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

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

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

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

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

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

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

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

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

$$\begin{tikzcd} X_1 \arrow[r] & Y \\ U \arrow[ur] & \end{tikzcd}$$

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

$$\begin{tikzcd} & X_1 \arrow[r] & Y \\ X_2 \arrow[ur] \arrow[urr] & & \\ & U \arrow[uur] & \end{tikzcd}$$

confounder

$$\begin{tikzcd} & X_1 \arrow[r] \arrow[dl] & Y \\ X_2 \arrow[urr] & & \\ & U \arrow[uur] & \end{tikzcd}$$

mediator

Linear combinations of coefficients

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

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

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

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

$SE[\hat{\beta}_1]^2$ $\quad$ $SE[\hat{\beta}_2]^2$

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

$$V[\hat{\beta}_1 - \hat{\beta}_2] = SE[\hat{\beta}_1]^2 + SE[\hat{\beta}_2]^2 - 2\sigma_{\hat{\beta}_1, \hat{\beta}_2}$$

$$Y = \beta_0^U + \beta_1^U \text{age} + U$$

$$Y = \beta_0^R + \beta_1^R \text{age} + U$$

$$V[\hat{\beta}_1^U - \hat{\beta}_1^R] = SE[\hat{\beta}_1^U]^2 + SE[\hat{\beta}_1^R]^2$$

$$t = \frac{\hat{\beta}_1^U - \hat{\beta}_1^R}{SE[\hat{\beta}_1^U - \hat{\beta}_1^R]}$$

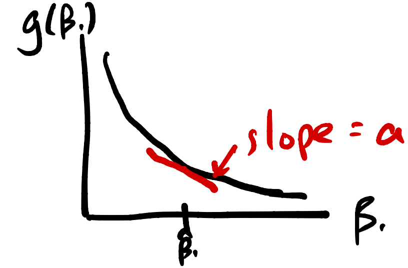
sqrt

Interactions:

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

Non-linear functions

$$g(\beta_1,\beta_2) = \frac{\beta_1}{\beta_2} \qquad g(\beta_1) = \frac{1}{\beta_1}$$



Delta method: 1st order Taylor approximation

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

$$Y = \beta_0 + \beta_1 \, white + u$$

$$\frac{\mu_W}{\mu_{NW}} = \frac{\beta_0 + \beta_1}{\beta_0} = g(\beta_0, \beta_1)$$

$$g_{\beta_0} = -\frac{\beta_1}{\beta_0} \qquad g_{\beta_1} = \frac{1}{\beta_0}$$

car::deltaMethod( )