

# 葛特曼量表之拒答插補研究

## 摘要

在抽樣調查的資料中，可能因為題意不清、關係到個人隱私，或是議題太過於敏感而導致受訪者「拒答」。透過存在拒答的樣本資料來做分析探討時，很可能會造成偏誤的研究結果，因此如何處理無反應的資料常常是一項研究結果是否可信的重要關鍵之一。常見的處理方式通常是設法對這些拒答資料進行插補。然而插補的好壞一直沒有一個判定準則，分析結果亦常因此受到質疑。

本研究將針對葛特曼量表的資料型態，利用「正確率」的概念，用不同的插補方式，包括社會科學研究常使用的簡易插補法，以及多重插補法與最鄰近插補法等方法，透過計算正確率來比較插補的好壞以及推論適用的時機。本研究以「台灣社會變遷基本調查」第四期第三次的調查資料中，有關性態度的題目做為例子，將其中符合葛特曼量表的資料視為「黃金標準」，並按照其中拒答部分的形態，從黃金標準中製造拒答資料。隨著拒答率的上升，每種拒答形態對應的個數將等量放大。

研究結果發現，簡易插補法的正確率可以透過公式推導求出。在這筆資料之下，不論何種簡易插補方法，其正確率都不超過 32%，但隨著拒答型態與社會開放程度的不同，拒答率會有很大的變化。多重插補法之下的結果比簡易插補法略好一些，有接近 33% 的正確率，但從便利性來看使用簡易插補法就比多重插補法來的高。最鄰近插補法的正確率是相對比較高的，最高可以達到約 47%，然而執行上比較花費時間，以及正確率有隨著拒答率的上升而下降的趨勢都是最鄰近插補法可能的問題。

關鍵詞：拒答、葛特曼量表、簡易插補、多重插補、最鄰近插補

# Imputation for Refusals under Guttman Scale

## Abstract

In a questionnaire survey, respondents may refuse to answer certain items since the questions themselves are unclear, sensitive, or relating to personal privacy. An analysis result using a data set containing refusal responses might be biased, how to deal with survey refusals have thus drawn much attention of late. One popular approach is through the use of imputation. However, lacking a criterion to evaluate its performances, there exist debates concerning the usefulness of this approach.

In this study, we compare Simple imputation Method, Multiple Imputation Method, and Nearest Neighbor Method to deal with refusals in a set of survey items forming a Guttman scale in terms of imputation accuracy. Data are taken from the 2002 Taiwan Social Change Survey (TSCS), and the items of interest are about sexual attitude. The parts of data that satisfy perfect Guttman scale are treat as “Gold Standard”, and refusals are generated according to the original refusal pattern appear in the data.

The result shows that the accuracy associated with Simple Imputation can actually be derived theoretically. No matter which version of Simple Imputation is applied, the accuracy is no more than 32%. Multiple Imputations performs slightly better than Simple Imputation, the accuracy is about 33%. However, it is less efficient in terms of computer time. Although Nearest Neighbor Method has the best performance the three, and its accuracy can reach as 47%, it requires much more computer time than the other two methods, and the accuracy would decrease as the refusal rate goes up.

*Key words:* Refusal, Guttman Scale, Simple Imputation, Multiple Imputation, Nearest Neighbor Imputation