國 立 政 治 大 學 應 用 數 學 系
數 學 教 學 碩 士 在 職 專 班

碩士學位論文

非固定權重因子之試題難易度模糊統計評估

Fuzzy Statistical Evaluation with Non-fixed Weighted Factors in the Item Difficult Parameter

碩專班學生：許家源撰

指導教授：吳柏林博士

中 華 民 國 一 百 年 六 月 十 三 日

中文摘要

在試題難易度分析上，傳統方式常以答對率或得分率的高低來認定試題的難易度。但答對率或得分率高低並不能真正表達應試者在答題時的難易感受程度。過去研究指出影響數學科試題難易度的主要因素有三：評量的數學內容、解題時的思考策略及解題所需的步驟數。

本論文針對影響數學科試題難易度的三個主要因素進行分析統計。以模糊統計的角度，提出非固定權重因子之二維模糊數。應用模糊數絕對距離的概念以摒除族群中異端值，最後針對不同族群進行難易指標分析與檢定。

關鍵字：難易度預估，模糊統計，數學試題，無母體數檢定

Abstract

On the basis of difficulty analysis, the hit rate or the scoring rate are considered as the index of the difficulty of the questions traditionally. However, these rates can not represent how test takers feel when taking tests. According to the previous papers on this subject, they note three factors to affect the difficulty analysis on math questions the content of the evaluation, the strategies of question solving, and the required steps to solve a question.

This research is focused on the three major factors which influence the difficulty of math examinations. With the angle of fuzzy statistics, a two-dimension fuzzy number will be presented with non-fixed weighted factors. By applying the absolute distance concept of fuzzy number, the extreme values are excluded. Consequently, the indices of difficulty from different groups can be tested and analyzed.

Keyword：Difficult forecast, Fuzzy Statistics, Mathematics question,
Nonparametric Tests

目次

圖目次

表目次

1. 前言

統計分析方法在社會科學調查中廣爲應用。在傳統的敘述統計中，平均數、眾數、中位數…等，常被作爲分析調查的工具，此類分析結果都屬單一表徵，然而在生活當中許多思維並非以單一面向而能言之。例如：人對天氣的冷熱感，什麼是「冷」？什麼是「熱」？假設用傳統方式問卷調查“你認爲天氣達多少溫度就是熱？”經傳統統計方式取問卷之平均數來分析，若結果爲 23 度。我們是否可以訂定達 23 度就稱爲「熱」?低於 23 度便稱爲「冷」呢？另外就大學畢業生每月薪水所得滿意度，若採用傳統方式做問卷調查“你認爲大學畢業生每月薪水應是多少元？”可能有人答 2.4 萬到 2.5 萬、有人答 2.3 到 2.6 萬、…等。那每人答的都是一區間，若以傳統方式統計則難以計算出“大學畢業生每月薪水應是多少？”因爲在人類社會生活中，許多事物難以用二元方式來論定，所以利用模糊統計的方式來分析事物，更能呈現人類真實的感受，更能貼近事物的真實面。

模糊理論由[11]Zadeh, L. A. (1965)提出，至今已漸漸從模糊數學理論研究及工程界之應用，推廣到社會科學界的傳統統計推論與方法論中。如[8]Lowen (1990), [9]Ruspini (1991), [7]Dubois 與 Prade (1991), [10]Tseng 與 Klein (1992) 分別提出對理念概似(approximate reasoning)的計量方法。在社會科學計量方法發展與應用亦有詳細的實例，如[6]黃仁德與吳柏林(1995)提出臺灣短期貨幣需求函數穩定性的模糊統計檢定方法；[4]吳柏林與楊文山(1997)對模糊統計在社會調查分析的應用；[3]吳柏林與曾能芳(1998)考慮以模糊迴歸參數估計於景氣對策信號之迴歸模型分析，這些研究均獲得不錯的結果。

在傳統教育測驗統計上，試題難易程度都以答對率或得分率的方式來評定。答對率或得分率高代表學生通過率高，因此推論試題較容易；反之，答對率或得分率低則代表學生通過率低，因此推論試題較難。但單純以答對率或得分率高低來評定試題的難易程度，則缺少人們真正的感受程度。假設某試題真的比較複雜而難以計算，但因大部分學生熟於此類型題目，造成答對率或得分率偏高，是否

將此題評定爲易呢？若某試題真實屬容易，但應答者可能某些原故造成答題上失誤，因此造成答對率或得分率偏低的結果，是否將此題評定爲難呢？所以一個試題的難易不應只由答對率或得分率高低來決定，需將人們對試題的感受程度納入評判項目，較爲客觀。

本論文期望以模糊軟計算方法代替傳統二元劃分方式，從大學學科能力測驗數學科試題中取材，針對高中學生、一般民眾、高、國中教師四個群組對試題的「難」、「易」認知進行研究分析。希望藉由本論文結果更能解讀各群組對試題難易程度的感受，以作爲未來教學上的參考。

2.糊模集合及軟運算

2.1 隸屬度函數

隸屬度函數是模糊理論的概念，其是由傳統集合中的特徵函數(characteristic function)所衍生出來，用來表示元素對模糊集合的隸屬度(membership grade)，其範圍在於 0 到 1 之間。[11]Zadeh (1965)在模糊集合論中提到，若某一元素屬於某一集合的程度愈大，其隸屬度值愈接近於 1，反之則愈接近於 0。

傳統上我們在定義一個人是否是「中年人」常以 45 歲做一個界限，然而一個 40 歲的人難道就不是「中年人」嗎？歲數 55 的人，難道也不能稱被作是中年人嗎？若利用隸屬度函數則能獲得較合理的答案。如果 45 歲絕對屬於「中年」，則其隸屬度自然屬於 1，而 40 歲的隸屬度函數值則爲 0.8，此表示 40 歲屬於「中年」的程度有 0.8；而 55 歲的隸屬度函數值爲 0.3，此表示 55 歲屬於「中年」的程度有 0.3。與傳統集合的特徵函數比較，隸屬度函數似乎將特徵函數平滑化，如圖 2.1。

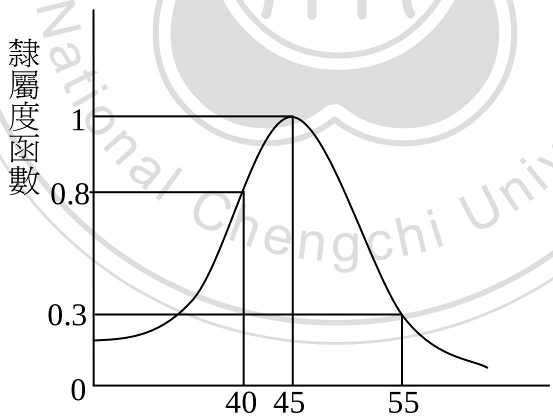

圖 2.1 模糊集合「中年」的隸屬度函數

在[12]Zimmermann (1991)提出隸屬度函數可分爲離散型(discretization)與連續型(continuous)兩種。離散型的隸屬度函數是直接給予有限模糊集合內每個元素的隸屬度，並以向量形式表現；而連續型隸屬度函數則有幾種常用的函數形式[S-函數、Z-函數、三角形函數、梯形函數、高斯(指數)函數]來描述模糊集點。函數

定義的表現，可以是無限模糊集合的元素及其隸屬度之間的關係，也可以是有限模糊集合的元素及其隸屬度之間的關係。

2.2 模糊數及軟計算

傳統問卷設計上常要求受訪者在問項上選填單一數值，而人類對於問項的感受程度具有糊模性及不確定性，因此傳統的問卷設計無法真實呈現受訪者的感受程度。若能讓受訪者根據自己的意識，利用隸屬度函數或區間值表達心中對於問項真正屬意的程度，則可更完整地傳達人類真實的思維。在考慮具有模糊特性的問項時，資料本身便具有不確定性與模糊性，故我們定義模糊數如下：

定義 2.1 模糊數 [1]吳柏林(2005)

設 U 為一論域，令 $\{L_1, L_2, ..., L_n\}$ 為論域 U 的因子集。u 為一對應到 $[0,1]$ 間的實數函數，即 $u: U \rightarrow [0,1]$。假若佈於論域 U 之一述句 X 其相對於因子集的隸屬度函數以 $\{\mu_1(X), \mu_2(X), ..., \mu_n(X)\}$ 表示，則在離散(discrete)的情形下，述句 X 的模糊數可表示成：

$$\mu_U(X) = \frac{\mu_1(X)}{L_1} + \frac{\mu_2(X)}{L_2} + ... + \frac{\mu_n(X)}{L_n}$$

其中+是或的意思，$\frac{u_i(X)}{L_i}$ 表示述句 X 隸屬於因子集 L_i 的程度。當 U 為連續時，述句 X 的模糊數可表示成：$\mu(X) = \int_{x \in X} \frac{u_i(X)}{L_i}$ 。

例 2.1 學測試題難度的模糊數表示

假設 X =學測試題難度的感覺，以模糊數表示爲 $\mu_U(X)$ 。假設論域 $U = \{L_1 = \text{很難}, L_2 = \text{偏難}, L_3 = \text{適中}, L_4 = \text{偏易}, L_5 = \text{容易}\}$，若 X =學測試題難度的感覺隸屬度函數爲

$$\{\mu_1(X) = 0.15, \mu_2(X) = 0.35, \mu_3(X) = 0.2, \mu_4(X) = 0.2, \mu_5(X) = 0.1\};$$

亦可以模糊數表示爲

$$\mu_U(X)=\frac{0.15}{L_1}+\frac{0.35}{L_2}+\frac{0.2}{L_3}+\frac{0.2}{L_4}+\frac{0.1}{L_5}$$

定義 2.2 模糊集合之基本運算 [1]吳柏林(2005)

設 A, B 為論域 U 中兩個模糊集合，其隸屬度表徵為 μ_A ，μ_B 。

(a)**補集：** $\mu_{A^c}(x)=1-\mu_A(x)$

(b)**交集：** $\mu_{A\cap B}(x)=\min(\mu_A(x),\mu_B)=\mu_A\wedge\mu_B$

(c)**聯集：** $\mu_{A\cup B}(x)=\min(\mu_A(x),\mu_B)=\mu_A\vee\mu_B$

(d)**距離：** d_1 **= 歐基里得(Euclidian)，** d_2 **= 漢民(Hamming)**

$$d_1(A,B)=\frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n}\left(\mu_A(x_i)-\mu_B(x_i)\right)^2}$$ （n =論域因子集個數）

$$d_2(A,B)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}1_{\{x_i:|\mu_A(x_i)-\mu_B(x_i)|>0\}}(x_i)$$ （n =論域因子集個數）

例 2.2

設 A, B 兩位學生對學習科目的隸屬度爲

$$A=\frac{0.2}{國文}+\frac{0.15}{英文}+\frac{0.2}{數學}+\frac{0.3}{社會}+\frac{0.15}{自然}$$

$$B=\frac{0.3}{國文}+\frac{0.25}{英文}+\frac{0.1}{數學}+\frac{0.2}{社會}+\frac{0.15}{自然}$$

則根據定義 2.2 可得以下之關係

$$A^c=\frac{0.8}{國文}+\frac{0.85}{英文}+\frac{0.8}{數學}+\frac{0.7}{社會}+\frac{0.85}{自然}$$

$$A\cap B=\frac{0.2}{國文}+\frac{0.15}{英文}+\frac{0.1}{數學}+\frac{0.2}{社會}+\frac{0.15}{自然}$$

$$A\cup B=\frac{0.3}{國文}+\frac{0.25}{英文}+\frac{0.2}{數學}+\frac{0.3}{社會}+\frac{0.15}{自然}$$

$$d_1(A,B)=\frac{1}{5}\sqrt{(0.1^2+0.1^2+0.1^2+0.1^2+0^2)}=0.04$$

$$d_2(A,B)=\frac{1}{5}\sum_{i=1}^{5}1_{\{x_i:|\mu_A(x_i)-\mu_B(x_i)|>0\}}=0.8$$

定義 2.3 離散型模糊樣本均數 [1]吳柏林(2005)

設 U 為一論域，$L=\{L_1=1, L_2=2, \ldots, L_k=k\}$ 為佈於論域 U 上的 k 個語言變數。若 $\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+\ldots\frac{m_{ik}}{L_k}, i=1,2,\ldots,n\}$，為一組模糊樣本（$\sum_{j=1}^{k} m_{ij}=1$）。則模糊樣本均數為

$$Fs=\frac{\frac{1}{n}\sum_{i=1}^{n} m_{i1}}{L_1}+\frac{\frac{1}{n}\sum_{i=1}^{n} m_{i2}}{L_2}+\ldots+\frac{\frac{1}{n}\sum_{i=1}^{n} m_{ik}}{L_k}$$

其中 m_{ij} 為第 i 個樣本相對於語言上變數 L_j 之隸屬度。

例 2.3 離散型模糊均數應用於餐廳服務滿意度調查

某餐廳想了解顧客對餐廳服務滿意程度，於是邀請五位顧客甲、乙、丙、丁、戊作調查，每位顧客對服務滿意度的隸屬度如表 2.1

表 2.1 五位顧客對餐廳服務滿意程度之隸屬度

滿意程度	L_1 很不滿意	L_2 不滿意	L_3 普通	L_4 滿意	L_5 很滿意
甲	0.4	0.5	0.1	0	0
乙	0	0.5	0	0.5	0
丙	0	0	0.4	0.4	0.2
丁	0.1	0.3	0	0	0.6
戊	0	0.1	0.4	0.4	0.1

則模糊樣本均數爲

$$Fs=\frac{\frac{1}{5}(0.4+0+0+0.1+0)}{\text{很不滿意}}+\frac{\frac{1}{5}(0.5+0.5+0+0.3+0.1)}{\text{不滿意}}+\frac{\frac{1}{5}(0.1+0+0.4+0+0.4)}{\text{普通}}$$

$$+\frac{\frac{1}{5}(0+0.5+0.4+0+0.4)}{\text{滿意}}+\frac{\frac{1}{5}(0+0+0.2+0.6+0.1)}{\text{很滿意}}$$

$$=\frac{\mathbf{0.1}}{\textbf{很不滿意}}+\frac{\mathbf{0.28}}{\textbf{不滿意}}+\frac{\mathbf{0.18}}{\textbf{普通}}+\frac{\mathbf{0.26}}{\textbf{滿意}}+\frac{\mathbf{0.18}}{\textbf{很滿意}}$$

3.多因子模糊數及模糊絕對距離

3.1 二維模糊數及樣本平均數

在傳統問卷設計上，常以單一變項做爲問卷調查的模式，而將單一變項的問卷結果做爲推論的唯一依據並不適合。本論文考慮人類的思考及感覺並非單一模式，故提出二因子(變項)的模糊問卷，期望在二因子(變項)的模糊問卷調查之下，所得問卷結果更能真實呈現受訪者的感受。首先本論文定義何謂二維模糊數。

定義 3.1 二維模糊數

設 U 為一論域，令 $\{C_1, C_2, \ldots, C_n\}$ 為論域 U 的因子集。μ 為一對應到 $[0,1]$ 間的實數函數，即 $\mu: U \to [0,1]$。假若佈於論域 U 之一述句 X 其相對於因子集的隸屬度函數以 $\{\mu_1(X), \mu_2(X), \ldots, \mu_n(X)\}$ 表示，另一述句 Y 其相對於因子集的隸屬度函數以 $\{\mu_1(Y), \mu_2(Y), \ldots, \mu_n(Y)\}$ 表示，則在離散(discrete)的情形下，述句 X 的模糊數可表示成：

$$\mu_U(X,Y) = \frac{[\mu_1(X), \mu_1(Y)]}{C_1} + \frac{[\mu_2(X), \mu_2(Y)]}{C_2} + \ldots + \frac{[\mu_n(X), \mu_n(Y)]}{C_n}$$

其中+是或的意思，$\frac{[\mu_i(X), \mu_i(Y)]}{C_i}$ 表示述句 X，Y 隸屬於因子集 C_i 的程度。

例 3.1a 顧客對某車商之汽車款式需求度及滿意程度的模糊數表示

某車商想了解顧客對汽車款式需求度及滿意程度，於是隨機邀請一位顧客作調查，設論域爲 $U = \{$休旅車、房車、跑車$\}$，$X =$ 顧客對車款的需求程度，則以 $\mu_U(X)$ 表示；$Y =$ 顧客對車款的滿意度，則以表示 $\mu_U(Y)$。請顧客在各車款的需求度填上 0~1 之間的數字，代表對此車款的“需求程度”，針對各車款的滿意度也填上 0~1 之間的數字，代表對此車款的“滿意程度”。則此顧客對購買車款車需求及滿意度的二維模糊數爲

$$\mu_U(X,Y) = \frac{(0.5,0.3)}{\text{休旅車}} + \frac{(0.4,0.2)}{\text{房車}} + \frac{(0.1,0.5)}{\text{跑車}}$$

例 3.1b 學生對學習科目喜愛程度及學習困難度的模糊數表示

設論域爲 $U=\{$國文、英文、數學、自然、社會$\}$，$X=$喜好程度，某學生針對各科喜好程度的隸屬度爲 $\{0.1, 0.1, 0.4, 0.3, 0.1\}$；$Y=$學習困難度，此學生對各科學習困難度的隸屬度爲 $\{0.4, 0.1, 0.1, 0.1, 0.3\}$。則此學生對學習科目喜愛程度及學習困難度的二維模糊數爲

$$\frac{(0.1,0.4)}{\text{國文}}+\frac{(0.1,0.1)}{\text{英文}}+\frac{(0.4,0.1)}{\text{數學}}+\frac{(0.3,0.1)}{\text{自然}}+\frac{(0.1,0.3)}{\text{社會}}$$

定義 3.2 離散型二維模糊樣本均數

設 U 為一論域，$L=\{L_1=1, L_2=2, ..., L_k=k\}$ 為佈於論域 U 上的 k 個語言變數。若 $\{x_i=\frac{(x_{i1}, y_{i1})}{L_1}+\frac{(x_{i2}, y_{i2})}{L_2}+...+\frac{(x_{ik}, y_{ik})}{L_k}, i=1,2,...,n\}$，為一組二維模糊樣本（$\sum_{j=1}^{k} x_{ij}=1$、$\sum_{j=1}^{k} y_{ij}=1$）。則模糊樣本均數為

$$Fs=\frac{\left(\frac{1}{n}\sum_{i=1}^{n}x_{i1}, \frac{1}{n}\sum_{i=1}^{n}y_{i1}\right)}{L_1}+\frac{\left(\frac{1}{n}\sum_{i=1}^{n}x_{i2}, \frac{1}{n}\sum_{i=1}^{n}y_{i2}\right)}{L_2}+...+\frac{\left(\frac{1}{n}\sum_{i=1}^{n}x_{ik}, \frac{1}{n}\sum_{i=1}^{n}y_{ik}\right)}{L_k}$$

其中 (x_{ij}, y_{ij}) 為第 i 個樣本相對於語言上變數 L_j 之隸屬度。

3.2 模糊樣本之絕對距離

在同一屬性的樣本群中，針對其屬性所得的統計資料應具有一致性。換句話說，同屬性的各樣本之間差異性不應過大，若某樣本與其他樣本之間差異性過大，即此樣本的代表性偏低。因此，將此代表性偏低之樣本納入統計資料，如此將會造成整體性的代表性降低。本論文爲了避免同屬性樣本中因差異性過大而造成群體代表性不足，將利用模糊樣本之絕對距離，統計各樣本之間的差異並摒除差異過大的樣本，如此資料更能具代表性。

定義 3.3 兩離散型模糊樣本之絕對距離 [1]吳柏林(2005)

設U為一論域，$L=\{L_1=1,L_2=2,...,L_k=k\}$為佈於論域$U$上的$k$個一組有序語言變數。若$\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+...+\frac{m_{ik}}{L_k}, i=p,q\}$, $\sum_{j=1}^{k}m_{ij}=1$為自論域中抽出的兩個模糊樣本。則兩離散型樣本x_p與x_q之距離定義為

$$d(x_p,x_q)=\frac{1}{k-1}\left|\sum_{j=1}^{k}(m_{pj}-m_{qj})\cdot j\right|$$

由定義可知距離d最大為 1，最小為 0。d值愈小，代表兩模糊樣本愈近似；d值愈大，代表兩模糊樣本愈不近似。值得注意的是此距離的定義不符合遞移性，但較符合我們日常生活的實際經驗。若「A與B的看法差不多」，「 B與C的看法差不多」，不見得就代表「A與C的看法差不多」，因為A、C的想法可能同中有異。

例 3.2 若有一組受訪者的模糊意見表示如下

表 3.1 五位顧客對商品滿意度之隸屬度

商品滿意程度	L_1 很不滿意	L_2 不滿意	L_3 普通	L_4 滿意	L_5 很滿意
甲	0.4	0.5	0.1	0	0
乙	0	0.5	0	0.5	0
丙	0	0	0.4	0.4	0.2
丁	0.1	0.3	0	0	0.6
戊	0	0.1	0.4	0.4	0.1

根據離散型模糊樣本相對距離定義，我們可算出各樣本間的相對距離。相對距離的模糊數值大小可看出各樣本在不同語意變數上的差異性，

如：

$$d(甲,乙)=\frac{|0.4-0|}{L_1}+\frac{|0.5-0.5|}{L_2}+\frac{|0.1-0|}{L_3}+\frac{|0-0.5|}{L_4}+\frac{|0-0|}{L_5}=\frac{0.4}{L_1}+\frac{0}{L_2}+\frac{0.1}{L_3}+\frac{0.5}{L_4}+\frac{0}{L_5}$$

$$d(丙,戊)=\frac{|0-0|}{L_1}+\frac{|0-0.1|}{L_2}+\frac{|0.4-0.4|}{L_3}+\frac{|0.4-0.4|}{L_4}+\frac{|0.2-0.1|}{L_5}=\frac{0}{L_1}+\frac{0.1}{L_2}+\frac{0}{L_3}+\frac{0}{L_4}+\frac{0.1}{L_5}$$

由 $d(甲,乙)$ 的模糊數值顯示甲、乙在「很不滿意」與「滿意」這二個選項差異性很大，在「普通」上差異性小，而其他部分則無差異。$d(丙,戊)$ 模糊數值顯示丙、戊整體上大致無差異，只在「不滿意」與「滿意」上有小差異。

由定義 3.2，我們可進一步算出各樣本之間的絕對距離值。絕對距離的數值大小則可看出各樣本整體之間的差異性。如下表 3.2 所示

表 3.2 五位顧客對商品滿意度之意見距離

d	甲	乙	丙	丁	戊
甲	0	0.325	0.525	0.5	0.45
乙	0.325	0	0.2	0.175	0.125
丙	0.525	0.2	0	0.025	0.075
丁	0.5	0.175	0.025	0	0.05
戊	0.45	0.125	0.075	0.05	0

由上表可發現：顧客丙的意見與顧客丁的意見最接近。而顧客丁的意見與顧客戊的意見次之，其中顧客甲的意見與顧客丙的意見最不接近。

定義 3.4 離散型模糊樣本之絕對距離總和

設 U 為一論域，$L=\{L_1=1,L_2=2,...,L_k=k\}$ 為佈於論域 U 上的 k 個一組有序語言變數。若 $\{x_i=\frac{m_{i1}}{L_1}+\frac{m_{i2}}{L_2}+...+\frac{m_{ik}}{L_k},i=1,2,...,n\}$, $\sum_{j=1}^{k}m_{ij}=1$ 為自論域中抽出的 n 個模糊樣本。則其中離散型樣本 x_p 之絕對距離總和為其與其他所有離散型絕對距離的總和，即表示為

$$S_p=d(x_p,x_1)+d(x_p,x_2)+...+d(x_p,x_n)=\sum_{i=1}^{n}d(x_p,x_i)$$

例 3.3 由定義 3.3 及例 3.2 中，我們可進一步計算出各樣本的絕對距離總和：

$S_{甲} = d(甲,乙) + d(甲,丙) + d(甲,丁) + d(甲,戊)$ =0.325+0.525+0.5+0.45=1.8

$S_{乙} = d(乙,甲) + d(乙,丙) + d(乙,丁) + d(乙,戊)$ =0.325+0.2+0.175+0.125=0.825

$S_{丙} = d(丙,甲) + d(丙,乙) + d(丙,丁) + d(丙,戊)$ =0.525+0.2+0.025+0.075=0.825

$S_{丁} = d(丁,甲) + d(丁,乙) + d(丁,丙) + d(丁,戊)$ =0.5+0.175+0.025+0.05=0.775

$S_{戊} = d(戊,甲) + d(戊,乙) + d(戊,丙) + d(戊,丁)$ =0.45+0.125+0.075+0.05=0.7

由上可得知，甲的絕對距離總和最大，代表丁的意見與其他樣本差異最大；而戊的絕對距離總和最小，代表戊的意見與其他樣本差異最小。

3.3 糊模樣本之異端值

針對同屬性群體進行屬性相關的統計調查，理論上所得統計資料的差異性不應過大。換句話說，樣本彼此之間的絕對距離應不大，如果有絕對距離過大的樣本我們便要稱之爲異端值，因此本論文依此訂定以下定義。

定義 3.5 單因子離散型模糊樣本之異端值

設 U 為一論域，$L = \{L_1 = 1, L_2 = 2, ..., L_k = k\}$ 為佈於論域 U 上的 k 個一組有序語言變數。若 $\{x_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + ... + \frac{m_{ik}}{L_k}, i = 1,2,...,n\}$, $\sum_{j=1}^{k} m_{ij} = 1$ 為自論域中抽出的 n 個模糊樣本。其中 S_p 代表某樣本 p 與其他樣本之絕對距離的總和，若 $S_p > Q_3$，Q_3 為此各樣本絕對距離的總和中的第 3 四分位數，則此樣本 p 為此樣本群的異端值。

例 3.4 由例 3.3 我們可計算所有樣本的絕對距離總和，並由總和值可計算出 $Q_3 = 0.825$，因爲甲與其他樣本絕對距離總和爲 1.8，即 $S_{甲} = 1.8 > Q_3$，所以樣本甲爲此樣本群中的異端值，它不具代表性，故可以摒除樣本甲。

3.4 試題難易指標

影響數學試題難易的主要三因素，分別是試題所評量的數學內容、以及思考策略與步驟數。本論文針對此三語言變數進行權重及難易度的二維模糊統計，設論域爲 $U = \{$數學內容、思考策略、步驟數$\}$，$X =$試題的權重程度，則以 $\mu_U(X)$ 表示；$Y =$試題的難易程度，則以 $\mu_U(Y)$ 表示。在問卷填寫上，針對論域的權重程度填入 0~1 之間的數字，代表對此題的“權重程度”，針對論域的難易度也填入 0~1 之間的數字，代表對此題“難易程度”。則此試題權重及難易程度的二維模糊數爲

$$\mu_U(X,Y) = \frac{[\mu_1(X), \mu_1(Y)]}{\text{數學內容}} + \frac{[\mu_2(X), \mu_2(Y)]}{\text{思考策略}} + \frac{[\mu_3(X), \mu_3(Y)]}{\text{步驟數}}$$

本論文爲考慮群組資料的代表性，除了去除異端值，另外在定義試題難易度指標上，因爲幾何平均數受極端值的影響較算術平均數小，故採用幾何平均作爲指標定義更具代表性。由此本論文定義出單一試題難易度指標。

定義 3.6 試題難易度指標

令 s^* 為一組異端值修正後之二維模糊樣本群，則 Fs^* 為此模糊樣本均數

$$Fs^* = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}, \frac{1}{n}\sum_{i=1}^{n} y_{i1}\right)}{L_1} + \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_{i2}, \frac{1}{n}\sum_{i=1}^{n} y_{i2}\right)}{L_2} + \dots + \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_{ik}, \frac{1}{n}\sum_{i=1}^{n} y_{ik}\right)}{L_k}$$

其中 (x_{ij}, y_{ij}) 為第 i 個樣本相對於語言上變數 L_j 之隸屬度，則試題難易度指標定義

為 $DT = \prod_{j=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n} y_{ij}\right)^{\frac{1}{n}\sum_{i=1}^{n} x_{ij}}$，$0 \le DT \le 1$。

由定義中可知 DT 最大爲 1，最小爲 0。 DT 值愈大代表試題難度較高； DT 值愈小代表試題較易。

例 3.5 如下爲一組異端值修正後之二維模糊樣本群，

$$甲=\frac{(0.4,0.5)}{數學內容}+\frac{(0.5,0.4)}{思考策略}+\frac{(0.1,0.1)}{步驟數}$$

$$乙=\frac{(0.3,0.4)}{數學內容}+\frac{(0.4,0.2)}{思考策略}+\frac{(0.3,0.4)}{步驟數}$$

$$丙=\frac{(0.2,0.3)}{數學內容}+\frac{(0.3,0)}{思考策略}+\frac{(0.5,0.7)}{步驟數}$$

則

$$Fs^{*}=\frac{\left(\frac{0.4+0.3+0.2}{3},\frac{0.5+0.4+0.3}{3}\right)}{數學內容}+\frac{\left(\frac{0.5+0.4+0.3}{3},\frac{0.4+0.2+0}{3}\right)}{思考策略}$$

$$+\frac{\left(\frac{0.1+0.3+0.5}{3},\frac{0.1+0.4+0.7}{3}\right)}{步驟數}$$

$$=\frac{(0.3,0.4)}{數學內容}+\frac{(0.4,0.2)}{思考策略}+\frac{(0.3,0.4)}{步驟數}$$

依定義 3.5，則此題難易度指標 $DT=0.4^{0.3}\times0.2^{0.4}\times0.4^{0.3}=0.30$

3.5 無母數統計檢定

在母體分配已知的假設下，進行母體如平均值、變異數、相關係數檢定等，我們稱爲母體統計檢定。一般統計檢定方法，都假定抽樣母體來自某一分配，最常見的是假設抽樣母體來自常態分配。但是如果母體來自某一未知分配時，我們可應用無母數統計方法，來解決當母體分配不是常態或母體分配未知，或樣本爲小樣本時的情形。本論文抽樣的對象爲高中學生、一般民眾、高中教師與國中教師，其樣本數較小且無法得知母體分配情形，故採此方法作爲檢定工具。

1. 威克生符號等級檢定(Wilcoxon Sign Rank Test)

假設我們由隨機成對樣本$\{(X_i,Y_i);i=1,\cdots,n\}$中，要檢定兩母體$X,Y$是否有差異。令$d_i=Y_i-X_i$為成對$(X_i,Y_i)$之差值，$i=1,\cdots,n$。對$d_i$取絕對值，再以等級 1 給$|d_i|$之最小值，等級 2 給$|d_i|$之次小值，…，依此類推，以等級 n 給$|d_i|$之最大值。若有數個$|d_i|$值相等，則取其在不同值下所對應等級的平均值。則易知全部$|d_i|$等級和$\frac{n(n+1)}{2}$。

令T^+為所有d_i為正數的等級和，T^-為所有d_i為負數的等級和。則易知T^+或T^-的最小值均可為 0，且T^+或T^-的最大值均可能為$\frac{n(n+1)}{2}$。因此在統計假設H_0：$M_X=M_Y$下，若T^+或T^-接近$\frac{n(n+1)}{4}$，則我們可接受H_0：X,Y兩母體沒有差異。反之若T^+或T^-接近 0 或$\frac{n(n+1)}{2}$，則我們拒絕H_0。

本論文將檢定三個群組彼此之間是否有差異性，因為問卷試題只有 20 個，屬於小樣本，且問卷對象無法確定是否為常態分配，故採用無母數檢定方式。其中威克生符號等級檢定(Wilcoxon Sign Rank Test)適合本研究資料上的檢定。

威克生符號等級檢定過程

1.資料：一組隨機成對樣本(X_i,Y_i)，記$d_i=Y_i-X_i$，$i=1,\ldots,n$

2.統計假設：X,Y兩母體分配的中位數一樣；即

$$H_0：M_X=M_Y；H_1：M_X\neq M_Y$$

3.統計量：$T=T^+$與T^-中較小的值

4.決策：在顯著水準α下的左尾檢定，查附錄表，若$T<T_{\alpha/2}$(臨界值)，則拒絕H_0

例 3.6a 針對學生對國文與英文喜好程度進行調查。經由調查 8 位學生，依喜好程度給予不同評分(1 至 10 分)。分數愈高代表愈喜歡。其結果如表 3.3，若以威

克生符號等級檢定，我們得到負號總和$T^{-}=13$，而在$\alpha=0.05$顯著水準下，用單尾檢定，$T_{0.05}=6<T^{-}=13$。故我們接受H_0，即學生在國文、英文喜好程度上無差異。

表 3.3 學生對國文與英文的喜好程度

學生	1	2	3	4	5	6	7	8
國文	7.5	4.0	6.5	3.4	4.1	5.7	6.2	8.9
英文	4.5	6.5	4.3	5.0	4.0	7.1	1.0	7.7
d_i	3.0	-2.5	2.2	-1.6	0.1	-1.4	5.2	1.2
$\|d_i\|$之等級	7	6	5	4	1	3	8	2

2.克洛斯−瓦力士檢定(Kruskal-Wallis Test)

克洛斯-瓦力士檢定，簡稱$K-W$檢定，乃用來檢定多組隨機樣本的母體分配是否相同的問題。當多組隨機樣本的母體分配來自常態分配時，我們可使用F檢定，來決定多組隨機樣本的母體分配是否相同。但是當母體分配不是來自常態分配時，我們就必須應用克洛斯-瓦力士檢定。

假設有k個相互獨立之母體，從每個母體抽出n_i個隨機樣本，$i=1,...,k$。將所有$n_1+n_2+n_3+\cdots n_k=N$個混合樣本，按觀測值大小排等級。最小的觀測值給 1，次小者給 2，…，以此類推，最大的觀測值給N；若有數個觀測值給相等，則取其在不同值下所對應等級的平均值。然後我們分別對每一組的隨機樣本計算等級總和，以R_i表示；即$R_i=$第i組樣本在混合樣本中的等級總和。

我們很容易知道R_i的期望值為$\dfrac{n_i(N+1)}{2}$。最後將R_i代入克洛斯可-瓦力士檢定量

$$K=\frac{12}{\mathrm{N(N+1)}}\sum_{i=1}^{k}\frac{1}{n}\left[R_i-\frac{n_i(N+1)}{2}\right]^2$$

$$= \frac{12}{\mathrm{N(N+1)}} \sum_{i=1}^{k} \frac{R_i^2}{n} - 3(N+1)$$

若各組的母體分配皆相同，那麼 R_i 值差異應不大，則 K 值將不會太大。在這種情況下，我們將接受 H_0 ：各母體分配相同的統計假設。但是若有某組等級總和 R_i ，相對過大或相對過小，則 K 值將會太大，這表示此 k 個母體的分配不完全相同，因此我們將拒絕 H_0 。

在統計假設 H_0 為真，且各組樣本數 $n_i \geq 5$ 的情況下， K 的統計量近似於自由度為 $(k-1)$ 的 χ_α^2 分配。因此，由附表找臨界值，若 K 的統計量值大於臨界值，則我們拒絕 H_0 。

克洛斯可-瓦力士檢定過程

1.資料：k 組相互獨立樣本，每組樣本有 n_i 個觀測值，$i = 1,2,....,k$

$n_1 + n_2 + n_3 + \ldots + n_m = N$

2.統計假設：該 k 組隨機樣本的母體分配是相同

H_0 ： $F_1 = F_2 = \ldots = F_k$

H_1 ：**至少有一組樣本其母體分配與其他組母體分配不同**

3.統計量： $K = \dfrac{12}{\mathrm{N(N+1)}} \sum_{i=1}^{k} \dfrac{R_i^2}{n} - 3(N+1)$

4.決策：在顯著水準 α 下的右尾檢定，如果 $n_i > 5$，則 K 近似自由度為 $k-1$ 之卡方分配，即若 K 值大於 $\chi_\alpha^2(k-1)$，則拒絕 H_0。

例 3.6b 某教師甄選想了解口試委員的評分標準是否一致。於是在三位口試委員所評分之考卷中，分別隨機抽取 7 位成績，其成績如表 3.4

表 3.4 三位口試委員的評分成績

口試委員 1	81	76	71	83	89	87	73
口試委員 2	75	82	79	88	70	78	86
口試委員 3	80	72	74	84	77	90	85

統計假設爲

H_0：每位口試委員評分分配相同

H_1：至少有一位口試委員評分分配與其他口試委員不同

將所有的 7+7+7=21 觀測值混合排序後我們得到

口試委員 1	71	73	76	81	83	87	89	$R_1 = 77$
口試委員 2	70	75	78	79	82	86	88	$R_2 = 75$
口試委員 3	72	74	77	80	84	85	90	$R_3 = 79$

$$K = \frac{12}{21(21+1)}\left[\frac{77^2}{7}+\frac{75^2}{7}+\frac{79^2}{7}\right]-3(21+1)=0.029$$

因爲 $\chi^2_{0.95}(2)=5.99>0.029$，所以我們接受 H_0；即口試委員在評分標準上有一致性。

3.隨機性檢定(Run Test)

在抽樣及統計分析的過程中，樣本之隨機性是一個相當重要的假設。若想要知道一組樣本是否隨機抽取，我們可應用連串法(run tes)來進行樣本之隨機性檢定。它是根據樣本資料中連串(run)的數目多寡，來決定此組樣本是否具隨機性之檢定方法。

所謂連串的意思是指一系列相同項或性質所組成，一直到下一個不同項出現

爲止。該連串之長度爲此相同項之數目。例如觀察最近 14 天加權股價指數的漲(+)跌(-)如下：<u>+ + + + - - - + - - - + + +</u>。則此組觀察值共有 5 個連串，第一個連串長度爲 4，第二個連串長度爲 3，…第五個連串長度爲 3。

當一組樣本其連串數目太多或太少時，我們便會懷疑此樣本可能不具隨機性。以 14 個觀察值而言，若出現順序爲<u>+ + + + + + + - - - - - - -</u> 則此組觀察值僅有兩個連串。若出現<u>+ - + - + - + - + - + - + -</u> 則此組觀察值共有 14 個連串。我們可以看出此觀察值的出現相當有規則性，應不具隨機性。

連串法檢定過程

1.資料：將 n_1 筆資料 $x_1, x_2, x_3, \cdots, x_n$ 按序記錄，且此 n 筆資料被區分為兩種型式。

n_1 為型式 1 出現的總數；n_2 為型式 2 出現的總數

2.統計假設：

H_0 **：此筆資料具隨機性**

H_1 **：此筆資料不具隨機性**

3.統計量：r = 連串的數目

4.決策：在顯著水準 $\alpha = 0.05$ 下的雙尾檢定，查附表得 r_L 和 r_U ，如果 $r_L \le r \le r_U$ ，則接受 H_0

例 3.6c 某百貨公司爲檢驗某商品每月的銷售量是否隨機出現，將二年內每個月銷售量製成下表 3.5，則在 $\alpha = 0.05$ 顯著水準下，是否每月的銷售量隨機出現?

表 3.5 商品每月銷售量

月份	1	2	3	4	5	6	7	8	9	10	11	12
銷售量	15	16	15	19	18	21	15	14	22	21	17	18

月份	1	2	3	4	5	6	7	8	9	10	11	12
銷售量	21	15	17	15	14	18	13	19	20	22	18	16

統計假設爲

H_0：此商品每月的銷售量是隨機出現

H_1：此商品每月的銷售量是隨機出現

由上表資料可得中位數是 17.5，再將觀察值與中位數比較，大於中位數者計 H，小於中位數者計 L。最後計算串連數：

$$\underline{\mathrm{L\,L\,L}}\ \underline{\mathrm{H\,H\,H}}\ \underline{\mathrm{L\,L}}\ \underline{\mathrm{H\,H}}\ \underline{\mathrm{L}}\ \underline{\mathrm{H\,H}}\ \underline{\mathrm{L\,L\,L\,L}}\ \underline{\mathrm{H}}\ \underline{\mathrm{L}}\ \underline{\mathrm{H\,H\,H\,H}}\ \underline{\mathrm{L}}$$

得知 $n_1 = n_2 = 12$，$r = 11$。由查表可知 $P(r \le 11) = 0.404 > 0.05$，因此我們接受 H_0，即此商品每月的銷售量是隨機出現。

4.實例探討

4.1 數學試題難易度變項

以 100 學年度學科能力測驗數學科試題難易度爲例。傳統試題的難易指標皆以學生的答對率或通過率做基準，難道答對率或通過率的高低就可以全然代表試題的難易嗎？畢竟每個人對試題的難易感覺程度不一定相同，這似乎不是能用單一的答對率或通過率呈現出來的。

影響數學試題難易的因素有許多，如試題所評量的數學內容與測驗目標、解題時的思考策略與步驟數、試題的文字敘述或附圖等。其相關研究大致可分爲三類，第一類爲分析試題難易層次；第二類是訂定數學科測驗目標；第三類則探討測驗目標與解題步驟兩者的關係。其中李白飛、林福來、林光賢(1991，1992)所提出影響數學試題難易的三種主要因素，最廣爲眾人採用。三位學者提出：影響數學試題難易的三種主要因素爲：步驟數、思考策略以及試題所評量的數學內容，其中步驟數是指解題時所涉及的概念個數與概念的難易。思考策略是指解題時的思考方法，如固定模式的解法、一題多種的解法，或是需關鍵的特殊解法等。數學內容是指高中數學的學科內容，即現行高中教材六冊的內容。本論文針對此三種因素進行二維權重及難易度模糊問卷調查。

4.2 抽樣調查與問卷設計

本論文採用四類樣本群，分別是高中學生、一般民眾、高中教師及國中教師。因爲學生在升高中時經過了基測篩選及志願選填，故在樣本取得上難以隨機抽取。本論文以新北市海山高中的高三某班 39 位學生作爲學生樣本群，同時隨機抽取海山高中的 10 位高中數學教師作爲高中教師群。國中教師群則抽取台北市立龍山國中 5 位數學教師及高雄市立立德國中 6 位數學教師作爲國中教師群。一般民眾則隨機採樣 12 位做爲樣本。

高中學生與一般民眾不易判別數學難易變項，因此在問卷設計上採用單一問項方式進行，將各試題的難易度分爲 9 等級(1 最易，9 最難)，請學生與民眾依

各題的難易感受程度分別填入 1 到 9 的數字，如附件 1。而一般的中學數學教師大都可針對試題的權重及難易程度進行判斷，故問卷上以此二因子作爲主要問項。但是考慮大部分教師少有接觸模糊問卷的經驗，可能不易回答，故在設計上只在權重部分採用模糊問答，即請教師在權重部分以 0 到 1 之間的數字，給予一隸屬度，且每題隸屬度總和爲 1；難易度部分仍採單一問項方式進行，將各試題的難易度分 9 等級(1 最易，9 最難)，請教師依各題的難易感受程度分別填入 1 到 9 的數字，問卷設計如附件 2。

4.3 各樣本群試題難易指標

針對問卷結果，爲了研究的統一性及方便性，再將問卷資料的難易度 9 等級轉換成 0 到 1 區間的數值。本文爲了求取各樣本群中難易指標的代表值，所以在分析學生樣本群及一般民眾樣本群的資料時，先將各題統計出標準差、平均數，且取正負 2 個標準差(約 95%)之內的樣本再進行平均以作爲此題難易指標，其統計結果如下表 4.1 及表 4.2。

表 4.1 高中學生群各題標準差、平均數及難易指標

題號	**1**	2	3	4	5	6	7	8	9	**10**	11	12	13	A	B	C	D	E	F	G
標準差	0.164	0.276	0.301	0.362	0.212	0.245	0.262	0.252	0.282	0.272	0.271	0.246	0.258	0.289	0.319	0.259	0.242	0.303	0.31	0.291
平均	0.093	0.231	0.516	0.542	0.231	0.218	0.49	0.5	0.397	0.599	0.474	0.439	0.465	0.282	0.343	0.269	0.24	0.442	0.455	0.397
難易指標	**0.064**	0.211	0.516	0.542	0.217	0.189	0.49	0.5	0.397	**0.632**	0.474	0.439	0.451	0.247	0.307	0.25	0.213	0.442	0.455	0.382

表 4.2 一般民眾群各題標準差、平均及難易指標

題號	**1**	2	3	4	5	6	7	8	9	10	11	12	13	A	B	C	D	E	**F**	G
標準差	0.235	0.193	0.217	0.178	0.161	0.283	0.135	0.185	0.152	0.078	0.229	0.241	0.151	0.178	0.191	0.269	0.249	0.258	0.146	0.216
平均	0.365	0.406	0.563	0.531	0.531	0.531	0.698	0.625	0.594	0.781	0.615	0.677	0.625	0.531	0.604	0.552	0.521	0.635	0.802	0.677
難易指標	**0.365**	0.406	0.523	0.531	0.531	0.531	0.67	0.659	0.625	0.795	0.659	0.727	0.625	0.531	0.604	0.602	0.568	0.635	**0.83**	0.727

由表 4.1 與表 4.2 中可得知，高中學生群在第 10 題難易指標最高(0.632)，第 1 題難易指標最低(0.064)；一般民眾群在第 F 題難易指標最高(0.83)，第 1 題難易指標最低(0.365)。

在高中教師群的統計上，爲去除群組內的異端值，首先將各題的樣本權重程度計算出絕對距離後再加總，由加總值統計出Q_3值，再去除大於Q_3值之資料，最後平均已去除異端值後的資料，作爲本題的難易指標。如表 4.3 及表 4.4 分別爲第一題的模糊問卷原始樣本資料及權重模糊絕對距離。由表 4.3 及 4.4 中可得知，去除異端值之後的樣本有 2、3、4、5、6、7、10。依定義計算出此題難易指標 $DT = 0.375^{0.47} \times 0.214^{0.31} \times 0.125^{0.21} = 0.249$。

表 4.3 第一題模糊樣本原始資料及取樣表

第一題	數學認知		思考策略		解題步驟	
	權重	難易	權重	難易	權重	難易
1	0.3	0.5	0.5	0.375	0.2	0.25
2	0.4	0.5	0.4	0.375	0.2	0.25
3	0.4	0.375	0.4	0.25	0.2	0.125
4	0.4	0.5	0.3	0.25	0.3	0.125
5	0.4	0.375	0.4	0.25	0.2	0.125
6	0.6	0.375	0.2	0.125	0.2	0
7	0.6	0.375	0.2	0.125	0.2	0.125
8	0.8	0	0.1	0	0.1	0
9	0.7	0	0.2	0	0.1	0.125
10	0.5	0.125	0.3	0.125	0.2	0.125
取樣平均	0.47	0.375	0.31	0.214	0.21	0.125

表 4.4 第一題權重模糊絕對距離

	1	2	3	4	5	6	7	8	9	10
1	0	0.15	0.15	0.4	0.15	0.45	0.45	0.8	0.65	0.3
2	0.15	0	0	0.25	0	0.3	0.3	0.65	0.5	0.15
3	0.15	0	0	0.25	0	0.3	0.3	0.65	0.5	0.15
4	0.4	0.25	0.25	0	0.25	0.35	0.35	0.7	0.55	0.2
5	0.15	0	0	0.25	0	0.3	0.3	0.65	0.5	0.15
6	0.45	0.3	0.3	0.35	0.3	0	0	0.35	0.2	0.15
7	0.45	0.3	0.3	0.35	0.3	0	0	0.35	0.2	0.15
8	0.8	0.65	0.65	0.7	0.65	0.35	0.35	0	0.15	0.5
9	0.65	0.5	0.5	0.55	0.5	0.2	0.2	0.15	0	0.35
10	0.3	0.15	0.15	0.2	0.15	0.15	0.15	0.5	0.35	0
總和	3.5	2.3	2.3	3.3	2.3	2.4	2.4	4.8	3.6	2.1
Q_3=3.45										

經此方式統計整份試題並統計各試題的難易指標如表 4.5。由表中得知在高中教師群的認知感受上，本份試題的第 10 題難易指標最高(0.474)，第 6 題難易指標最低(0.199)。

在國中教師群部分，亦採取與高中教師群一致的統計方法，所得國中教師群在各試題的難易指標結果如表 4.6。由表 4.6 得知，國中教師群在第 G 題難易指標最高(0.557)，第 1 題難易指標(0.201)最低。由表 4.5 與表 4.6 相比較，發現國中教師群與高中教師群的認知感受上，最難與最易的題目全然不同。

表 4.5 高中教師群在各試題的難易指標

取樣平均 / 題號	數學認知 C_1		思考策略 C_2		解題步驟 C_3		難易指標
	權重 x_1	難易 y_1	權重 x_2	難易 y_2	權重 x_3	難易 y_3	$DT = y_1^{x_1} y_2^{x_2} y_3^{x_3}$
1	0.47	0.375	0.31	0.214	0.21	0.125	0.249
2	0.33	0.281	0.36	0.344	0.31	0.156	0.238
3	0.35	0.469	0.39	0.469	0.26	0.438	0.388
4	0.44	0.422	0.34	0.391	0.23	0.328	0.323
5	0.4	0.321	0.36	0.321	0.24	0.357	0.304
6	0.475	0.172	0.2625	0.141	0.275	0.156	**0.199**
7	0.39	0.464	0.41	0.446	0.2	0.411	0.358
8	0.4	0.411	0.29	0.357	0.31	0.429	0.386
9	0.41	0.339	0.31	0.393	0.27	0.429	0.342
10	0.36	0.482	0.3	0.536	0.34	0.571	**0.474**
11	0.33	0.328	0.35	0.406	0.33	0.391	0.347
12	0.413	0.438	0.31	0.328	0.25	0.469	0.391
13	0.33	0.469	0.38	0.469	0.3	0.453	0.406
A	0.43	0.393	0.343	0.268	0.23	0.25	0.294
B	0.31	0.429	0.46	0.482	0.23	0.339	0.326
C	0.33	0.393	0.46	0.429	0.21	0.304	0.3
D	0.35	0.458	0.35	0.604	0.30	0.438	0.387
E	0.38	0.528	0.34	0.514	0.28	0.417	0.398
F	0.4	0.482	0.34	0.429	0.26	0.304	0.335
G	0.29	0.554	0.43	0.554	0.26	0.518	0.455

表 4.6 國中教師群在各試題的難易指標

題號 \ 取樣平均	數學認知 C_1		思考策略 C_2		解題步驟 C_3		難易指標
	權重 x_1	難易 y_1	權重 x_2	難易 y_2	權重 x_3	難易 y_3	$DT = y_1^{x_1} y_2^{x_2} y_3^{x_3}$
1	0.54	0.234	0.26	0.234	0.2	0.109	**0.201**
2	0.35	0.297	0.23	0.344	0.43	0.391	0.345
3	0.39	0.563	0.31	0.531	0.30	0.531	0.543
4	0.45	0.484	0.26	0.422	0.29	0.422	0.449
5	0.38	0.406	0.29	0.406	0.34	0.359	0.39
6	0.34	0.297	0.40	0.5	0.26	0.359	0.385
7	0.475	0.578	0.25	0.453	0.28	0.578	0.544
8	0.46	0.531	0.25	0.469	0.29	0.531	0.515
9	0.40	0.438	0.29	0.422	0.31	0.438	0.433
10	0.43	0.578	0.25	0.531	0.33	0.516	0.545
11	0.31	0.516	0.36	0.469	0.33	0.453	0.478
12	0.38	0.578	0.31	0.484	0.31	0.563	0.542
13	0.39	0.5	0.38	0.516	0.28	0.531	0.501
A	0.36	0.359	0.24	0.359	0.40	0.375	0.366
B	0.36	0.469	0.30	0.484	0.34	0.469	0.473
C	0.33	0.359	0.38	0.438	0.30	0.375	0.392
D	0.31	0.438	0.35	0.516	0.34	0.453	0.469
E	0.29	0.453	0.36	0.563	0.35	0.531	0.518
F	0.40	0.5	0.28	0.531	0.33	0.547	0.523
G	0.39	0.531	0.35	0.594	0.26	0.547	**0.557**

4.4 不同樣本群對試題難易度分析

一般而言，一個試題難易可分爲 5 等級，分別是極易、偏易、適中、偏難、很難。本論文將依各題難易(DT)分至此 5 等級，若 $0 \le DT \le 0.2$，則此題的難易度屬「極易」；$0.2 < DT \le 0.4$，則此題的難易度屬「偏易」；$0.4 < DT \le 0.6$，則此題的難易度屬「適中」，$0.6 < DT \le 0.8$，則此題的難易度屬「偏難」；$0.8 < DT \le 1$，則此題的難易度屬「很難」。以此劃分各試題在各樣本群中難易等級，結果如下表 4.7。

表 4.7 各試題在各樣本群中的難易等級

	高中學生 S	一般民眾 AP	高中老師 T_1	國中老師 T_2
極易	**1、6**		**6**	
偏易	**2、5、9、A、B、C、D、G**	**1**	**1、2、4、5、9、11、A、C**	**1、2、5、6、A、C**
適中	**3、4、7、8、11、12、13、E、F**	**2、3、4、5、6、A、D**	**3、7、8、10、12、13、B、D、F、G**	**3、4、7、8、9、10、11、12、13、B、D、E、F、G**
偏難	**10**	**7、8、9、10、11、12、13、B、C、E、G**		
很難		**F**		

由 4.7 表中明顯看見，一般民眾(AP)在試題難易度的整體認知上，試題難易度多分布於偏難，總數 20 題的考題中，就高達 11 題。而另外三個樣本群對試題難易度的整體認知上，試題難易度多分布於適中。

由於各群組在試題難易題數分布狀況不相同，本論文利用克洛斯-瓦力士檢定，檢驗此四個樣本群在試題難易認知上是否有相同。

統計假設爲

H_0 ：四個樣本群在試題難易度看法彼此無差異

H_1 ：四個樣本群在試題難易度看法彼此有差異

Kruskal-Wallis Test: ALL versus NO

Kruskal-Wallis Test on ALL

NO	N	Median	Ave Rank	Z
1	20	0.4180	28.0	-2.78
2	20	0.6145	64.2	5.27
3	20	0.4055	28.7	-2.62
4	20	0.4755	41.0	0.12
Overall	80		40.5	

H = 31.78 DF = 3 P = 0.000

H = 31.79 DF = 3 P = 0.000 (adjusted for ties)

經檢定結果 P=0.000<0.05，因此我們拒絕 H_0 。換句話說，此四個樣本群中至少有一個樣本群跟其他樣本群在大學學測數學科試題難易度的看法上不一致。爲探討各樣本群之間在難易度的看法是否一致，本論文再利用威克生符號等級檢定檢視各群組彼此之間的試題難易度看法。

統計假設爲

H_0 ：兩樣本群在試題難易度看法彼此無差異

H_1 ：兩樣本群在試題難易度看法彼此有差異

經檢驗，結果如表 4.8。

表 4.8 四樣本群在威克生符號等級檢定結果

*	高中學生(S)	一般民眾 (AP)	高中教師(T_1)	國中教師(T_2)
高中學生(S)	*	拒絕 H_0 P=0.000<0.05	接受 H_0 P=0.444>0.05	拒絕 H_0 P=0.001<0.05
一般民眾 (AP)	拒絕 H_0 P=0.000<0.05	*	拒絕 H_0 P=0.000<0.05	拒絕 H_0 P=0.000<0.05
高中教師(T_1)	接受 H_0 P=0.444>0.05	拒絕 H_0 P=0.000<0.05	*	拒絕 H_0 P=0.001<0.05
國中教師(T_2)	拒絕 H_0 P=0.001<0.05	拒絕 H_0 P=0.000<0.05	拒絕 H_0 P=0.001<0.05	*

由表 4.8 得知，只有高中學生(S)與高中教師(T_1)在試題難易度感受程度較一致，而其他樣本群的看法上彼此不一致。同時我們也發現，高中教師(T_1)國中教師(T_2)雖皆屬中學教師，但試題難易感受程度並不一致。

在國中基本學力測驗(簡稱基測)中，數學科試題難易的分布經過特殊規畫，一般常以前面試題偏易，後面試題偏難的方式進行試題安排。而大學學力基本測驗的試題難易分布是否也以此模式安排？傳統以答對率或得分率(p)的高低來判別試題的難易，答對率或得分率高代表試題較易；反之較難。本論文應用連串法隨機性檢定，檢驗此四個樣本群是否覺得試題在難易分布上亦是隨機出現，另外由大考中心所提供本學度(100 學年度) 的數學科各試題答對率，將答對率轉成難易指標($DT = 1 - P$)，在 $\alpha = 0.05$ 顯著水準下，檢定試題難易度是否亦是隨機出現。

統計假設爲

H_0：試題難易感受程度分布是隨機出現

H_1：試題難易感受程度分布不是隨機出現

表 4.9 五樣本群的隨機檢定結果

隨機檢定 / 群組	中位數	r	P-value	檢定結果
高中學生	0.418	9	0.358>0.05	接受 H_0
一般民眾	0.615	4	0.001<0.05	拒絕 H_0
高中教師	0.406	12	0.646>0.05	接受 H_0
國中教師	0.476	8	0.168>0.05	接受 H_0
大考中心	0.535	10	0.646>0.05	接受 H_0

由表 4.9 發現，只有一般民眾在對試題難易感受程度分布屬不隨機出現，其他群組皆認爲是隨機出現。因此我們推論，一般民眾的觀念可能受到國中基測影響，認知上會覺得試題難易安排是由簡易到困難。但在其他群組及大考中心的資料，卻呈現試題難易安排是隨機出現的。

5. 結論

在社會科學調查中，最想被了解的部分是人類的思維運作。以往科學研究希望用簡易的方式探討複雜的事物，因此常用二元邏輯分析多變的思維。但人類的思維是模糊的，單用二元邏輯方式進行分析會造成一定程度的失真。本文提出二維模糊數將單一維模糊數擴展出去，由此可將人類思維得以更深入且完善的方式表達出來。另外本論文所應用單因子離散型模糊樣本之異端值，可將樣本群之統計分析更具代表性，在未來統計分析上能有更佳的數據表現。

由本論文實例中，我們發現高中學生群與高中教師群在對試題難易感受程度上無差異，但其他群組在難易度感受程度上彼此是有差異的。在教學上，高中老師對高中教材的熟悉度相對比其他群組高，因此高中老師在判別試題的難易度時，會以近似學生的角度來思考。而高中教師在指導學生時，常會因個人對試題難易感受而間接影響了學生對試題難易度的看法；高中教師在教導試題的過程中，若以個人感覺引領學生此題難易時，往後學生可能會推類此型試題屬難或屬易。如此造成學生在作答時的難易感覺受到影響。

一般而言，同屬中學數學科教師，在同樣試題下的難易感受程度應差異不大。但本論文發現不盡如此，在難易感受程度上，高中教師與國中教師並無顯著的一致性，推估其原因，在教學上，高中教材的熟悉度以高中教師較爲了解，國中教師對高中教材接觸上可能偏少，因此造成難易感受上有所差異。同理，我們也發現一般民眾對試題難易感受程度上有著更大的差異性。

在本論文中，我們發現除了一般民眾認爲試題難易非隨機性，其他群組認知上都屬隨機性。而在大考中心的答對率中也發現，大學學力測驗數學科試題難易安排也屬隨機性。推估原因，一般民眾的思維上，認爲國中基測試題難易的安排是由簡易漸近到困難，同樣在大學學力測驗上也應如此安排，所以在認知上覺得試題難易是有跡可循。在教學現場上，教師需引導學生了解大學學力測驗在試題難易分佈是採隨機性，所以在應答整份試題時，最好先流覽過各題，再依個人難

易感受由易到簡逐題進行作答，如此不會因前面偏難的試題而影響作答的心情。進而增加得分率。

由於本論文之參與人數並非龐大，未來的研究可針對更大量的學生與教師做調查。另一方面，未來的研究可針對學生答試題結果與其難易感受程度作更深入的探討。如此當我們身臨教學現場時，可以啟發更多學生的學習效果。

6. 參考文獻

[1] 吳柏林(2005)，模糊統計導論：方法與應用，臺北：五南。

[2] 吳柏林(2002)，現代統計學，臺北：全程。

[3] 吳柏林、曾能芳(1998)，模糊回歸參數估計及在景氣對策信號之分析應用，中國統計學報，36(4)，399-420。

[4] 吳柏林、楊文山(1997)，模糊統計在社會調查分析的應用，社會科學計量方法發展與應用，楊文山主編：中央研究院中山人文社會科學研究所，289-316。

[5] 李白飛、林福來、林光賢(1994)。大學入學考試數學科試題分析與命題研究(三)。大學入學考試中心。

[6] 黃仁德、吳柏林(1995)，臺灣短期貨幣需求函數穩定性的檢定，模糊時間數列方法之應用，臺灣經濟學會年會論文集，169-190。

[7] Dubois, D. and Prade, H. (1991), Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions, Fuzzy Sets and Systems, 40, 143-202.

[8] Lowen, R. (1990), A fuzzy language interpolation theorem. Fuzzy Sets and Systems, 34, 33-38.

[9] Ruspini, E. (1991), Approximate Reasoning: past, present, future. Information Sciences, 57, 297-317.

[10] Tseng, T. and Klein, C. (1992), A new Algorithm for fuzzy multicriteria decision making. International Journal of Approximate Reasoning, 6, 45-66.

[11] Zadeh, L. A. (1965), Fuzzy Sets. Information and Control, 8, 338-353.

[12] Zimmermann, H. J. (1991), Fuzzy Sets Theory and Its Applications. Boston: Kluwer Academic.