Written Report – 6.419x Module 1

Name: (NAI CHUN CHEN)

Problem 1.1 The Salk Vaccine Field Trial

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. (We recommend 100 words. Maximum 200 words)
Solution: To determine the effectiveness of the vaccine using a randomized controlled double-blind experiment, follow these steps:

1. Recruitment: Gather a large, diverse sample of participants.
2. Randomization: Randomly assign participants to the treatment group or the control group to prevent selection bias.
3. Blinding: Ensure that neither participants nor researchers know which group participants are in, achieved through coded labeling of treatments.
4. Administration: Administer the vaccine to the treatment group and the placebo to the control group under identical conditions.
5. Monitoring: Track both groups over a predetermined period, noting the incidence of polio cases.
6. Data Collection: Systematically collect and accurately record data on polio cases.
7. Analysis: Compare polio incidence rates between the treatment and control groups using statistical methods to determine if there is a significant difference, indicating vaccine effectiveness.

This approach ensures unbiased results, with randomization and blinding preventing selection and observation bias, respectively.

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.(*We recommend 100 words. Maximum 200 words*)

Solution: NFIP Study:

- Grade 2 (vaccine): 25 per 100,000
- Grade 1 and 3 (no vaccine): 54 per 100,000
- Grade 2 (no consent): 44 per 100,000

The lower polio rate in the vaccinated Grade 2 group (25 per 100,000) compared to the unvaccinated groups (54 and 44 per 100,000) suggests the vaccine is effective.

Randomized Controlled Double-Blind Experiment:

- Treatment (vaccine): 28 per 100,000
- Control (salt injection): 71 per 100,000
- No consent: 46 per 100,000

The polio rate in the treatment group (28 per 100,000) is significantly lower than in the control group (71 per 100,000), indicating the vaccine's effectiveness. The "no consent" group's rate (46 per 100,000) further supports this conclusion, as it is lower than the control but higher than the vaccinated group.

3. Both studies show a clear reduction in polio incidence among those who received the vaccine, demonstrating its effectiveness.
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.

- (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.

  (We recommend 100 words. 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.
  (We recommend 50 words. Maximum 200 words)

Solution:

(a) Differences Between Grade Levels

Yes, differences between Grade 1, Grade 3, and Grade 2 students can influence the results of the NFIP experiment. For example, if younger children (Grade 1) are more susceptible to polio than older children (Grade 3), the higher polio rate in these groups could be due to age differences rather than the lack of vaccination.

Experimental Design to Prevent This: Use a randomized controlled trial (RCT) where students from all grades are randomly assigned to either the vaccine or control group. This

ensures that any age-related susceptibility to polio is evenly distributed between the groups, making the comparison of polio rates more reliable.

(b) Bias from Non-Blind Study

Knowing whether they received the vaccine or not could bias the results. For example, vaccinated children might engage in riskier behavior, assuming they are protected, potentially exposing themselves to infection sources. Conversely, unvaccinated children might take more precautions.

Preventing This Bias: Implement a double-blind design where neither the participants nor the administrators know who received the vaccine and who received a placebo. This prevents behavior changes based on vaccination status and ensures that differences in polio rates are due to the vaccine's effectiveness, not behavioral changes.

(c) Influence of "Getting Vaccine" Itself

Even if receiving the vaccine reduces infection rates, it might not be the vaccine itself causing the reduction. For instance, children who get the vaccine might receive more overall medical care or take additional health precautions.

Eliminating Non-Vaccine Biases: Include a placebo group in the experimental design. Randomly assign participants to receive either the vaccine or a placebo injection. Ensure that both groups receive identical care and instructions, isolating the effect of the vaccine itself from other potential influences on infection rates.

2. (**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:

The lower polio rate in the no-consent group compared to the control group can be attributed to several factors:

1. **Selection Bias:** Parents who opted out might be more health-conscious or have better access to healthcare and hygiene, reducing their children's risk of polio.
2. **Herd Immunity:** If many people are vaccinated, it reduces the virus's circulation, indirectly protecting unvaccinated individuals through herd immunity.
3. **Behavioral Differences:** The no-consent group might take more precautions, knowing they are unvaccinated, such as practicing better hygiene or avoiding crowds, leading to lower polio rates.

4. **Reporting Bias:** Differences in healthcare access or awareness might affect how polio cases are reported or diagnosed in the no-consent group versus the control group.

These factors underscore the importance of considering potential biases and designing experiments carefully to ensure accurate and reliable results.

The parents' conclusion that participating in the trial increased the risk of polio is not necessarily correct. The slightly higher polio rate (49 per 100,000) among children whose parents consented compared to the no-consent group (46 per 100,000) could be due to random variation or other factors, not the trial itself.

Potential Consequences:

1. **Selection Bias**: If many parents refuse participation, it creates a biased sample, making the study less representative of the general population and skewing results.
2. **Reduced Statistical Power**: Fewer participants mean the study might lack the statistical power to detect significant differences, making it harder to assess the vaccine's effectiveness.
3. **Herd Immunity Impact**: Lower participation could reduce herd immunity, potentially leading to higher overall polio rates and undermining public health efforts.

5. Problem 1.3

(a-1) Using a large set of variables and selecting those with statistically significant results to inform policy can be problematic. This approach risks overfitting, where the model captures noise instead of true relationships, leading to unreliable predictions. It can also result in p-hacking, where researchers inadvertently find significant results by chance.

(a-2) Even with more data, including many variables and selecting significant ones can still mislead due to overfitting and multiple comparisons. Larger datasets improve precision but don't eliminate these issues. To find true effects, use robust methods like pre-registering hypotheses, cross-validation, and corrections for multiple testing. Combining these with theoretical guidance ensures reliable results for policy-making.

(b-2) No, the lab should not conclude that chocolate consumption directly affects intelligence based on the correlation found. Correlation does not equal causation, and other factors like socioeconomic status or education systems could influence the results. Controlled experiments or longitudinal studies are necessary to establish a direct causal relationship and account for these confounding variables

(b-3) To study the relationship between chocolate consumption and intelligence, the lab can:

1. Conduct Controlled Experiments: Randomly assign participants to consume varying amounts of chocolate and monitor their intelligence levels over time.
2. Perform Longitudinal Studies: Track individuals' chocolate consumption and intelligence development over an extended period.
3. Control for Confounding Variables: Account for factors like socioeconomic status, education, and health to isolate the effect of chocolate.
4. Use Diverse Samples: Ensure the sample represents different demographics to generalize findings.

These approaches will provide more reliable evidence of any causal link between chocolate consumption and intelligence

(b-4) No, the lab should not immediately conclude that chocolate consumption improves cognitive power in mice based solely on a p-value below 0.05. While this indicates a statistically significant difference, it doesn't establish causation. The lab should verify that the experiment is robust, replicable, and accounts for other variables. Further studies are needed to confirm these findings and understand the mechanisms involved before making definitive conclusions about chocolate's effect on cognitive abilities

(b-5) No, this approach is not entirely correct. Reporting only statistically significant results while ignoring others can lead to biased conclusions and p-hacking. It's crucial to report all findings, including non-significant ones, to ensure a complete and transparent analysis. This maintains the study's integrity and avoids misleading interpretations.

(c) The title "New trial shows strong effect of drug X on curing disease Y" is somewhat appropriate but should specify statistical significance. The title "New drug proves over 95% success rate of drug X on curing disease Y" is misleading, as a p-value below 0.05 does not indicate a 95% success rate. Accurate reporting should highlight statistical significance and avoid overstating results or implying causation without robust evidence.

(d) No, his reasoning is not entirely correct. A lack of statistical significance does not mean there is no effect; it could be due to insufficient data or variability. The large effect size suggests potential impact worth investigating further. Decisions should consider practical significance, cost-benefit analysis, and additional research to gather more data.

(e) Even if a test shows significant results through replication, it isn't sufficient for a definitive scientific claim. Scientific claims require multiple evidence lines, theoretical support, robustness in various contexts, and elimination of confounding factors. Replication is crucial but must be combined with thorough analysis and validation through diverse methods and independent studies for reliability.

(f) No, it is not okay to report only the statistically significant results. This practice, known as "p-hacking" or "selective reporting," leads to biased conclusions and overestimates the true effect sizes. It undermines the integrity of the research and can mislead readers. All results, including non-significant ones, should be reported to provide a complete and transparent view of the study's findings.

(g) True. A significant p-value means the data are unlikely under the null hypothesis, but it doesn't prove the null hypothesis is false. The statistical model might not match reality, leading to a significant p-value even if the null hypothesis is true. Mis-specification or unaccounted confounders can cause this.

7. Problem 1.5

(8) To show that repeated independent testing by different teams can reduce the probability of research being true, we start with the positive predictive value (PPV):

$$\text{PPV} = \frac{\mathbb{P}(\text{relation exists} \mid \text{at least one of the } n \text{ repetitions finds significant})}{\mathbb{P}(\text{at least one of the } n \text{ repetitions finds significant})}$$

With repeated independent testing, the probability of at least one significant result increases, even by chance, thereby decreasing the overall PPV and the probability of the research being true if results are only significant once out of many tests.

(9) Addressing Bias and Multiple Testing Impact on PPV

Bias: Mitigate bias by pre-registering study protocols, blinding, and ensuring transparency in methods. Reducing bias maintains higher PPV.
Multiple Testing: For multiple testing by different teams, use meta-analysis, apply corrections for multiple comparisons (like Bonferroni), and uphold rigorous replication standards. This helps maintain stable PPV despite multiple testing.

(10) Modeling with Unanimous Replication Requirement

If multiple teams must unanimously replicate a finding for acceptance, adjust the PPV formula:

$$\mathrm{PPV} = \frac{\mathbb{P}(\text{relation exists} | \text{all } n \text{ teams find significance})}{\mathbb{P}(\text{all } n \text{ teams find significance})}$$

This increases PPV, as unanimous replication by all teams indicates higher reliability and robustness, making the findings more likely to be true.

(11) Even without bias or misconduct, publications can still be more likely to be false due to factors like small sample sizes, low statistical power, and multiple testing. These issues lead to false positives and overestimated effects. Therefore, rigorous methodologies, larger sample sizes, and replication are essential to enhance the reliability of published research findings.

(12) Basing scientific claims on p-values affects the significance level, α. If scientists explore random relations and rely on p-values alone, the PPV decreases. This approach increases false positives, as p-values do not consider prior probability RRR or power β, leading to many findings being significant by chance and less likely to be true.