Written Report – 6.419x Module 1

<div align="right">

Name: ptoche
</div>

Long calculations and code snippets are in the appendix at the end of this report.

Problem 1.1: The Salk Vaccine Field Trial

Introduction

The first polio epidemic hit the United States in 1916. By the 1950s several vaccines against the disease had been discovered. The one developed by Jonas Salk seemed the most promising in laboratory trials. By 1954, the National Foundation for Infantile Paralysis (NFIP) was ready to try the vaccine in the real world. They ran a controlled experiment to analyze the effectiveness of the vaccine. The data is shown in the first table below (grade refers to educational stage).

Table 1: NFIP Study

		Size	Polio rate per $100,000$
Grade 2	(vaccine)	225,000	25
Grade 1 and 3	(no vaccine)	725,000	54
Grade 2	(no consent)	125,000	44

From this table, you interpret that the experiment was run the following way:

- *Vaccines were offered to Grade 2 students, but some Grade 2 students did not consent and opted out of the offered vaccine.*

- *Vaccines were not offered to Grade 1 and Grade 3 students.*

The experiment was later repeated as a randomized controlled double-blind experiment. The data is shown in the second table below.

Table 2: Randomized Controlled Double-Blind Experiment

		Size	Polio rate per $100,000$
Treatment	(vaccine)	200,000	28
Control	(salt injection)	200,000	71
Other	(no consent)	350,000	46

In this problem, you will compare these two studies.

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)

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)

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. (Maximum 200 words)

(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. (Maximum 200 words)

(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. (Maximum 200 words)

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? (Maximum 200 words)

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? (Maximum 200 words)

Problem 1.3: The American Statistical Association on p-values

Read the statement by the American Statistical Association about p-values (Wasserstein and Lazar: The ASA's statement on p-values: context, process, and purpose). Limit your answers for each part to 300 words.

(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~~ many 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?

(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?

(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?

(b-2)

(2 points) A neuroscience lab is interested in how consumption of sugar and cocoa may ~~effect~~ affect 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?

(b-3)

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

(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 than 0.05. Should they conclude that chocolate consumption leads to improved cognitive power in mice?

(b-5)

(3 points) The lab collects individual level data on 50,000 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?

(c)

(3 points) A lab just finishes a randomized controlled trial on 10,000 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"?

(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 facts 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 correct?

(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?

(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?

(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?

Problem 1.5: Why Most Published Research Findings Are False

Read the paper by Ioannidis on why most published research findings are false (PLoS Medicine, 2005). By answering the questions below you will summarize the paper in your own words.

1-9.

autograded section omitted

8.

(3 points) Show that the extent of repeated independent testing by different teams can reduce the probability of the research being true. Start by writing the PPV as

$$PPV = \frac{I\!\!P[\text{relation exists, at least one of the } n \text{ repetitions is significant}]}{I\!\!P[\text{at least one of the } n \text{ repetitions is significant}]}$$

(Note that this does not include a bias term and you will not need one to answer this question.)

9.

(2 points) What would make bias or increasing teams testing the same hypothesis not decrease PPV? (Assuming $\alpha = 0.05$.)

10.

(5 points) Read critically and critique! Remember the golden rule of science, replication? For the third table in the paper, if researchers work on the same hypothesis but only one team finds significance, the other teams are likely to think the results is not robust, since it is not replicable. In light of this, how would you model the situation when multiple teams work on the same hypothesis and the scientific community requires unanimous replication? What would be the PPV? (You do not need to include a bias term for this question.)

11.

(3 points) Suppose there is no bias and no teams are racing for the same test, so there is no misconduct and poor practices. Will publications still be more likely to be false than true?

12.

(2 points) In light of this paper, let's theoretically model the problem of concern in Problem 1.3! Suppose people base the decision to making scientific claim on p-values, which parameter does this influence? R, α, or β? Describe the effect on the PPV if scientists probe random relations and just look at p-value as a certificate for making scientific conclusion.