

Unit 1.4 Drill 4

Evaluate an Experiment Analysis

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.
 - a. Flaw(s): this experiment fails to take into account the effect of the recruiters delivery. The experiment was also conducted on droids that had already been recruited and it might have been better to target droids that had not already been converted.
 - b. Correction: Capturing the demographic and psychological profile of the typical droid using Base sampling plan then matching the slogan to those that would be more susceptible based on the demographic/psych criteria.
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.
 - a. Flaw(s): any method of self-reporting has a high chance of introducing bias. There can also be some lurking variables in the data such as key-talking points made by the representatives.
 - b. Correction: giving both envoys similar talking points to deal with planets that are friendlier and another set of talking points for those that have a more negative perception of the Jedi, then measure perception change indirectly instead of relying on self-reporting.

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.**
 - a. Flaw(s): the issue of self-reporting arises again in this example and thus bias is more likely to be introduced. Additionally, averaging will not account for functional differences that are likely to appear on different worksites and the workload impact. There are also other country related factors that could play a role in employee satisfaction.
 - b. Correction: it might be best to measure employee satisfaction indirectly. It might also be best to compare employee satisfaction relative to each department and focusing on worksite and country related factors.
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.**
 - a. Flaw(s): this is another example of self-selection bias; the users that are opting into the data collection scheme may not be representative of all users. Another thing to consider is that the users behavior may be driven by self-consciousness - knowing that they are being monitored, some users may be driven to do more physical activity.
 - b. Correction: in order to avoid self-selection bias, the company can make data collection a precondition instead of relying on people to opt in. This way, the entire populations behavior is monitored and more representative of the whole user base instead of a select few.
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.**
 - a. Flaw(s): If we make the assumption that all three versions are unique and have scrambled answers, this may be an example of clustering. Perhaps those that had version B sat close enough to each other and were able to cheat of each other. It is
 - b. Correction: To correct this issue the seating arrangements should be more randomized. The tests should also be more randomized so that A, B, and C tests are not all grouped together.