Lecture 5: Multivariate Regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + U$$

$$= \beta_0 + \sum_{k=1}^{K} \beta_k X_k + U$$

$$= X'\beta + U$$

Omitted variable bias

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U \qquad \text{(long)}$$

$$Y = \alpha_0 + \alpha_1 X_1 + V \qquad \text{(short)}$$

$$X_2 = \gamma_0 + \gamma_1 X_1 + \varepsilon \qquad \text{(aux)}$$

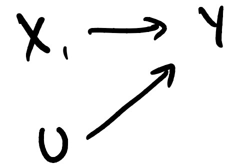
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2(\gamma_0 + \gamma_1 X_1 + \varepsilon) + U$$

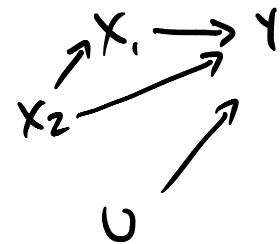
$$Y = \underbrace{(\beta_0 + \beta_2 \gamma_0)}_{\substack{\text{new} \\ \text{constant} \\ \alpha_0}} + \underbrace{(\beta_1 + \beta_2 \gamma_1)}_{\substack{\text{new} \\ \text{slope}}} X_1 + \underbrace{(U + \beta_2 \varepsilon)}_{\substack{\text{new} \\ \text{error}}}$$

$$\alpha_1 = \beta_1 + \underbrace{\beta_2 \gamma_1}_{\text{OVB}}$$

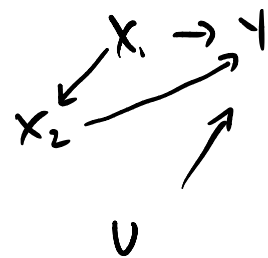
$$Y = \beta_0 + \beta_1 X_1 + U$$



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$



- confounding
- X_2 is a confounder



- mediating
- X_2 is a mediator

Linear combinations of coefficients

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

$$a\hat{\beta}_1 + b\hat{\beta}_2 \leftrightarrow a\beta_1 + b\beta_2$$

$$V[a\hat{\beta}_1 + b\hat{\beta}_2 | \mathbf{X}] = a^2 V[\hat{\beta}_1] + b^2 V[\hat{\beta}_2] + 2ab\,\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2)$$

$$= a^2 \sigma^2_{\hat{\beta}_1} + b^2 \sigma^2_{\hat{\beta}_2} + 2ab\,\sigma_{\hat{\beta}_1,\hat{\beta}_2}$$

null: $\beta_1 = \beta_2 \Rightarrow \beta_1 - \beta_2 = 0 \Rightarrow a = 1,\ b = -1$

$$V[\hat{\beta}_1 - \hat{\beta}_2] = \underbrace{\sigma^2_{\hat{\beta}_1}}_{SE[\hat{\beta}_1]^2} + \underbrace{\sigma^2_{\hat{\beta}_2}}_{SE[\hat{\beta}_2]^2} - 2\sigma_{\hat{\beta}_1,\hat{\beta}_2} \leftarrow ?$$

$$Y = \beta_0^U + \beta_1^U X_1 + U$$

$$Y = \beta_0^R + \beta_1^R X_1 + U$$

$$V[\hat{\beta}^U - \hat{\beta}^R] = SE[\hat{\beta}^U]^2 + SE[\hat{\beta}^R]^2 + \cancel{\mathrm{cov}}$$

$$SE[\hat{\beta}^U - \hat{\beta}^R] = \sqrt{V[\cdot]} \longleftrightarrow t = \frac{\hat{\beta}^U - \hat{\beta}^R}{SE}$$

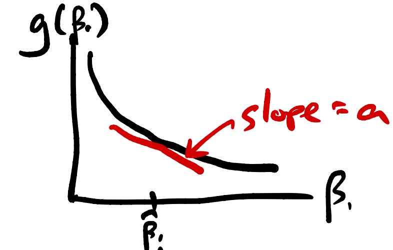
Interaction:

$$Y = \beta_0 + \beta_1 \text{age} + \beta_2 \text{urban} + \beta_3 \text{age} \times \text{urban}$$

diff: $\hat{\beta}_1^U - \hat{\beta}_1^R$

Non-linear functions of coefs

$$g(\beta_1, \beta_2) = \frac{\beta_1}{\beta_2} \quad \text{or} \quad g(\beta_1) = \frac{1}{\beta_1}$$



Delta method: 1st-order Taylor approximation

$$V[a\hat{\beta}_1 + b\hat{\beta}_2 | X] = \overset{g_{\beta_1}^2}{a^2}\, V[\hat{\beta}_1] + \overset{g_{\beta_2}^2}{b^2}\, V[\hat{\beta}_2] + 2\,\overset{g_{\beta_1} \cdot g_{\beta_2}}{ab}\, cov(\hat{\beta}_1, \hat{\beta}_2)$$

$$= a^2 \sigma_{\hat{\beta}_1}^2 + b^2 \sigma_{\hat{\beta}_2}^2 + 2ab\, \sigma_{\hat{\beta}_1, \hat{\beta}_2}$$

---

car::deltaMethod( )