**STAT 5340 Statistical Analysis I**

**Fall 2023 – Mid-Term Exam**

1. The data below is the number of detected external attempts to breach a company’s 4 major servers for 20 days.

|  |
| --- |
| Detected External Breach Attempts |
|  |

1. Construct a histogram and a box plot of the entire set of 80 results. Comment on what these plots suggest about the data. (4 pts)
2. Calculate the sample average and sample variance for the entire set of 80 results, then calculate the sample averages and sample variances for each day of operations. (4 pts)
3. Calculate the variance of the 20 daily sample averages and the average of the 20 daily variances. (4 pts)
4. Compare the values calculated in c) with the entire sample variance calculated in b). Comment on these results. (5 pts)
5. Plot the 20 daily sample average, minimum, and maximum results (y-axis) by day (x-axis) together on the same chart. (4 pts)
6. What does this plot suggest about sources of variation in the external breach attempts? (4 pts)

1. Detected external breach attempts are classified according to potential severity as Low, Medium, or High. Among all detected breach attempts, it is known that approximately 2/3 are of Low severity, 20% are of Medium concern, with the remainder being of High severity. The method of classifying individual breaches involves the evaluation of their penetration, potential source, any known history, etc. and is subject to error. The approximate error probabilities are:

0.2% for classifying truly Low severity attempts as of Medium severity

0.01% for classifying a truly Low severity attempt as of High severity

5% for classifying a truly Medium severity attempt as of High severity

0.3% for classifying a truly Medium severity attempt as one of Low severity

4% for classifying a truly High severity attempt as of only Medium severity

0.05% for classifying a truly High severity attempt as of merely Low severity

1. Find the probability of correctly classifying a specific detected external attempted breach. (9 pts)
2. For each type of attempt, find the probability that if a single attempt evaluation has classified an attempt as a specific type, that the attempt actually is an attempt of that type. (6 pts)
3. Suppose that in order to reduce probability of incorrectly classifying a specific attempt, the cybersecurity manager proposes the following process:
4. Obtain 3 independent evaluations of the attempted breach’s level of severity generating three independent classifications of the attempt
5. If two or more agree, then classify the attempt accordingly
6. If all three disagree, then obtain one more evaluation and classification with this result determining the ultimate classification of the attempt

With this process, what is the probability of incorrectly classifying an attempted breach? How much reduction has occurred over a single classification? (10 pts)

1. You manage an information systems group at a major computer hardware and software company and have to make a decision between two cybersecurity contractor companies (A and B) you want to manage cybersecurity for your organization and promote to your customers. You know that currently each company has 8 active high-level cybersecurity contracts with the Department of Defense (DOD), but that five new such contracts have recently been awarded by the DOD with each contract going to one or the other of these companies. A source at the Defense Department informed you prior to the contract awards that Company B had a slightly higher probability of being awarded any individual contract (ie, probability any one of the five contracts being awarded to B was approximately 55%), but the contract award decisions were independent among the contracts. The award decisions were not made public knowledge due to their high-level nature; however, you want to work with the company that received the majority of these contract awards.
2. Define a random variable NA = Number of these 5 high-level cybersecurity contracts

awarded to company A, and find the probability mass function for this random variable. What is E[NA]? What is the probability company A received a majority of these new cybersecurity contracts? Which cybersecurity contractor would you choose? (10 pts)

1. Suppose you have a source in company A that just this week informed you that she was working on two DOD contracts, one of which was a newly awarded contract. Now, find the probability mass function for the random variable NA|Info = Number of new contracts awarded to company A given that two working contracts included one such contract. What is E[NA|Info]? What is the probability company A received a majority of these contracts given the inside information? Would this change your decision from part a)? (15 pts)
2. Extra Credit: What if your insider had been working on 3 contracts, of which, again, just one was a new contract? What would E[NA|Info] be now? What would be the probability company A received a majority of these contracts? How would you choose to spend the firm’s money with this information? (4 pts)

4. Verify your solutions for the following via simulation:

a) Problem 2, Part a). (5 pts)

b) Problem 2, Part b). (5 pts)

c) Problem 2, Part c). (5 pts)

d) Problem 3, Part a), E[NA]. (5 pts)

e) Problem 3, Part b), E[NA|Info]. (5 pts)

Team Assignments for Mid-Term Exam:

|  |  |  |
| --- | --- | --- |
| Team 1 | Team 2 | Team 3 |
| Jaslynn G | Steven H | Viktor E |
| Sido S | Vivek G | Sneha M |
| Jaliyah H | Job A | Clark O |