Case Study 9: Cheating

Lindsey, Emily, Felipe, Rachel, Garrett

Case Recap

A student was suspected of cheating off another student's paper in a standardized examination. The suspect had exhibited unusual and suspicious behavior indicative of cheating. The school and the student have taken this matter to court. A testing company used the cheater's examination to perform an "agreement analysis" which could be presented in court as evidence.

Is the analysis done by the testing company appropriate?

Our group felt that the analysis done in this situation was somewhat minimal and did not provide the degree of evidence needed to present in a court situation.

What evidence does it provide for or against the student?

The test provided evidence against the student by indicating a very extreme p-value. Because the choice of a very small type I error likely limits the power of the test to actually detect when someone has cheated, if we believe in this methodology, there is strong evidence that the student cheated.

Are there issues or problems with this analysis?

Agreement analysis uses only those test items that both examinees in the pair answered incorrectly which may be lacking in validity. The agreement analysis document explicitly states that "the analysis does not indicate what factor or factors led to an agreement", and that the results are "not sufficient by themselves for arriving at a conclusion."

Essentially, this kind of analysis can only provide further evidence of cheating once it is already suspected, as in this case. However, as with all statistics, there are rare cases where a small p-value may occur by chance.

It is also unknown how the rest of the group fared on the questions the student got wrong. We'd like to have seen a breakdown of the answers chosen by all students. This could significantly effect how we looked at the questions that were jointly incorrect with the same incorrect answer.

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

We located several other plausible and perhaps, more reasonable, styles of analyses for this situation. One statistic that may be better is something that not only focuses on the number of matching incorrect scores, but also takes the length of the longest string of identical answers into account. There are windows of opportunity for cheating, thus, using a string of identical answers is likely more indicative of cheating than similar wrong answers alone.

Another that we found in our literature search is a statistic that uses both the number of matching incorrect answers in the longest string of identical answers and a statistic based on the number of items, the number of identical responses, and the number of identical incorrect responses.

A final technique we thought about would be to use the probability that the copier selects the same alternative as the source, given the popularity of each alternative and the copiers' total score.