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, 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

Flaws and corrections:

* 1. Should these speeches be given to droids, as opposed to humans?
  2. The samples should be formed by first mixing together the droids and then dividing randomly into two groups
  3. Sample size seems a bit small.
  4. The speech should be given by the same person. It is possible that Vader just isn’t that charismatic.
  5. Do these slogans come in the context of bigger speeches? These should be the same

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.

Flaws and corrections:

* 1. Is the difference between 65% and 75% meaningful?
  2. It seems that we have a case of Simpson’s Paradox and I would check the planet sizes to see what populations we are talking about
  3. How were they friendly and unfriendly planets divided among the two and what were their populations? Ideally you want to send them to the same mix of friendly and unfriendly (in terms of proportions) and similar population sizes and similar societal structure (I am not sure this absolutely has to be, but social phenomena like favorability can be impacted by group size and group structure).

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.

Flaws and corrections:

* 1. You have to control for the job satisfaction by country and by worksite (if your worksites have different cultures or processes, etc.).
  2. You need to compare HR and IT at the same work sites, in the same countries.

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.

Flaws and corrections:

* 1. There is likely to be some selection bias here. People who choose to “opt in” are more likely to care about monitoring their fitness so are more likely to make use of the app.

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.

Flaws and corrections:

* 1. It is likely we are talking about very small numbers (maybe 10 per Version) so spurious effects happen very easily.
  2. It is also unclear to what extent the order of arrival of student impacted this experiment. Where were the students coming? Where they coming primarily from a certain activity that impaired or helped them do better or from a different class that would have been an indicator of poorer performance (e.g. most of the B group was part of the drama team and the test was a mandatory computer science test)