# Quantitative Sociological Analysis

# Inferential Statistics
# Multivariate Statistics

## Part 8

April 22-24, 2025

# Multivariate LRM

$$Y = \beta_0 + \beta_1 X + \beta_2 X \ldots + \beta_k X + \varepsilon$$

- When predicting an outcome ($Y$) as a function of an independent variable ($X$), this only
  - tells us how $X$ alone is associated with $Y$

- What if another variable ($Z$) is associated with $X$ and $Y$?
  - recall spuriousness (PPT 3)

- Multivariate modeling is a tool that can help address this issue
  - improve specification, the accuracy of the model
    - recall importance of theory

Let's consider this by expanding some previous simple LRM examples…

# Multivariate LRM: example

$$\hat{Y}_{edu} = b_0 + b_1 white + b_2 polit$$

- We are interested in the association between edu and political party affiliation, but we also
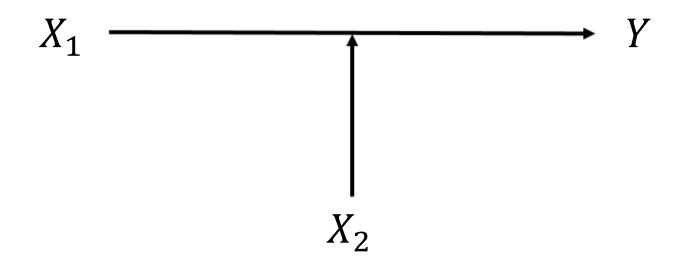  - know that race is associated with both

|  | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| intercept | 12.403*** | 12.851*** | 12.42*** |
| white | 0.712*** |  | 0.603*** |
| republican |  | 0.508*** | 0.382*** |
| independent |  | - 0.278*** | - 0.318*** |
| $R^2$ | 0.008 | 0.008 | 0.014 |

- For each model can you interpret the:
  - intercept ($\beta_0$), slope coefficient(s) ($\beta_j$), and $R^2$
    - when more than 1 X, this is holding all other Xs "constant"

We are also concerned the association between edu and political affiliation may differ by race…

# Moderation

- The effect of one predictor $X_1$ depends on the value of another $X_2$
  - in terms of strength and/or direction

$$X_1 \longrightarrow Y$$

$$\uparrow$$

$$X_2$$

- Can test with interaction terms
  - $Y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_{1*2} + \varepsilon$

# Multivariate LRM: example

$$\hat{Y}_{edu} = b_0 + b_1 white + b_2 polit + b_{1*2} white \times polit$$

- We are interested in whether the association between edu and political party differs by race

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| intercept | 12.403*** | 12.851*** | 12.42*** | 12.44*** |
| white | 0.712*** | | 0.603*** | 0.559*** |
| republican | | 0.508*** | 0.382*** | 0.046 |
| white*rep | | | | 0.371*** |
| independent | | - 0.278*** | - 0.318*** | -0.269*** |
| white*ind | | | | -0.058 |
| $R^2$ | 0.008 | 0.008 | 0.014 | 0.014 |

- The inclusion of interaction terms alters the interpretation of slope coefficients
  - regarding their respective reference group
    - Notice how the coefficients change, and the association between edu and republican is no longer statistically significant

Let's further consider how interpret the results from this model with interaction terms…

# Multivariate LRM: example

$$\hat{Y}_{edu} = b_0 + b_1 white + b_2 polit + b_3 white \times polit$$

- We are interested in whether the association between edu and political party differs by race

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| intercept | 12.403*** | 12.851*** | 12.42*** | 12.44*** |
| white | 0.712*** | | 0.603*** | 0.559*** |
| republican | | 0.508*** | 0.382*** | 0.046 |
| white*rep | | | | 0.371*** |
| independent | | - 0.278*** | - 0.318*** | -0.269*** |
| white*ind | | | | -0.058 |
| R² | 0.008 | 0.008 | 0.014 | 0.014 |

**Interpretation guide**

- Intercept: predicted years of edu for non-white dem
- White: among dem, white vs non-white edu
- Rep: among non-white, rep vs dem
- WRep: among white, rep vs dem
- Ind: among non-white, ind vs dem
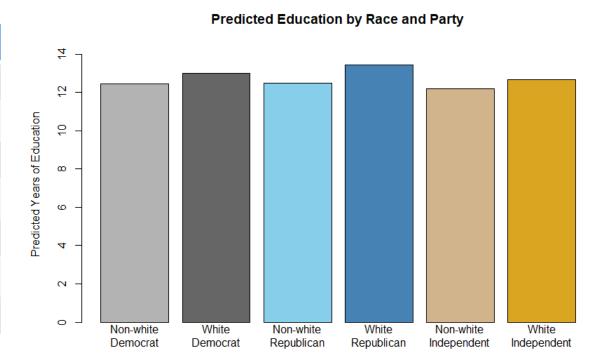- WInd: among white, ind vs dem

It may be helpful to consider predicted values visually…

# Multivariate LRM: example

$$\hat{Y}_{edu} = b_0 + b_1 white + b_2 polit + b_3 white \times polit$$

- We are interested in whether the association between edu and political party differs by race

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| intercept | 12.403*** | 12.851*** | 12.42*** | 12.44*** |
| white | 0.712*** | | 0.603*** | 0.559*** |
| republican | | 0.508*** | 0.382*** | 0.046 |
| white*rep | | | | 0.371*** |
| independent | | - 0.278*** | - 0.318*** | -0.269*** |
| white*ind | | | | -0.058 |
| $R^2$ | 0.008 | 0.008 | 0.014 | 0.014 |

**Predicted Education by Race and Party**



- See here for additional resources regarding [interaction terms in LRM](interaction terms in LRM)

Let's explore some LRM models with our Netflix survey data…

# Netflix survey data

- I augmented our dataset with 1,000 synthetic responses
  - retaining same data structure

  Week 15: Summarize learning goals     Descriptive_Table_Netflix_Augmented.pdf

- Download this augmented dataset, and
  - save in location where you can retrieve again

  Netflix_1000.RData

- Download the new Rscript, and

  RScript_Netflix_Augmented.R

  - save in a location where you can retrieve again

- Recode comedy preference to better reflect a continuous variable, so we can
  - assume slope coefficients $\beta_j$ from LRM reflect $Y$ units

  `10    comedy_pref<-6-comedy`

Let's run some LRM models with these augmented Netflix survey data…