多元回歸的基本假設

1. $e_i \sim N(0,\sigma^2)$

i.i.d.為 **independently**（**獨立**）, **identically**（**同質**）, **distributed**（**分佈**）成以 0 為平均數，以σ^2為變異數的常態分配。

但是，下列兩種情況是違反此假設的：

（1）growth model 或 panel data

（2）nested data (cluster sampling)

2. 線性關係（linearity）

效標變項 y 和預測變項 x 之間，呈現直線關係。

3. 無嚴重的多元共線性（multicollinearity）問題

亦即，我們要求r_{yx_i}愈大愈好，但$r_{x_ix_j}$愈小愈好。

（1）當變異數膨脹因子（variance inflation factor, VIF）>10 時，即表示預測變項間具有嚴重的共線性問題，導致回歸係數的估計不易達顯著。

VIF＝$VIF = \frac{1}{1-R_j^2}$， R_j^2係指以x_j當作效標變項，而以其餘的預測變項 x 來預測它所獲得的R^2值。

（2）當條件數（conditional number, CN）>30 時，即表示預測變項間具有嚴重的共線性問題，導致回歸係數的估計不易達顯著。

$CN = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$，即針對預測變項矩陣 **X** 進行求解特徵值（eigenvalue）和特徵向量（eigenvector），所獲得的最大特徵值和最小特徵值的比值，再開根號（root）。

4. **預測變項沒有測量誤差存在**

$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$，每一個預測變項值都是精確的，不含測量誤差。即，不再有$x = t + e$，只有$x = t$。

（本項假設，直到結構方程式模型（structural equation modeling, SEM）的方法學出現後，已經被克服取代了；換句話說，在 SEM 模型之下，預測變項是被允許有測量誤差存在的。）