Case Study 9: Cheating

Group 2 - Jenny, Xiaoyi, Sarah, Yu-Feng, and Bryan

Recap of the Situation

- A student is suspected of cheating off of another student on a standardized multiple choice test after a proctor observed suspicious behavior

- An agreement analysis was performed to look at the incorrect answers of the two students in question and compare them to the wrong answers of 67 other students who took the same test (19 from the same school, and 48 from other schools)

- Our task is to provide statistical expertise by interpreting the agreement analysis

Details of the Agreement Analysis

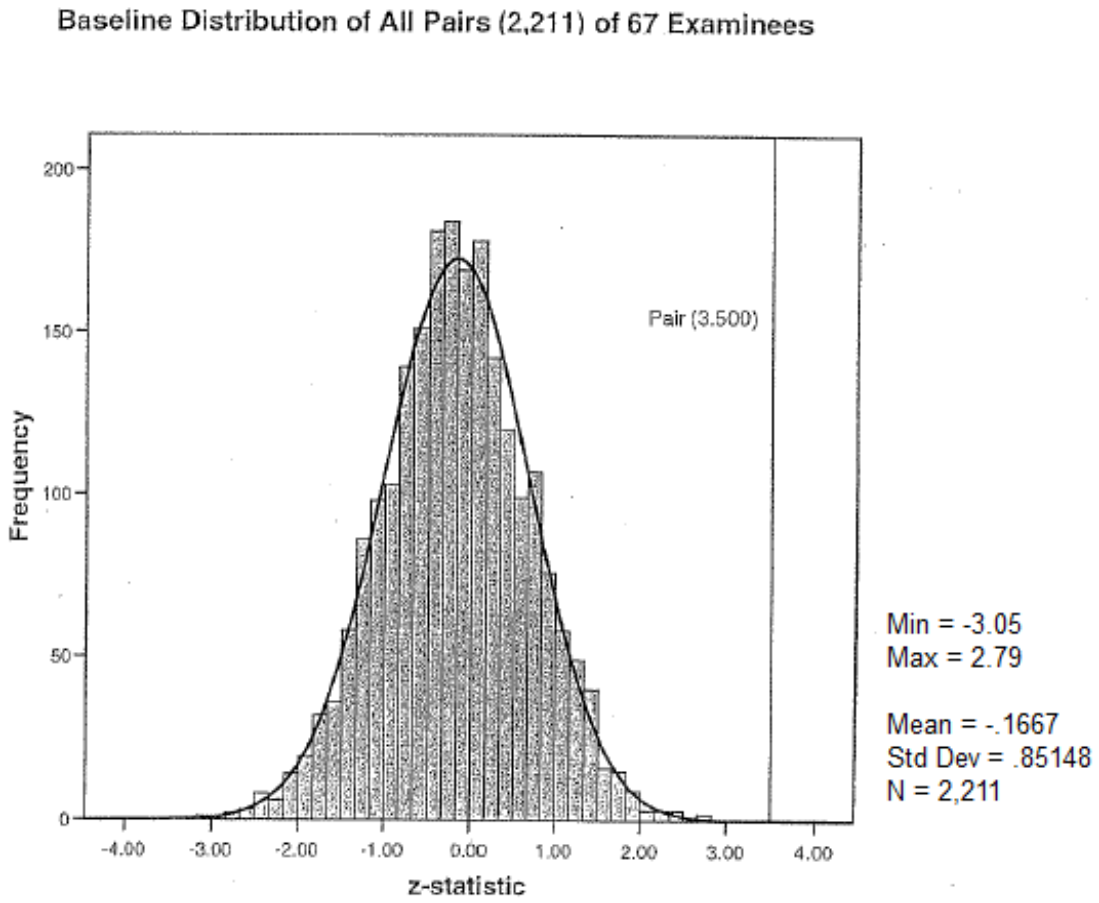
- The incorrect answers of each pair of examinees in the comparison group are analyzed
  - In our case, there are 67 students, offering 2211 possible pairs
  - Only questions that both students got wrong are considered
  - The empirical level of agreement (giving the same incorrect answer) is observed among all pairs in the comparison
- The incorrect answers the suspected examinees are recorded and the level of agreement is measured
- Standardized Z scores are calculated for each pair, comparing the actual level of agreement of wrong answers to the expected level, based on data from the comparison group

Results of Agreement Analysis

Incorrect Answers	# Qs Answered Incorrectly by Both	Same Incorrect Answers by Both	Obvserved Agreement	Expected Agreement	Calculated Z	p-value
28/21	14	13	93%	46%	3.5	0.00023

- In our case: $Z = (13 - 14*.46)/\sqrt{(14*.46*.54)} = \mathbf{3.5}$
- The Z scores in the comparison group ranged from -3.05 to 2.79 and follows an approximate normal distribution
- The suspected cheater is obviously on the extreme end of our distribution

Results of Agreement Analysis



Conclusions on Analysis

- The Z score for the suspected cheater was found to be 3.5, which gives us a p-value of .00023
- This means that based on the empirical results from the comparison group, it is extremely unlikely that the suspect randomly and independently matched 13 out of 14 incorrect answers with the other student
- There could be other underlying reasons for some of these matches, however

Was analysis appropriate?

- Overall, we feel like the agreement analysis was appropriate and is a good tool to use in the investigation
  - It is noted that finding evidence of more widespread collusion through the analysis is much more difficult and would require an adjustment for multiple comparisons, but in this case
  - Given the evidence that the student was behaving suspiciously during the test leading the proctor to suspect cheating, and that the agreement analysis backed up this claim, things don't look promising for the suspect

Issues and Additional Analyses

- The assumption of independence is questionable
  - We would like to see an additional analysis only among students within the same medical school to see if there are certain questions they are more likely to get wrong and more likely to give the same incorrect answer to
- The student in question had 28 total questions wrong, and 14 of these were answered correctly by the other student; we would be interested to know if there is some kind of pattern here