

大量資料遺失情形下缺失資料處理方法效用之研究-以「青少年家庭中的親職壓力與親子衝突」研究資料為例

江文楷*、王鴻龍
國立臺北大學統計學研究所

摘要

社會科學研究常需整合兩種以上問卷資料，本研究旨在探討整合資料有大量資料遺失情形下，使用不同缺失資料處理方法對資料填補的成效。我們以 1998 年到 1999 年「青少年家庭中的親職壓力與親子衝突」研究計畫中的學生、父親、母親整合問卷為實證資料，探討缺失處理方法對青少年家庭親職壓力與親子衝突分析結果的影響。

我們透過判別資料缺失與否的虛擬變數，利用其與問卷變項進行 T 檢定與卡方檢定，分別找出顯著影響家長未填答的連續變數和類別變數，並以虛擬變數為依變數配適邏輯斯迴歸篩選出影響缺失的主要顯著變數。接著我們以完整學生、家長樣本資料作為基準(baseline)，依據原有的缺失機制，建構家長問卷的缺失資料集。為解決資料集組合數過多的問題，採用群集分析中的二階段群集法，運用缺失機制模型中的顯著變數按其特性分成 3 個群集，使同一群集內具有高度相似性。再根據分群後原有的缺失結構建構出缺失資料集，依不同群集缺失比例隨機抽樣建構各 30 組缺失資料集來進行家長缺失資料之處理，比較使用三種缺失資料處理方法(個體刪除法、迴歸逐步填補法、蒙第卡羅—馬可夫鏈單一填補法)分別探討使用變異數分析時顯著因子個數，以及在迴歸分析中的係數影響，並與資料集的結果進行驗證比較，最後將最佳缺失處理方法引進，對原始資料進行填補與分析。本研究結果可以做為有多類型大量缺失資料型態時，如何進行缺失處理的參考。

關鍵詞：不完整資料、缺失處理、插補法、親職壓力