Applied Statistical Analysis I/ Quantitative Methods I
POP77003/77051
Fall 2024
Week 8

Yao (Sara) HAN

hany3@tcd.ie

Department of Political Science

Trinity College Dublin

30 October, 2024

Today's Agenda

(1) Lecture recap

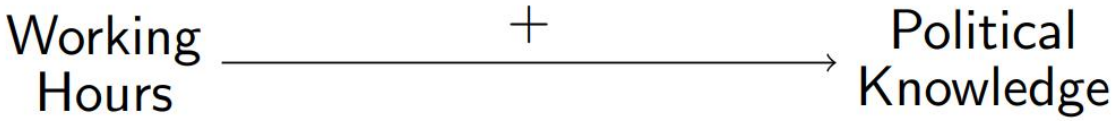
(2) Tutorial exercises

Multiple linear regression

Why do we need multiple linear regression? And what is a multiple linear regression model?

Multiple linear regression

Why do we need multiple linear regression?




```
## Call:
## lm(formula = polknow ~ work_hours, data = samp)

## Residuals:
##    Min     1Q Median     3Q    Max
## -7.686 -1.760 -0.061  1.683 10.385

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.59166    1.09142  13.369   <2e-16 ***
## work_hours   0.06791    0.02640   2.572   0.0103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 2.565 on 998 degrees of freedom
## Multiple R-squared:  0.006585,  Adjusted R-squared:  0.00559
## F-statistic: 6.615 on 1 and 998 DF,  p-value: 0.01025
```

How convincing is this finding?

Multiple linear regression

Why do we need multiple linear regression?

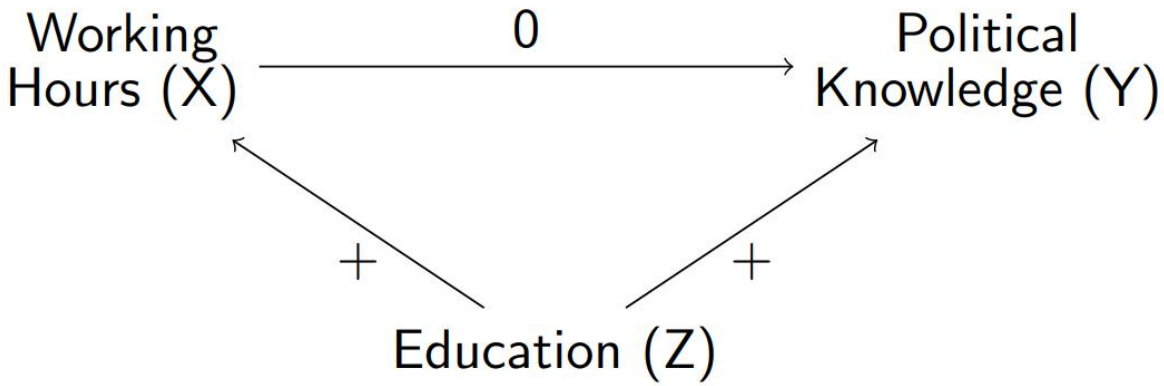


Figure: Education as confounder—Controlling for education is relevant, because it might drive both working hours and political knowledge. Education is causally prior to working hours.

$\rightarrow$ Avoid omitted variable bias. Include relevant control variables (Z) which are correlated with both X and Y, and causally prior to X.

Multiple linear regression

Why do we need multiple linear regression?

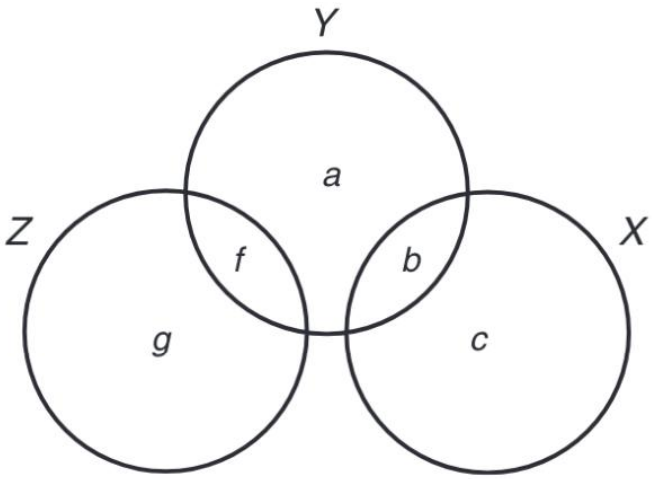


Figure 9.2. Venn diagram in which X and Z are correlated with Y, but not with each other.

"In that case – which, we have noted, is unlikely in applied research – we can safely omit consideration of Z when considering the effects of X on Y. In that figure, the relationship between X and Y – the area b – is unaffected by the presence (or absence) of Z in the model" (Kellstedt and Whitten 2018, 213).

Multiple linear regression

Why do we need multiple linear regression?

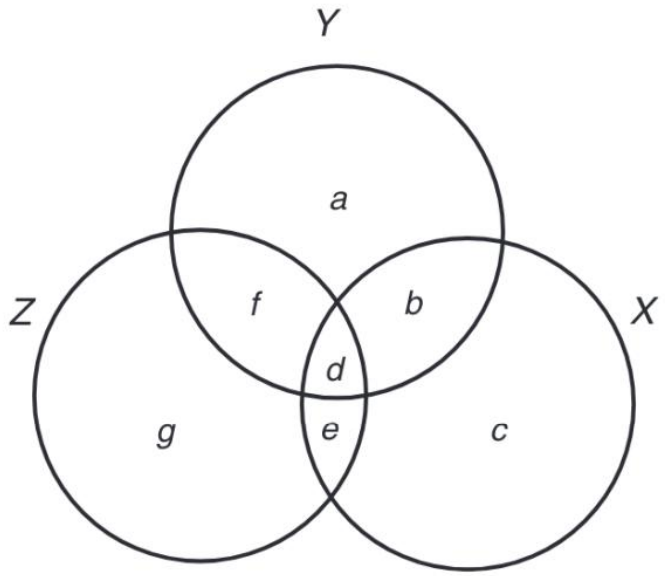


Figure 9.1. Venn diagram in which X, Y, and Z are correlated.

"If, hypothetically, we erased the circle for Z from the figure, we would (incorrectly) attribute all of the area b + d to X, when in fact the d portion of the variation in Y is shared by both X and Z. This is why, when Z is related to both X and Y, if we fail to control for Z, we will end up with a biased estimate of X's effect on Y" (Kellstedt and Whitten 2018, 212).

Multiple linear regression

What is a multiple linear regression model?

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \epsilon_i$$

- α (intercept): expected value of Y when $X_1 = 0$, ..., $X_k = 0$.
- β_1 (coefficient): expected change in Y when X_1 increases by one unit, while controlling for the remaining explanatory variables in the model.
- ...
- β_k (coefficient): expected change in Y when X_k increases by one unit, while controlling for the remaining explanatory variables in the model.

Multiple linear regression

What is a multiple linear regression model?

```
## Call:
## lm(formula = polknow ~ work_hours + edu, data = samp)

## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7835 -1.6733  0.0035  1.5941 10.6778

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.854461   1.368601   3.547 0.000408 ***
## work_hours  0.006205   0.025623   0.242 0.808714
## edu         0.767650   0.070797  10.843  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 2.427 on 997 degrees of freedom
## Multiple R-squared:  0.1114,  Adjusted R-squared:  0.1096
## F-statistic: 62.48 on 2 and 997 DF,  p-value: < 2.2e-16
```

The effect of working hours *disappears*.

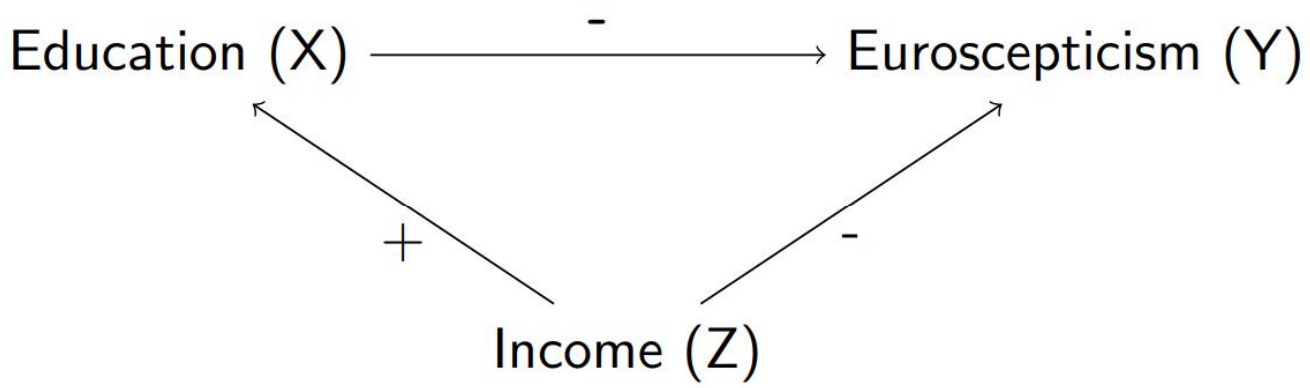
$\rightarrow$ Controlling for working hours, with every additional year of education, the political knowledge increases by 0.76765 scale points.

What is the relationship between education and Euroscepticism?

$$\text{Education (X)} \xrightarrow{\quad - \quad} \text{Euroscepticism (Y)}$$

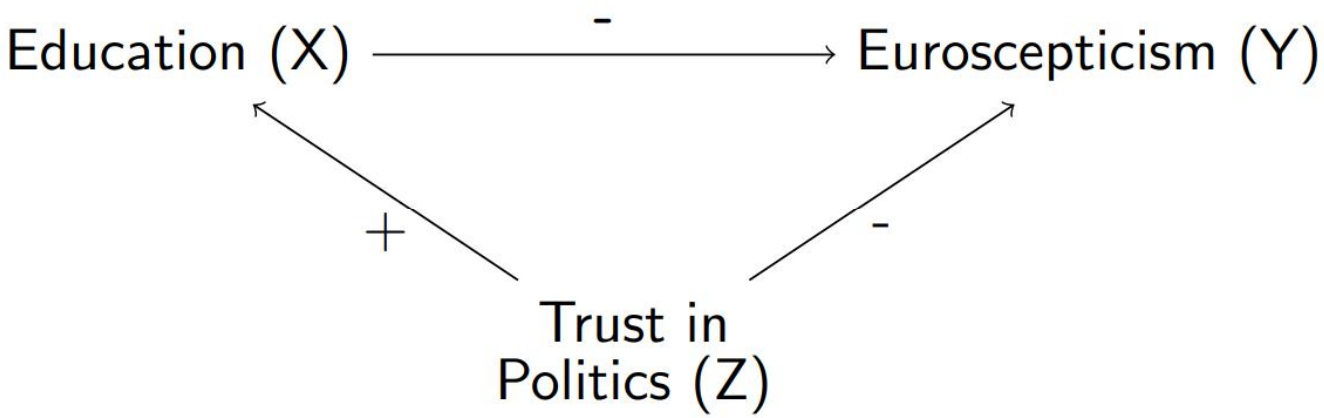
Hypothesis$_1$: The higher the years of education, the lower the level of Euroscepticism.

What is the relationship between education and Euroscepticism?



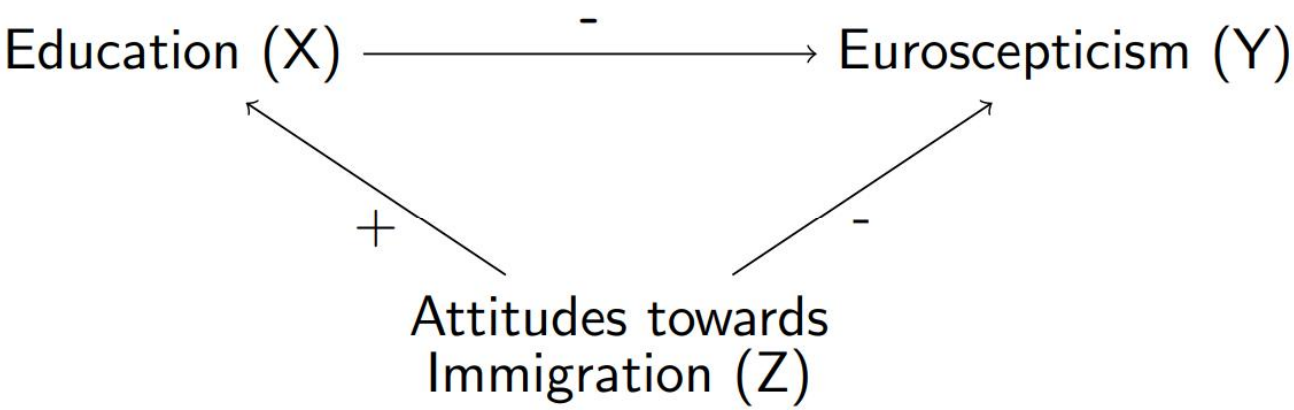
Hypothesis$_2$: The higher the income, the lower the level of Euroscepticism. $\rightarrow$ Economic dimension

What is the relationship between education and Euroscepticism?



Hypothesis$_3$: The higher the trust in politics, the lower the level of Euroscepticism. $\rightarrow$ Political dimension

What is the relationship between education and Euroscepticism?



Hypothesis$_3$: The more positive attitudes towards immigration, the lower the level of Euroscepticism. $\rightarrow$ Cultural dimension

References I


Kellstedt, Paul M., and Guy D. Whitten. 2018. *The fundamentals of political science research.* Cambridge: Cambridge University Press.