Blessy Chinthapalli
HW1

Exercise 1: Game Fun: Customer Acquisition through Digital Advertising

In the A/B testing framework, especially when assessing the effectiveness of interventions like digital advertising campaigns, This step aims to verify that the random assignment of users to the test and control groups has been executed correctly, ensuring that both groups are probabilistically equivalent at the start of the experiment.

Question 1: Before evaluating the effect of an experiment, it is important to make sure that the experiment was executed correctly. To whether the test and control groups are probabilistically equivalent on their observables -

To ensure that any observed differences in outcomes (e.g., purchase rates) between the test and control groups post-experiment are due to the intervention (ad campaign) and not due to pre-existing differences.

a. Computing statistical significance:

```
         income    gender     gamer
test
0     55.166012  0.647905  0.601823
1     54.938236  0.647289  0.601331
income   -0.412890
gender   -0.095049
gamer    -0.081720
dtype: float64
TtestResult(statistic=-1.5238055388794196, pvalue=0.12757124641315226, df=22666.753090524868)
TtestResult(statistic=-0.11806188043361146, pvalue=0.9060196778523593, df=22568.271654365355)
TtestResult(statistic=-0.09200032581416154, pvalue=0.9266985974352482, df=22564.381288798413)
```

b. Comment on what these metrics tell you about probabilistic equivalence for this experiment.

Means Comparison:
The average values for income, gender, and gamer status are very close between the test (1) and control (0) groups, with differences less than 1% in most cases.

- This suggests that both groups were likely similar at the start of the experiment regarding these characteristics.

Percentage Differences

- These very small percentage differences further support the notion that the groups are statistically comparable for these key variables.

Statistical Significance (T-tests): Attributing any differences in performance metrics (like purchase rate) to the effect of the experimental manipulation (displaying Game Fun ad vs. irrelevant ad)

Null Hypothesis (H0): The default assumption(avg income, prop of M/F, gamer/ non - gamer) that there is no effect or no difference between two groups with respect to the variable under investigation.

Alternative Hypothesis (H1): The assumption (avg income, prop of M/F, gamer/ non - gamer) that there is an effect or a difference between the groups.

- Income: p-value = 0.1276; Since the p-value is greater than 0.05, we fail to reject the null hypothesis that there is no difference in income distributions between the two groups. This indicates that any observed differences in income are likely due to chance.

- Gender: p-value = 0.9060; Similarly, the high p-value here suggests no significant difference in gender proportion between the groups

- Gamer: p-value = 0.9267; Again, the high p-value indicates no statistically significant difference in the proportion of gamers between test and control groups.

c. If there's large difference between test and control groups

If the pre-experiment analysis reveals a significant difference between the test and control groups on income, gender, or gamer status, it suggests that the randomization process may have failed to create equivalent groups. This imbalance can introduce bias into the experimental results, making it difficult to attribute any differences in outcomes (like purchase rates) solely to the effects of the intervention (the advertising campaign, in this case).

If the initial randomization fails to produce equivalent groups, we can consider re-randomizing the entire sample. We can also use stratified randomization to ensure that these variables are balanced across the groups. This involves dividing participants based on income levels or age brackets and then randomly assigning participants within each stratum to test and control groups.

d. Statistical tests for big data

In scenarios involving "big data" with millions of consumers, traditional statistical significance tests can often yield misleading results. A common approach for big data analysis involves using hypothesis testing frameworks based on resampling techniques. These techniques include:

- Effect size estimation: This method helps quantify the strength or importance of a relationship between variables, or the size of a difference between groups. By using standardized metrics, effect size estimation can reveal impactful variables that traditional tests might miss in big data scenarios.

- Permutation testing: This technique involves shuffling the labels of observations (like switching test/control group labels) randomly. The test statistic is then recalculated for each shuffle. Repeating this process many times creates a "null distribution" of the statistic, which allows for p-value calculation, a key element in hypothesis testing.

Both effect size estimation and permutation testing are effective for big data analysis. While computationally intensive, these methods can be efficiently implemented with modern computing resources.

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.

Comparison 1: Average purchase rate all customers
Average purchase rate for the test group (all customers): 0.0768
Average purchase rate for the control group (all customers): 0.0362
Absolute difference between the test and control purchase rates (all customers): 0.0406

Comparison 2: Male vs Female customers

Average purchase rate for the test group (male customers): 0.0746
Average purchase rate for the control group (male customers): 0.0372
Absolute difference between the test and control purchase rates (male customers): 0.0374

Average purchase rate for the test group (female customers): 0.0809
Average purchase rate for the control group (female customers): 0.0344
Absolute difference between the test and control purchase rates (female customers): 0.0465

Comparison 3: Gamers vs Non-Gamers Customers

Average purchase rate for the test group (gamer customers): 0.1045
Average purchase rate for the control group (gamer customers): 0.0354
Absolute difference between the test and control purchase rates (gamer customers): 0.0691

Average purchase rate for the test group (non-gamer customers): 0.0351

Average purchase rate for the control group (non-gamer customers): 0.0374
Absolute difference between the test and control purchase rates (non-gamer customers): 0.0023

Comparison 4: Female Gamers vs Male Gamers

Average purchase rate for the test group (female gamer customers): 0.1101
Average purchase rate for the control group (female gamer customers): 0.032
Absolute difference between the test and control purchase rates (female gamer customers): 0.0781

Average purchase rate for the test group (male gamer customers): 0.1014
Average purchase rate for the control group (male gamer customers): 0.0373
Absolute difference between the test and control purchase rates (male gamer customers): 0.0641

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

Comparison 1: **All Customers**
Expected total net revenue for test group (all customers): $ 26975.0
Expected total net revenue for control group (all customers): $ 5412.5
Net revenue difference (test - control, all customers): $ 21562.5

Comparison 2: **Female and Male Gamers**
Expected total net revenue for test group (female gamers): $ 8250.0
Expected total net revenue for control group (female gamers): $ 1012.5
Net revenue difference (test - control, female gamers): $ 7237.5

Expected total net revenue for test group (male gamers): $ 13812.5
Expected total net revenue for control group (male gamers): $ 2175.0
Net revenue difference (test - control, male gamers): $ 11637.5

4. 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.

Based on the promotional campaign analysis for Game-Fun, the financial outcomes indicate substantial gains in net revenue across different customer segments. The test groups, which were offered a $25 signing-up bonus to promote gaming subscription packages, demonstrated significantly higher revenue generation compared to the control groups.

- All Customers: The test group generated $26,975 in net revenue, compared to $5,412.50 in the control group, resulting in a net revenue difference of $21,562.50. This suggests a robust overall effectiveness of the campaign in attracting new subscribers across the entire customer base.

- Female Gamers: Among this targeted subgroup, the test group's net revenue was $8,250, compared to $1,012.50 in the control group, with a revenue difference of $7,237.50. Female gamers responded exceptionally well to the promotion, indicating high potential for targeted marketing strategies.

- Male Gamers: The test group for male gamers brought in $13,812.50 in net revenue, while their control counterparts generated $2,175, showing a difference of $11,637.50. While the impact is also positive, it is slightly less pronounced compared to female gamers.

Given these results, it is advisable for **Game-Fun to not only continue this promotional strategy but also consider optimizing it by focusing more on targeted segments, particularly female gamers.**

This group has shown the highest relative increase in revenue contribution, highlighting their greater susceptibility to such marketing incentives. Implementing targeted promotions could further enhance efficiency, maximizing the return on marketing expenditures while fostering higher engagement and loyalty within this segment. Continuous monitoring and tweaking of the campaign based on detailed performance analytics would ensure sustained growth and profitability.

Exercise 2: Non-Compliance in Randomized Experiments

1. The first data scientist advised that one should compare the survival rate of babies whose mothers were offered Vitamin A shots to the survival rate of babies whose mothers were not offered a Vitamin A shot.

Percent of babies whose mothers were offered Vitamin A shots for their babies and died: 0.38%
Percent of babies whose mothers were not offered Vitamin A shots for their babies and died: 0.64%
Difference in mortality: -0.26%

> **Assumptions:** The validity of using the difference in mortality rates between the Vitamin A shot and control groups as an estimate of the causal impact of the treatment on survival depends on the assumption that there are no confounding variables. This means that the only difference between the two groups is whether the mothers were offered the shot or not. If there are other factors that affect both the treatment assignment and the outcome variables, such as mother's education, socio-economic status, or past health history, the estimate may be biased and the method may not be appropriate.

Insights: The difference in mortality between babies whose mothers were offered Vitamin A shots and those who were not offered was -0.258%, indicating that the babies whose mothers were offered Vitamin A shots faced a lower mortality rate than those who were not offered. To ensure that this difference is a valid estimate of the causal impact of receiving Vitamin A shots on survival, we need to assume that the groups are comparable, all mothers who were offered Vitamin A shots actually took them, there were no spillover effects, SUTVA held, and there was no missing data.

2. The second data scientist advised that one should compare the survival rates of babies who received Vitamin A shots to babies who did not receive Vitamin A shots

Percent of babies who received Vitamin A shots and died: 0.12%
Percent of babies who did not receive Vitamin A shots and died: 0.77%
Difference in mortality: -0.65%

Assumptions: As discussed earlier, a valid estimate of the causal impact of Vitamin A shots on survival can be obtained under the assumption that there are no confounding factors. However, in the context of this particular dataset, there is a one-sided compliance issue, where some mothers in the treatment group did not receive the Vitamin A shot as intended. This non-compliance may lead to biased estimates and therefore, the difference in mortality rates between the treatment and control groups may not be the appropriate approach for this analysis. Alternative methods, such as instrumental variable analysis, may be required to account for the compliance issue and provide more reliable estimates of the causal effect of Vitamin A shots on survival.

Insights: The difference in mortality between those who received Vitamin A shots and those who did not is -0.644%, suggesting that babies who received Vitamin A shots experienced a lower mortality rate than those who did not receive the shot. To consider this estimate valid, we must assume randomization of subjects, no spillover effects, SUTVA, and no missing data.

3. The third data scientist advised that one should consider only babies whose mothers were offered Vitamin A shots, and compare babies who received shots to babies who did not receive shots

Difference in mortality: -1.28%
The difference in mortality between those who received Vitamin A shots and those who did not receive them is -1.28%

Assumptions: The validity of the estimate of the causal impact of Vitamin A shots on survival depends on the assumption that the only difference between the treatment and control groups is the receipt of the shot...

Insights: Comparing the mortality rates of babies who received Vitamin A shots versus those who did not, under the assumption that only mothers who were offered Vitamin A shots were considered, the difference in mortality is -1.28%...

4. The fourth data scientist suggested the following Wald estimator for the effect of Vitamin A shots on mortality

Wald estimate is -0.003228 percent

Assumptions: The Wald estimator can provide a valid estimate of the causal impact of vitamin A shots on survival under the assumption of no confounding and that the treatment assignment is independent of potential outcomes, given the covariates used in the analysis. However, if any of these assumptions are violated, such as the presence of unmeasured variables associated with both the treatment and outcome, the estimate may be biased and unreliable. Therefore, caution must be exercised, and alternative methods like IV's must be considered

The standard error for Intent-To-Treat estimate is 0.0009278269110806206
The standard error for Wald estimate is 0.0008211357141670643

The Wald standard error is typically higher than the ITT error due to potential weak instrument problems or heterogeneous treatment effects across the population...

Exercise 3: Causal Inference in Observational Studies

Donald Rubin's paper on the design and analysis of observational studies for causal effects offers invaluable insights into the complexities of conducting observational studies. After reading the paper, I have reflected on the following points:

- **Minimal Bias:** Rubin emphasizes the importance of designing observational studies to minimize bias in the estimation of causal effects, which is in line with the design principles of randomized trials. This insight helps us understand that randomized trials and observational studies have more in common than we might think.
- **Perfect Designs:** Rubin focuses on the importance of objectivity in study design prior to examining outcome data. This practice ensures that subsequent model adjustments produce consistent point estimates, a vital aspect of the study design process.
- **Stratified sampling** and using propensity scores to create sub-groups with similar covariate distributions in both test and control groups are critical for the design of observational studies. These techniques are as close to random treatment assignment, enhancing the study's integrity.
- The tobacco example effectively illustrates the value of stratified sampling in observational studies. Rubin's focus on the principal stratification framework is also insightful, as it allows researchers to identify treatment effect heterogeneity and assess the validity of assumptions in the analysis.

• **Confounding factors** can be a potential challenge in observational studies, and it may not be possible to account for them in the study design. However, researchers should be aware of this limitation and address it in the analysis.

Overall, Rubin’s paper excellently explains the design and analysis methodologies of observational studies for causal effects, reinforcing the necessity to minimize bias in causal inference. This work has significantly enhanced my understanding of these critical concepts and their application in real-world research contexts.