

AUA Fall 2023

DS 110 Statistics 2

Homework 3

Problem 1 (10 points)

Suppose you want to evaluate the effectiveness of a job-training program. The goal is to determine whether participating in the program increased income compared to those who did not participate. You have access to a dataset (Income.xlsx) that includes the following variables for a sample of individuals:

Participation: A binary variable (1 if the individual participated in the job-training program, 0 otherwise).

Income: The income of the individual after a certain period (measured in dollars).

Estimate the effect of participating in the job training program on income, meaning how does participating in the program affect an individual's income compared to not participating? (10 points)

Problem 2 (30 points)

Imagine you are an economist working for a government agency responsible for economic policy analysis. The government is considering a change in tax policy that would affect small businesses. The goal is to understand the potential impact of this policy change on small businesses' profitability and investment decisions.

The investment.xlsx dataset contains information on small businesses, including the following variables:

Profitability: The annual profitability (net income) of each small business.

Tax Policy: A binary variable indicating whether the small business is subject to the new tax policy (1 for affected, 0 for unaffected).

Investment: The amount of capital investment made by each small business during the year.

Estimate the causal effect of the new tax policy on both profitability and investment decisions of small businesses. Specifically, address the following questions:

1. Interpret the result of Linear Regressions. (5 points)
2. Are coefficients significant? (5 points)
3. What is the average treatment effect on profitability (ATE)? In other words, how does the new tax policy impact the profitability of small businesses that are subject to it compared to those that are not affected? (10 points)
4. What is the average treatment effect on investment (ATE)? How does the new tax policy influence the investment decisions of small businesses? (10 points)

Problem 3 (60 points, 10 points for each question)

A car renting company discovered that a defective product was used in the car manufacturing process. It took seven months to correct the defects, during which time approximately 14 cars from the 100-unit park were taken out of service for one month at a time. The company lost profits due to this. Use the cars.csv dataset to answer the following questions.

- The occupancy rate of cars of that company is CAR_PCT , and the competitor occupancy rate is $COMP_PCT$. On the same graph, plot these variables against $TIME$. Which had the higher occupancy before the repair period? Which had the higher occupancy during the repair period?
- Compute the average occupancy rate for the company and competitors when the repairs were not being made (\overline{CAR}_0 and \overline{COMP}_0) and when they were being made (\overline{CAR}_1 and \overline{COMP}_1). During the nonrepair period, what was the difference between the average occupancies, $\overline{CAR}_0 - \overline{COMP}_0$? Assume that the company's occupancy rate would have maintained the same relative difference in occupancy if there had been no repairs. That is, assume that the company's occupancy would have been $\overline{CAR}^*_1 = \overline{COMP}_1 + (\overline{CAR}_0 - \overline{COMP}_0)$. Compute the "simple" difference estimate of lost occupancy $\overline{CAR}^*_1 - \overline{CAR}_1$. Compute the amount of revenue lost during the seven-month period (215 days) assuming an average car rent rate of \$56.61 per day.
- Draw a plot that explains the calculation in part (b).
- Alternatively, consider a regression approach. A model explaining car renting occupancy uses as explanatory variables the competitors' occupancy, the relative price ($RELPRICE$) and an indicator variable for the repair period ($REPAIR$).

$$CAR_PCT_t = \beta_0 + \beta_2 COMP_PCT_t + \beta_3 RELPRICE_t + \beta_4 REPAIR_t + e_t$$

Obtain the least squares estimates of the parameters. Interpret the estimated coefficients, as well as their signs and significance.

- Using the coefficient of $REPAIR$ from part (d), estimate the revenue lost by the damaged car renting company during the repair period. Compare this value to the "simple" estimate in part (b). Construct a 95% interval estimate for the estimated loss. Is the estimated loss from part (b) within the interval estimate?
- Plot the least squares residuals against $TIME$. Are there any obvious patterns?