are *statistical* in nature, and as appropriately stated in the report, should not be viewed as the sole or even the most important evidence. At the very least, these analyses do not take the case beyond any reasonable doubt. Corroboration from more behavioral-based evidence is more fundamental.