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.**

There is an incorrect assumption that Darth Vader and Emperor Palpatine are equally effective in converting droids to the dark side. The same person should deliver both slogans. If desired, in separate groups (3 and 4), the other person should deliver both slogans for a comparison of effectiveness.

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.**

|  |
| --- |
| The populations of the planets which Mace Windu visited are less favorable for the Jedi when compared to populations of planets where Jar Jar Binks visited. This information is important because Jar Jar Binks had an easier population to convince. Even though we are measuring improvement, this bias is in Jar Jar Binks favor. The best way to address this bias is by finding out the population differences in Jedi preferences between Mace Windu & Jar Jar Binks planets before their efforts. |
|  |

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.**

The report will be flawed in that it will not tell the whole story. If each worksite/country has different satisfaction rates, the data will get lost if it is only split based on job type.

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.**

The data will be skewed because before the opt-in, people show no activity. Once they are enrolled in the app they steadily build up data and therefore have physical activity. Also, it is possible that the population that opts-in is more likely to exercise more.

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.**

|  |
| --- |
| The difficulty of questions may vary with different versions of test. Also, more intelligent or more prepared students could have received test Version B. This can depend on the sample size. In order to fix this problem, versions can be the same questions but in different orders. |
|  |

Discuss your answer to each of these questions with your mentor. For a deep and irreverent dive into the critical evaluation of experiments (among other sources of information), check out [Calling Bullshit in the Age of Big Data](http://callingbullshit.org/" \t "_blank).