Drill: Evaluate an experiment analysis

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.

   Flaws: There is a bias in assignment conditions – 2 different people are delivering the slogan. Delivery should come from just one Sith to eliminate this bias.

2. 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.

   Flaws: There is a bias in assignment conditions – 2 different people are delivering the slogan. Delivery should come from just one Sith to eliminate this bias. Also, there is sampling bias in that the planets in each sample all either all friendly or all unfriendly.

   To eliminate these biases, have the same Jedi visit each group of planets, and have the planets visited have the same ratio of friendly to unfriendly planets.

3. 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.

   This is an example of sampling bias, in that job satisfaction may be biased in relation to geographic location. An average taken across different geographic locations would not be representative for the HR workers, which are more concentrated in three of the countries. To remedy this, limit the analysis scope to one country at a time for HR because of the workers distribution, or better yet, for both groups. This may provide secondary outcomes with additional information for each country that could be remedied locally.

4. 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.

   There is a sampling bias because those people that decide to 'opt in' may be more motivated to exercise, either because the nature of 'opt in' people or because they opt in and may assume they are being measured.

   Also, there is not a baseline from which to measure relative activity. I would track activity level before asking people to opt in, thus working from an objective baseline. I would also limit the conclusion of effectiveness to the demographic of 'opt in'-type people.

5. 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 is a bias in assignment of conditions with respect to time and order. Students who arrive early and get Version C may not know material well and arrive early because they know they need more time. Those that arrive a little later and pick up Version B that scored higher may be more comfortable with the material and score higher because of that. The last group to arrive and pick up Version A may have been hurried or not had time to complete the whole exam.

   The better process would be to randomly stack the copies so that students are picking them up randomly, eliminating condition or time-based biases.