第壹章 研究背景與動機

在抽樣調查的資料中，拒答(refusal)是一種常見的無反應(nonresponse)，可能的原因很多，包括問題題意不清、關係到個人隱私，或是太過於敏感性的議題。透過存在無反應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。先前的研究中，比較多的文獻在討論不知道(don't know)以及無意見(no opinion)這些類型的無反應，面對拒答的資料，通常是當成遺漏值(missing value)而將其刪除。然而拒答在某種程度上也是一種意見的表達，刪除這些資料對於後續的分析或多或少會有些影響。很多學者也意識到這樣的問題，進而設法對這些拒答資料進行插補(陳信木、林佳瑩，1997)。常見的作法是，將拒答與某些回答結果合併，例如問題屬於是非題類型，即回答只有兩種選項，則可能將拒答歸類到正向回答(positive response)、負向回答(negative response)、多數回答(majority response)，及少數回答(minority response)。但這樣的作法似乎太過於主觀，歸類結果的差異也缺乏一個標準來進行比較，因而難以判斷好壞。

有鑑於此，本文希望提供一些可行的方法來插補拒答的資料，利用「正確率」，即插補成正確答案的比例，來比較多種插補方法的優缺點以及適用的時機，提供研究者在面對存在拒答資料時，能夠選用適合的插補方法。

本文的研究動機主要來自於中央研究院社會學研究所與調查研究專題中心於 2002 年合作執行之臺灣地區社會變遷基本調查第四期第三次的問卷資料。該份問卷以性別議題為主要範疇，對隨機抽取之臺灣地區十八歲以上民眾進行面對面訪問，其中與性態度及行為相關的敏感問題則由受訪民眾自行填寫。相關問題列出如下：

(Q1) 對於未婚男女可不可以和偶然認識的人牽手 (1)絕對不可以(2)不可以(3)看情況(4)可以
(Q2) 對於未婚男女可不可以和偶然認識的人接吻 (1)絕對不可以(2)不可以(3)看情況(4)可以
(Q3) 對於未婚男女可不可以和偶然認識的人愛撫 (1)絕對不可以(2)不可以(3)看情況(4)可以
(Q4) 對於未婚男女可不可以和偶然認識的人發生性關係 (1)絕對不可以(2)不可以(3)看情況(4)可以

從問卷內容可以看出以上四題在一般認知上有明顯的層次強弱關係。基本上如果受訪者在前面的題項勾選了不贊同的答案,後面的題項理應也會選擇不同意的選項。針對這種特性的量表，本文將藉由「葛特曼量表模型(Guttman Scaling Model)」來進行資料分析，針對不同「拒答」比例的資料進行多種不同方法的插補，比較討論其優缺點。

我們將在第貳章探討過去學者的研究，以及數種與插補相關的理論；第參章介紹研究資料的特性及合適的量表模型；第肆章將進行不同方法的插補，並計算其正確率；第伍章為結論與建議事項。

第貳章 文獻探討

第一節 葛特曼量表模型(Guttman Scaling Model)

「葛特曼量表」是一種在社會學或是心理學的研究中，常常使用的一種量表模型。Guttman(1950)提出了此種量表，並且介紹如何對於此種量表下的資料作進一步的分析。葛特曼量表的分析方法，是將受試者對一組題目的回答，按照困難度排列，根據其回答的狀況來找出受試者能力定位的方法。受試者的人數一般而言為題目個數的五倍以上。題目的困難度則被假定為單一向度的排列，也就是說一旦受試者的能力被定位成會答錯某個題目，此受試者對於難度更高的題目應該也會答錯；若是受試者的回應沒有符合此性質，則稱之為「葛特曼錯誤(Guttman error)」，並被視為錯誤回答，出現錯誤回答過多的時候，表示此量表很可能不符合葛特曼量表的要求。

檢定一量表是否符合葛特曼量表的特性，一般常用的指標包括：「再生係數(Coefficient of Reproducibility, CR) 」與「量表係數 (Coefficient of Scalability, CS) 」。「再生係數」的概念是用來衡量研究者能將受試者定位正確的程度(Guttman，1950)；也就是說再生係數可以檢測問題是否確實具備難度有單一向度的特性。此值應該至少要達到0.9，算法為{1－(錯誤個數/總回答個數)}。Menzel(1953)對此指標提出了質疑，認為若是受試者不論任何題目，皆回答負向答案，此時雖然符合了單一向度的特性，但這卻是虛假的，並不能真正表示出受試者的特質，於是提出了「量表係數」的概念。欲計算量表係數必須先算出「最小再生係數(Minimal Marginal Reproducibility, MMR) 」，首先計算各個問題正向答案與負向答案的個數之中較小者，並將其加總，然後計算{1－(個數較小者加總/總回答個數)}，這就是所謂的最小再生係數，此值表示受試者特質強度的最低基準。量表係數則定義為{CS＝(CR－MMR)/(1－MMR)}。實例與計算過程詳見本文第參章。若再生係數大於0.9，且量表係數大於0.6時，則這些問題組合可以視為構成葛特曼量表。

第二節 過去相關研究

實證研究常將問題中的無反應，包括不知道、無意見與拒答，以遺漏值、插補法或與其他回答類別合併的方式來處理，例如將遺漏值之外的無反應與正向回答、負向回答或佔多數回答合併來進行分析(陳信木、林佳瑩，1997；Yamaguchi，2000)。以遺漏值的方式來處理無反應可能會造成樣本的偏誤，而降低分析結果的可信度，由於合併的作法通常是由研究者主觀認定其意涵，因此分析合併後樣本的可信度也是討論的議題。

當資料符合葛特曼量表的形式時，若考慮「拒答」為實質答案， 分別以葛特曼量表模型、潛藏距離模型與 Goodman 量表模型來檢測不同的「拒答」的處理策略(Liao and Tu，2006)，分析結果指出無論以何種策略來處理比例約佔百分之五至六的拒答，量表指標的計算結果都會顯示不違反葛特曼量表的假設。但是可以想見的是，百分之五至六的拒答並不算很多，對於資料的影響可能有限，插補後的完整資料是否符合葛特曼量表的判定似乎不能真正判斷插補的好壞。因此本研究將從此著手，改為考慮插補拒答的正確率來作為插補結果好壞的指標，以免除拒答比例較低時，前述測量指標不太能夠反應出資料變動的疑慮。

第三節 遺漏值的定義與機制

遺漏值有時稱為「無反應(nonresponse)」，一般定義為調查資料中出現遺漏或是不完整的資料，也就是收集到樣本資料中有一部份或是全部資料有缺失，通常包含拒答(refusal)、不知道(don’t know)以及無意見(no opinion)，可能因為研究設計不當或是資料蒐集、建檔過程的疏忽造成。

Rubin(1976)探討有關資料的遺漏機制，遺漏值可以分為完全隨機遺漏(Missing Completely at Random, MCAR)、隨機遺漏(Missing at Random, MAR)及非隨機遺漏(Missing Not at Random, MNAR)。

令欲觀察的變數為 $Y = (y_{ij})$，其中 $i = 1, 2,\cdots, n$，$j = 1, 2,\cdots, M$，而 y_{ij} 表示第 j

個變數的第 i 個觀察值；又令 $Z=(z_{ij})$為一指標函數，若 y_{ij} 為遺漏值，則 $z_{ij}=1$，否則 $z_{ij}=0$。令觀察變數 Y 和指標函數 Z 的條件分配為 $f(Z|Y)$ 。定義 Y_{obs} 表示變數有觀察到的部份，Y_{mis} 表示變數遺漏的部分，則三種遺漏值的定義分述如下

(1)完全隨機遺漏：$f(Z|Y)=f(Z)$，表示遺漏值的產生和整個資料無關。

(2)隨機遺漏：$f(Z|Y)=f(Y_{obs})$，遺漏值的產生與資料觀察到的部份有關，而與其遺漏部分無關。

(3)非隨機遺漏：$f(Z|Y)=f(Y_{mis})$，遺漏值的產生與其他遺漏值有關，而與觀察到的部份無關。

一般在社會學上的討論中，拒答較容易因為題目的敏感度高，受訪者回答的意願降低而產生(Shoemaker et al.，2002)，因此被歸屬於「隨機遺漏」，也就是說拒答的部份與資料其他有回答的部份有關係，因此可以考慮以不含拒答部份的訊息來插補拒答部份。

第四節 多重插補法(Multiple Imputation Method)

學者們在對有遺漏的資料進行分析之前，考量到遺漏可能會造成資料的偏誤，通常會對其進行「插補(Imputation) 」。Rubin(1976)提出插補是為了資料的完整性，針對遺漏的部份填補一些合理的數值，以求儘可能減少偏誤。將每一筆遺漏的資料填補一個合理的數值稱為「單一插補(Single Imputation)」；而將每一筆遺漏的資料填補多個合理的數值稱為「多重插補(Multiple Imputation)」(Rubin，1987)。多重插補法利用蒙地卡羅法(Monte Carlo Method)的概念，將遺漏的資料填補m個數值(見圖2.1)。Rubin(1987)指出，若無限多次的插補可以使效率最大，則有限的m次插補後的估計相對於無限多次插補後的估計的相對效率為$(1+\lambda/m)^{-1}$，其中λ為遺漏資料比例(missing data proportion)，例如當遺失部分為50%時，若插補2次($m=2$)，此時相對效率為80%；但若插補3次($m=3$)，此時相對效率提高為85.7%，增加了5.7個百分點；又若插補10次($m=10$)，則此時相對效率

已可達到95%；若插補11次($m = 11$)，相對效率則約為95.7%，僅增加0.7個百分點，可知再增加插補次數，相對效率不會有太多提升，因此m通常必須大於3，但是不需要超過10。

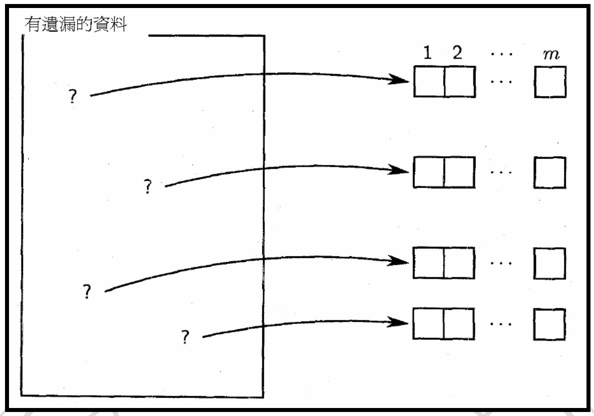

圖2.1 多重插補法圖示，m表示插補次數

假設一組欲觀察的資料$Y = (Y_{obs}, Y_{mis})$服從一個參數為θ的分配$P(Y|\theta)$，其中Y_{obs}為資料不含遺漏的部分，Y_{mis}為資料遺漏的部分，且其遺漏是屬於隨機遺漏；此外參數θ有其先驗分配(prior distribution)。因為

$$P(Y_{mis} \mid Y_{obs}) = \int P(Y_{mis} \mid Y_{obs}, \theta) P(\theta \mid Y_{obs}) d\theta$$

首先，利用未知參數θ的後驗分配隨機模擬出θ^*，即$\theta^* \sim P(\theta \mid Y_{obs})$，接下來插補的值${Y_{mis}}^*$則需滿足${Y_{mis}}^* \sim P(Y_{mis} \mid Y_{obs}, \theta^*)$，如此重覆$m$次即可完成插補。此方法理論上雖然可行，但事實上要找出未知參數θ的後驗分配在很多時候是難以進行的，或者是說找出的分配不易進行模擬，Rubin(1987)因此提出了一些可以執行的策略。但是不久後，隨著較簡易方法－MCMC(Markov Chain Monte Carlo)的理論發展，取代了Rubin的策略，也成為最常用的方法。

遺漏值問題常使用的MCMC方法為Tanner及Wong(1987)提出的資料擴增法(data augmentation)。考慮上面提及的兩階段插補方法，先從遺漏部份模擬出資料${Y_{mis}}^{(t)}$，其中${Y_{mis}}^{(t)} \sim P(Y_{mis} \mid Y_{obs}, \theta^{(t-1)})$，此時便擁有一組模擬出的資料，且沒有遺漏值，而未知參數θ的後驗分配成了$\theta^{(t)} \sim P(\theta \mid Y_{obs}, {Y_{mis}}^{(t)})$，如此即可進行模擬。

一旦給定起始值 $\theta^{(0)}$ 後，理論上可以証實 $\{(Y_{mis}{}^{(t)}, \theta^{(t)}), t = 1,2,\cdots\}$ 構成了馬可夫鏈(Markov Chain)，並滿足會收斂到穩定分配(Stationary distribution) $P(Y_{mis}, \theta \mid Y_{obs})$ 。m次執行此動作之後，也可以完成插補。

Buuren 及 Oudshoorn(2000)依據遺漏資料的類型不同，列出一些在使用統計軟體 R 時的建議分配，即 $P(Y|\theta)$，列於表 2.1。

遺漏資料的類型	R 之程式碼	方法描述
Numeric	impute.norm	Bayesian linear regression
	impute.pmm	Predictive mean matching
	impute.mean	Unconditional mean imputation
2 categories	impute.logreg	Logistic regression
≧2 categories	impute.polyreg	Polytomous logistic regression
	impute.lda	Linear discriminant analysis
Any	impute.sample	Random sample from the available observed values

表 2.1 執行多重插補法的建議分配

若以本研究欲插補的變數來看，是屬於兩種類別的變數(同意、不同意)，因此在 R 中的設定會利用邏輯斯迴歸的模型來進行插補。

第五節 最鄰近插補法(Nearest Neighbor method)

Fix及Hodges(1951)提出了「最鄰近法則(nearest neighbor rule)」。給定一組包含n個樣本(x_1,θ_1), (x_2,θ_2), …, (x_n,θ_n)的集合，其中x_i為第i筆資料的觀察值，θ_i為第i筆資料的所屬類別指標。若有另一筆資料(x,θ)，其中x是可以觀察到的，我們試圖利用前n筆資料來將x作正確的分類。

令

$$x^* \in \{x_1, x_2, \cdots, x_n\}$$

且滿足

$$\min d(x_i, x) = d(x^*, x) \text{ ，} i=1, 2, \cdots, n$$

其中，$d(x_i, x)$ 表示觀察值 x_i 與觀察值 x 的距離，則 x^* 稱為最鄰近 x 的資料點。也

就是說最鄰近法則將x分類到最鄰近的x^*所屬的類別θ^*。若事實上$\theta^* \neq \theta$，則此時就產生了分類錯誤。此外最鄰近法只是將資料分類到與其最鄰近的一類，剩下的$n-1$類資料並沒有列入考慮，故又稱為「1-最鄰近決策法則(1-nearest neighbor rule)」。

由於在資料歸類時，涉及到兩筆資料距離(distance)的遠近，或是相異度(dissimilarity)的大小。常用的距離定義為歐氏距離(Euclidean distance)，假設在歐氏空間中有兩點$x=(x_1,x_2,\cdots,x_n)$與$y=(y_1,y_2,\cdots,y_n)$，則此二點距離定義為

$$d(x,y)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}$$

然而歐氏距離基本上適用於連續型的資料，用來定義類別型資料的相異度並不恰當。Kaufman 與 Rousseeuw(1990)提出一種適用於混合各種不同測量尺度的資料下，相異度的定義。假定每一個觀測值由多個屬於不同類型的變數所構成時，兩個觀測值之間的相異度定義為

$$d(i,j)=\frac{\sum_{f=1}^{p}\delta_{ij}^{(f)}d_{ij}^{(f)}}{\sum_{f=1}^{p}\delta_{ij}^{(f)}}$$

其中$d(i,j)$表示第i筆與第j筆觀測值之間的相異度，其範圍介於 0 到 1 之間。$d_{ij}^{(f)}$的意義為在第f個變數之中，第i筆觀測值與第j筆觀測值間的相異程度，$f=1,2,\cdots,p$。依據變數類型的不同，$d_{ij}^{(f)}$的計算方式有很大差異，設定如下：

(1) 若變數f屬於二元(binary)或名目(nominal)資料，則當$x_{if}=x_{jf}$時，$d_{ij}^{(f)}=$ 0，否則$d_{ij}^{(f)}=1$。

(2) 若變數f屬於區間(interval)資料，則$d_{ij}^{(f)}=\frac{|x_{if}-x_{jf}|}{\max_h x_{hf}-\min_h x_{hf}}$，其中$x_{hf}$為變數$f$的一個非遺漏的觀察值。

(3) 若變數f屬於次序(ordinal) 或比率(ratio)資料，則先計算每筆資料的排名

r_{if}，則 $d_{ij}^{(f)} = \dfrac{|r_{if} - r_{jf}|}{M_f - 1}$，其中，$M_f$ 為變數 f 中的最大排名。

又第 f 個變數的指標函數 $\delta_{ij}^{(f)}$ 在下列兩種情形為 0，一是當第 f 個變數的第 i 筆觀測值 x_{if} 或是第 j 筆觀測值 x_{jf} 為遺漏，二是當第 f 個變數為非對稱二元變數 (asymmetric binary)，且第 f 個變數的第 i 筆觀測值與第 j 筆觀測值都被標記為 0 的時候；其餘情況下，$\delta_{ij}^{(f)}$=1。因此可以看出 $d(i,j) = \dfrac{\sum_{f=1}^{p} \delta_{ij}^{(f)} d_{ij}^{(f)}}{\sum_{f=1}^{p} \delta_{ij}^{(f)}}$ 是利用類似權數的概念，來計算整體第 i 筆與第 j 筆觀測值之間的相異度 $d(i,j)$。

如此定義之下，即可以計算出資料由不同類型變數組成時，兩筆資料的相異度。以下將舉一例作說明。

假設五筆資料：

編號	性別	年齡	婚姻	教育年數	收入
1	1	25	1	12	0
2	2	33	3	12	34999
3	2	28	1	16	44999
4	1	33	3	16	44999
5	1	22	1	14	14999

其中性別屬於對稱二元變數，婚姻為四類別的名目變數，其餘三者為區間變數。根據定義可知 $\delta_{ij}^{(f)}$ 在此皆為 1，接著再將 $d_{ij}^{(f)}$ 分別求出，即可算出兩兩距離 $d(i,j)$，結果列於下表 2.2。

編號	1	2	3	4	5
1	0	.	.	.	.
2	0.7010	0	.	.	.
3	0.6545	0.5354	0	.	.
4	0.7455	0.4444	0.4909	0	.
5	***0.2212***	***0.7889***	0.5424	0.6333	0

表 2.2 五筆資料距離計算結果

根據上表可以看出編號 2 與 5 的差異最大，1 與 5 的差異最小。因此，以我們的資料為例，若人口變項已知，但性態度的四個問題卻出現拒答，則此時就可以找

出人口變項與其他完整資料中，距離最短的一筆，以該筆資料的回答來插補之。

第參章 資料分析

第一節 變數介紹

本研究將問題一：對於未婚男女可不可以和偶然認識的人牽手，以 H 標記；問題二：對於未婚男女可不可以和偶然認識的人接吻，以 K 標記；問題三：對於未婚男女可不可以和偶然認識的人愛撫，以 C 標記；問題四：對於未婚男女可不可以和偶然認識的人發生性關係；以 S 標記。另外，人口變項包括：性別(gender)、年齡(age)、婚姻狀況(marriage)、教育程度(education)，及平均每月收入(income)。

第二節 反應變數說明

根據問卷的四個問題，可以看出有強烈的順序關係，通常一般人不致於會有容許與偶然認識的人發生性行為，但是卻不容許牽手的態度反應，因此形成了一個葛特曼量表的型態。為了簡化分析起見，本研究將問題選項中的絕對不可以與不可以合併為「不容許」，並標記為 2；合併問題選項中的看情況與可以為「容許」，並標記為 1；若是對該問題選擇「拒答」，則標記為 8。依據 1983 份回收問卷，這四道問題的回答類型可以歸納如下：

(1)五種符合葛特曼量表的類型(共 1785 筆)

(H,K,C,S)	(1,1,1,1)	(1,1,1,2)	(1,1,2,2)	(1,2,2,2)	(2,2,2,2)
個數	399	140	296	484	466

(2)九種不符合葛特曼量表的類型(共 36 筆)

(H,K,C,S)	(1,1,2,1)	(1,2,1,1)	(1,2,1,2)	(1,2,2,1)	(2,1,1,1)
個數	5	5	7	7	2

(H,K,C,S)	(2,1,1,2)	(2,1,2,2)	(2,2,1,2)	(2,2,2,1)
個數	4	4	1	1

以上為不含拒答類型，總共有 1821 筆，其所佔比例列於下表 3.1。可以看出四題回答依序為(1,1,1,2)者最少，大約佔 7.7%，回答為(1,2,2,2)與(2,2,2,2)者最多，都佔 25%以上，另外錯誤回答的比例相當少，只有不到 2%。

量表類型之比例	不含「拒答」
1 (1, 1, 1, 1)	.2191
2 (1, 1, 1, 2)	.0769
3 (1, 1, 2, 2)	.1625
4 (1, 2, 2, 2)	.2658
5 (2, 2, 2, 2)	.2559
錯誤回答(不符合葛特曼量表)	.0198

表 3.1 不含拒答類型資料比例(共 1821 筆)

(3)十二種包含拒答的類型(共 121 筆)

(H,K,C,S)	(1,1,8,8)	(8,8,8,8)	(1,8,8,8)	(8,8,1,8)	(2,8,8,8)	(1,1,8,2)
個數	1	90	15	1	3	3

(H,K,C,S)	(1,8,2,2)	(1,8,8,2)	(8,8,8,2)	(8,2,2,2)	(1,2,8,8)	(2,2,8,8)
個數	1	1	2	1	1	2

包含拒答的類型之中，以四題回答為(8,8,8,8)者最多，比例約為 74%，回答為(1,8,8,8)者次之，但也只佔不到 13%，其他拒答類型所佔比例都非常的少，不超過 2%。此外，若整份問卷都沒有回答，則視為遺漏值，標記為 0；若回答「不知道」則標記為 7。四道問題回答為(0,0,0,0)共 38 筆，(7,7,7,7)及(1,1,7,7))共 3 筆，這 41 筆觀測值因為與本研究的主題不符合，故將其刪除。因此本文中所用來進行分析的資料總計有 1942 筆。以下將運用葛特曼量表的一些指標來判斷這組資料是否具有葛特曼量表的特性。

將不含拒答的部分(共 1821 筆)套入再生係數(CR)、最小再生係數(MMR)、量表係數(CS)的公式，計算如下：

(a) CR

想要計算 CR 首先要計算錯誤個數，錯誤回答的定義為：若不同意前面問題，卻同意後面問題，則算後面問題是錯誤回答。例如回答為(1,1,2,1)，根據葛特曼量表的特性，第四題應該為 2，因此錯誤個數計為 1，若五個人回答(1,1,2,1)則有五個錯誤，又例如兩個人回答為(2,1,1,1)，則錯誤個數計為 6；另外總回答個數為 7284(4*1821)。全部結果列於下表 3.2。

(H,K,C,S)	(1,1,2,1)	(1,2,1,1)	(1,2,1,2)	(1,2,2,1)	(2,1,1,1)	(2,1,1,2)	(2,1,2,2)	(2,2,1,2)	(2,2,2,1)
個數	5	5	7	7	2	4	4	1	1
錯誤個數	5	10	7	7	6	8	4	1	1

表 3.2 錯誤個數計算

此時可得 CR = $1-[(5+10+7+\cdots+1)/(4*1821)]=0.9933$

(b) MMR

根據 MMR 的定義，先找出各個問題同意與不同意的個數之中較小者。列於下表 3.3。

H		K		C		S	
同意	不同意	同意	不同意	同意	不同意	同意	不同意
1343	478	850	971	558	1263	419	1402

表 3.3 各問題同意與不同意個數計算

此時可得 MMR $=1-[(478+850+558+419)/(4*1821)]=0.6836$。

(c) CS

CS = (CR－MMR)/(1－MMR) = 0.9788。

全部結果整理於下表 3.4。

葛特曼量表指標	
再生係數（CR）	***0.9933***
最小再生係數（MMR）	0.6836
量表係數（CS）	***0.9788***

表 3.4 葛特曼量表指標計算結果

根據定義，再生係數若是大於 0.9，且量表係數大於 0.6 時，這樣的資料就符合了葛特曼量表的形態。上表顯示再生係數值為 0.9933，且量表係數值為 0.9788，表示此組資料並沒有違反葛特曼量表的假設。

第三節 人口變項分析

(1)性別

性別

	次數	百分比	有效百分比	累積百分比
男(1)	956	49.2	49.2	49.2
女(2)	986	50.8	50.8	100.0
總和	1942	100.0	100.0	

表3.5 性別的次數分配表

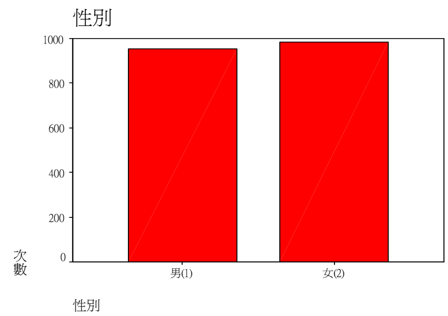

圖3.1 性別的次數分配圖

男性比例佔49%，女性則佔51%，兩者所佔比例相當接近。

(2)年齡

資料蒐集時，調查的是出生年次，為了資料分析起見，將年次轉換成在2002年的年齡。其中，平均年齡為43.7歲，標準差為16.57，最大值為90歲，最小值為19歲。資料結構可以看出多半集中在50歲以下，佔了66%。

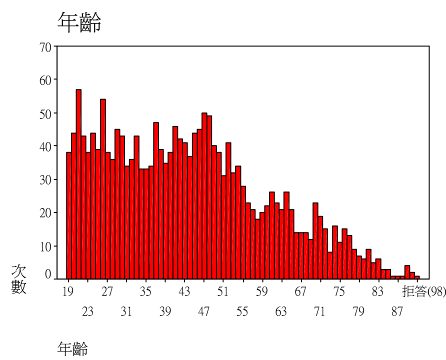

圖3.2 年齡的次數分配圖

(3)婚姻狀況

婚姻狀況

	次數	百分比	有效百分比	累積百分比
單身且從未結過婚(1)	497	25.6	25.6	25.6
同居(2)	7	.4	.4	26.0
已婚有偶(3)	1232	63.4	63.4	89.4
離婚(4)	38	2.0	2.0	91.3
分居(5)	29	1.5	1.5	92.8
喪偶(6)	136	7.0	7.0	99.8
其它(7)	3	.2	.2	100.0
總和	1942	100.0	100.0	

表3.6 婚姻狀況的次數分配表

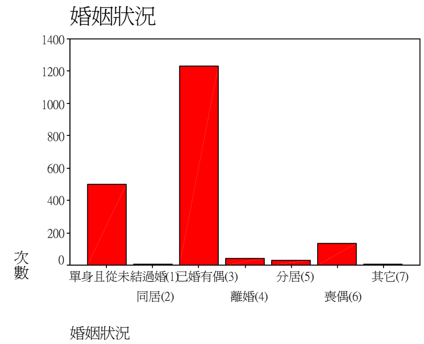

圖3.3 婚姻狀況的次數分配圖

婚姻狀況的分佈，以已婚有偶的受訪者佔最多數，有63.4%，其次為單身且從未結婚，佔25.6%，喪偶的比例也有佔7%。

(4)教育程度

教育程度

	次數	百分比	有效百分比	累積百分比
無(1)	146	7.5	7.5	7.5
自修(2)	8	.4	.4	7.9
小學(3)	397	20.4	20.4	28.4
國（初）中(4)	251	12.9	12.9	41.3
初職(5)	5	.3	.3	41.6
高中普通科(6)	114	5.9	5.9	47.4
高中職業科(7)	130	6.7	6.7	54.1
高職(8)	285	14.7	14.7	68.8
士官學校(9)	6	.3	.3	69.1
五專(10)	82	4.2	4.2	73.3
二專(11)	161	8.3	8.3	81.6
三專(12)	21	1.1	1.1	82.7
軍警專修班(13)	3	.2	.2	82.9
軍警專科班(14)	8	.4	.4	83.3
空中行專(15)	5	.3	.3	83.5
軍警官學校(16)	14	.7	.7	84.2
大學(17)	253	13.0	13.0	97.3
碩士(18)	44	2.3	2.3	99.5
博士(19)	7	.4	.4	99.9
其他(20)	2	.1	.1	100.0
總和	1942	100.0	100.0	

表3.7 教育程度的次數分配表

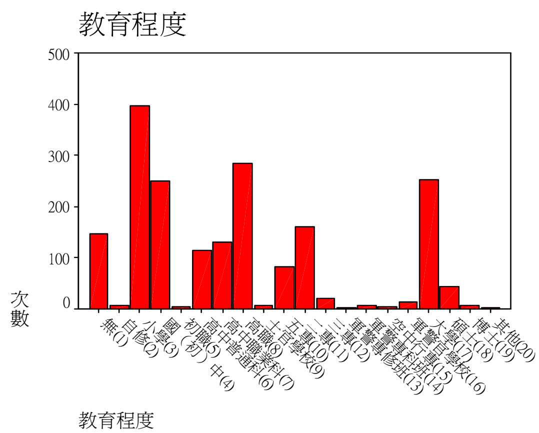

圖3.4 教育程度的次數分配圖

教育程度的選項分的非常詳細，總共有20種類別，從未受教育一直到博士程度。此筆資料中，數量最多的是小學，佔了20%，其次為高職，佔14.7%，國(初)中以及大學的比例差不多，大約為13%。

為了後面的分析起見，將教育程度轉換成教育年數，例如填答無(1)、自修(2)及其他(20)的受訪者表示教育年數為0，填答小學(3)的受訪者表示教育年數為6，以此類推。教育年數的分配列出如下：

教育年數

	次數	百分比	有效百分比	累積百分比
0	156	8.0	8.0	8.0
6	397	20.4	20.4	28.5
9	256	13.2	13.2	41.7
12	535	27.5	27.5	69.2
14	259	13.3	13.3	82.5
15	21	1.1	1.1	83.6
16	267	13.7	13.7	97.4
18	44	2.3	2.3	99.6
23	7	.4	.4	100.0
總和	1942	100.0	100.0	

表3.8 教育年數的次數分配表

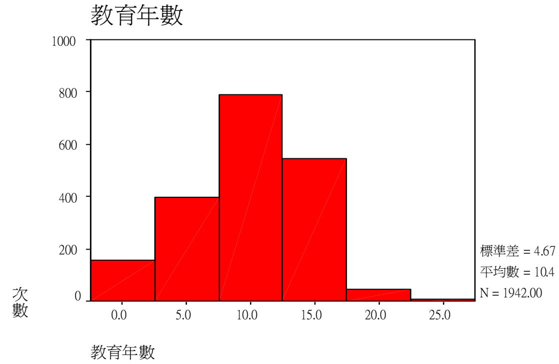

圖3.5 教育年數的次數分配圖

轉換成教育年數後，可以看出平均數為10.4(年)，標準差為4.67，眾數是12(年)。

(5)平均每月收入

每月平均收入

	次數	百分比	有效百分比	累積百分比
無收入(1)	612	31.5	31.5	31.5
1萬元以下(2)	130	6.7	6.7	38.2
1-2萬元以下(3)	234	12.0	12.0	50.3
2-3萬元以下(4)	300	15.4	15.4	65.7
3-4萬元以下(5)	252	13.0	13.0	78.7
4-5萬元以下(6)	171	8.8	8.8	87.5
5-6萬元以下(7)	103	5.3	5.3	92.8
6-7萬元以下(8)	41	2.1	2.1	94.9
7-8萬元以下(9)	24	1.2	1.2	96.1
8-9萬元以下(10)	18	.9	.9	97.1
9-10萬元以下(11)	18	.9	.9	98.0
10-11萬元以下(12)	14	.7	.7	98.7
11-12萬元以下(13)	1	.1	.1	98.8
12-13萬元以下(14)	1	.1	.1	98.8
13-14萬元以下(15)	1	.1	.1	98.9
15-16萬元以下(17)	1	.1	.1	98.9
16-17萬元以下(18)	1	.1	.1	99.0
17-18萬元以下(19)	1	.1	.1	99.0
19-20萬元以下(21)	2	.1	.1	99.1
20-30萬元以下(22)	1	.1	.1	99.2
30萬元以上(23)	1	.1	.1	99.2
不知道(97)	3	.2	.2	99.4
拒答(98)	12	.6	.6	100.0
總和	1942	100.0	100.0	

表3.9 每月平均收入的次數分配表

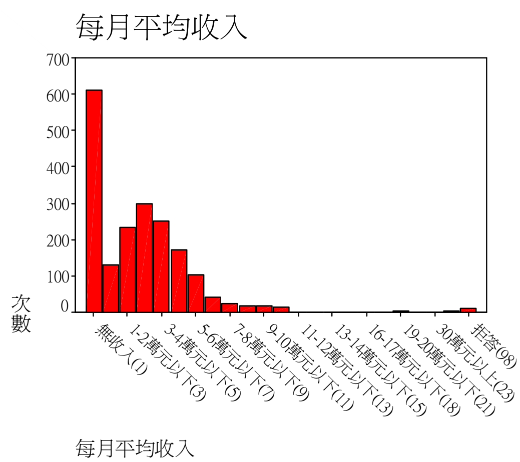

圖3.6 收入的次數分配圖

平均月收入的選項也劃分的非常詳細，總共有 23 種類型，從無收入一直到 30 萬元以上，這其中最多數的受訪者為無收入，佔了 31.5%，其次為 2 萬~3 萬元以下，佔 15.4%，1 萬~2 萬元以下，以及 3 萬~4 萬元以下則各佔了 12%以及 13%。由此可以看出平均月收入的分布大致集中在 0~4 萬元以下，比例為 78.7%。

第肆章 實證分析

這一章我們將就拒答的情況進行處理。值得注意的是，既然已經認定資料符合了葛特曼量表的假設，之後的插補勢必也要滿足這樣的要求，也就是說，當問題一已經回答不允許的時候，後面的問題一定要插補為不允許，否則就產生了葛特曼錯誤。然而，插補之後的好壞很難去實際比較，先前的研究有利用插補後的資料是否依然符合葛特曼量表來評斷插補的結果(Liao and Tu，2006)。本文採用的指標則是一正確率，也就是插補成原本答案的比例。

我們的作法是將符合葛特曼量表的 1785 筆資料視為「黃金標準」，從中去製造出一些拒答的反應，接著透過各種插補方式來比較看看能否插補成原本答案。此處特別要求四題答案必須完全正確才算差補正確，例如原本是(1,1,2,2)的反應類型一定要插補成(1,1,2,2)，若是其中一題答案插補錯誤，例如(1,2,2,2)，在這個定義之下並不算差補正確。拒答的製造是按照資料的原始組成的比例來進行，並且僅由可能產出該拒答形式的反應中來產生。表 4.1 列出拒答型態對應可能的回答形式，其中，第一行表示可能的拒答型態，第一列表示回答的形式，*部分表示左邊的拒答形式有可能由對應的上方回答製造出來，例如(1,8,8,8)可能從該列對應的*位置，依序為(1,1,1,1)，(1,1,1,2)，(1,1,2,2)及(1,2,2,2)等回答之中產生；又例如(1,1,8,2)可能從(1,1,1,2)以及(1,1,2,2)等回答來產生。由於我們也想要了解不同拒答比例對於插補正確率的影響，因此本研究會假設各種不同拒答比例，從5%、10%，一直到 30%，不同比例的拒答會導致各型態個數會有所變化。舉例來說，假設 10%的拒答率，表示要製造出 179 個(即 1785*0.1)拒答，故必須從(1,1,1,1)中隨機選出 40 筆(即 179*399/1785)，又(8,8,8,8)約佔 0.74(即 90/(1+90+⋯+2))，所以這 40 筆中有 74%，也就是大約 30 筆要製造成(8,8,8,8)的拒答型式，因此將隨機抽出 30 筆原本回答(1,1,1,1)的資料，將其製造成(8,8,8,8)。依此類推，可以計算得出其餘類型的拒答個數，結果列於表 4.2。如此從原本有答案的資料製造拒答，再進行插補，就可以知道插補的正確率。本章之後將使用

數種插補法進行分析。

回答型式 / 拒答型態	(1,1,1,1)	(1,1,1,2)	(1,1,2,2)	(1,2,2,2)	(2,2,2,2)
(1,1,8,8)	*	*	*	0	0
(8,8,8,8)	*	*	*	*	*
(1,8,8,8)	*	*	*	*	0
(8,8,1,8)	*	*	0	0	0
(2,8,8,8)	0	0	0	0	*
(1,1,8,2)	0	*	*	0	0
(1,8,2,2)	0	0	*	*	0
(1,8,8,2)	0	*	*	*	0
(8,8,8,2)	0	*	*	*	*
(8,2,2,2)	0	0	0	*	*
(1,2,8,8)	0	0	0	*	0
(2,2,8,8)	0	0	0	0	*

表 4.1 拒答型態對應可能的回答形式，*部分表示左邊的拒答形式有可能由對應的上方回答製造出來(原始資料)

回答型式 / 拒答型態	(1,1,1,1)	(1,1,1,2)	(1,1,2,2)	(1,2,2,2)	(2,2,2,2)
(1,1,8,8)	1	0	0	0	0
(8,8,8,8)	30	11	22	36	35
(1,8,8,8)	7	2	5	9	0
(8,8,1,8)	1	0	0	0	0
(2,8,8,8)	0	0	0	0	5
(1,1,8,2)	0	2	3	0	0
(1,8,2,2)	0	0	0	0	0
(1,8,8,2)	0	0	0	1	0
(8,8,8,2)	0	0	1	1	1
(8,2,2,2)	0	0	0	1	0
(1,2,8,8)	0	0	0	1	0
(2,2,8,8)	0	0	0	0	3

表 4.2 拒答率為 10%之下，各種回答須對應產生的個數

第一節 簡易插補法

(一) 正確率推導

參考先前文獻中的作法，將拒答以正向回答(容許)、負向回答(不容許)、多數回答、少數回答來作插補，例如將拒答以正向回答(容許)作插補，則將拒答(8)的部分全數以容許(1)來取代，並分別計算其正確率。由於簡易插補法並不在意其人口變項資訊，因此實務上很容易進行。此外由於此筆資料中拒答的製造是按照資料原本的拒答結構等量放大，故在拒答結構固定不變的情況下，簡易插補法的插補正確率在與拒答率無關。

此筆資料中，若採用多數回答的方式進行插補，由於四題的多數回答依序是1、2、2、2，故若拒答形式為(8,8,8,8)，插補後將得到(1,2,2,2)，又若拒答形式為(1,1,8,8)，插補後將得到(1,1,2,2)；若插補原則採用少數回答，依序是 2、1、1、1。由於資料的拒答形式以(8,8,8,8)為最多數，若採用少數回答的方式進行插補，會使得歸類到(2,1,1,1)的個數很多，由於這並不符合葛特曼量表的要求，可以想見如此歸類一定會使正確率很低，因此用少數回答來插補應該是最不適合的。另外若是強調插補必須符合葛特曼量表的要求，如果拒答出現有確定答案的形式，例如(1,2,8,8)、(2,2,8,8)等，一定只能分別插補成(1,2,2,2)及(2,2,2,2)；或是(8,8,1,8)這種拒答的形式出現時，應將其與(1,1,1,8)歸在同類。因此可以針對這些類型的拒答形式先插補，剩下不一定有確定答案的形式再利用容許、不容許、多數及少數回答等方法插補。不過本文中也將提供不考慮葛特曼量表之下的插補推導，並且比較兩者的優劣(見附錄一)。可以想見的是，不考慮葛特曼量表下的插補正確率勢必不可能高於考慮葛特曼量表的情況。詳細正確率計算推導如下：

令 $N_{(i,j,k,l)}$表示回答(i,j,k,l)的個數，其中$(i,j,k,l) = (1,1,1,1)$，$(1,1,1,2)$，$(1,1,2,2)$，$(1,2,2,2)$及$(2,2,2,2)$，$N = N_{(1,1,1,1)}+N_{(1,1,1,2)}+N_{(1,1,2,2)} +N_{(1,2,2,2)} +N_{(2,2,2,2)}$。$n_{(a,b,c,d)}$表示拒答形式為$(a,b,c,d)$的個數，$a,b,c,d$=1,2,8 且 a,b,c,d 至少有一個為 8。又令 $T_{(i,j,k,l)}$表示可以插補成(i,j,k,l)的拒答形式之總數，例如 $T_{(1,1,1,1)}= \sum_{a,b,c,d=1,8} n_{(a,b,c,d)}$。拒答率為 r。

在此筆資料中，有確定答案的拒答形式包括(1,2,8,8)、(2,2,8,8)及(2,8,8,8)，故此時必定正確的個數為$r\{N_{(1,2,2,2)}*\frac{n_{(1,2,8,8)}}{T_{(1,2,2,2)}}+N_{(2,2,2,2)}*\frac{n_{(2,2,8,8)}+n_{(2,8,8,8)}}{T_{(2,2,2,2)}}\}$。若代入資料數值可求得 28.1359$r$，也就是說至少會有 0.0158(即 28.1359r/1785r)的正確率。

除了這些拒答反應之外的類型，再以容許、不容許、多數回答及少數回答等方法插補。加上前述必定正確的個數之後，各方法的插補正確率敘述如下：

(1)將拒答以容許插補，即以 1 取代所有的 8，則插補的正確率為

$$r\{N_{(1,1,1,1)}+N_{(1,1,1,2)}*\frac{n_{(1,1,8,2)}+n_{(1,8,8,2)}+n_{(8,8,8,2)}}{T_{(1,1,1,2)}}+N_{(1,1,2,2)}*\frac{n_{(1,8,2,2)}}{T_{(1,1,2,2)}}$$
$$+N_{(1,2,2,2)}*\frac{n_{(8,2,2,2)}+n_{(1,2,8,8)}}{T_{(1,2,2,2)}}+N_{(2,2,2,2)}*\frac{n_{(2,2,8,8)}+n_{(2,8,8,8)}}{T_{(2,2,2,2)}}\}/rN^{1} \qquad \textbf{(1)}$$

其中，$N_{(1,1,1,1)}=N_{(1,1,1,1)}*(T_{(1,1,1,1)}/T_{(1,1,1,1)})$，因為分子分母相同而可約分。將拒答以容許插補，則會發現原本回答為(1,1,1,1)的部份全部插補正確，所以正確個數為拒答比例乘以回答為(1,1,1,1)的個數，即$r*N_{(1,1,1,1)}$；回答為(1,1,1,2)的部分之中，拒答形式為(1,1,8,2)、(1,8,8,2)，及(8,8,8,2)等三種將 8 以 1 取代後會插補正確，故正確個數為$r*N_{(1,1,1,2)}*\frac{n_{(1,1,8,2)}+n_{(1,8,8,2)}+n_{(8,8,8,2)}}{T_{(1,1,1,2)}}$；回答為(1,1,2,2)的部分之中，將拒答形式為(1,8,2,2)中的 8 以 1 取代後會插補正確，故正確個數為$r*N_{(1,1,2,2)}*\frac{n_{(1,8,2,2)}}{T_{(1,1,2,2)}}$；回答為(1,2,2,2)的部分之中，將拒答形式為(8,2,2,2)中的 8 以 1 取代後會插補正確，再加上葛特曼量表下必定正確的插補個數，故正確個數為$r*N_{(1,2,2,2)}*\frac{n_{(8,2,2,2)}+n_{(1,2,8,8)}}{T_{(1,2,2,2)}}$；最後回答為(2,2,2,2)的部分之中，只有葛特曼量表下必定正確的插補個數$r*N_{(2,2,2,2)}*\frac{n_{(2,2,8,8)}+n_{(2,8,8,8)}}{T_{(2,2,2,2)}}$。將上述總個數相加，再除以拒答比例乘以總個數 rN 即可求出以容許插補之正確率。[2]

(註 1：r 表示拒答率，分子分母同時存在理應約分，但為了表達計算之方式，先予以保留。)

(註 2：其他插補方式的計算過程相同，所以不再贅述。)

實際代入資料計算如下：

$$r\{399+140*\frac{3+1+2}{1+90+15+1+3+1+2}+296*\frac{1}{1+90+15+3+1+1+2}$$
$$+484*\frac{1+1}{90+15+1+1+2+1+1}+466*\frac{2+3}{90+3+2+1+2}\}/1785r$$
$$=0.2473$$

(2)將拒答以不容許插補，即以 2 取代所有的 8，則插補的正確率為

$$r\{N_{(1,1,2,2)}*\frac{n_{(1,1,8,8)}+n_{(1,1,8,2)}}{T_{(1,1,2,2)}}+N_{(1,2,2,2)}*\frac{n_{(1,8,8,8)}+n_{(1,8,2,2)}+n_{(1,8,8,2)}+n_{(1,2,8,8)}}{T_{(1,2,2,2)}}+N_{(2,2,2,2)}\}/rN \quad \textbf{(2)}$$

，其中，$N_{(2,2,2,2)}=N_{(2,2,2,2)}*(T_{(2,2,2,2)}/T_{(2,2,2,2)})$，因為分子分母相同而可約分。

(3)將拒答以多數回答插補，四題的多數回答依序為 1、2、2、2，分別插補拒答對應位置，插補的正確率為

$$r\{N_{(1,1,2,2)}*\frac{n_{(1,1,8,8)}+n_{(1,1,8,2)}}{T_{(1,1,2,2)}}+N_{(1,2,2,2)}+N_{(2,2,2,2)}*\frac{n_{(2,8,8,8)}+n_{(2,2,8,8)}}{T_{(2,2,2,2)}}\}/rN \quad \textbf{(3)}$$

，其中，$N_{(1,2,2,2)}=N_{(1,2,2,2)}*(T_{(1,2,2,2)}/T_{(1,2,2,2)})$，因為分子分母相同而可約分。

(4)將拒答以少數回答插補，四題的少數回答依序為 2、1、1、1，分別插補拒答對應位置，插補的正確率為

$$r\{N_{(1,1,1,1)}*\frac{n_{(1,1,8,8)}+n_{(1,8,8,8)}+n_{(8,8,1,8)}{}^{3}}{T_{(1,1,1,1)}}+N_{(1,1,1,2)}*\frac{n_{(1,1,8,2)}+n_{(1,8,8,2)}}{T_{(1,1,1,2)}}+N_{(1,1,2,2)}*\frac{n_{(1,8,2,2)}}{T_{(1,1,2,2)}}$$
$$+N_{(1,2,2,2)}*\frac{n_{(1,2,8,8)}}{T_{(1,2,2,2)}}+N_{(2,2,2,2)}*\frac{n_{(2,2,8,8)}+n_{(2,8,8,8)}}{T_{(2,2,2,2)}}\}/rN \quad \textbf{(4)}$$

(註 3：(8,8,1,8)在葛特曼量表之下已插補成(1,1,1,8)，故必定會從(1,1,1,1)製造出此拒答形式，以少數回答插補時，會插補成(1,1,1,1)，而使得插補正確。)

以上四個式子解釋了簡易插補方法的結果其實是不需要經過模擬，就可以用數學來推導出來，也就是將原始資料的值代入以上四個式子即可知道正確期望個數，或是正確率。實際帶入結果如下：

	容許	不容許	多數	少數
正確率	0.2473	***0.3107***	***0.2901***	0.0555

表 4.3 正確率計算(原始資料)

實際計算結果發現，以不容許以及多數回答插補會有比較高的正確率，最高可達到 31.07%，此外若用少數回答來插補，正確率會非常低，只有 5.56%，其原因在前面曾經討論過，因為黃金標準中並沒有(2,1,1,1)的答案且拒答形式最多為(8,8,8,8)，在此可驗證此說法。

透過前面四個式子，其實可以看出在我們所用製造拒答方法之下，正確率的高低與拒答率的大小沒有關係。此外如果拒答類型以(8,8,8,8)為主時，插補正確率的高低主要是根據社會保守或自由的程度，也就是跟 $N_{(1,1,1,1)}$，$N_{(1,1,1,2)}$，$N_{(1,1,2,2)}$，$N_{(1,2,2,2)}$，$N_{(2,2,2,2)}$的大小有關。本文欲探討在哪些現象之下，用哪種方法插補會有最好結果。但是在四個問題之下可能的情況有太多，而不易討論其性質，故以下先從問題只有兩題的情形來討論。

(二) 兩問題性質

假設只有兩個問題，則符合葛特曼量表之回答有(1,1)，(1,2)，(2,2)

則拒答的形態如下：

回答型式 / 拒答型態	(1,1)	(1,2)	(2,2)
(8,8)	*	*	*
(1,8)	*	*	0
(2,8)	0	0	*
(8,1)	*	0	0
(8,2)	0	*	*

表 4.4 拒答型態對應可能的回答形式，*部分表示左邊的

拒答形式有可能由對應的上方回答製造出來(當問題數為二時)

令 $N_{(1,1)}$、$N_{(1,2)}$，$N_{(2,2)}$分別表示回答(1,1)、(1,2)，(2,2)的個數，$N=N_{(1,1)}+N_{(1,2)}+N_{(2,2)}$；$n_{(a,b)}$表示拒答為$(a,b)$的個數，$a,b$ =1,2,8 且 a,b 至少一個為 8；$T_{(1,1)}$、$T_{(1,2)}$、$T_{(2,2)}$分別表示可以插補成(1,1)、(1,2)，(2,2)的拒答形式之總數。又令拒答率為 r。根據葛特曼量表的特性，有確定答案的拒答形式包括(2,8)及(8,1)，應分別插補為(2,2)及(1,1)，故此時必定正確的個數 $r(N_{(1,1)}*\frac{n_{(8,1)}}{T_{(1,1)}}+N_{(2,2)}*\frac{n_{(2,8)}}{T_{(2,2)}})$

除了這些拒答反應之外的類型再以容許、不容許、多數及少數回答等方法插補。加上必定正確的個數之後，各方法的插補正確率敘述如下：

(1)將拒答以容許插補，即以 1 取代所有的 8，則正確率為

$$r\{N_{(1,1)}+N_{(1,2)}*\frac{n_{(8,2)}}{T_{(1,2)}}+N_{(2,2)}*\frac{n_{(2,8)}}{T_{(2,2)}}\}/rN \tag{5}$$

，其中，$N_{(1,1)}=N_{(1,1)}*(T_{(1,1)}/T_{(1,1)})$

(2)將拒答以不容許插補，即以 2 取代所有的 8，則正確率為

$$r\{N_{(1,1)}*\frac{n_{(8,1)}}{T_{(1,1)}}+N_{(1,2)}*\frac{n_{(1,8)}}{T_{(1,2)}}+N_{(2,2)}\}/rN \tag{6}$$

，其中，$N_{(2,2)}=N_{(2,2)}*(T_{(2,2)}/T_{(2,2)})$

(3)將拒答以多數回答插補，若兩題的多數回答依序為 1、2[4]，以其插補拒答對應位置，插補的正確率為

$$r\{N_{(1,1)}*\frac{n_{(8,1)}}{T_{(1,1)}}+N_{(1,2)}+N_{(2,2)}*\frac{n_{(2,8)}}{T_{(2,2)}}\}/rN \tag{7}$$

，其中，$N_{(1,2)}=N_{(1,2)}*(T_{(1,2)}/T_{(1,2)})$

(註 4：當多數回答依序為 2、1，則計算出的葛特曼量表指標很可能不滿足要求，故不考慮多數回答為 2、1 的可能；又若多數回答為 1、1 或 2、2，則分別與容許、不容許插補結果相同。)

(4)將拒答以少數回答插補，若兩題的少數回答依序為 2、1，以其插補拒答對應位置，插補的正確率為

$$r\{N_{(1,1)}*\frac{n_{(1,8)}+n_{(8,1)}}{T_{(1,1)}}+N_{(2,2)}*\frac{n_{(8,2)}+n_{(2,8)}}{T_{(2,2)}}\}/rN \qquad \textbf{(8)}$$

（推論一）

當拒答部份以兩問題都拒答的形式最多時，會造成(8,8)個數很多而其他形式相對很小。將 $n_{(8,8)}$以外的形式以 0 代入上述四式，代入**(5)**式可得 $N_{(1,1)}/N$，代入**(6)**式可得 $N_{(2,2)}/N$，代入**(7)**式可得 $N_{(1,2)}/N$，代入**(8)**式則為 0。假設在比較自由的社會中，可知回答(1,1)的個數會最多，即 $N_{(1,1)}$ 最大，則此時以容許插補會使正確率最高；同理若是在比較保守的社會中，回答(2,2)個數最多，即 $N_{(2,2)}$ 最大，此時以不容許插補會使正確率最高；若是在一般非極端社會中，以多數回答插補會使正確率最高。以下將舉兩個簡單例子說明：

(例 1.1)

	$N_{(i,j)}$	$N_{(1,1)}$=200	$N_{(1,2)}$ =350	$N_{(2,2)}$ =500
$n_{(a,b)}$		(1,1)	(1,2)	(2,2)
$n_{(8,8)}$ =80	(8,8)	*	*	*
$n_{(1,8)}$ =0	(1,8)	*	*	0
$n_{(2,8)}$ =0	(2,8)	0	0	*
$n_{(8,1)}$ =0	(8,1)	*	0	0
$n_{(8,2)}$ =0	(8,2)	0	*	*

假設一個比較保守的社會，且拒答只有(8,8)這種形式，則此時若以容許插補，代入**(5)**式可得到正確率為 19.05%；以不容許插補，代入**(6)**式可得到正確率為 46.72%；以多數回答(假設依序為(1,2))插補，代入**(7)**式可得到正確率為 33.33%，以少數回答(假設依序為(2,1))插補，代入**(8)**式則為 0；則可知以不容許插補會有較高正確率。

(例 1.2)

	$N_{(i,j)}$	$N_{(1,1)}$=500	$N_{(1,2)}$ =350	$N_{(2,2)}$ =200
$n_{(a,b)}$		(1,1)	(1,2)	(2,2)
$n_{(8,8)}$ =80	(8,8)	*	*	*
$n_{(1,8)}$ =0	(1,8)	*	*	0
$n_{(2,8)}$ =0	(2,8)	0	0	*
$n_{(8,1)}$ =0	(8,1)	*	0	0
$n_{(8,2)}$ =0	(8,2)	0	*	*

又若假設一個比較自由的社會，且拒答只有(8,8)這種形式，則此時若以容許插補，代入***(5)***式可得到正確率為 46.72%；以不容許插補，代入***(6)***式可得到正確率為 19.05%；以多數回答(假設依序為(1,2))插補，代入***(7)***式可得到正確率為 33.33%，以少數回答(假設依序為(2,1))插補，代入***(8)***式則為 0；則可知以容許插補會有較高正確率。

(推論二)

當拒答部份以第二題拒答的形式比較多時，會造成(1,8)及(2,8)個數很多而其他形式相對很小。將 $n_{(1,8)}$及 $n_{(2,8)}$以外的形式以 0 代入上述四式，代入***(5)***式可得 $\{N_{(1,1)}+N_{(2,2)}\}/N$，代入***(6)***式可得 $\{N_{(1,1)}+N_{(2,2)}\}/N$，代入***(7)***式可得 $\{N_{(1,2)}+N_{(2,2)}\}/N$，代入***(8)***式可得 $\{N_{(1,1)}+N_{(2,2)}\}/N$。首先將(2,8)插補成(2,2)後，若是在比較自由的社會，$N_{(1,1)}$ 比 $N_{(1,2)}$ 大，則此時(1,8)應插補成(1,1)；反之，若 $N_{(1,2)}$ 比 $N_{(1,1)}$ 大，則此時(1,8)應插補成(1,2)。以下將舉兩個簡單例子說明：

(例 2.1)

	$N_{(i,j)}$	$N_{(1,1)}$=200	$N_{(1,2)}$ =350	$N_{(2,2)}$ =500
$n_{(a,b)}$		(1,1)	(1,2)	(2,2)
$n_{(8,8)}$ =0	(8,8)	*	*	*
$n_{(1,8)}$ =40	(1,8)	*	*	0
$n_{(2,8)}$ =40	(2,8)	0	0	*
$n_{(8,1)}$ =0	(8,1)	*	0	0
$n_{(8,2)}$ =0	(8,2)	0	*	*

假設一個比較保守的社會，且拒答有(1,8)及(2,8)兩種形式，先將(2,8)插補成(2,2)後，剩下的(1,8)若插補成(1,1)，代入***(5)***或***(8)***式，均可得到正確率為 66.67%，若插補成(1,2)，代入***(6)***或***(7)***式，均可得到正確率為 80.95%，則可知(1,8)插補成會(1,2)有較高正確率。

(例 2.2)

	$N_{(i,j)}$	$N_{(1,1)}$=500	$N_{(1,2)}$ =350	$N_{(2,2)}$ =200
$n_{(a,b)}$		(1,1)	(1,2)	(2,2)
$n_{(8,8)}$ =0	(8,8)	*	*	*
$n_{(1,8)}$ =40	(1,8)	*	*	0
$n_{(2,8)}$ =40	(2,8)	0	0	*
$n_{(8,1)}$ =0	(8,1)	*	0	0
$n_{(8,2)}$ =0	(8,2)	0	*	*

又若假設一個比較自由的社會，且拒答有(1,8)及(2,8)兩種形式，先將(2,8)插補成(2,2)後，剩下的(1,8)若插補成(1,1)，代入***(5)***或***(8)***式，均可得到正確率為 80.95%，若插補成(1,2)，代入***(6)***或***(7)***式，均可得到正確率為 66.67%，則可知(1,8)插補成會(1,1)有較高正確率。

(推論三)

當拒答部份以第一題拒答的形式比較多時，會成(8,1)及(8,2)個數很多而其他形式相對很小。將 $n_{(8,1)}$及 $n_{(8,2)}$以外的形式以 0 代入上述四式，代入***(5)***式可得 $\{N_{(1,1)} + N_{(1,2)}\}/N$ ，代入***(6)***式可得 $\{N_{(1,1)} + N_{(2,2)}\}/N$ ，代入***(7)***式可得

$\{N_{(1,1)}+N_{(1,2)}\}/N$，代入**(*8*)**式可得$\{N_{(1,1)}+N_{(2,2)}\}/N$。首先將(8,1)插補成(1,1)後，若是在比較保守的社會，$N_{(2,2)}$ 比 $N_{(1,2)}$ 大，則此時(8,2)應插補成(2,2)；反之，若 $N_{(1,2)}$ 比 $N_{(2,2)}$ 大，則此時(8,2)應插補成(1,2)。以下將舉兩個簡單例子說明：

(例 3.1)

	$N_{(i,j)}$	$N_{(1,1)}$=200	$N_{(1,2)}$ =350	$N_{(2,2)}$ =500
$n_{(a,b)}$		(1,1)	(1,2)	(2,2)
$n_{(8,8)}$ =0	(8,8)	*	*	*
$n_{(1,8)}$ =0	(1,8)	*	*	0
$n_{(2,8)}$ =0	(2,8)	0	0	*
$n_{(8,1)}$ =40	(8,1)	*	0	0
$n_{(8,2)}$ =40	(8,2)	0	*	*

假設一個比較保守的社會，且拒答有(8,1)及(8,2)兩種形式，先將(8,1)插補成(1,1)後，剩下的(8,2)若插補成(1,2)，代入**(*5*)**或**(*7*)**式，可得到正確率為 52.38%，若插補成(2,2)，代入**(*6*)**或**(*8*)**式可得到正確率為 66.67%，則可知(8,2)插補成會(2,2)有較高正確率。

(例 3.2)

	$N_{(i,j)}$	$N_{(1,1)}$=500	$N_{(1,2)}$ =350	$N_{(2,2)}$ =200
$n_{(a,b)}$		(1,1)	(1,2)	(2,2)
$n_{(8,8)}$ =0	(8,8)	*	*	*
$n_{(1,8)}$ =0	(1,8)	*	*	0
$n_{(2,8)}$ =0	(2,8)	0	0	*
$n_{(8,1)}$ =40	(8,1)	*	0	0
$n_{(8,2)}$ =40	(8,2)	0	*	*

假設一個比較保守的社會，且拒答有(8,1)及(8,2)兩種形式，先將(8,1)插補成(1,1)後，剩下的(8,2)若插補成(1,2)，代入**(*5*)**或**(*7*)**式，可得到正確率為 66.67%，若插補成(2,2)，代入**(*6*)**或**(*8*)**式可得到正確率為 52.38%，則可知(8,2)插補成會(1,2)有較高正確率。

下表 4.5 列出問題數為二，出現拒答時建議的簡易插補方法，其中第一行為沒有確定答案的拒答型態：

社會風氣 / 拒答型態	自由	中庸	保守
(8,8)	(1,1)	(1,2)	(2,2)
(8,2)	(1,2)	(1,2)	(2,2)
(1,8）	(1,1)	(1,2)	(2,2)

表 4.5 問題數為二，出現拒答時建議的簡易插補方法

(三) 三問題性質

假設有三個問題，則符合葛特曼量表之回答有(1,1,1)，(1,1,2)，(1,2,2)，(2,2,2)則拒答的形態如下：

回答型式 / 拒答型態	(1,1,1)	(1,1,2)	(1,2,2)	(2,2,2)
(8,8,8)	*	*	*	*
(1,8,8)	*	*	*	0
(2,8,8)	0	0	0	*
(8,1,8)	*	*	0	0
(8,2,8)	0	0	*	*
(8,8,1)	*	0	0	0
(8,8,2)	0	*	*	*
(1,1,8)	*	*	0	0
(1,2,8)	0	0	*	0
(2,2,8)	0	0	0	*
(1,8,1)	*	0	0	0
(1,8,2)	0	*	*	0
(2,8,2)	0	0	0	*
(8,1,1)	*	0	0	0
(8,1,2)	0	*	0	0
(8,2,2)	0	0	*	*

表 4.6 拒答型態對應可能的回答形式，*部分表示左邊的

拒答形式有可能由對應的上方回答製造出來(當問題數為三時)

其中 $N_{(i,j,k)}$表示回答(i,j,k)的個數，$i,j,k = 1,2$ 且 $i \leqq j \leqq k$，且 $N = N_{(1,1,1)}+N_{(1,1,2)}+N_{(1,2,2)}+N_{(2,2,2)}$；$n_{(a,b,c)}$表示拒答為$(a,b,c)$的個數，$a,b,c = 1,2,8$ 且 a,b,c 至少一個為 8；$T_{(i,j,k)}$表示可以插補成(i,j,k)的拒答形式之總數。又令拒答率為 r。根據葛特曼量表的特性，有確定答案的拒答形式包括(2,8,8)，(8,8,1)，(1,2,8)，(2,2,8)，(1,8,1)，(2,8,2)，(8,1,1)及(8,1,2)，應分別插補為(2,2,2)，(1,1,1) ，(1,2,2)，(2,2,2)，(1,1,1)，(2,2,2)，(1,1,1)及(1,1,2)；此外，拒答形式為(8,1,8)應視為(1,1,8)，(8,2,8) 應視為(8,2,2)，故此時必定正確的個數為

$$r\{N_{(1,1,1)}*\frac{n_{(8,8,1))}+n_{(1,8,1)}+n_{(8,1,1)}}{T_{(1,1,1)}}+N_{(1,1,2)}*\frac{n_{(8,1,2)}}{T_{(1,1,2)}}+N_{(1,2,2)}*\frac{n_{(1,2,8)}}{T_{(1,2,2)}}+N_{(2,2,2)}*\frac{n_{(2,8,8)}+n_{(2,2,8)}+n_{(2,8,2)}}{T_{(2,2,2)}}\}$$

除了這些拒答反應之外的類型再以容許、不容許、多數及少數回答等方法插補。加上必定正確的個數之後，各方法的插補正確率敘述如下：

(1)將拒答以容許插補，即以 1 取代所有的 8，則正確率為

$$r\{N_{(1,1,1)}+N_{(1,1,2)}*\frac{n_{(8,8,2)}+n_{(1,8,2)}+n_{(8,1,2)}}{T_{(1,1,2)}}+ N_{(1,2,2)}*\frac{n_{(8,2,2)}+n_{(1,2,8)}}{T_{(1,2,2)}}+N_{(2,2,2)}*\frac{n_{(2,8,8)}+n_{(2,2,8)}+n_{(2,8,2)}}{T_{(2,2,2)}}\}/rN \qquad \textbf{(9)}$$

，其中，$N_{(1,1,1)} = N_{(1,1,1)}*\frac{T_{(1,1,1)}}{T_{(1,1,1)}}$

(2)將拒答以不容許插補，即以 2 取代所有的 8，則正確率為

$$r\{N_{(1,1,1)}*\frac{n_{(8,8,1)}+n_{(1,8,1)}+n_{(8,1,1)}}{T_{(1,1,1)}}+N_{(1,1,2)}*\frac{n_{(1,1,8)}+n_{(8,1,2)}}{T_{(1,1,2)}} +N_{(1,2,2)}*\frac{n_{(1,8,8)}+n_{(1,2,8)}+n_{(1,8,2)}}{T_{(1,2,2)}}+N_{(2,2,2)}\}/rN \qquad \textbf{(10)}$$

，其中，$N_{(2,2,2)} = N_{(2,2,2)}*\frac{T_{(2,2,2)}}{T_{(2,2,2)}}$

(3)將拒答以多數回答插補，若三題的多數回答依序為 1、1、2，以其插補拒答對應位置，插補的正確率為

$$
\begin{aligned}
& r\{N_{(1,1,1)} * \frac{n_{(8,8,1)}+n_{(1,8,1)}+n_{(8,1,1)}}{T_{(1,1,1)}} + N_{(1,1,2)} + \\
& N_{(1,2,2)} * \frac{n_{(8,2,8)}+n_{(1,2,8)}+n_{(8,2,2)}}{T_{(1,2,2)}} + N_{(2,2,2)} * \frac{n_{(2,8,8)}+n_{(2,2,8)}+n_{(2,8,2)}}{T_{(2,2,2)}}\}/rN
\end{aligned}
\tag{11}
$$

，其中， $N_{(1,1,2)} = N_{(1,1,2)} * \frac{T_{(1,1,2)}}{T_{(1,1,2)}}$

(4)將拒答以多數回答插補，若三題的多數回答依序為 1、2、2，以其插補拒答對應位置，插補的正確率為

$$
\begin{aligned}
& r\{N_{(1,1,1)} * \frac{n_{(8,8,1)}+n_{(1,8,1)}+n_{(8,1,1)}}{T_{(1,1,1)}} + N_{(1,1,2)} * \frac{n_{(8,1,8)}+n_{(1,1,8)}+n_{(8,1,2)}}{T_{(1,1,2)}} + \\
& N_{(1,2,2)} + N_{(2,2,2)} * \frac{n_{(2,8,8)}+n_{(2,2,8)}+n_{(2,8,2)}}{T_{(2,2,2)}}\}/rN
\end{aligned}
\tag{12}
$$

，其中， $N_{(1,2,2)} = N_{(1,2,2)} * \frac{T_{(1,2,2)}}{T_{(1,2,2)}}$

(5)將拒答以少數回答插補，若三題的多數回答依序為 2、1、1，以其插補拒答對應位置，插補的正確率為

$$
\begin{aligned}
& r\{N_{(1,1,1)} * \frac{n_{(1,8,8)}+n_{(8,1,8)}+n_{(8,8,1)}+n_{(1,1,8)}+n_{(1,8,1)}+n_{(8,1,1)}}{T_{(1,1,1)}} + N_{(1,1,2)} * \frac{n_{(1,8,2)}+n_{(8,1,2)}}{T_{(1,1,2)}} + \\
& N_{(1,2,2)} * \frac{n_{(1,2,8)}}{T_{(1,2,2)}} + N_{(2,2,2)} * \frac{n_{(2,8,8)}+n_{(8,2,8)}+n_{(2,2,8)}+n_{(2,8,2)}+n_{(8,2,2)}}{T_{(2,2,2)}}\}/rN
\end{aligned}
\tag{13}
$$

(6)將拒答以少數回答插補，若三題的多數回答依序為 2、2、1，以其插補拒答對應位置，插補的正確率為

$$
\begin{aligned}
& r\{N_{(1,1,1)} * \frac{n_{(8,1,8)}+n_{(8,8,1)}+n_{(1,1,8)}+n_{(1,8,1)}+n_{(8,1,1)}}{T_{(1,1,1)}} + N_{(1,1,2)} * \frac{n_{(8,1,2)}}{T_{(1,1,2)}} \\
& + N_{(1,2,2)} * \frac{n_{(1,2,8)}+n_{(1,8,2)}}{T_{(1,2,2)}} + N_{(2,2,2)} * \frac{n_{(2,8,8)}+n_{(8,2,8)}+n_{(8,8,2)}+n_{(2,2,8)}+n_{(2,8,2)}+n_{(8,2,2)}}{T_{(2,2,2)}}\}/rN
\end{aligned}
\tag{14}
$$

(推論一)

當拒答部份以三題都拒答的形式比較多時，會成(8,8,8)個數很多而其他形式相對很小。將 $n_{(8,8,8)}$以外的形式以 0 代入上述六式，代入(***9***)式可得 $N_{(1,1,1)}/N$，代入(***10***)式可得 $N_{(2,2,2)}/N$，代入(***11***)式可得 $N_{(1,1,2)}/N$，代入(***12***)式可得 $N_{(1,2,2)}/N$，代入(***13***)及(***14***)式則為 0。假設在比較自由的社會中，可知回答(1,1,1)的個數會最多，即 $N_{(1,1,1)}$ 最大，則此時以(1,1,1)插補對應位置會使正確率最高；同理若是在比較保守的社會中，回答(2,2,2)的個數會最多，即 $N_{(2,2,2)}$ 最大，此時以(2,2,2)插補對應位置會使正確率最高；其餘以此類推。

(推論二)

當拒答部份以第三題拒答的形式比較多時，會成(a,b,8)的個數特別多而其他形式相對很小，a,b =1,2 且 $a \leqq b$。將 $n_{(a,b,8)}$以外的形式以 0 代入上述六式，代入(***9***)、(***13***)及(***14***)式可得$\{N_{(1,1,1)} + N_{(1,2,2)} + N_{(2,2,2)}\}/N$，代入(***10***)、(***11***)及(***12***)式可得$\{N_{(1,1,2)} + N_{(1,2,2)} + N_{(2,2,2)}\}/N$。首先將(1,2,8)，(2,2,8)分別插補成(1,2,2)與(2,2,2)後，此時要插補的僅剩(1,1,8)，若是在比較自由的社會，$N_{(1,1,1)}$ 比 $N_{(1,1,2)}$ 大，則此時(1,1,8)應插補成(1,1,1)；反之，若 $N_{(1,1,2)}$ 比 $N_{(1,1,1)}$ 大，則此時(1,1,8)應插補成(1,1,2)。

(推論三)

當拒答部份以第一題拒答的形式比較多時，會成(8,b,c)的個數特別多而其他形式相對很小，b,c =1,2，且 $b \leqq c$。將 $n_{(8,b,c)}$以外的形式以 0 代入上述六式，代入(***9***)、(***11***)及(***12***)式可得$\{N_{(1,1,1)} + N_{(1,1,2)} + N_{(1,2,2)}\}/N$，代入(***10***)、(***13***)及(***14***)式可得$\{N_{(1,1,1)} + N_{(1,1,2)} + N_{(2,2,2)}\}/N$。首先將(8,1,1)，(8,1,2)分別插補成(1,1,1)與(1,1,2)後，

此時要插補的僅剩(8,2,2)，若是在比較保守的社會，$N_{(2,2,2)}$ 比 $N_{(1,2,2)}$ 大，則此時(8,2,2)應插補成(2,2,2)；反之，若 $N_{(1,2,2)}$ 比 $N_{(2,2,2)}$ 大，則此時(8,2,2)應插補成(1,2,2)。

(推論四)

當拒答部份以第二題及第三題拒答的形式比較多時，會成(a,8,8)的個數特別多而其他形式相對很小，a =1,2。將 $n_{(a,8,8)}$以外的形式以 0 代入上述六式，代入***(9)***及***(13)***式可得$\{N_{(1,1,1)}+N_{(2,2,2)}\}/N$，代入***(10)***及***(12)***式可得$\{N_{(1,2,2)}+N_{(2,2,2)}\}/N$，代入***(11)***式可得$\{N_{(1,1,2)}+N_{(2,2,2)}\}/N$，代入***(14)***式可得$N_{(2,2,2)}/N$。首先將(2,8,8)插補成(2,2,2)後，若是在比較自由的社會，$N_{(1,1,1)}$ 會最大，則此時(1,8,8)應插補成(1,1,1)；又若是在稍微保守一點的社會，$N_{(1,1,2)}$ 會最大，則此時(1,8,8)應插補成(1,1,2)；又若是在更保守的社會，$N_{(1,2,2)}$ 會最大，則此時(1,8,8)應插補成(1,2,2)。

下表 4.7 列出問題數為三，出現拒答時建議的簡易插補方法，其中第一行為沒有確定答案的拒答型態：

社會風氣 / 拒答型態	極自由	自由	保守	極保守
(8,8,8)	(1,1,1)	(1,1,2)	(1,2,2)	(2,2,2)
(8,2,2)	(1,2,2)	(1,2,2)	(1,2,2)	(2,2,2)
(1,1,8)	(1,1,1)	(1,1,2)	(1,1,2)	(1,1,2)
(1,8,8)	(1,1,1)	(1,1,2)	(1,2,2)	(1,2,2)

表 4.7 問題數為三，出現拒答時建議的簡易插補方法

(四) 四問題推廣

假設有四個問題，則符合葛特曼量表之回答有(1,1,1,1)，(1,1,1,2)，(1,1,2,2)，(1,2,2,2)及，(2,2,2,2)。根據以上兩問題及三問題性質，可歸納得到四問題之下的結果：

(推論一)

當拒答部份以四題都拒答的形式比較多時，會成(8,8,8,8)個數很多而其他形式相對很小，則插補正確率為 $N_{(i,j,k,l)}/N$，(i,j,k,l)為上述的五種可能回答。此時應根據社會保守或自由程度來決定插補方式，例如較自由社會則以(1,1,1,1)插補對應位置最好；較保守社會則以(2,2,2,2)插補對應位置最好。

(推論二)

當拒答部份以第四題拒答的形式比較多時，會成$(a,b,c,8)$的個數特別多而其他形式相對很小，$a,b,c=1,2$ 且 $a\leqq b\leqq c$。在此情況下，先依據葛特曼量表的特性插補有確定答案者，例如(1,2,2,8)勢必要插補成(1,2,2,2)，則事實上欲插補的形式僅剩(1,1,1,8)，插補正確率分別為 $(N_{(1,1,1,1)}+N_{(1,1,2,2)}+N_{(1,2,2,2)}+N_{(2,2,2,2)})/N$ 及 $(N_{(1,1,1,2)}+N_{(1,1,2,2)}+N_{(1,2,2,2)}+N_{(2,2,2,2)})/N$。此時只需比較(1,1,1,1)與(1,1,1,2)的回答個數即可決定要插補的值。

(推論三)

當拒答部份以第一題拒答的形式比較多時，會成$(8,b,c,d)$的個數特別多而其他形式相對很小，$b,c,d=1,2$ 且 $b\leqq c\leqq d$。在此情況下，先依據葛特曼量表的特性插補有確定答案者，例如(8,1,1,2)勢必要插補成(1,1,1,2)，則事實上欲插補形式僅剩(8,2,2,2)，插補正確率分別為 $(N_{(1,1,1,1)}+N_{(1,1,1,2)}+N_{(1,1,2,2)}+N_{(1,2,2,2)})/N$ 及

$(N_{(1,1,1,1)}+N_{(1,1,1,2)}+N_{(1,1,2,2)}+N_{(2,2,2,2)})/N$。此時只需比較(1,2,2,2)與(2,2,2,2)的回答個數即可決定要插補的值。

(推論四)

當拒答部份以第三題及第四題拒答的形式比較多時，會成(a,b,8,8)的個數特別多而其他形式相對很小，$a,b=1,2$ 且 $a \leqq b$。在此情況下，先依據葛特曼量表的特性插補有確定答案者，例如(1,2,8,8)勢必要插補成(1,2,2,2)，則事實上欲插補形式僅剩(1,1,8,8)，插補正確率分別為$(N_{(1,1,1,1)}+N_{(1,2,2,2)}+N_{(2,2,2,2)})/N$、$(N_{(1,1,1,2)}+N_{(1,2,2,2)}+N_{(2,2,2,2)})/N$及$(N_{(1,1,2,2)}+N_{(1,2,2,2)}+N_{(2,2,2,2)})/N$。此時只需比較(1,1,1,1)、(1,1,1,2)與(1,1,2,2)的回答個數即可決定要插補的值。

(推論五)

當拒答部份以第二題、第三題及第四題拒答的形式比較多時，會成(a,8,8,8)的個數特別多而其他形式相對很小，$a=1,2$。在此情況下，先依據葛特曼量表的特性插補有確定答案者，也就是(2,8,8,8)勢必要插補成(2,2,2,2)，則事實上欲插補形式為(1,8,8,8)，插補正確率分別為$(N_{(1,1,1,1)}+N_{(2,2,2,2)})/N$、$(N_{(1,1,1,2)}+N_{(2,2,2,2)})/N$、$(N_{(1,1,2,2)}+N_{(2,2,2,2)})/N$及$(N_{(1,2,2,2)}+N_{(2,2,2,2)})/N$。此時需比較(1,1,1,1)、(1,1,1,2)、(1,1,2,2)與(1,2,2,2)的回答個數才可決定要插補的值。

下表 4.8 列出問題數為四，出現拒答時建議的簡易插補方法，其中第一行為沒有確定答案的拒答型態：

社會風氣 / 拒答型態	極自由	自由	中庸	保守	極保守
(8,8,8,8)	(1,1,1,1)	(1,1,1,2)	(1,1,2,2)	(1,2,2,2)	(2,2,2,2)
(8,2,2,2)	(1,2,2,2)	(1,2,2,2)	(1,2,2,2)	(1,2,2,2)	(2,2,2,2)
(1,1,1,8)	(1,1,1,1)	(1,1,1,2)	(1,1,1,2)	(1,1,1,2)	(1,1,1,2)
(1,1,8,8)	(1,1,1,1)	(1,1,1,2)	(1,1,2,2)	(1,1,2,2)	(1,1,2,2)
(1,8,8,8)	(1,1,1,1)	(1,1,1,2)	(1,1,2,2)	(1,2,2,2)	(1,2,2,2)

表 4.8 問題數為四，出現拒答時建議的簡易插補方法

(簡易插補法結論)

本研究之拒答部份是屬於受試者四題都拒絕回答，也就是(8,8,8,8)最多，大約佔了拒答資料的 74%，(1,8,8,8)其次，約佔 12%。由本節前面述推論，應根據已知回答類型之個數來進行插補。此份問卷為敏感性議題的調查，而台灣對於性態度方面，一般認為是偏向保守的，而回收問卷中也以回答(1,2,2,2)與(2,2,2,2)者最多，分別佔了不含拒答資料的 27%與 26%，但這兩者數量差異非常小，因此依序以 1、2、2、2 或 2、2、2、2 插補對應位置均會有最高的正確率。

第二節 多重插補法

(一) 模式一

利用多重插補法來插補遺漏值，同樣根據先前表 4.1 的方式製造拒答。不同的是，要使用多重插補法來進行插補時必須要有解釋變數，本研究在此使用先前提及的五個人口變項(性別、年齡、婚姻狀況、教育程度，及收入)來作為解釋變數。執行多重插補法時，收入是以組中點代入，例如收入若是回答 4，表示是在 2~3 萬之間，則組中點為 24999。教育程度則是轉換成教育年數，例如回答 17 表示是教育程度為大學，而教育年數即為 16。另外婚姻狀況則合併為四類：未婚、已婚、離婚及其他(分居、喪偶等)。

製造拒答時，會因為抽出的人口變項不同而影響結果，所以將模擬 500 次，每次插補以後，計算其平均數及標準差。插補的進行，是利用五個人口變項來插補第一個問題 H，再利用五個人口變項以及插補後 H 的資料來插第二個問題 K，以此類推。此外，根據葛特曼量表的特性，一旦較簡單的問題出現了負向答案時，之後較難的問題也應該都是負向答案，這個特性也要考量在多重插補法的過程當中，也就是說本研究使用的多重插補法在某些情況下是有些限制的，並不完全仰賴機率。本研究欲插補的變數是屬於二元變數，根據 Buuren 及 Oudshoorn(2000)的建議模型，應採用邏輯斯迴歸模型，利用統計軟體 R 來進行，部份程式碼請參閱附錄(見附錄二)。

欲插補的變數總共有四個，而每筆資料將插補三次($m = 3$)。因為我們的目的是要求正確率，可以在模擬之前先將原本是回答(1,1,1,1)的編號為 1，回答為(1,1,1,2)的編號為 2，以此類推，最後將插補後所有的資料混合起來，總共 144,585 筆(即 $1785*3^4$)，看看插補前後是否會有相同的回答。表 4.9 將列出多重插補法之下，一些常用統計量的結果比較。

拒答比例	5%	10%	15%	20%	25%	30%
最小值	0.2637	0.2878	0.2876	0.2971	0.2900	0.2845
第一四分位數	0.3119	0.3202	0.3179	0.3207	0.3191	0.3200
中位數	0.3249	0.3293	0.3257	0.3251	0.3240	0.3261
第三四分位數	0.3391	0.3376	0.3342	0.3319	0.3310	0.3393
最大值	0.3813	0.3747	0.3547	0.3425	0.3453	0.3487
平均數	**0.3255**	**0.3289**	**0.3260**	**0.3250**	**0.3253**	**0.3257**
標準差	**0.0201**	**0.0141**	**0.0114**	**0.0096**	**0.0086**	**0.0062**

表 4.9 模式一多重插補法之下正確率的常用統計量比較

由表 4.6 可以看出在不同拒答率之下，執行多重插補法的正確率差異很小，大約在 32.5%到 32.9%之間。以此筆資料來說，多重插補法的正確率比簡易插補法中正確率最高的一以不容許插補(正確率為 31.07%)高出了大約 1 個百分點。標準差方面，拒答率為 5%的標準差比較大一些，為 0.0201，其他的標準差則不超過 0.015。此外，由上表可以看出標準差隨著拒答率上升有略為下降的趨勢，可能的原因是，拒答率比較低的時候，拒答個數比較少，插補正確個數，也就是正確率分子的變化對於正確率的影響會比較大，因此可能會使得正確率的變異比較大；而隨著拒答率上升，拒答個數增加，正確率分子的變化對於正確率的影響就相對比較小，而使得正確率的變異比較小。

另外本研究也提供另外三種模式的多重插補法結果比較，但因為這些模式之下多重插補法的正確率相對來說都比較低一些，因此詳細過程放在附錄參考(見附錄三)。以下僅列出全部四種模式多重插補法的正確率來做比較。

拒答比例	5%	10%	15%	20%	25%	30%
模式一	***0.3255***	***0.3289***	***0.3260***	***0.3250***	***0.3253***	***0.3257***
模式二	0.3020	0.3086	0.3031	0.2988	0.3024	0.3134
模式三	0.2934	0.3042	0.3015	0.2983	0.2998	0.3029
模式四	0.3183	0.3173	0.3139	0.3166	0.3153	0.3203

表 4.10 四種多重插補法的正確率比較

(多重插補法結論)

比較第一種到第四種模式的多重插補，從正確率看來，第一種模式的正確率會最高，可以達到 32%以上，其他模式相對都低了 1%到 2%左右，因此若以正確率的觀點會比較建議採用第一種模式。但實際上其他模式的正確率也只是稍微低一些，其中又以第二種模式的變數個數最少，耗費時間也比較少，故如果考慮到效率的問題，第二種模式也是一種選擇。另外不同拒答率之下，正確率的變化不大，推測原因與拒答形式有關，因為(8,8,8,8)佔了大多數，要完全插補正確並不容易，而拒答率增加同時也使得(8,8,8,8)增加，因此造成正確率的變化不大。

第三節 最鄰近插補法

使用最鄰近插補法來進行遺漏值的插補，是根據人口變項的距離來插補，也就是當 H 等變數出現拒答時，找出與其人口變項最接近的資料，以其答案來插補之。在執行的過程中，最重要的一點莫過於怎樣才能稱為「最鄰近」，距離如何定義將會影響整個結果。一般最常用的距離定義首推「歐式距離(Euclidean distance)」。但是由於本研究中的人口變項包括性別、婚姻狀態等類別型變數，標記之數字並沒有實際的意義，並不符合歐式距離的概念。此外還有年齡、收入等連續型變數，因此考慮距離時要特別注意。考慮到上述原因，本研究採用混合型距離公式。

跟多重插補法相同，將收入以組中點代入。教育程度也是轉換成教育年數而成為連續型資料。婚姻狀態的分類也是分為未婚、已婚、離婚及其他。

插補的進行同樣要先根據表 4.1 的型態製造一定比例的拒答。將拒答資料與完整資料的人口變項依照混合型距離公式一一求出距離，找出最小值，則以該筆資料之回答插補之，此程序重覆 500 次，同樣使用統計軟體 R 來執行。平均正確率以及一些常用統計量的計算結果列於下表 4.8。

拒答比例	5%	10%	15%	20%	25%	30%
最小值	0.3708	0.4045	0.3933	0.4062	0.4036	0.4012
第一四分位數	0.4382	0.4494	0.4279	0.4370	0.4369	0.4318
中位數	0.4719	0.4691	0.4494	0.4566	0.4531	0.4476
第三四分位數	0.4944	0.4874	0.4747	0.4650	0.4665	0.4619
最大值	0.5618	0.5281	0.5393	0.5370	0.5256	0.5027
平均數	**0.4690**	**0.4673**	**0.4542**	**0.4532**	**0.4508**	**0.4480**
標準差	**0.0445**	**0.0315**	**0.0320**	**0.0231**	**0.0201**	**0.0188**

表 4.11 最鄰近插補法之下的常用統計量比較

使用最鄰近插補法的正確率是目前各種插補法之中最高的，都有超過 45%以上，最高可達到 46.9%。根據表 4.11 可以發現正確率隨著拒答率的增加有略微下降的趨勢。如此的結果其實是合理的，因為最鄰近插補法是去計算拒答資料人口變項與有完整回答人口變項的距離，找出最小者而與該筆資料回答相同。可以想見的是可比較的資料愈多的時候，找到的最小距離愈可信，而隨著可比較的資料變少，找到的最小距離可能可信度愈低，因而使得正確率下降。標準差的部份也是隨著拒答率的上升，有略為降低的趨勢。原因可能跟多重插補法部分一樣，跟計算正確率時，分母的變化有關。

雖然最鄰近插補法的正確率比較高，此方法所需要的時間是最多的，例如拒答率為 5%時，1785 筆資料中有 89 筆為有拒答部分，1696 筆為完整回答，則此時總共須計算 150,944(即 89*1696)次距離，找出每筆拒答資料對應的最小距離看其是否插補正確並計算正確率。而當拒答率為 30%時，距離的計算次數增加到 669464(即 536*1249)次，可以想見在執行上相當耗時，也因此本研究在拒答率較高的部分重覆的次數相對少了很多。

第伍章 結論與改進事項

第一節 結論

對於存在遺漏的資料，研究者通常會選擇不同方法對其進行插補，但是插補方法的決定常常沒有一個準則。本研究的目標著重於在葛特曼量表中，用不同方法來插補拒答資料的「正確率」比較。

本文使用的資料為，「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。根據文中的討論及模擬結果，對此筆資料而言，選用最鄰近插補法會有較高的正確率，大約可以達到 45%以上，而且不同拒答率的變化對於正確率並不會有太大的影響；用多重插補法的第一種模式計算出來的正確率次之，大約在 32%到 33%左右。求出的正確率乍看之下似乎不盡理想，但由於本文定義的正確率是要四題都插補正確才算，且拒答形態以(8,8,8,8)佔最多數，若假設每一題插補正確的機率為 0.7，且四題互相獨立的角度來看，四題皆插補正確的機率僅有 0.2401(即 0.7^4)；雖然這四題並非互相獨立，但對於全部插補正確的機率也不會有太多提升，而本文使用一些方法插補的正確率達到 32%甚至 45%，已相當理想。

這兩種方法有一個共同特色，就是使用較多的資訊來進行插補。最鄰近插補法是將拒答資料的人口變項與完整資料的人口變項比較，與最相似者的回答歸在同類，因此用到的變數最多；多重插補法的第一種模式也是不斷放入變數做為插補的依據。這也說明了遺漏的機制如果是屬於隨機遺漏，插補的進行是有其必要性的。

然而，實務上對資料進行插補只是調查研究的其中一步，若是在此耗費太多的時間，對研究者來說可能是一大負擔。從這筆資料來看，上述兩種插補法雖然因為使用變數較多而可得到較高的正確率，插補花費的時間也是最多的，加上其餘插補方法所求得的正確率並沒有低太多，例如簡易插補法中的以不容許或多數回答插補，因此若在時間有限的情況下，研究者也可考慮以不容許或多數回答插

補。

此外，最鄰近插補法與多重插補法的第一種模式會有最高正確率的前提是在我們所考慮的這筆資料之下，而此筆資料的一個特色是拒答形態為(8,8,8,8)的個數佔最多，若是其他拒答的形態佔最多時，正確率的結果可能會有不小的變化。本文在簡易插補法的討論中有提及當某種拒答形態出現最多時，簡易插補法的正確率會提高很多，例二就可以驗證此說法，正確率最高可達 80.95%。因此在進行插補時，觀察拒答形態也是一個重要關鍵。

第二節 改進事項

本研究主要針對完美的葛特曼量表之下的拒答進行插補。因為完美的葛特曼量表的假設比較強烈，也就是說一旦較簡單的問題回答負向答案，比較困難的問題也要回答負向答案。因此在插補時，這樣的條件假設必須要先行設定。但是這樣的假設可能太過強烈，根據問卷內容可以看出，事實上還是有些人會選擇同意後面的問題，卻不同意前面的問題，例如同意與剛認識的人發生性行為，卻不同意牽手；而當插補進行時，卻完全不考慮這些反應。因此也許可以試著將假設放寬一些，改用其他量表的模式來進行插補研究。

再者，本研究將問卷試題的四個回答合併為兩個，但是在葛特曼量表其實並沒有要求一定要二元的題目，合併只是為了便於後續的插補討論，因此當問題真的有四個回答，或甚至更多回答的時候，該如何進行插補也是值得討論的議題。

正確率的定義在本研究中是要求一定要四題都插補正確，但事實上插補的進行是一題一題逐項插補，所以其實可以考慮逐題來計算正確率。例如當原本回答為(1,1,2,2)，製造成拒答形態為(1,8,8,8)，再經過插補後回答為(1,2,2,2)，如此需要插補的三題之中，插補正確了兩題，因此可以定義為 0.67 個正確。這樣定義有其好處，以此問卷為例，回答(1,1,2,2)的人，對於性觀念的開放程度與回答(1,2,2,2)的人某種程度上滿接近的，若是完全視為插補錯誤似乎不是那麼合理。

另外，拒答的製造方式是按照資料的原始結構，個數隨著拒答率增加而放大，故在拒答結構固定不變的情況下，簡易插補法的插補正確率在與拒答率無關，甚至因為資料的拒答形態以(8,8,8,8)佔最多數，使得多重插補法與最鄰近插補法在不同拒答率之下的插補正確率變化也不大，因此可能可以改用其他的方式來製造拒答，或許更能比較不同拒答率之下的插補好壞。