3、模糊無母數檢定與應用

3.1 模糊排序法應用於 Kruskal-Wallis 檢定

Wilcoxon 兩樣本檢定法是利用兩個獨立模本來比較兩個母體是否一致，後來由 Kruskal 與 Wallis 加以推廣，是用來檢定多組隨機樣本之母體分配是否相同的問題，它是一種最常被用來檢定多個獨立母體間，是否有差異的方法。

假設我們有 k 個獨立母體，而每個獨立母體之模糊觀察值分別有 n_i 個，$i=1,....,k$ 。將所有 $n_1+n_2+n_3+.....+n_k=N$ 個樣本混合，依據前面描述模糊樣本排序方法，最小的觀測值給 1，次小者給 2，依此類推，最大的模糊觀測值給 N ；若有數個觀察值大小相同時，則取其在該順序下所對應的秩平均值。

接下來，分別將每組模糊樣本所對應到的等級做總和，令 R_i=表示第 i 個模糊母體在混合樣本中所對應到的等級總和。此時若假設 H_o ：各模糊母體分配無差異，則若各組的模糊母體分配一致時，R_i 的值應該不會差異太大，但若 R_i 有過大或過小的情形時，則我們會拒絕 H_o。

Fuzzy Kruskal-Wallis 檢定過程：

1 資料：k 個相互獨立模糊母體，每個母體有 n_i 個模糊觀測值，$i=1,....,k$ ，

$n_1+n_2+n_3+.....+n_k=N$ ，根據模糊排序規則加以排序。

2.假設： H_o ：各模糊母體分配無差異。

3.統計量： $H=\dfrac{12}{N(N+1)}\sum_{i=1}^{k}\dfrac{1}{n_i}\left[R_i-\dfrac{n_i(N+1)}{2}\right]^2$

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

，其中R_i的期望值爲$\frac{n_i(N+1)}{2}$。

4.決策：(1)當k=3，n_1，n_2，$n_3 \le 5$，查 Kruskal-Wallis 檢定附表，決定機率 p，若$p < \alpha$，則拒絕H_o。

(2)若$n_i > 5$時，H分配接近χ^2分配，在α顯著水準下，自由度 $v = k-1$，如果$H > \chi^2_{1-\alpha}(k-1)$，則拒絕$H_o$。

例 3.1：離散型模糊樣本

由於高速公路屬於高速行駛的狀態，一有交通事故，往往是較嚴重的情形，而且波及到的事故車輛也相對較密集，因此，高速公路交通安全是值得重視的地方。94 年十二月期間，高速公路內線車道發生交通事故，緣由來自於對向車道的車車速過快，失控撞護欄並衝撞到內線車道之轎車，致使二輛轎車駕駛身受重傷，車體亦嚴重變形。基於此，對於高速公路三線道：內線、中線及外線車道之失事狀況，引起高度的興趣。

若針對 94 年一整年來看，將之分成四等份，也就是以三個月爲一個計算單位，計算三個月內平時白天、平時晚上、假日白天及假日晚上各線道失事之模糊數，表 3.1 提供了 94 年高速公路三線道交通事故發生之反模糊化值。在α=0.05 顯著水準下，欲判斷這三線車道之交通事故發生是否有不同。

表 3.1 高速公路三線道交通事故發生之反模糊化值

	1 ~ 3 月	4 ~ 6 月	7 ~ 9 月	10 ~12 月
內線車道	31	29	27	32
中線車道	26	25.5	27	24.5
外線車道	25	27	28	26.5

統計假設爲

H_o：高速公路三線道交通事故分配相同

H_1：至少有一線道交通事故分配不同

表 3.2 排序後的模糊資料

內線車道	11	10	7	12	R_1=40
中線車道	4	3	7	1	R_2=15
外線車道	2	7	9	5	R_3=23

$$H=\frac{12}{12(12+1)}\left[\frac{196}{4}+\frac{121}{4}+\frac{9}{4}\right]=6.27$$

在 α=0.05 顯著水準下，當 $n_1=n_2=n_3=4$，$k=3$ 時，查附表得到當 $H=7.53$ 時，$p=0.001$，當 $H=5.69$ 時，$p=0.049$，而本例之 $H=6.27$，故可得到 $0.001<p<0.049$，則 $p<\alpha$。

根據 Kruskal-Wallis 的檢定方法，所以我們拒絕 H_o；即高速公路三線道之交通事故失事率分配不同。

例 3.2：連續型模糊樣本

由於現今的潮流及趨勢，女性的社會地位，已不再像以往傳統的女性。由於教育的普及與知識流動的便捷，晚婚的人口比例逐年增高，適婚年齡也隨著環境及女性個人意識抬頭，而有不同的見解。於是針對亞洲國家中，台灣、泰國及大陸廣東省做研究，研究六年級女性認爲自己的適婚年紀是大約幾歲到幾歲，以了解同樣身爲東方女性，對於適婚年紀的看法，是否會因生長環境及背景的不同而有所差異。

我們隨機抽樣三個地區各 10 組樣本，得到三個地區共三十位女生，她們心目中理想適婚年紀的範圍，並記錄下來。在顯著水準 α=0.05 下，我們想要了解三個地區之六年級女生，對於適婚年紀的看法是否相同。

表 3.3 三個國家之六年級女生連續型模糊資料

台灣	28-30	30-35	26-30	26-28	26-30	33-35	24-28	31-33	28-32	26-32
泰國	28-32	21-25	22-24	19-25	26-30	18-24	22-26	20-24	26-28	23-27
廣東	26-30	28-32	26-28	23-27	26-30	25-33	27-29	28-32	30-34	23-27

統計假設為

H_o：三個國家六年級女生對於適婚年紀看法一致。

H_1：至少有一個國家六年級女生對於適婚年紀看法不同。

表 3.4 三個國家混合排序後的模糊資料

台灣	20	29	17	11.5	17	30	10	27	24.5	21	R_1=207
泰國	24.5	5	4	3	17	1	6	2	13	8	R_2=83.5
廣東	17	24.5	11.5	8	17	22	14	24.5	28	8	R_3=174.5

統計量： $H = \frac{12}{30(30+1)}\left[\frac{207^2}{10} + \frac{83.5^2}{10} + \frac{174.5^2}{10}\right] - 3(30+1) = 10.576$

因為 $\chi^2_{0.05}(2) = 5.99 < 10.576$ ，所以我們拒絕 H_o。

也就是根據上述的檢驗方法我們得知，雖然同樣身為東方女性，但在婚姻的理念上，仍對適婚年紀的看法有所不同。

3.2 模糊排序法應用於 run test 檢定

在做抽樣調查時，常會討論到樣本是否隨機，因此隨機性檢定是相當重要的議題。若欲知一組模糊樣本是否由隨機取得，可用模糊連串法檢定。所謂連串是指由一串相同的符號或相同的性質所組成，直到下一個不同符號出現為止，則連

串的長度是根據有幾個相同符號的數量而定。例如，觀察近一個星期加拿大氣溫之高低，高於 0 度為＋，低於 0 度為－，則得到如下資料：＋＋　－　＋　－－－　。則上例觀察值共有四個連串，第一個連串長度為 2，第二個連串長度為 1，第三個連串長度為 1，第四個連串長度為 3。

如何得知此模糊樣本是否由隨機抽取，直觀來說，若一組模糊樣本的連串數目太多或太少時，則其隨機性就值得我們去懷疑。就上述例題而言，若所得資料如下：－－－－　＋＋＋　，則其僅有兩個連串。若所得資料為：＋　－　＋　－　＋　－　＋　，則此模糊樣本共有七個連串。如此有規則的排列方式，此模糊樣本應該不具有隨機性。

連串法檢定過程：

1. 資料：將 n 筆模糊資料按照順序記錄，將 n 筆資料分為兩類。令 n_1 為第一類的總數，n_2 為第二類的總數。
2. 假設：H_o：此模糊資料具有隨機性。
   H_1：此模糊資料不具有隨機性。
3. 統計量：r 表示連串的數目。
4. 決策：在顯著水準 α =0.05 下利用雙尾檢定，查表得 r_L 和 r_U ，如果 $r_L \le r \le r_U$ 則接受 H_o。

例 3.3：離散型模糊樣本

對於近日政局較不穩定，股市上也因此受到影響，而股市投資人最在意的往往是股市的漲跌狀況。於是中央開始調查這段時間股市的波動是否是隨機性的，相信這是所有投資人都相當關心的話題。

針對股市的漲跌情況，若全以收盤的價格當作當天的參考，並不那麼恰當，若以當天的漲跌情況來看，會比較能了解整個股盤的走勢。於是，觀察五月份最後十天有交易的日子，記錄每日的波動情形，即記錄每日從開盤到收盤過程中的漲、跌或平盤狀況，然後將每日波動狀況資料依漲、跌、平盤作為三個語言變數。

則在α=0.05 顯著水準下，欲了解股市波動情形，表 3.5 提供每日關於三種波動情形之隸屬度數據，可得資料如下：

表 3.5　十筆資料對於三種股市波動情形之隸屬度

	漲	跌	平盤	反模糊化值
1	0.4	0.5	0.1	-0.1
2	0.6	0.2	0.2	0.4
3	0.2	0.3	0.5	-0.1
4	0.2	0.7	0.1	-0.5
5	0.5	0.2	0.3	0.3
6	0.4	0.2	0.4	0.2
7.	0.1	0.7	0.2	-0.6
8	0.3	0.4	0.3	-0.1
9	0.7	0.2	0.1	0.5
10	0.8	0..2	0.2	0.6

統計假設爲：

H_o：股市的漲跌波動走勢是隨機性的。

H_1：股市的漲跌波動走勢是非隨機性的。

根據樣本的反模糊化值，可得反模糊化值之中位數爲 1.75。令大於 1.75 爲＋，小於 1.75 爲－，則可得：

$$\underline{-\,-}\ \ \underline{+\,+\,+\,+}\ \ -\,+\,\underline{-\,-}$$

由上可知$n_1 = n_2 = 5$，$r = 5$。查附錄表可得$r_L = 2$，$r_U = 10$，則$r_L \le r \le r_U$，即可得知近日股市漲跌情形是一個隨機過程。

例3.4　連續型的模糊樣本例子

近日正巧蹚上梅雨季節，陸續在各地傳出豪雨所造成的災情，氣溫的變化也相當地大，有時早上下雨氣溫較低，但到了晚上反而天氣變晴朗，這樣的氣候使

得氣溫變化落差很大，不少人因此而不適應而感冒，因此開始關心近日來的氣溫變化情形。

分別記錄 5/29 ~ 6/11 共計兩週每日的氣溫範圍，資料如下表：

表 3.6　5/29 ~ 6/11 每日氣溫狀況

	氣　溫	樣本中點		氣　溫	樣本中點
5/29	18~28	23	6/5	18~25	21.5
5/30	19~28	23.5	6/6	18~24	21
5/31	19~25	22	6/7	17~23	20
6/1	20~30	25	6/8	19~27	23
6/2	21~29	25	6/9	18~25	21.5
6/3	20~28	24	6/10	17~25	21
6/4	18~24	21	6/11	22~30	26

統計假設為：

H_o：氣溫的變動是隨機現象。

H_1：氣溫的變動不是隨機現象。

根據模糊區間樣本的中點值，求出樣本中點值的中位數為 22.5。令大於 22.5 的氣溫值為＋，小於 22.5 氣溫值為－，則可得：

$$\underline{++}-\underline{+++}\ \underline{----}+\underline{--}+$$

由上可知$n_1=n_2=7$，$r=7$。查附錄表可得$r_L=3$，$r_U=13$，則$r_L \le r \le r_U$，即可得知近日氣溫的變化情形是一個隨機過程。