University of Oxford: MPhil in Politics

1090063

2025-02-08

Contents

1 Problem 1

We will examine how randomized experiments work by creating an imaginary experiment. Use the dataset `a` from the file called `experiment.Rda`. For each individual unit (i) in our sample, the dataset contains the potential outcome under control (Y_i^0 or $Y_0 i$) and the potential outcome under treatment (Y_i^1 or $Y_1 i$) in the columns `a$y0` and `a$y1`, respectively. This is a purely hypothetical scenario. In reality, we never observe potential outcomes under both treatment and control for the same units: we can only observe one of them (the fundamental problem of causal inference). By creating a randomized experiment with this dataset, we'll demonstrate how experiments overcome this fundamental problem.

1.1 Question 1

Find the "true" Average Treatment Effect across all units. **[5 points]**

The true Average Treatment Effect (ATE) is the value obtained if we were able to observe both the potential outcomes (treatment and control) for each individual. The true ATE is `20.995` and is calculated by taking the mean difference between the potential outcomes under treatment ($Y_1 i$) and control ($Y_0 i$) for each unit. However, this is a purely hypothetical scenario as we can never observe both potential outcomes for the same unit in reality.

1.2 Question 2

Next, we'll implement a randomized experiment on this sample of 100 units. We randomly assign half of the units to treatment and half to control by creating a new variable indicating treatment status (D_i).

Conduct a test to assess whether the treatment and control groups have the same average potential outcomes under control. Has randomization succeeded in creating treatment and control groups with equivalent potential outcomes under control? Why? **[5 points]**

To assess whether the randomisation succeeded in creating treatment and control groups with equivalent potential outcomes under control, we compare whether the average potential outcomes under control for the treatment and control groups are statistically different. The mean potential outcomes under control for the treatment group is `78.649` and for the control group is `80.909`. A t-test comparing the average potential outcomes under control for the treatment and control groups gives a p-value of `0.106`.

Therefore, randomisation has succeeded in creating treatment and control groups with equivalent potential outcomes under control. The null hypothesis is that the average potential outcomes under control are equal between the treatment and control groups. Given that `p(0.106) > 0.05`, we cannot reject the null hypothesis as the treatment and control groups are not statistically different from one another.

1.3 Question 3

Estimate the Average Treatment Effect based on your experiment. How similar is it to the "true" Average Treatment Effect? Explain. **[10 points]**

Since we have now randomised the distribution of the treatment and control groups, we can calculate the estimated Average Treatment Effect (ATE) to give the expected difference in outcomes between the treated and the 'comparable control'.

The estimated ATE is `19.695`. This is calculated by taking the mean difference between the potential outcomes under treatment ($Y_1 i$) of the treatment group and the potential outcomes under control ($Y_0 i$) of the control group. The esimated ATE is lower than the true ATE `19.695 < 20.995` by `6.194`% because the treatment and control groups are not perfectly balanced. However, the estimated ATE is still close to the true ATE because the randomisation has created treatment and control groups with equivalent potential outcomes under control, as shown in `1.2`.

1.4 Question 4

Now, let's see how the experimental procedure performs over repeated randomizations.

What is the average estimated ATE across your 10,000 experiments? Does this suggest that your estimator is unbiased? Why? [**10 points**]

The mean estimated ATE across the 10,000 experiments is `20.989`. This suggests that the estimator is unbiased because the average estimated ATE across the repeated randomisations is very close to the true ATE, `20.995`, a difference of `0.007`. The repeated randomisations improve our mean estimate of the ATE significatly from the single ransomisation sample done in `1.3`.

1.5 Question 5

Repeat Task (4), calculating the mean difference in potential outcomes under control (a$y0) between the treatment and control groups instead of the ATE. What is the mean difference from your 10,000 experiments? What does this signify? [**10 points**]

The mean difference in potential outcomes under control between the treatment and control groups across the 10,000 experiments is `0.004`. This signifies that the randomisation has succeeded in creating treatment and control groups with equivalent potential outcomes under control. The mean difference in potential outcomes under control is close to zero, indicating that the treatment and control groups are not statistically different from one another. This is consistent with the results from `1.2`.

2 Problem 2

Past research suggests that ballot secrecy influences turnout. A recent field experiment sent emails to a random group of nonvoters around the 2014 election in Mississippi, reminding them that their vote was secret.

For this exercise, we wish to establish if the results of this experiment hold when we focus on the female subsample (instead of the full sample).

2.1 Question 1

Confirm that the randomization process was successful by making sure that women in treatment and control groups are similar in all relevant aspects, e.g., age, ethnicity, non-voting habits. Show your results either using a figure or by producing a publishable table. [**10 points**]

To test whether the women in treatment and control groups are similar in all relevant aspects, t-tests were conducted to see whether the groups were statostically significantly different from one another on each

variable of interest. Table 1 shows the results of t-tests comparing the characteristics of women in the treatment and control groups. The null hypothesis is that the average values of the variables are equal between the treatment and control groups. The p-values are all greater than 0.05 indicating we cannot reject the null hypothesis. This suggests that the randomisation process was successful in creating treatment and control groups with similar characteristics as there are no statistically significant differences between groups. Although differences between observed variables across groups is found, there could still be selection bias from unobserved variables; however, this is unlikely if randomisation was successful.

Table 1: t-test and p-value Results for Variables

Variable	t-test	p-value
d_age	-0.354	0.723
d_race_blk	0.712	0.477
d_race_hsp	-0.279	0.780
d_race_oth	0.275	0.783
never_voted	0.695	0.487
vote_year	-0.444	0.657

Note: significance values * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.2 Question 2

Estimate the Average Treatment Effect and test whether the effects you calculated are robust to the inclusion of covariates. Report all results in a single table. Is it necessary to include covariates when calculating the ATE? [**10 points**]

Although randomisation appears to have been successful, we still want to test for the effects of covariates on the estimated Average Treatment Effect (ATE). Firstly, a simple model is estimated for the ATE without covariates. Secondly, to estimate the ATE and test whether effects are robust to the inclusion of covariates, a regression model is compared to ATE without the inclusion of `d_age`, `d_race_blk`, `d_race_hsp`, `d_race_oth`, `never_voted`, and `vote_year`. Results of these two models are shown in Table 2.

The ATE without covariates is `-0.005`. This tells us the difference in probability that someone voted between the treatment and control groups. Therefore, as the `vote_2014` is binary, those who received mail reminding them their vote was a secret saw a decreased probability of voting of `-0.532` percentage points compared to the control group,a very small amount. The t-test for the ATE without covariates gives a p-value of `0.203` meaning the results are also not statistically significant. Whilsy ballot secrecy may theoretically influence voting turnout, the effects seen are neither large nor statistically significant.

The second model accounts for the covariates which were shown to be similar across treatment and control groups in `2.1`. The inclusion of covariates does not change the estimated ATE significantly, with the ATE with covariates being `-0.005`. The p-value for the ATE with covariates is `0.096`, showing signs of weak significance, unlike when estimated without covariates. The inclusion of covariates controlling for black voters and those who have never voted before are the most significant. The model with covariates also has a higher R-squared value of `(0.563 > 0.000)` compared to the model without covariates. This suggests

Table 2: Estimated ATE for Mail Treatment (With and Without Covariates)

	ATE without Covariates	ATE with Covariates
Mail Treatment	−0.005	−0.005*
Age		0.000
Black		−0.016***
Hispanic		−0.010
Other Race		−0.011
Never Voted		0.361***
Years Since Last Vote		0.000***
Num. Obvs	7969	7969
R-squared	0.000	0.563
Adj. R-squared	0.000	0.563

* p <0.1, ** p <0.05, *** p <0.01
Note: Standard errors are in parentheses. Both models are estimated using OLS regressions.

that the inclusion of covariates improves the model's explanatory power, but whilst the estimated ATE's are similar and neither is significantly robust, the inclusion of covariates is not necessary when calculating the ATE in this case, but should be included to ensure a more robust model.