

Problem Set 1

Tate Mason

2025-09-14

Problem 1: Conceptual Problem

Part 1: What is the effect of whether a child watches TV on their math skill?

Part 2: Is working around retirement age good for the person's health?

Part 3: Is the racial wage gap in part due to discrimination against racial minorities?

- Note: Consider only two groups: racial minorities vs. racial majority.

Part 4: What is the effect of whether the mother receives welfare money support while the child is young on the child's future income (by age 40)?

Problem 2: Coding

Part 1: Creating dataset

1. Set random seed to 1
 2. $N = 10000$
 3. Draw ϵ_i^D / *perp* $\epsilon_i^Y \sim N(0, 1)$
 4. Draw $U_i \sim N(0, 0.5)$ (Note: s.d. is 0.5)
 5. Create $Z_i = \mathbb{1}(z_i > 0.5)$ where z_i is randomly drawn from a uniform distribution on $[0, 1]$
 6. Create $D_i = \mathbb{1}(\alpha_0 + \alpha_Z Z_i + \alpha_U U_i + \epsilon_i^D > 0)$ such that $\alpha_0 = -4, \alpha_Z = 5, \alpha_U = 4$
 7. Create $Y_i = \beta_0 + \beta_D D_i + \beta_Z Z_i + \beta_U U_i + \epsilon_i^Y$ such that $\beta_0 = 3, \beta_D = 2, \beta_Z = 0, \beta_U = 6$
- $\beta_Z Z_i + \beta_U U_i + \epsilon_i^Y = \epsilon_i$

```
df <- data.frame(
  id = 1:10000,
  epsilon_D = rnorm(10000, mean = 0, sd = 1),
  epsilon_Y = rnorm(10000, mean = 0, sd = 1),
  U_i = rnorm(10000, mean = 0, sd = 0.5)
)

z_i <- runif(10000, min = 0, max = 1)
df <- df %>%
  mutate(
    Z_i = as.numeric(z_i > 0.5),
    D_i = as.numeric(-4 + 5 * Z_i + 4 * U_i + epsilon_D > 0),
    Y_i = 3 + 2 * D_i + 0 * Z_i + 6 * U_i + epsilon_Y
  )
```

Part 2: Estimating the effect of D on Y with OLS

```
OLS <- lm(Y_i ~ D_i, data = df)
summary(OLS)
```

Call:

```
lm(formula = Y_i ~ D_i, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-10.2473	-2.0084	-0.1544	1.9659	9.2807

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.02336	0.03642	55.56	<2e-16 ***
D_i	4.72039	0.06021	78.40	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.9 on 9998 degrees of freedom

Multiple R-squared: 0.3807, Adjusted R-squared: 0.3807

F-statistic: 6146 on 1 and 9998 DF, p-value: < 2.2e-16

Part 3: Estimating the effect of D on Y with IV

Part 4: Estimating the effect of D on Y with 2SLS