Evaluate an experiment analysis

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, Dark 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.

* Are the two groups of 50 captured droids identical?
* Understand Profile of droids groups
* Is this result statically robust? A/B test
* Can we extrapolate this result with a good level of confidence?

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.

* Have the planets the same level of unfriendliness?
* Could we compare
* What was the size to the population of these 4 planets
* How different the 4 planets were? (deviation)
* Are the results statically robust? A/B test
* Need to investigate potential outliers planets

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

* Need to put these results in context of the overall satisfaction level for each country
* Need to understand whether these results over/under index performance of the country
* The groups being very different, we need to present the profile of the different responders (age, ..) and ensure it is similar to avoid facing Simpson’s Paradox
* Recommend to present the results by countries and by department relative to overall performance
* Need to investigate potential outliers employees

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.

* Is there a way to understand activity level before the installation of the app
* People installing the app are more likely to have already a high activity level
* Ideally, we will need to understand how this compares with people with similar profile that are not installing the app
* Need to understand for each group the app has been the most beneficial
* Simpson paradox could bias the results
* Need to investigate potential outliers subscribers

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.

* The test B could have been taken students having more intellectual faculties
* Need to understand which students took each version of the test
* Need to run a significance test (ANOVA) to validate that the results of test B significantly higher
* Need to investigate potential outliers students