Exercise 1: Game Fun: Customer Acquisition through Digital Advertising

Q1A: Comparing averages of income, gender and gamer variables in test and control groups. Report the % difference in the averages and compute statistical significance.

Answer: 1

part a - Yes, it is important to ensure that the test and control groups are probabilistically equivalent on their observables before evaluating the effect of an experiment. One way to check this is by comparing the distribution of the demographic variables (gender, income, and gamer) and the website that a customer visited between the test and control groups. If there are significant differences in the distributions of these variables, it may indicate that the randomization process was not effective and that there may be biases in the experiment results. To check for probabilistic equivalence, you can use statistical tests such as t-tests or chi-squared tests to compare the means or proportions of the variables between the test and control groups. If the p-values of these tests are greater than the significance level (usually 0.05), then it indicates that there are no significant differences between the groups and that they are probabilistically equivalent.

Percentage difference in means:

Gender difference = - -.095%
Gamer difference = -.081%
Income difference = -.41%

T tests proves that the difference is not statistically significant as we fail to reject the null hypothesis.

Q1B Comment on metrics that tell you about probabilistic equivalence for this experiment

Ans: The metrics calculated (i.e., the means and percentage differences) suggest that there are no significant differences between the test and control groups with respect to the demographic variables of gender, gamer, and income. Specifically, the percentage differences are quite small, ranging from -0.095% to -0.41%, indicating that the means of the variables are nearly identical across the two groups. These findings suggest that the randomization process used in the experiment was successful in creating two equivalent groups that are comparable with respect to their demographic characteristics. This is important because it ensures that any differences in the purchase rates between the two groups can be attributed to the treatment (i.e., the display ad campaign) and not to any pre-existing differences between the groups

Q1C If you had run this type of analysis before executing the experiment and found a large difference between test and control groups, what should you do?

Ans: If a large difference between test and control groups is found before executing an experiment, it indicates that the two groups are not probabilistically equivalent. In such a case, it is crucial to investigate and identify the reasons for the imbalance. Possible reasons for the imbalance may include incorrect randomization, sample selection bias, or other factors that may have influenced the allocation of subjects to the test and control groups. #In such a situation, it is essential to address the issue and try to balance the groups to ensure that any difference in the outcome is attributable to the treatment and not to any confounding variables. One way to address the issue is to increase the sample size, which may help in balancing the groups. Alternatively, it may be necessary to adjust for any confounding variables by using statistical techniques such as regression analysis or propensity score matching. It is important to note that identifying and addressing issues with the experimental design before executing the experiment is critical to ensure that the results are reliable and unbiased.

Q1D If you had millions of consumers, your “classic” statistical significance tests would not work (this is because the number of samples is used to compute those classic statistical tests). Do some research online and propose what significance test would you do in case you had big data.

Ans: Traditional statistical significance tests assume that the sample size is relatively small compared to the population size. These tests rely on the central limit theorem, which states that the distribution of sample means approaches a normal distribution as the sample size increases. This allows for the use of techniques such as the t-test or ANOVA to test for differences between groups or the correlation between variables. However, when dealing with "big data" where the sample size is very large, traditional statistical significance tests may not be appropriate for several reasons:

Statistical significance does not necessarily imply practical significance: With large sample sizes, even small differences between groups or correlations between variables can be statistically significant, but the effect size may be so small that it has no practical significance. In other words, the difference or correlation may be statistically significant, but it may not be meaningful in the real world.

Type I errors: With large sample sizes, the probability of a Type I error (rejecting the null hypothesis when it is true) increases. This is because even small deviations from the null hypothesis can lead to statistically significant results, which may not reflect a true effect.

Computational limitations: Traditional statistical significance tests may become computationally impractical or impossible with very large sample sizes, as they require a lot of memory and processing power. Assumptions may be violated: With very large sample sizes, assumptions such as normality, homogeneity of variance, and independence may be violated, which can lead to biased or incorrect results. Therefore, alternative methods may be more appropriate for analysing big data, such as machine learning algorithms or resampling techniques like bootstrapping or permutation tests. These methods do not rely on the assumptions of traditional statistical tests and can handle very large sample sizes more effectively.

Question 2

Evaluate the average purchase rates in the test and control for the following groups. For each comparison, report the average purchase rate for the test, average purchase rate for the control and the absolute difference (not the % difference) between the test and control.
a. Comparison 1: All customers
b. Comparison 2: Male vs Female customers
c. Comparison 3: Gamers vs Non-Gamers Customers
d. Comparison 4: Female Gamers vs Male Gamers

Ans:

All customers :
Average purchase rate for the test group: 0.0768
Average purchase rate for the control group: 0.0362
Absolute difference between the test and control purchase rates 0.0406

Male vs Female customers:

Comparison 2: Male vs Female customers
Avg. purchase rate for males in test group: 0.0746
Avg. purchase rate for males in control group: 0.0372
Absolute difference for males: 0.0374
Avg. purchase rate for females in test group: 0.0809
Avg. purchase rate for females in control group: 0.0344
Absolute difference for females: 0.0465

Gamer vs non-Gamer customers:

Comparison 3: Gamers vs Non-Gamers Customers
Avg. purchase rate for gamers in test group: 0.1045
Avg. purchase rate for gamers in control group: 0.0354
Absolute difference for gamers: 0.0691
Avg. purchase rate for non-gamers in test group: 0.0351
Avg. purchase rate for non-gamers in control group: 0.0351

Absolute difference for female gamers: 0.0781
Absolute difference for male gamers: 0.0641

Q3: Assess the expected revenue in the test vs. control for the following comparisons:

Comparison 1:

For all customers:

Average Test Revenue per user: 0.960272
Average Control Revenue per user: 1.357991
On an average Total Test Revenue: 26975
On an average Total Control Revenue: 16237.5

Female Gamers vs Male Gamers:

On an average Test Revenue per customer for male gamers: 1.267551
On an average Test Revenue per customer for female gamers: 1.376147
On an average Control Revenue per customer for male gamers: 1.397815
On an average Control Revenue per customer for female gamers: 1.201543

On average Total Test Revenue for male gamers: 13812.5
On average Total Test Revenue for female gamers: 8250
On average Total Control Revenue for male gamers: 6525
On average Total Control Revenue for female gamers: 10436.7

Q4: Based on your previous answers, provide a brief recommendation to your management team summarizing the expected financial outcome for Game-Fun. a. Should Game-Fun run this promotion again in the future? If no, explain why. If yes, should Game-Fun offer it to all customers or a targeted segment

Ans: Based on the analysis, the promotion appears to have a positive impact on revenue for both the test and control groups. The average revenue per customer in the test group was lower than the control group, but the total revenue for the test group was higher.
#In terms of gender, female gamers had a higher average revenue per customer in the test group, while male gamers had a higher average revenue per customer in the control group. However, the total revenue for female gamers was higher in the control group, while the total revenue for male gamers was higher in the test group.
#Overall, it seems that Game-Fun should run this promotion again in the future, but it should be offered to a targeted segment instead of all customers. Based on the analysis, female gamers in the control group appeared to be the most lucrative segment, with the highest average revenue per customer and total revenue. Therefore, Game-Fun could consider offering the promotion to this segment in the future to maximize revenue.

Exercise 2: Non-Compliance in Randomized Experiments

Question 1A: What percent of babies whose mothers were offered Vitamin A shots for their babies died?
Ans: The percent of babies whose mothers were offered Vitamin A shots for their babies that died is 0.38%.

Question 1B: What percent of babies whose mothers were not offered Vitamin A shots for their babies died?
Ans: The percent of babies whose mothers were not offered Vitamin A shots for their babies that died is 0.64%.
Question 1C: What is the difference in mortality, and under what assumptions is the difference between these two percentages a valid estimate of the causal impact of receiving vitamin A shots on survival?
Ans: Difference in mortality: -0.26 %

Assumptions to be made:

The difference in mortality is the difference between the percentage of babies who received the shot and did not survive and the percentage of babies who did not receive the shot and did not survive. Under the assumption that there are no other systematic differences between the two groups, the difference in mortality is a valid estimate of the causal impact of receiving Vitamin A shots on survival. In other words, if we assume that the only difference between the two groups is whether they received the shot, then any difference in outcomes must be due to the treatment. However, it is possible that there are other differences between the two groups that could affect the outcome, such as differences in socioeconomic status, access to healthcare, or environmental factors. In that case, the difference in mortality may not be a valid estimate of the causal impact of receiving Vitamin A shots on survival.

In the Sommer and Ziegler (1991) study, the Vitamin A supplements were randomly assigned to villages, but some of the individuals in villages who were assigned to the treatment group failed to receive them. This means that the study suffered from one-sided non-compliance, which occurs when some individuals in the treatment group do not receive the treatment. In this case, none of the individuals in the control group received the supplements.
One-sided non-compliance can introduce bias into the analysis because it can make the treatment group more like the control group than it would be if everyone in the treatment group had received the treatment. This can lead to an underestimate of the treatment effect, which means that the difference in mortality between the treatment group and the control group may be smaller than it would be if everyone in the treatment group had received the Vitamin A supplements.
Therefore, although the difference in mortality of -0.26% suggests that receiving Vitamin A shots for babies may reduce their risk of mortality, we need to be cautious in interpreting this result due to the one-sided non-compliance issue in the study.

The difference in mortality of -0.26% means that the percentage of babies who died among those whose mothers were offered a Vitamin A shot was 0.26% lower than the percentage of babies who died among those whose mothers were not offered a Vitamin A shot. In other words, the evidence suggests that receiving Vitamin A shots for babies may reduce their risk of mortality. However, we need to be cautious in interpreting this result because of the one-sided non-compliance issue in the study.

Question 2A: What percent of babies who received Vitamin A shots died?
Ans: The percent of babies who received Vitamin A shots that died is 0.12%.
Question 2B: What percent of babies who did not receive Vitamin A shots died?
Ans: The percent of babies who did not receive Vitamin A shots that died is 0.77%.
Question 2C: What is the difference in mortality, and under what assumptions is the difference between these two percentages a valid estimate of the causal impact of receiving vitamin A shots on survival?

Difference in mortality: -0.647012 %

Assumptions: The difference in mortality is -0.64%. This means that the mortality rate was lower among babies who received a Vitamin A shot compared to those who did not receive a shot.

Assuming that the only systematic difference between the two groups is whether or not they received Vitamin A shots, the difference in mortality can be interpreted as the causal impact of receiving vitamin A shots on survival. This is known as the "Intention-to-Treat" (ITT) estimate of treatment effect. However, this assumption may not always hold true in practice. There may be other factors that influence both whether a baby receives Vitamin A shots and their likelihood of survival, such as socioeconomic status, access to healthcare, or underlying health conditions. If these factors are not accounted for, the difference in mortality may be biased and not a valid estimate of the causal impact of receiving Vitamin A shots on survival. In this case, more sophisticated statistical methods, such as regression analysis, may be needed to estimate the causal impact of the treatment while controlling for these confounding factors.
However, as mentioned earlier, we need to be cautious in interpreting this result due to the non-compliance issue in the study.

Question 3A: What percent of babies who received Vitamin A shots died?
Ans: 0.124%
Question 3B: What percent of babies whose mothers were offered Vitamin A shots, but the mothers did not accept them, died?
Ans: 1.40%
Question 3C: What is the difference in mortality, and under what assumptions is the difference between these two percentages a valid estimate of the causal impact of receiving vitamin A shots on survival?
Ans: Difference in mortality: -1.28 %

Assumptions: The difference in mortality is -1.28%, meaning that the mortality rate is lower for babies who received Vitamin A shots compared to those whose mothers were offered the shots but did not accept them.
The validity of this estimate depends on the assumption that the only difference between the two groups is whether or not they received Vitamin A shots. If there are other factors that differ between the two groups, such as differences in health status or access to healthcare, then the difference in mortality may be biased and not accurately reflect the causal impact of receiving Vitamin A shots on survival. Additionally, the assumption of no

hidden bias needs to be satisfied. To ensure the validity of our estimate, we assume comparability of the groups, random decision-making for babies receiving the shot, no spill over effects, SUTVA, and no missing data.

Question 4A: Compute the above Wald estimate for the given dataset.
Ans: the Wald estimate is -0.003228 percent.

Question 4B: Under what assumptions is this estimate a valid estimate of the causal impact of vitamin A shots on survival?
Ans: The Wald estimate assumes that there is no selection bias, which means that the groups receiving and not receiving Vitamin A shots are comparable and any differences in mortality rates are due solely to the effect of the Vitamin A shots. The validity of the Wald estimate as an estimate of the causal impact of vitamin A shots on survival is based on certain assumptions. Firstly, it assumes that the assignment of the vitamin A shots to villages was done randomly, which ensures that there are no systematic differences between the treatment and control groups. Secondly, the treatment status of each baby should be consistent with the treatment assignment of its village. This means that if a village was assigned to receive the treatment, all babies in that village should have received it, and if a village was assigned to the control group, no baby in that village should have received the treatment. Thirdly, the effect of receiving the treatment should only operate through the variable being manipulated (in this case, the vitamin A shot). There should be no other variables that are affected by the treatment that could influence the outcome. Finally, the treatment status of one baby should not affect the outcome of another baby, which means there should be no spill over effects from the treatment.

Question 4C: What is the standard error for the intent-to-treat estimate recommended by the first data scientist?
What is the standard error for the Wald estimate recommended by the fourth data scientist?

i. Which one is larger and why?
ii. Why might these standard errors be biased? What information would you ideally want to have to address this bias?

Ans:

The standard error for the intent-to-treat estimate is 0.1882078
The Standard error for the Wald estimate recommended by the fourth data scientist is 0.008573505

Part 1:
The larger standard error observed in the instrumental variable (IV) estimate compared to the intent-to-treat estimate is due to the two-step regression process used in the IV method. This process introduces additional uncertainty, which is not accounted for in the intent-to-treat method that only compares the effect of receiving the treatment or not. Specifically, in the first stage of the IV method, the effect of the IV on the treatment is

estimated, which further increases the uncertainty and hence results in a larger standard error compared to the SE of the intent-to-treat estimator.

Part 2:

Unobserved confounding variables such as socioeconomic status, access to healthcare, environmental factors, and maternal beliefs in medication could potentially bias the standard errors of the intent-to-treat and two-stage least squares (2SLS) estimators if not taken into account. However, gathering information about all these factors is difficult in practice, and therefore modelling assumptions and matching may be employed to mitigate potential bias. By adjusting for these factors, we can obtain more accurate standard errors and reduce bias in our estimates.

Exercise 3: Causal Inference in Observational Studies

After reading Donald B. Rubin's article titled "The Design versus the Analysis of observational studies for causal effects: Parallels with the design of randomized trials," I gained a deeper appreciation for the importance of appropriate study design in enhancing the accuracy of causal conclusions in both randomized trials and observational studies. The article covers several thought-provoking ideas, including:

- The author highlights the common problem of confounding and selection bias in observational studies, which can lead to biased estimates of causal effects. To mitigate these issues, researchers should pay close attention to study design and subject selection and consider using methods such as propensity score matching and inverse probability weighting to adjust for potential confounding.
- The paper discusses the concept of "principal stratification," which involves analyzing the causal effects of treatments in situations where some subjects may not comply with the treatment assignment. This highlights the importance of considering the potential for non-compliance and other types of missing data in observational studies.
- Pre-specifying a causal model is emphasized as important in guiding the selection of covariates and analysis of the data. This underscores the importance of carefully considering the causal relationships among variables to make valid causal inferences.
- The paper emphasizes sensitivity analysis as essential to establish the credibility of study outcomes by evaluating the study findings' resilience to various assumptions and potential sources of bias.
- The paper draws attention to the difference in design between randomized trials and observational studies. Participants in randomized trials are randomly selected to ensure that the treatment and control groups have similar baseline characteristics. In contrast, observational studies may suffer from selection bias if the investigator selects participants who do not represent the larger population being studied.
- The author points out that in randomized trials, the treatment is assigned and measured objectively, whereas observational studies rely on self-reported data or other measures that may be subject to bias.

- While the article stresses the importance of mimicking the design of randomized controlled trials in observational studies, it is acknowledged that these two study designs are fundamentally different, and that the methods for analyzing and interpreting the data will necessarily differ between the two.

In conclusion, the article provides an insightful introduction to the complexities involved in designing and analyzing observational studies to establish causal relationships. Although the article focuses primarily on the comparison between observational studies and randomized controlled trials, most of the ideas discussed are relevant to any study design with a focus on causal inference.