

多點膨脹截斷卜瓦松迴歸模型

徐小鈞*、林定香、蔡旻曉
國立臺北大學統計學系

摘要

計數資料(count data)廣泛存在於各個領域中，實際搜集到的資料經常包含過多的「零值」與某些「非零正整數」，例如一個月內有幾天覺得身體不適，受試者的回答集中於0天、30天、5與7倍數的天數。通常稱此過多的「零值」與某些「非零正整數」發生的現象為多點膨脹值(multiple-inflation)。例如，當問卷調查是關於兩星期內有幾天事件發生的題目，延伸零膨脹卜瓦松分配的概念，將事件的發生的天數分成膨脹值與非膨脹值：前者包含「0」、「14」、「5」、「7」、「10」值，並假設它們發生的機率分別為 ϕ 、 ψ 、 ζ 、 ξ 與 θ 的多項分配；後者為假設事件發生數服從截斷卜瓦松分配(truncated Poisson)，在此假設下可推導出多點膨脹截斷卜瓦松模型。再分別利用邏輯斯模型(logistic model)與數連結的(log link)卜瓦松模型，即可建立多點膨脹截斷卜瓦松迴歸模型。最後以一個實例比較多點膨脹截斷卜瓦松迴歸模型與其他模型配適實際資料的能力。

關鍵詞：膨脹資料、零膨脹、截斷卜瓦松分配