

迴歸共線性問題

余民寧 教授
國立政治大學教育學系

一、基本迴歸模式

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi}$$

$i = 1 \sim N$ (受試者), p 個預測變項, 1 個效標變項

二、共線性問題的定義

對任何 i 或 j 預測變項而言,

$$\text{當 } r_{x_i x_j} = 0 \text{ 時, } R^2_{Y.X_1 X_2 \cdots X_p} = R^2_{Y.X_1} + R^2_{Y.X_2} + \cdots + R^2_{Y.X_p}$$

當 $r_{x_i x_j} > .90$ 時, 即為共線性變項

三、共線性問題的影響

1. 對迴歸係數的影響:
2. 對偏決定係數的影響:
3. 對迴歸係數的估計標準誤 (S_{b_j}) 的影響:
4. 對適配值 (fitted values) 和預測值 (predictions) 的影響:
5. 對迴歸係數的同時區間考驗 (simultaneous tests) 的影響:

四、診斷方法

Variance inflation factor (變異數膨脹因子, 簡稱 VIF)

1. 若 $VIF_j = \frac{1}{1 - R_j^2} > 10$, 即表示第 j 個預測變項即為共線性變項, $j = 1 \sim p$

R_j^2 是以第 j 個 X 變項當效標 Y , 其餘 X 變項對它進行預測所得之 R^2 值

其中, $(1 - R_j^2)$ 亦稱作「容忍度」(tolerance), 一般設定為 .01、.001、或 .0001

2. 若 $\overline{VIF} = \frac{\sum_{j=1}^p VIF_j}{p} > 10$, 即表示具有很嚴重的共線性問題

五、補救措施

1. 刪除共線性變項。

2. 使用標準分數化的迴歸分析法。

3. 使用「偏差迴歸分析法」(Biased Regression Analysis, 簡稱 **BRA**)，如：使用 Ridge regression analysis (山脊型迴歸分析法)。

在 LS (least square) 估計法下， $\beta_{LS} = (X'X)^{-1} X'y$ ，具有 **BLUE** (best linear unbiased estimator) 特性，即不偏性和有效性。

若犧牲一點不偏性，但堅持仍具有有效性的話，則可用 ridge estimator 取代如下：

$$\beta_r = (X'X + cI)^{-1} X'y \quad , \quad \text{其中，} c \text{ (稱 ridge trace) 為一個微小的純數}$$

凡是能讓 β_r 估計值趨於平穩的 c 值即為我們要尋找的「跡」；再將其數值代入上述公式，即可求得「山脊型迴歸係數」(ridge estimator)。不過，該「山脊型迴歸係數」是有偏差的估計值 (biased estimators)，但具有較 β_{LS} 估計值還小的估計標準誤，因此，對改善預測精確度問題具有相當大的幫助。

註：目前，SAS/SUGI Supplemental Library User's Guide 中有 proc ridge; 模組程式可資應用。

SPSS for Windows 10.0 以後的版本中，有 Ridge regression 語法程式 (Syntax) 可供使用。

Advanced readings (建議讀物)：

Belsley, D.A. (1991). *Conditioning diagnostics: Collinearity and weak data in regression*. New York: John Wiley & Sons.

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed.). Belmont, CA: Duxbury Press.

Neter, J., Wassermann, W., & Kutner, M. H. (2004). *Applied linear regression models* (4th ed.). Homewood, IL: Irwin.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Yu, M. N. (1989) (余民寧). Detection and remedies of multicollinearity in multiple regression analysis. *Journal of Education and Psychology* (政大教育與心理研究), 12, 385-407.