Now it's time to flex your critical evaluation skills. Read the following descriptions of an experiment and its analysis, identify the flaws in each, and describe what you would do to correct them.

1. The Sith Lords are concerned that their recruiting slogan, "Give In to Your Anger," isn't very effective. Darth Vader develops an alternative slogan, "Together We Can Rule the Galaxy." They compare the slogans on two groups of 50 captured droids each. In one group, Emperor Palpatine delivers the "Anger" slogan. In the other, Darth Vader presents the "Together" slogan. 20 droids convert to the Dark Side after hearing Palpatine's slogan, while only 5 droids convert after hearing Vader's. The Sith's data scientist concludes that "Anger" is a more effective slogan and should continue to be used.

The person who delivers the slogan may have a convincing skill or tone or the way the slogan was delivered is different and how much time did they use? An Emperor who used longer time will have a greater chance to attract more droids. The group of droids captured should be similar and must be from the same population. A sampling where Palpatine fought a war to capture 50 droids while Vader is in a different battle to capture 50 droids is biased.

The first thing I do is take a representative sample by randomly selecting droids. The selection should not be related to the Emperors, like the wars they fought and personal choice. All 100 droids shall be representative of the population. I will give equal amount of time for each slogan. I will make Palpatine deliver Anger slogan to one group and Together slogan to the other and eliminate the person bias, in fact both Vader and Palpatine can deliver both slogans to two samples each, and the total number of droids who converted for each slogan will give a good comparison.

See the table below

|  |  |  |
| --- | --- | --- |
|  | Palpatine | Vader |
| Anger | 15 | 11 |
| Together | 8 | 9 |

So, an average of (15 + 11)/2 = 13 people converted by the “Anger” slogan while only 8.5 people converted by the “Together” slogan. Now I can safely say the Anger slogan is more effective.

1. In the past, the Jedi have had difficulty with public relations. They send two envoys, Jar Jar Binks and Mace Windu, to four friendly and four unfriendly planets respectively, with the goal of promoting favorable feelings toward the Jedi. Upon their return, the envoys learn that Jar Jar was much more effective than Windu: Over 75% of the people surveyed said their attitudes had become more favorable after speaking with Jar Jar, while only 65% said their attitudes had become more favorable after speaking with Windu. This makes Windu angry, because he is sure that he had a better success rate than Jar Jar on every planet. The Jedi choose Jar Jar to be their representative in the future.

Not Clear!! Please give me suggestions on how to approach this question

1. A company with work sites in five different countries has sent you data on employee satisfaction rates for workers in Human Resources and workers in Information Technology. Most HR workers are concentrated in three of the countries, while IT workers are equally distributed across worksites. The company requests a report on satisfaction for each job type. You calculate average job satisfaction for HR and for IT and present the report.

I think this has the Simpson’s Paradox because averaging the job satisfaction of all countries will hide the job satisfaction in each country. Large number of workers might be positively or negatively correlated with Job satisfaction. What I will do is: Calculate the average Job satisfaction for each country for both HR and IT and do the overall average, see if there is lurking variable.

1. When people install the Happy Days Fitness Tracker app, they are asked to "opt in" to a data collection scheme where their level of physical activity data is automatically sent to the company for product research purposes. During your interview with the company, they tell you that the app is very effective because after installing the app, the data show that people's activity levels rise steadily.

A major bias point is “opt in”. A data containing only voluntary participants cannot represent the whole population. In fact, those who like to participate may be sport enthusiasts who decided to install the app to track their activity with numerical figures. They may be following a common routine where they start slow and steadily increase activity in 6 months and again take a break and repeat, so the increased activity may not be because of the app but by the routine they follow. So, it will bias the conclusion about a population because it represents particular type of people.

The solution is to enable auto tracking and let users know that their activity is being monitored for research purposes, it can be included in an agreement form that people read before installing the app. Then a time series analysis can be done and make a conclusion on new results.

1. To prevent cheating, a teacher writes three versions of a test. She stacks the three versions together, first all copies of Version A, then all copies of Version B, then all copies of Version C. As students arrive for the exam, each student takes a test. When grading the test, the teacher finds that students who took Version B scored higher than students who took either Version A or Version C. She concludes from this that Version B is easier, and discards it.

There are many possibilities for a bias. If she makes totally different questions on the three versions, Version B might of course contain easy questions or questions from a well-covered chapter during class. But if the same questions are just shuffled, the conclusion is in question because a mere shuffling will not have this much impact on the score of a student although some might argue that doing easy questions first will give time to answer harder ones later. The other bias I see is on the way she stacked the exams. The top students may not arrive first or last but right in the middle of the sequence, where most get Version B.

I propose the following changes. Instead of stacking them in order it is better to shuffle the exam papers like we do for a new card game so that there is no relationship between arrival time and which exam version a student gets. Using different questions for same class students, same course is not fair because it will create unfair competition where some are lucky to get easy questions while others fight with the hard questions. I will make same questions and shuffle them and create three versions, similar to the way done in national examinations. After doing all these changes and doing the experiment again if there is a high scoring Version, then it may be the case that specific arrangement of questions is helping students score higher results or it is by chance. From my experience, when an exam contains questions from each chapter sequentially placed like the paragraphs I read it will help somewhat to complete the exam fast and correctly, but only if the student read before the exam.