

Omitted Variable Bias (遺漏變數偏誤)

應有而未有一解釋變數
會使迴歸結果產生偏誤

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (\text{正確})$$

$$Y = \beta_0 + \beta_1 x_1 + \underline{u'} \quad (\text{少了 } x_2)$$

此時估計結果 $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$ \rightarrow bias

原因: 令 $x_2 = \delta_0 + \delta_1 x_1 + v$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 (\delta_0 + \delta_1 x_1 + v) + u$$

$$= \boxed{} + \boxed{(\beta_1 + \beta_2 \delta_1)} x_1 + \boxed{}$$

\uparrow
 x_2 被省略, 它的效果會跑到 x_1 , 造成 x_1 係數偏誤

例:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

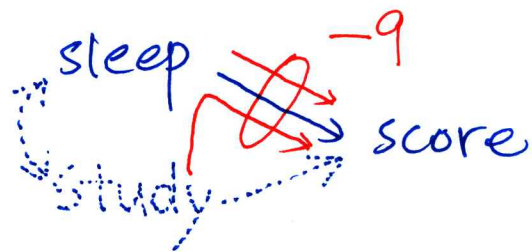
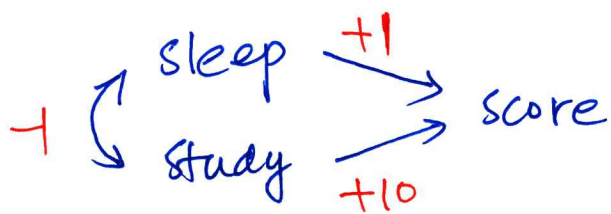
\downarrow score \downarrow +1 \downarrow sleep \downarrow +10 \downarrow study

$$\text{若 } x_2 = \delta_0 + \delta_1 x_1 + v$$

(正確) score = $\dots + (1) \text{ sleep} + (10) \text{ study} + \dots$

(偏誤) score = $\dots + (-9) \text{ sleep} + \dots$

\downarrow
(-9 中帶有 study 的效果)



Collinearity (共线性)

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + u$$

$$se(\hat{\beta}_1) \leftrightarrow Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1 (1 - R_1^2)}$$

标准误 (standard error)

VIF

$$(\sqrt{Var(\hat{\beta}_1)})$$

把 x_1 对 $x_2 \dots x_k$ 跑回归
得到的 R_1^2

$$\text{变量被放大因子 } VIF = \frac{1}{1 - R_1^2}$$

(variance inflation factor)

- 如果 $x_1, x_2, x_3 \dots$ 存在一线性关系
会使 x_1 对 x_2, x_3 跑回归时, $R_1^2 \rightarrow 1, VIF \rightarrow \infty$
也就是 $se(\hat{\beta}_1) \rightarrow \infty \Rightarrow$ 回归模型失效
- 如果 R_1^2 太高, 使 $VIF > 10$ (multicollinearity)
表变量被放大 10 倍, 并不可接受
(\therefore 一般可接受标准 $\Rightarrow VIF < 10$)