Written Report – 6.419x Module 1

Name: RPiconAlvaro

- **Problem 1.1**

1. (2 points) How would you run a randomized controlled double-blind experiment to determine the effectiveness of the vaccine? Write down procedures for the experimenter to follow. (Maximum 200 words)

Solution: I'd group students into 6 distinct groups: Grade 1/2/3, Vaccine/Placebo in all their combinations (Grade 1 with Vaccine, Grade 1 with Placebo, Grade 2 with Vaccine…). Then compare the numbers for every student grade or grouping all the results for Vaccinated and Placebo patients **but** not making conclusions from crossed groups (for example Group 1 Vaccine with Group 2 Placebo).

2. (3 points) For each of the NFIP study, and the Randomized controlled double blind experiment above, which numbers (or estimates) show the effectiveness of the vaccine? Describe whether the estimates suggest the vaccine is effective.(Maximum 200 words)

Solution: <u>NFIP</u>: Grade 2 (vaccine) versus Grade 2 (no consent) shows the greatest effectiveness in this study. We don't know if there are any other underlying conditions behind the different with Grade 1/3 students. <u>Randomized controlled double blind:</u> Treatment (Vaccine) versus Control (Salt Injection) shows the greatest effectiveness. There might be some underlying conditions behind the people who did not give consent (such as taking their health less seriously)

3. Let us examine how reliable the estimates are for the NFIP study. A train of potentially problematic but quite possible scenarios cross your mind:

(a) (2 points) Scenario: What if Grade 1 and Grade 3 students are different from Grade 2 students in some ways? For example, what if children of different ages are susceptible to polio in different degrees?

Can such a difference influence the result from the NFIP experiment? If so, give an example of how a difference between the groups can influence the result. Describe an experimental design that will prevent this difference between groups from making the estimate not reliable.

Solution: Yes, if Grade 1/3 students are more susceptible to polio then the effectiveness of the vaccine would be overstated. It could be solved by grouping all students into only two groups: vaccinate & unvaccinated or comparing the groups separately: Group 1 vaccinated vs Group 1 unvaccinated…

(b) (2 points) Polio is an infectious disease. The NFIP study was not done blind; that is, the children know whether they get the vaccine or not. Could this bias the results? If so, Give an example of how it could bias the results. Describe an aspect of an experimental design that prevent this kind of bias.

Solution: Yes, children getting the vaccine might think they are more 'immune' to the disease because of this hence lowering the precautions. To eliminate the bias we could use a placebo (as in the second experiment) or let them think we're testing something completely different (such as sleeping medicine).

(c) (2 points) Even if the act of "getting vaccine" does lead to reduced infection, it does not necessarily mean that it is the vaccine itself that leads to this result. Give an example of how this could be the case. Describe an aspect of experimental design that would eliminate biases not due to the vaccine itself.

Solution: People who are willing to get the vaccine may be more inclined to having healthy lives and be more careful with diseases as a whole, therefore there may be a correlation but not necessarily a causal relationship. Using a placebo and comparing people who are vaccinated with the real shot and with the placebo.

4. (2 points) In both experiments, neither control groups nor the no-consent groups got the vaccine. Yet the no-consent groups had a lower rate of polio compared to the control group. Why could that be?

Solution*:* It could be down to behavioural reasons, for example people who don't consent to the vaccine may be less trustworthy of other people and be more lonely in general because of this. Given that polio is a contagious disease this may lower the rate of infection in these people.

5. (3 points) In the randomized controlled trial, the children whose parents refused to participate in the trial got polio at the rate of 46 per 100000, while the children whose parents consented to participate got polio at a slighter higher rate of 49 per 100000 (treatment and control groups taken together). On the basis of these numbers, in the following year, some parents refused to allow their children to participate in the experiment and be exposed to this higher risk of polio. Were their conclusion correct? What would be the consequence if a large group of parents act this way in the next year's trial?

Solution: Although the numbers might have been slightly lower in the unvaccinated children, a statistical study would be necessary to determine whether this was down to random variance or if it was actually due to the vaccine. It may be that there is no reason at all to think that the vaccine leads to a higher rate of polio. The consequence could be that the experiment could not be repeated due to a lack of numbers in the vaccinated group and therefore it'd be hard to validate the effectiveness.

- **Problem 1.3**

(a-1). (2 points) Your colleague on education studies really cares about what can improve the education outcome in early childhood. He thinks the ideal planning should be to include as much variables as possible and regress children's educational outcome on the set. Then we select the variables that are shown to be statistically significant and inform the policy makers. Is this approach likely to produce the intended good policies?

Solution: No, the larger the number of variables the larger the number of false rejections we will be getting. Therefore, although we might be able to find most variables contributing to the end result, we will also be picking a lot of fake ones (depending on the alpha level).

(a-2). (3 points) Your friend hears your point, and think it makes sense. He also hears about that with more data, relations are less likely to be observed just by chance, and inference becomes more accurate. He asks, if he gets more and more data, will the procedure he proposes find the true effects?

Solution: With more data we will still be finding a percentage alpha of false positives, so it might not help the results at all even. Only if we were to reduce the alpha level thanks to a more accurate model would we be getting better results.

(b-1). (2 points)

A economist collects data on many nation-wise variables and surprisingly find that if they run a regression between chocolate consumption and number of Nobel prize laureates, the coefficient to be *statistically significant. Should he conclude that there exists a relationship between Nobel prize and chocolate consumption?*

Solution: We should also consider other variables (highest level of education, growing up conditions) which may affect the result. It is possible that there is a high correlation between them but that the underlying causal effect is different (such as coming from a wealthier background).

(b-2). (2 points)

A neuroscience lab is interested in how consumption of sugar and coco may effect development of intelligence and brain growth. They collect data on chocolate consumption and number of Nobel prize laureates in each nation, and finds the correlation to be statistically significant. Should they conclude that there exists a relationship between chocolate consumption and intelligence?

Solution: Again, more variables should be taken into account to account for the nations' differences but averaging over a whole nation already reduces the variance in comparison to the last experiment (b-2).

(b-3). (1 point)

In order to study the relation between chocolate consumption and intelligence, what can they do?

Solution: Have a large number of people with similar characteristics participate in a years long study evaluating the performance of both groups. The more similar both groups are the more trustworthy the results will be.

(b-4). (3 points)

The lab runs a randomized experiment on 100 mice, add chocolate in half of the mice's diet and add in another food of the equivalent calories in another half's diet. They find that the difference between the two groups time in solving a maze puzzle has p-value lower then 0.05. Should they conclude that chocolate consumption leads to improved cognitive power in mice?

Solution: In this case groups can be thought to be large enough and the experiment to be quite isolated (ie the only thing that can really impact is the food difference) so the results might be close to the underlying truth. Further tests could be run with different alternatives to make sure it is not the alternative food which is giving mice a worse performance at performing the puzzle.

(b-5). (3 points)

The lab collects individual level data on 50000 humans on about 100 features including IQ and chocolate consumption. They find that the relation between chocolate consumption and IQ has a p-value higher than 0.05. However, they find that there are some other variables in the data set that has p-value lower than 0.05, namely, their father's income and number of siblings. So they decide to not write about chocolate consumption, but rather, report these statistically significant results in their paper, and provide possible explanations.

Is this approach correct?

Solution: Yes, from the results we can conclude that there is a correlation between all the variables mentioned but we cannot be sure which is the causal variable between them. It may be that father's income directly impacts both chocolate consumption and IQ hence meaning that chocolate actually is not significant.

(c). (3 points)

A lab just finishes a randomized controlled trial on 10000 participants for a new drug, and find a treatment effect with p-value smaller than 0.05. After a journalist interviewed the lab, he wrote a news article titled "New trial shows strong effect of drug X on curing disease Y." Is this title appropriate? What about "New drug proves over 95% success rate of drug X on curing disease Y"?

Solution: Second title is more appropriate as the strong effect on curing disease Y might not be applicable to all people taking the medicine. Instead, saying that 95% of the people will be cured is closer to reality as not all people will have a strong reaction to the medicine.

(d). (1 point)

Your boss wants to decide on company's spending next year. He thinks letting each committee debates and propose the budget is too subjective a process and the company should learn from its past and let the fact talk. He gives you the data on expenditure in different sectors and the company's revenue for the past 25 years. You run a regression of the revenue on the spending on HR sector, and find a large effect, but the effect is not statistically significant. Your boss saw the result and says "Oh, then we shouldn't increase our spending on HR then".

Is his reasoning right?

Solution: Yes, although some might think its still worth it if increasing the budget is not very risky and at least a slightly positive relationship has been found.

(e). (1 point)

Even if a test is shown as significant by replication of the same experiment, we still cannot make a scientific claim.

True or False?

Solution: False, if the test is replicable and still significant we can make the scientific claim.

(f). (2 points)

Your lab mate is writing up his paper. He says if he reports all the tests and hypothesis he has done, the results will be too long, so he wants to report only the statistical significant ones.

Is this OK? If not, why?

Solution: It is not okay because if he's tested a large number of hypothesis then by definition an alpha percentage of them will be false positives. It is important to know the rate of statistical significance as well.

(g). (2 points)

If I see a significant p-values, it could be the case that the null hypothesis is consistent with truth, but my statistical model does not match reality.

True or False?

Solution: True, p values allow a level alpha of false positives. It could also be that my measurements are not accurate enough and therefore I'm getting results which do not reflect the real situation.

- **Problem 1.5**

(8). (3 points) - Use http://asciimath.org/ *for equation conversion if necessary.*

Solution:

Relation exists & found significant (RE&FS): R*(1-beta)

Found significant (regardless of existing relation) (FS): (R*(1-beta)+alpha)

PPV when test once: (RE&FS)/FS

PPV when testing n times: (1-(RE&FS))^n/FS

Given that the numerator (1-(RE&FS))^n will always be smaller than or equal to 1, it will always decrease the whole PPV value.

(9). (2 points)

Solution: If all hypothesis were to be true then PPV would not decrease even if results were biased of increased in teams.

(10). (5 points)

Solution: If unanimous (n teams) replication is required, the PPV would be: *(R*(1-beta)+alpha)^n.* Ie, all teams would have to find the relationship statistically significant, regardless of whether it is true or not that there is a relationship.

(11). (3 points)

Solution: Yes, without replication from other teams we are more likely to have false positive results therefore it is more likely that publications will be false rather than true.

(12). (2 points)

Solution: It influences beta, hence increasing the PPV value as more relationships will be found significant.