

Delphi分析發展、 應用問題與改進(下)



葉連祺/國立嘉義大學教育系教育行政與政策發展碩士班教授

摘要

Delphi分析廣受教育研究採用，屢見創新分析實務做法，但存在一些應用問題影響其分析效益。對此，本研究說明有關Delphi分析理念、分析取向、運作流程、應用實務等發展成果，進而討論涉及專業評估不足、分析嚴謹性偏弱、新法應用落後、採用係數簡略、運作生態薄弱等八項問題，並提出強化學理和方法研究、建立品管機制、結合多元統計、設計專用程式等改善做法。另設計適用傳統Delphi分析的程式適合SPSS作業環境，其操作簡易，提供豐富分析資訊，能大幅降低分析成本。

關鍵字：德懷術、共識評估、變異係數

投稿日期：2026年1月5日

完成修稿日期：2026年1月28日

DOI: 10.6423/HHHC.202605_(163).0007



Development, Application Problems and Improvement of Delphi Analysis (II)

LAIN-CHYI YEH¹

Abstract

Delphi method are used prevail among education study. Various innovative analytic practices about Delphi analysis have been proposed, but on the other hand several application problems also impacted its benefits. The study illustrated analytic ideas, analysis approaches, operational procedures, and practices of application on Delphi method. Moreover using Delphi method encountered with eight problems that related with insufficient professional assessment, weak rigidness in analysis, lag of adopting new methods, using oversimple consensus measurement coefficient, and frail operational ecology. Therefore, the study proposed some resolution strategies: enforce research on theory and method for Delphi analysis, build quality control mechanism, use various statistical methods, and design specific programs. Finally several specific programs for classic Delphi analysis worked on SPSS platform that conduct easy and rapidly, provide rich analytic information, as well as reduce most cost of analysis.

Keywords: Delphi method, consensus assessment, coefficient of variation, CV

¹ Master's Program of Educational Administration and Policy Development, Department of Education, National Chiayi University

二、研議完善全面檢視Delphi分析各階段運作品質的方法和標準

1. 釐清 Delphi 分析階段工作品質重點

要確保 Delphi 分析品質，需要從全面檢視角度實施品保措施，觀察 RAM 法和 ROMPER 法採行 GRADE 法和 ACCOR 規範去加強確保執行品質，前述也提及一些論者強調必須檢視其信度和效度，因此全面檢視運作品質殆無疑義。參考 Delphi 分析流程和綜觀研究論述，能夠分成實施前、實施時和實施後三個階段品質管控，並細分七項工作，見表 30 說明。其中 1. 評估團隊組成工作涉及評估者專業品質，已被關注但是仍被普及採用，2. 檢視評估項目工作涉及評估工具品質，也被提及但仍未受到普遍重視，而 4. 檢視評估共識工作涉及 Delphi 分析品質，是最核心的關鍵工作，討論甚多、提出做法甚多，但是總體看來多元紛雜，由前述討論就可得知，有待進一步整合。還有此表揭示出邀請諮詢者和評估者參與品質工作的可行性和參與時機，其包括評估項目信度和效度、檢視回饋意見和後續處理做法的看法、以及參與評估分析報告和最終方案的適宜性等，觀諸既有研究文獻多半缺少這方面的作為，頗值得思考是否納入。

表 30
 Delphi 分析各階段品質管檢視焦點和工作項目

階段	工作項目	工作焦點	工作項目	參與者 ^a
實施前	1. 評估團隊組成	查核專家條件和認知	綜合調查專家條件和專家自評	主持者/諮詢者
	2. 檢視評估項目	查核評估項目品質	檢視評估項目的信度(內部一致性),項目的效度(內容效度,構念效度)	評估者
	3. 檢視共識標準	提出共識評估標準	檢視共識(異議)評估標準	主持者/諮詢者
實施時	4. 檢視評估共識	1. 查核單項共識性	1. 檢視單項評估共識性(一致性)	評估者
		2. 查核多項共識性	2. 檢視多項評估共識性(一致性)	評估者
		3. 提出項目適宜性回饋	3. 檢視項目適宜性並回饋	主持者
		4. 整合項目評估結果	4. 彙整共識性評估和其他回饋看法	主持者
	5. 提出項目調整(增項,刪項,調項,修項敘述)和評定做法調整、處理異議做法	5. 彙整項目調整、續評估和回饋、異議處理做法	主持者	
6. 查核項目評估跨樣本變化和跨輪回變化	6. 檢視評估結果跨樣本和跨輪回一致性	主持者		
5. 確認評估終止	1. 確認單多項共識達成 2. 確認評估結果可用 ^b	檢視項目共識性評估結果和後續做法合理性	主持者/評估者/諮詢者	
實施後	6. 報告分析結果	提出分析報告	檢視分析報告	主持者/諮詢者
	7. 提出最終方案 ^b	提出最終方案	檢視最終方案	主持者/諮詢者/評估者

註：^a“主持者/諮詢者”指主要由主持 Delphi 分析案者參與，亦可邀請該案的諮詢者協同參與處理；“主持者/評估者/諮詢者”指主持者主導，或邀請評估者、諮詢者、評估者和諮詢者協同參與處理。^b該項工作視需要而定，可執行或省略。

2. 建立檢視 Delphi 分析共識性和信、效度的合理決策流程和規範

前已敘及檢視共識性是 Delphi 分析的重點工作，而評估共識性與評估一致性 (agreement)、關聯性 (association, correlation)、信度 (reliability)、效度 (validity) 有關，因此可見不少檢視信度和效度的統計方法被應用於 Delphi 分析。針對前述表 所提 Delphi 分析各階段品管檢視焦點和工作項目，可思考確保各階段工作品質的流程見圖 11 所示，主要確保評者素質、評值品質、評項 (工具) 品質、評估品質和成果品質，其也揭示可進行的各項分析及其大致可用資料分析做法，而標示*為當前未被普遍採用，**為分析方法尚待深入討論，因為當前可用做法尚少，未來值得深究。基本上，進行 Delphi 分析應該實施圖未被特別標示的確保品質工作，而標示*分析工作宜盡量採用，至於各品管工作的規範可參考 Khodyakov 等 (2023) 提出的觀點。

圖 11
Delphi 分析確保品質工作流程



其次，共識性評估奠基於評估者依循專業自覺自主判斷，然而已見一些論述指出各輪回評估可能存在認知偏見、評值行為偏差等人為問題；也可能評者受次團體特性影響，產出不屬於專業判斷偏差的共識性評估差異現象，如不同性別者、職務者對於項目評估有明顯相異看法，即生態效度不佳，這是自然現象，係

實施前未注意，卻會衝擊後續提出整體評者共識性分析結果的合理性，亦即需要一併呈現整體和各子群體的評估結果，並視需要進行評者分群影響分析，提出客觀佐證資訊。

又上述流程揭露三個可思考課題，即在實施階段是否需要藉由調整項目、調整評者、分類評者做法，促進達成高共識性。就調整項目，包括增加項、刪除項、調動項、潤飾項等處理做法選擇；就調整評者，指選擇增加評者、排除評者、調整評者、選擇評者等可用處理做法，其中已見極少數研究採行，在各輪迴調查階段皆調整採用新一批評者；就分類評者，若是進行政策Delphi、爭論Delphi，則需要區別異議者，有效辨識異議者是有必要，以便進一步調查了解堅持異議的理由。

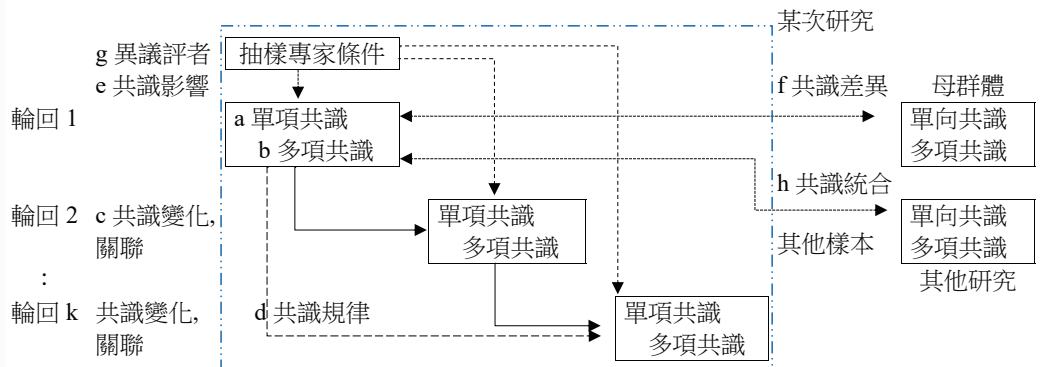
三、結合多元統計考驗，提升共識評估指數的分析品質和效益

針對目前Delphi分析頗多採用簡單描述統計方法，無法獲取豐富和多元的共識性評估資訊，改善策略之一就是嘗試結合應用多元化的統計考驗，提升分析品質，可行做法簡述如下：

1. 根據分析目的，思考和選用合適的統計分析共識性方法

關於Delphi分析可探討聚焦如圖12所示，觀察已知Delphi分析幾乎聚焦於單次研究的單輪回單項共識評估（圖12a部分），或再聚焦在單次研究的單輪回多項共識評估（圖12b部分），少數觸及分析單次研究的前後輪回單項共識評估差異變化或關聯檢視（圖12c部分）。再細思，若有單次多輪回的單項共識資料，則能進行共識規律變化的分析（圖12d部分），另可發現每輪回出現異議的評估者及其評估意見的影響（圖12g部分），尤其是進行政策Delphi分析時發現異議實屬重要。又Delphi分析是基於單次研究調查樣本專家看法，那麼檢視過去類似或相同研究的共識結果是否有差異（圖12f部分）、應用後設分析（meta-analysis）進行統合的看法是否存在差異（圖12h部分）？還有該次研究結果能否代表具類似條件或特性專家母群體的看法（圖12f部分）？都是可資探討的課題，因此這能構成一個較為完整和全面的Delphi分析共識研究體系，見圖12說明八項探討焦點。

圖 12
Delphi分析共識研究體系及探討焦點



進一步細緻化比較圖 12 的 Delphi 分析共識探討焦點內容，整理於表 31，必須注意該表不適用於如 fuzzy set 等區間型資料。該表透露各項探討焦點的分析對象、可分析事項及若干可用資料分析技術舉例，能供按圖索驥時參考。

表 31
Delphi 分析共識研究探討焦點之比較

探討焦點	分析對象	分析事項	資料分析技術舉例 ^a
a. 單項共識	單次研究, 單項, 專家樣本	1. 共識性程度 2. 共識性差異/關聯	1. 集中和離散量數(M, SD, CV), 評估指數(CDI)及分類, 內容效度考驗(CVI, CVR) 2. 差異考驗(t考驗), 關聯考驗(r)
b. 多項共識	單次研究, 多項, 專家樣本	1. 共識性程度 2. 一致性程度 3. 共識性群聚	1. 變異量數(MCV), 關聯考驗(kappa係數, Cochran Q) 2. 一致性考驗(kappa係數), 信度考驗(Cronbach α), 構念效度考驗(主成分分析, 因素分析) 3. 集群分析
c. 共識變化/關聯	單次研究, 單項, 多項, 專家樣本	1. 共識性變化 2. 共識性關聯	1. 差異考驗(t考驗), 差異比較(CV, MCV) 2. 關聯考驗(χ^2 考驗, QAP係數, 積差相關)
d. 共識規律	單次研究, 單項, 專家樣本	1. 共識性變化 2. 共識性關聯	1. 變化分析(McNemar考驗) 2. 關聯考驗(r)
e. 共識影響	單次研究, 單項, 多項, 專家樣本	1. 共識性差異 2. 對共識性影響	1. 差異考驗(t考驗, CHAID分析), CV _w 2. 影響考驗(多元迴歸), 關聯考驗(積差相關)
f. 共識差異	單次研究, 單項, 多項, 專家樣本和母群體	共識性差異	差異比較(CV, MCV), 差異考驗(t考驗)
g. 異議評者	單項研究, 專家樣本	1. 共識性程度 2. 異議者偵測	1. 關聯考驗(Cochran Q考驗) 2. outlier考驗, 集群分析
h. 共識統合	多項研究, 單項, 多項, 專家樣本	共識性整合	後設分析(lnCVR)、整合分析(CV _w)

註：^a以適用於Likert多點量尺值和其他非區間資料型態資料值（指fuzzy set等類資料）為主。

觀察上述八個探討焦點中，頗少見針對 h. 共識統合部分的 Delphi 分析研究，檢視文獻已見論者提出加權 CV（weighted coefficient of variation, CV_w）

(Sheret, 1984)、Williamson CV (Williamson coefficient of variation, CV_w) 或稱 Williamson index (Williamson, 1965), 用於綜合各分群樣本變異資料以說明整體母群體變異情形。Sheret (1984) 提出加權的 CV (CV_w), 計算

公式為 $CV_w = \frac{\sqrt{\frac{\sum_{i=1}^k w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^k w_i}}}{\bar{x}_w}$, $\bar{x}_w = \frac{\sum_{i=1}^k w_i x_i}{\sum_{i=1}^k w_i}$, w_i 是各子群體的加權值, 由分析者設定, 通常可設定 $w_i = \frac{n_i}{\sum n_i} = \frac{n_i}{N}$, n_i 是子群的樣本數, N 是母群體數;

相對地是未加權 CV (CV_{uw}), $CV_{uw} = \frac{\sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x}_{uw})^2}{N}}}{\bar{x}_{uw}}$, $\bar{x}_{uw} = \frac{\sum_{i=1}^k x_i}{N}$, 此指數其實等於 CV。而 Williamson CV 常被使用, 本質上就是一種加權 CV, 分析

公式為 $CV_w = \frac{\sqrt{\sum_{i=1}^k (x_i - \bar{x}_w)^2 n_i}}{\bar{x}_w}$, $\bar{x}_w = n_1 x_1 + \dots + n_k x_k$, n_i 是各群樣本數, x_i 是各群資料值, k 是群數; 另有論者 (Gluschenko, 2018) 將 CV_w 擴展為國家 CV ('national' coefficient of variation, CV_{nat}), 以分析國家此母群體內子族群 (如地區或省市) 樣本間的不均等 (national inequality) 程度,

$CV_{nat} = \frac{\sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2}{N}}}{\bar{x}} = \frac{\sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x}_w)^2 n_i}{N}}}{\bar{x}_w}$, $\bar{x}_w = n_1 x_1 + \dots + n_k x_k$, n_i 是各群樣本數,

x_i 是各群資料值, k 是群數。CV_w 係數應用於實際資料的結果見表 32, 顯示全體資料的 CV 值會折衷呈現所有子群體的變異差異, 反觀 CV_w 值較能表明子群體 CV 值差異的情形。

簡言之, CV_w 可顯示傳統上分析整體 CV 值隱藏了子群體 CV 值差異的資訊, 更確切掌握整體變異情形的真實樣貌, 補充 CV 係數的不足。但可惜的是前述兩個 CV_w 似乎能視為相依樣本的綜合多個 CV 值方法, 故不見得適用於分析綜合多個獨立互斥樣本的 CV 值, 要進行整合不同獨立樣本的 CV 值, 似乎可參考既有後設分析的整合資料做法為基礎, 進行研議。

檢視前述兩個 CV_w 是以 CV 為分析基礎, 且是整合各分群體的 CV 值, 而 CV 值並非侷限於 [0,1], 因此要整合多個不同群體的 CV 值得思考使用 CV_s, 其值被規範於 [0,1], 且係數值分配近似常態分配, 所以加總 CV_s 成為新的 CV_s 似乎可行。據此可提出綜合 CV (CV for synthesis data, CV_{sd}), 用於總合多個 CV 值, $CV_{sd} = \frac{\sum CV_s}{k}$, CV_{sd} 係數值也被規範在 [0,1], 數值越大越好, 其顯著性考驗和信賴區間考驗有待後續探討。

表 32

使用實際資料分析 CV、CV_w、lnCVR 和 CV_{trim} 之比較

	項目1	項目2	項目3	項目4
CV				
子群體1 (n ₁ =4)	0.1333	0.1176	0.2041	0.1333
子群體2 (n ₂ =8)	0.2402	0.1336	0.2714	0.2926
全體 (N=12)	0.2132	0.1261	0.2510	0.2421
CV _w				
全體 (N=12)	0.3143	0.2619	0.2222	0.2712
lnCVR	0.4936	0.0322	0.1898	0.6907
CV _{trim} (N=11)	0.1442	0.0967	0.2109	0.2542
與全體 CV 差異	0.0690	0.0294	0.0401	0.0121
差異占全體 CV% (CVR _{trim})	32.364%	23.315%	15.976%	4.998%

註：採用 Sheret 提出的 CV_w。設定群體 1 為實驗組，群體 2 為控制組，分析 lnCVR。分析 CV_{trim} 是依據項目 1 的 1 個樣本被偵測為 outlier 而刪去該樣本，使用 boxplot 法 (Q₁-1.5IQR, Q₃+1.5IQR) 協助偵測。

其次，要比較不同研究所得 CV 值，對此論述提出對數轉換 CV 比率 (log-transformed CV ratio, lnCVR) (Nakagawa *et al.*, 2015; Senior, Viechtbauer, & Nakagawa, 2020) 為比較兩個 CV 值的效果量係數， $\ln CVR = \ln \left(\frac{CV_E}{CV_C} \right) + \frac{1}{2(N_E-1)} - \frac{1}{2(N_C-1)}$ ， $CV_E = \frac{SD_E}{M_E}$ ， $CV_C = \frac{SD_C}{M_C}$ ，E 為實驗組，C 為控制組；若 CV_E 值 > CV_C 值則 lnCVR 值 > 0，表示實驗組的分數變異差異高於控制組，反之 lnCVR 值 < 0，即控制組評分變異高於實驗組，就 Delphi 分析的兩輪回 CV 值變化而言，設輪回 2 為實驗組，則希望 lnCVR 值 < 0 且越小，接近 0 為佳，表示第二輪回評者看法達共識性高於第一輪回時。表 32 說明使用實際資料分析的 lnCVR 值，觀察兩子群體 CV 值可見群體 2 的 CV 值大於群體 1，其中項目 4 的 lnCVR 值最大，達 0.6907。綜觀之上述諸係數漸被採用，但是可惜在臺灣地區尚未見進行 Delphi 分析有關的整合分析和後設分析，也頗少見探討和使用 CV_w、CV_{nat}、lnCVR 係數的研究，後續值得推廣。

再者，CV 係數分析考量了平均數和標準差兩個統計量數，這兩者都易受 outlier 的影響，若考慮去除 outlier 則能構成刪減 CV (trimmed coefficient of variation, CV_{trim}, tCV)， $CV_{trim} = \frac{SD_{trim}}{\bar{x}_{trim}}$ ，SD_{trim} 是刪去 outlier 樣本的 SD， \bar{x}_{trim} 是刪去 outlier 樣本的平均數，此可使用 IQR、MAD、z 分數、盒狀圖 (box plot) 等傳統方法偵測 outlier (Rousseeuw & Hubert, 2011)。表 32 說明分析 CV_{trim} 的實例，不難看出刪去 outlier 樣本後的 CV 值普遍變小，表示剩下評估者看法的共識性更高，可據以了解異議者 (指 outlier) 對全體共識性的影響程度 (ratio of CV_{trim} to CV, CV_{Rtrim}) 達 4.998%~32.364%，顯然不可小覷，此處 $CV_{Rtrim} = \frac{|CV - CV_{trim}|}{CV} \times 100$ 。目前 Delphi 分析僅針對出現異議項進行刪去該項的做法，並未刻意排除異議者，為增進分析品質，或許可偵測 outlier 者，並檢視其

對整體共識性的影響程度，當造成影響度超過某特定值如 $CVR_{diff} > 30\%$ ，或是排除異議者前後的兩個 CV 值差異考驗 ($H_0: CV = CV_{trim}$ 或 $H_0: \frac{CV}{CV_{trim}} = 1$) 達統計顯著 ($p < .05$)，則不妨思考是否排除 outlier 者。

2. 結合使用多元統計方法，提供豐富和多向度檢視共識性評估資訊

計量 Delphi 分析為當前主流，質性 Delphi 分析則頗少，前已簡要說明 Delphi 分析應用統計方法概況，不難得知需要善用適當統計方法才能提供有效評估看法共識性和穩定性的資訊。前述表 9 和表 31 摘要簡述一些 Delphi 分析可應用統計分析量數和統計方法，能進一步充實見圖 13。儘管 Delphi 分析多數使用 Likert 量尺值（係連續量尺），然而實際用於共識性評估時卻採用適合次序量尺值和類別量尺值的 Kendall 和諧係數、Kappa 一致性係數，顯然實際分析時能將輸入的評定值轉換成非連續量尺值，因而適用 Delphi 分析的統計方法就顯得眾多了。基本上這些統計方法見諸一般統計教科書，且多數統計套裝軟體或程式提供協助分析功能，至於特定評估指數（如 CDI）和係數（如 CV、 CV_2 ）需要另行計算。還有進行多項共識分析時，宜視需要做事後比較或進階比較，如進行 Cronbach α 分析時多會進行刪項後的 Cronbach α 分析，以了解哪一項的評者看法異於其他項；類似於此，也能針對單項共識分析結果，進行配對任兩項的 CV 值差異顯著性考驗，這些進階比較是當前 Delphi 分析研究尚欠缺思考之處。

圖 13
 適用 Delphi 分析各類量尺評定資料之多元統計方法和統計量數彙總



註：適用分析專家評定看法計量資料。

進一步觀察前述圖 13 所臚列的 Delphi 分析可用資料分析方法，似乎忽視了關於 CV 係數的眾多研究成果涵蓋了單項共識、多項共識、共識變化、共識差異、共識統合等方面適用性，彙總見表 33；而非 CV 係數的諸多資料分析方法其實是聚焦於一致性（agreement）考驗，如常用於 Delphi 分析的 Kappa 係數、Kendall 和諧係數等，均屬於此類分析。這是共識性考驗和一致性考驗有時能視為相同，故 Delphi 分析就將一致性考驗方法納入採用，此表示圖 13 能夠繼續擴充，只要是檢視一致性程度的分析方法，便能考慮引用。另一方面，一些檢核不均等（inequality）或均等（equality）、變異性（Variation）等方法因為探討焦點與分析共識和一致性有相仿之處，似乎能夠研議將其引進處理 Delphi 分析的共識性問題。

簡言之，關於 Delphi 分析可用資料分析法，若將 CV 係數及其衍生係數視為一個分析方法族群，即 CV 係數族（CV group），而其他適用的統計方法則可歸類為單變項統計族（univariable statistics group），一般教育統計專書對於前者介紹甚少，多著墨於討論後者；兩者具互補作用，前者已漸成體系，並具應用效益，有待推廣和深入探討。

表 33
適用 Delphi 分析之 CV 係數族分析方法彙總

分析焦點	分類	適用 CV 及衍生係數
單項共識檢視	CV 係數及考驗, 衍生係數	CV, CV _s , CV ₂ , CV'', WCV, ICV, CV _{pop} , CV _{QME} , %CV, GS, %M, CV ² , CV _{trim} , CVR _{trim}
	CV 係數值標準及分類	CV 值標準及分類 (Pearson, English & Keran, Costa, Lorenzo 等, Garcia, Pimentel & Gomes, Yeh)
	CV 衍生調整係數	CDI, CDIs, CDI _A , DC, DCs
多項共識檢視	CV 係數比較及考驗	CV _{QME} , CV _{diff} , CV _{CG}
	MCV 係數及考驗	MCV

更進一步觀察表 33，其從單項和多項共識檢視角度思考，若納入思考分析的樣本為單一樣本或多群體樣本，則可見表 34，顯示已知係數集中於處理單一樣本的單項共識性評估問題，而處理多樣本的多項共識性評估問題，則似乎尚未見可資採用的 CV 族係數，有待後續研討。

表 34
CV 及衍生係數之適用情境

	單一樣本	多(群體)樣本
單項	CV 係數	CV, CV _s , CV ₂ , CV'', %CV, ICV
	衍生係數	lnCVR, CVR _{trim} , CV _{trim} , CV _{diff} , CV _{CG} , CV _w , CV _{SD} , CV _{nat}
多項		CDI, CDIs, CDI _A , DC, DC _s , GS
		MCV, CV _{diff} , CV _{SD}

四、就採用bootstrap法角度，建立適合評估CV、CDI等係數值的評估標準和顯著性考驗方法，提升共識評估結果的精準性

1. 研議採用 bootstrap 法的 CV 係數評估標準及信賴區間考驗，增進使用效益

當前關於採用bootstrap法分析CV係數的論述已談及信賴區間估計、多CV值比較等課題，唯獨欠缺探討針對CV係數評估標準和分類一事。觀察bootstrap法產出模擬資料所得CV值分布型態如圖14a，可知CV值接近常態分配型態，偏向高狹峰分布，常態分配考驗結果顯示 $KS = 0.069$ ， $p < .001$ ， $SW = 0.988$ ， $p < .001$ ，乃非常態分配。參考Yeh三分類構想，以 $M \pm SD$ 和 $M \pm 1.96SD$ 為評估基準值，若 $CV \leq (M - SD)$ 或 $CV \leq (M - 1.96SD)$ 屬於低度共識， $CV > (M + SD)$ 或 $CV > (M + 1.96SD)$ 屬於高度共識，餘歸屬為中度共識，整理分析結果見表35。其顯示四項實際CV值在bootstrap法分析1000次所得CV值的PR值為50.4~54.5，表示高於中數，以Carcia、Lorenzo標準分類為中度（編碼2）或低度（編碼1）。若採取1SD和1.96SD為分類基準，都被歸類為中度（編碼2），需要關注的是這兩個分類的分類區別正確率隨項目資料而異，多數是採取1SD為基準時的分類區別正確率高於使用1.96SD為基準，皆逾81%，表示應用 $M \pm 1SD$ 似乎是較佳的做法。

其次，要了解所得CV值與母群體CV值差異，即考驗 $H_0: CV = CV_{pop}$ ，可以藉助比較bootstrap法所得P5和P95，此兩百分位數大致顯示母群體CV值的90%信賴區間，若CV值落入該區間內，表示接受 $CV = CV_{pop}$ ，CV值 $< P5$ 或CV值 $> P95$ 就表示該CV值屬於偏離值，即CV值極佳或極差。表35指出各項CV值都落入P5和P95構成的區間內，其PR值也是介於5~95之間，顯示各項CV值並無明顯極佳或極差。

表35
使用不同評估標準針對以bootstrap法模擬所得CV值之評定結果

項目	分析值	PR ^a	評估標準		bootstrap法 ^a							
			Carcia	Lorenzo	P5	P95	最小值	最大值	M	SD	1SD ^b	1.96SD ^b
CV												
1	0.213	54.5	2	2	0.106	0.297	0.070	0.380	0.2011	0.0606	2 (82.8%)	2 (75.4%)
2	0.126	50.4	2	1	0.071	0.185	0.070	0.250	0.1244	0.0384	2 (91.5%)	2 (98.1%)
3	0.251	53.3	2	2	0.131	0.325	0.070	0.380	0.2400	0.0579	2 (81.1%)	2 (80.5%)
4	0.242	53.0	2	2	0.107	0.325	0.070	0.380	0.2309	0.0621	2 (86.4%)	2 (81.3%)
CVs												
1	0.064	54.5	2	2	0.032	0.090	0.021	0.116	0.0606	0.0183	2 (82.8%)	2 (75.4%)

項目	分析值	PR ^a	評估標準		bootstrap法 ^a								
			Garcia	Lorenzo	P5	P95	最小值	最大值	M	SD	1SD ^b	1.96SD ^b	
2	0.038	50.4	1	2	0.021	0.056	0.021	0.075	0.0375	0.0116	2 (91.5%)	2 (98.1%)	
3	0.076	53.3	2	2	0.040	0.098	0.021	0.114	0.0724	0.0175	2 (81.1%)	2 (80.1%)	
4	0.073	53.0	2	2	0.032	0.098	0.021	0.114	0.0696	0.0187	2 (87.3%)	2 (81.4%)	
CDI													
1	0.171	55.4	2	2	0.085	0.226	0.060	0.280	0.1584	0.0427	2 (84.9%)	2 (71.9%)	
2	0.103	49.3	2	1	0.058	0.148	0.060	0.200	0.1012	0.0307	2 (87.8%)	2 (97.6%)	
3	0.180	56.1	2	2	0.103	0.223	0.060	0.250	0.1694	0.0344	2 (88.0%)	2 (82.8%)	
4	0.178	56.3	2	2	0.085	0.217	0.060	0.260	0.1667	0.0385	2 (82.8%)	2 (84.1%)	
CV ₂													
1	0.209	54.5	2	2	0.106	0.285	0.070	0.360	0.1961	0.0570	2 (82.7%)	2 (75.4%)	
2	0.125	50.4	2	1	0.071	0.182	0.070	0.240	0.1231	0.0371	2 (91.5%)	2 (98.1%)	
3	0.244	53.3	2	2	0.130	0.309	0.070	0.350	0.2324	0.0533	2 (78.7%)	2 (80.3%)	
4	0.235	53.0	2	2	0.106	0.309	0.070	0.350	0.2239	0.0577	2 (86.4%)	2 (81.8%)	

註：^a使用實證資料(N=12)以bootstrap法進行1000次產生模擬資料所得CV、CV_s、CDI、CV₂值計算PR值。^b括號前數值為採用M±SD和M±1.96SD為評估基準值所得CV、CV_s、CDI、CV₂值評定等級，數值1~4分別表示低、中、高和極高，括號內%為區別分析所得分類正確率。

2. 研議 CV_s 係數評估標準和顯著性考驗方法 增進應用效益

觀察文獻未見提出關於 CV_s 值的評估標準和顯著性考驗做法，對此可思考採用 bootstrap 法以取得信賴區間估計值，可用考驗方法見表 36。使用 bootstrap 法產出模擬資料進行分析 CV_s 值分布型態如圖 14b 所示，不難得知 CV_s 值接近常態分配型態，但略微偏向高狹峰分布，常態分配考驗結果為 KS = 0.069， $p < .001$ ，SW = 0.988， $p < .001$ ，屬於非常態分配。

至於 CV_s 值如何評估優劣，觀察過去論述已提出對於 CV 值的判斷優劣標準，但是受到 CV_s 值值域被規範於 [0,1] 與 CV 值值域不同，故該標準不見得適用於 CV_s。思路之一是採取 ±1 個標準差 (SD) 為裁決依據，則 $CV_s < (M_{cv_s} - SD_{cv_s})$ 為極佳 (excellent)， $(M_{cv_s} - SD_{cv_s}) \leq CV_s \leq (M_{cv_s} + SD_{cv_s})$ 為良好 (good)， $CV_s > (M_{cv_s} + SD_{cv_s})$ 為不佳 (worst)，此處的 M_{cv_s} 和 SD_{cv_s} 是 bootstrap 法分析所得整體 CV_s 值的平均數和標準差，另一個寬鬆做法是採用 ±1.96 個標準差為裁決出三分類的依據，或者思考採用既有有關 CV 值評估標準的分類方法。表 35 說明採用 1SD 和 1.96SD 為標準的 CV_s 值分類結果皆為中等，區別正確率部分多數是採取 1SD 者高於採用 1.96SD 者，此與前面對 CV 值檢視結果相似。

再者，亦能以 bootstrap 法所得 CV_s 的 P5 和 P95 去判斷 CV_s 值的優劣，表 35 指出 CV_s 值落於 P5 和 P95 之間，CV_s 值的 PR 值 >5 且 <95，顯示其屬於中等，也表示

所得樣本 CV_s 與母群體 CV_s 無異，即接受 $H_0: CV_s = CV_{s\ pop}$ 。簡言之，此處所提出考驗方法和評估 CV_s 值做法應屬可行。

表36

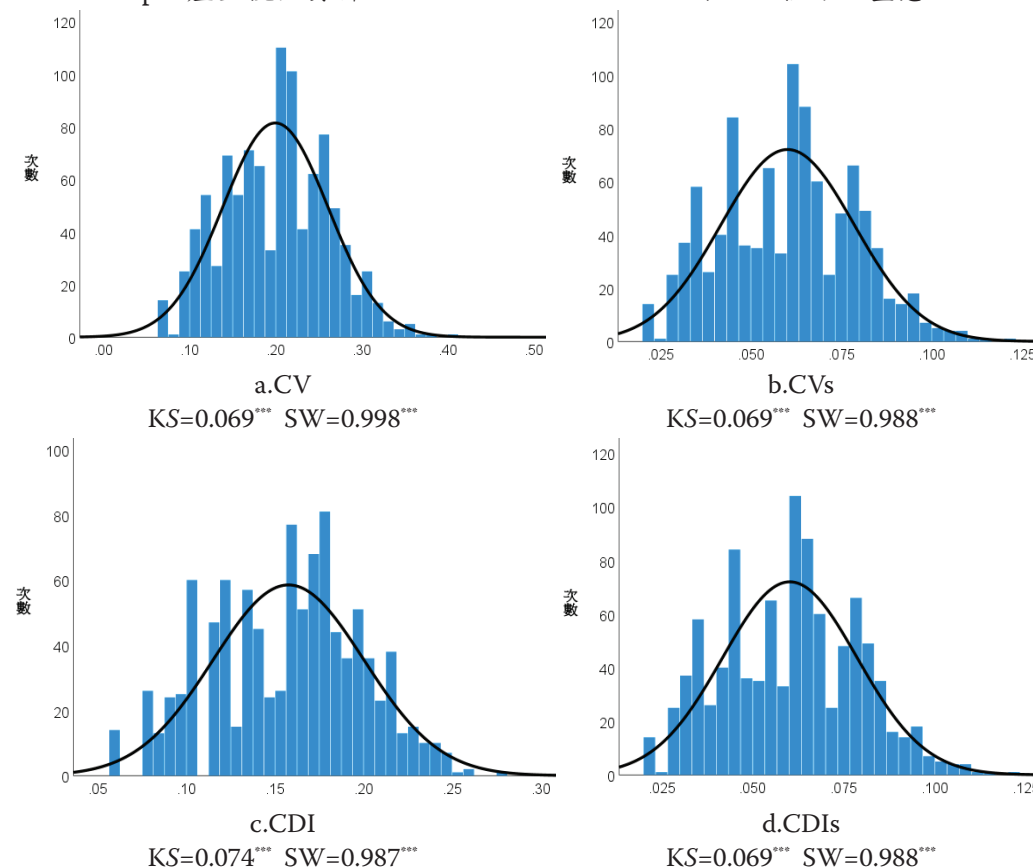
CVs值和CV2值信賴區間考驗方法

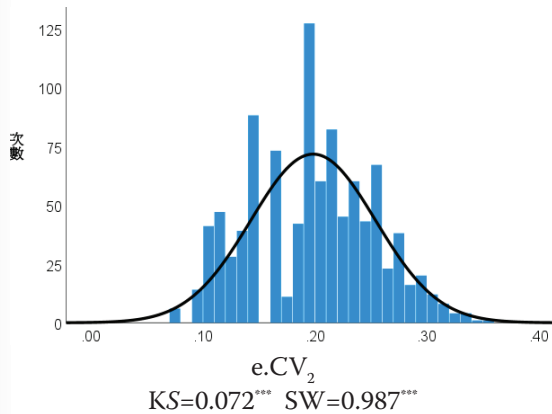
方法	信賴區間下界(LCI)	信賴區間上界(UCI)
CDI		
非參數 bootstrap 法 (NP bootstrap)	$LCI_{CDI} = \widehat{CDI}_{(\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDI}_{\alpha/2}$ 是 P2.5 的模擬值	$UCI_{CDI} = \widehat{CDI}_{(1-\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDI}_{1-\alpha/2}$ 是 P97.5 的模擬值
Bootstrap z 法	$LCI_{CDI} = \widehat{CDI} - Z_{1-\alpha/2} \times \widehat{SD}_{CDI}$	$UCI_{CDI} = \widehat{CDI} + Z_{1-\alpha/2} \times \widehat{SD}_{CDI}$
CDIs		
非參數 bootstrap 法 (NP bootstrap)	$LCI_{CDIs} = \widehat{CDIs}_{(\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDIs}_{\alpha/2}$ 是 P2.5 的模擬值	$UCI_{CDIs} = \widehat{CDIs}_{(1-\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDIs}_{1-\alpha/2}$ 是 P97.5 的模擬值
Bootstrap z 法	$LCI_{CDIs} = \widehat{CDIs} - Z_{1-\alpha/2} \times \widehat{SD}_{CDIs}$	$UCI_{CDIs} = \widehat{CDIs} + Z_{1-\alpha/2} \times \widehat{SD}_{CDIs}$

註：通常設定模擬至少500次，1000次較佳。可設定 $\alpha=.05$ 、.01或.10，通常設定 $\alpha=.05$ 。SDCVs是全部模擬資料產生 CV_s 值的標準差。

圖14

以bootstrap法產生模擬資料之CV、CVs、CDI、CDIs和CV2值分配型態





註：使用實證資料(N=12)以bootstrap法進行1000次產生模擬資料所得CV、CVs、CDI、CDIs和CV₂值。圖中顯示常態分配曲線。KS是Kolmogorov-Smirnov常態考驗統計量值及顯著性考驗結果，SW為Shapiro-Wilk常態考驗統計量值及顯著性考驗結果。*** $p < .001$

3. 研議 CV₂ 係數評估標準和顯著性考驗方法 增進應用效益

前所介紹的 CV₂ 有數值侷限於 0~1 便利解讀的優點，值得應用，惟文獻少見討論評估標準和顯著性考驗做法。觀察使用 bootstrap 法產出模擬資料的 CV₂ 值分布如圖 14e，類似常態分配，因此思考採用 bootstrap 法取得信賴區間估計值，考驗方法見表 36。又如何評估 CV₂ 值優劣，並未見討論。參考前述提及關於 CV、CVs 等係數值評估標準的思路，大致能採取 ±1 個標準差或 ±1.96 個標準差設計係數值評定等級的參考依據，區分出 $CV_2 < (M_{CV_2} - SD_{CV_2})$ 為極佳， $(M_{CV_2} - SD_{CV_2}) \leq CV_2 \leq (M_{CV_2} + SD_{CV_2})$ 為良好， $CV_2 > (M_{CV_2} + SD_{CV_2})$ 為不佳等三類，而 M_{CV_2} 和 SD_{CV_2} 是 bootstrap 法分析所得整體 CV₂ 值的平均數和標準差。

對此採取以 Bootstrap 法所得 CV₂ 值資料進行相關分析，表 35 顯示四項 CV₂ 值的 PR 值為 50.4~54.5，使用 Carcia、Lorenzo 標準評估屬於中度或低度，採取 ±1SD 或 ±1.96SD 為評估分類標準，皆落於中度一類，觀察區別正確率，多數項也是呈現使用 1SD 者高於採用 1.96SD 者。而四項 CV₂ 值都大於 P5 值且 < P95 值，顯然得接受 $H_0: CV_2 = CV_{2pop}$ 。大體言之，可藉助 bootstrap 法設定 CV₂ 係數評估標準，及進行與母群體 CV₂ 值差異的顯著性考驗。

4. 研議 CDI 指數評估標準和顯著性考驗方法，增進應用效益

如前所述，CDI 值的評估標準和顯著性考驗方法付之闕如，研究者多主觀選定評估 CDI 值達共識性的參考標準值，缺乏客觀統計學理支持證據，其也不能取代為考驗 CDI 值為合理的統計顯著性做法，故亟待研議。據此能考慮採用 bootstrap 法取得信賴區間估計值，考驗方法如表 37。使用 bootstrap 法產出模擬資料進行分析 CDI 值分布型態見圖 14c，發現 CDI 值接近常態分配型態，偏向略微高狹峰分布，常態分配考驗結果為 $KS=0.074$ ， $p < .001$ ， $SW=0.987$ ， $p < .001$ ，屬

於非常態分配。至於CDIs值分布型態，圖14d顯示CDIs值接近常態分配型態，偏向略微高狹峰分布，常態分配考驗結果指出 $KS=0.069$ ， $p<.001$ ， $SW=0.998$ ， $p<.001$ ，係屬非常態分配，此與CVs值的考驗結果（見圖14b）相仿。

再者評估CDI值和CDIs值優劣，可行思路是採取±1個標準差為裁決依據，即 $CDI < (M_{CDI} - SD_{CDI})$ 為極佳（excellent）， $(M_{CDI} - SD_{CDI}) \leq CDI \leq (M_{CDI} + SD_{CDI})$ 為良好（good）， $CDI > (M_{CDI} + SD_{CDI})$ 為不佳（worst），此處的 M_{CDI} 和 SD_{CDI} 是bootstrap法分析所得整體CDI值的平均數和標準差，另一個較寬鬆做法是採用±1.96個標準差來裁決出三分類。而此分類標準做法也適合於判斷CDIs值的優劣。

上述構想以bootstrap法模擬產生的CDI值資料進行考驗，結果見表35。其揭露出四項目CDI值的 $PR>50$ ，就Carcia、Lorenzo標準觀之係屬中度或低度，檢視採取1SD和1.96SD標準評估的正確區別率多數逾80%，表示兩標準適合採用。而CDIs值分布型態與CVs值幾乎相同，故其分析結果應與前面討論的CVs值相同。綜觀之，此處的考驗方法及評估CDI值和CDIs值標準做法應屬可行。

表37

CDI值和CDIs值信賴區間考驗方法

方法	信賴區間下界(LCI)	信賴區間上界(UCI)
CDI		
非參數 bootstrap 法 (NP bootstrap)	$LCI_{CDI} = \widehat{CDI}_{(\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDI}_{\alpha/2}$ 是 P2.5 的模擬值	$UCI_{CDI} = \widehat{CDI}_{(1-\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDI}_{1-\alpha/2}$ 是 P97.5 的模擬值
Bootstrap z 法	$LCI_{CDI} = \widehat{CDI} - Z_{1-\alpha/2} \times \widehat{SD}_{CDI}$	$UCI_{CDI} = \widehat{CDI} + Z_{1-\alpha/2} \times \widehat{SD}_{CDI}$
CDIs		
非參數 bootstrap 法 (NP bootstrap)	$LCI_{CDIs} = \widehat{CDIs}_{(\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDIs}_{\alpha/2}$ 是 P2.5 的模擬值	$UCI_{CDIs} = \widehat{CDIs}_{(1-\alpha/2)B}$ 若 $\alpha=.05$ ， $\widehat{CDIs}_{1-\alpha/2}$ 是 P97.5 的模擬值
Bootstrap z 法	$LCI_{CDIs} = \widehat{CDIs} - Z_{1-\alpha/2} \times \widehat{SD}_{CDIs}$	$UCI_{CDIs} = \widehat{CDIs} + Z_{1-\alpha/2} \times \widehat{SD}_{CDIs}$

註：通常設定模擬至少500次，1000次較佳。B是模擬次數。可設定 $\alpha=.05$ 、.01或.10，通常設定 $\alpha=.05$ 。 SD_{CDI} 和 SD_{CDIs} 是全部模擬資料產生CDI值和CDIs值的標準差。

5. 善用多變量變異係數(MCV)提供更多共識性評估資訊

觀察CV屬於單變量變異係數（univariate CV），說明單一變項資料的變異程度，欲應用於綜合檢視多個項目的CV值是否顯示評估達共識性，一個粗略做法是檢視多項的CV值信賴區間是否出現交疊現象，若無就表示有某些項與其餘項的評估變異程度有明顯差異，此可稱為信賴區間交疊查核（check up the overlap of item confidence interval of CV, checkCVCI）。分析流程是計算各項CV的95%信賴區間，將其繪製成CVCI圖（CVCI plot），觀察該圖中如有某項目的CI區域不和其他項的CI區域交疊，就能識別為無共識，此為最簡略做法。大致上將出現四類CI區域交疊情形見圖15，包括兩種完全無交疊（none overlap, NO），部分

無交疊 (part of overlap, PO) 和完全交疊 (complete overlap, CO), 分別表示毫無共識 (none consensus, NC)、不穩定共識 (unstable consensus, USC)、高度共識 (high consensus, HC), 其中NC和HC兩類無爭議地可明確認定, 而USC如圖 15c, 不難看出僅有項目2和項目3的CI區域沒交疊, 其餘項目則彼此交疊, 即6組項目CV的CI交疊情形中, 只有1組情形是沒交疊, 其餘5組皆出現交疊現象, 如此情形該怎麼判斷整體項目是否達共識有待思考。

對此, 需要確定一個容忍的未達共識比率, 做為比較基準如60% (屬於寬鬆基準)、80% (為較高基準) 或90% (係嚴格基準), 若達比較基準則能判斷為可接受共識 (acceptable consensus, AC), 反之可視為不合理共識 (unreasonable consensus, URC)。更明確地說, 以二分法而論, NC和URC為未達共識, 不能接受, 而HC和AC為達到共識。圖 16 係實際資料分析所得CVC1圖, 可知圖 16a 屬於高度共識, 圖 16b 是不穩定共識狀態。

圖 15
四類CV信賴區間構成之共識辨識類型

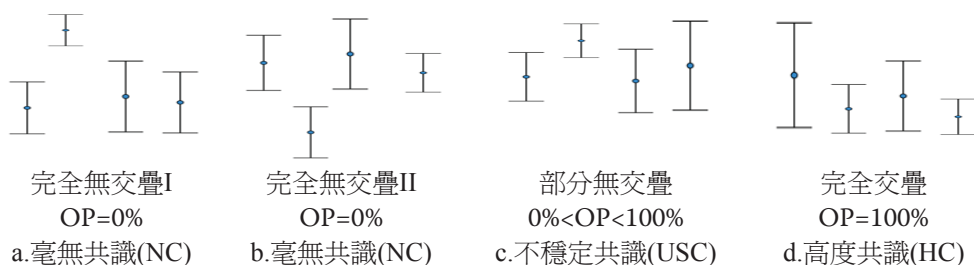
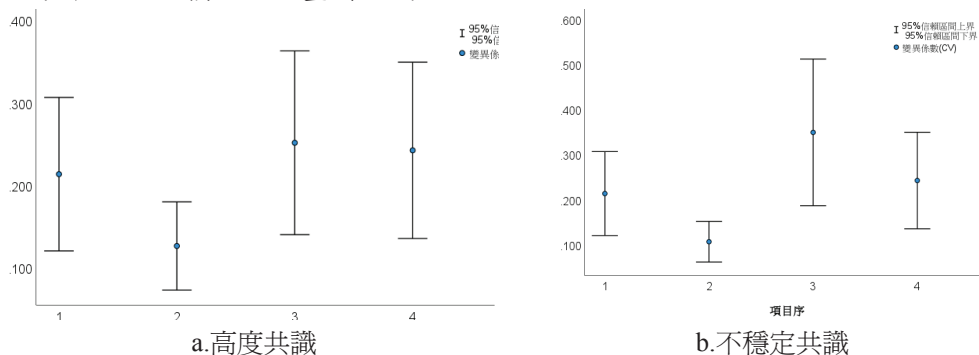


圖 16
實證資料之CV信賴區間交疊圖示例



進一步可分析能指出各項評估達一致的比率, 即共識比率 (the ratio of consensus, RC), 做為客觀佐證資訊, 若項目數為 k , 項目配對比較次數為 $k(k-1)/2$, CV 信賴區間未交疊次數為 u , 則 $RC = 1 - \frac{u}{k(k-1)/2} = 1 - \frac{2u}{k(k-1)}$, $0 \leq RC \leq 1$, 可以 $RC \geq 0.8$ 為判斷標準, 表示 80% 以上配對項目的 CV 無顯著差異。就圖 15c 和圖 16b 來看, $RC = 1 - 1/6 = 1 - 0.1667 = 0.8333$, 因此能算是可接受

共識(AC)。做法二是進行 $k(k-1)/2$ 次不重複的項目CV值配對比較，即比較 CV_i 值和 CV_j 值的差異是否達統計的顯著性認定，乃採用考驗兩個CV值差異方法如Miller法，但必須注意這做法會如做多次 t 考驗將遭遇類型I錯誤（type one error）增加的風險。

另一個較為精準做法是採用可綜合說明多項變異一致性的係數，對此針對多個變項資料闡釋其變異情形，已見論述（Aerts, Haesbroeck, & Ruwet, 2015, 2018; Albert & Zhang, 2010; Banerjee, 2018, 2019; Jayakumar, Kan, & Wilfred, 2021）提出適用連續量尺資料的多變量變異係數（multivariate coefficients of variation, multidimensional coefficients of variation, MCV），見表38，而Anderson（2023）另提出適用類別量尺資料的MCV，Aerts和Haesbroeck（2017）則提出考驗多個MCV均等的方法。這些MCV若能應用於Delphi分析，將能提供關於多個變項資料變異程度的資訊，可惜少見使用，也少見分析程式和統計軟體提供分析功能，後續有待突破。

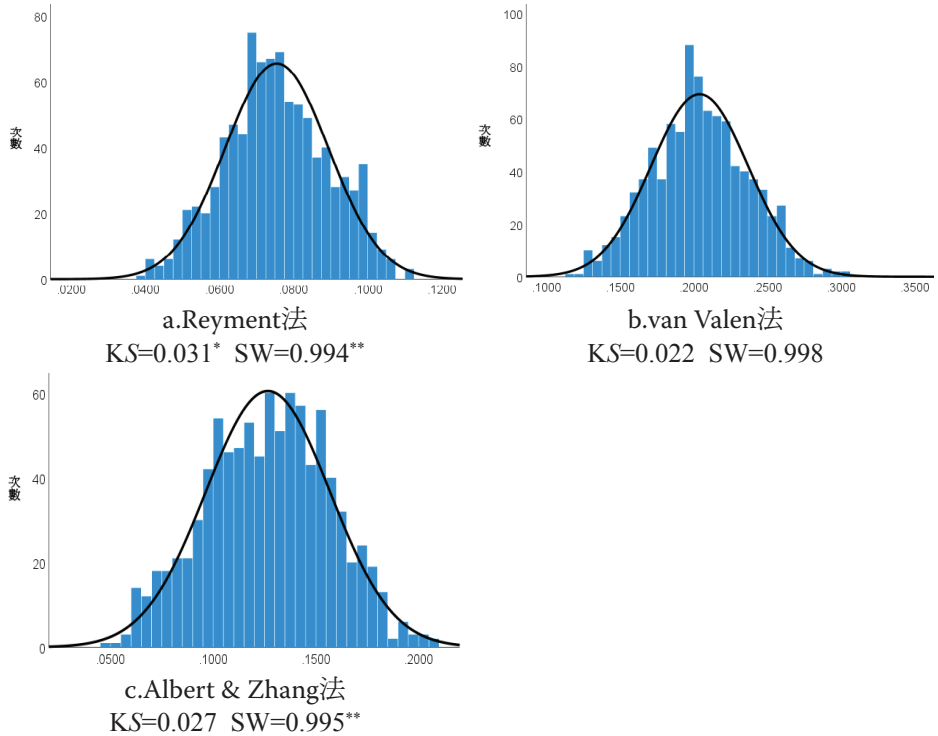
表38
 適用連續量尺資料之多變量變異係數計算方法彙總

提出者	計算公式	標準誤	上限值
Reyment (1960)	$MCV_R = \sqrt{\frac{\det(\Sigma)^{1/k}}{\mu^T \mu}}$		
van Valen (1974,2005)	$MCV_{VV} = \sqrt{\frac{\text{trace}(\Sigma)}{\mu^T \mu}}$		
Voinov & Nikulin (1996)	$MCV_{VN} = \sqrt{\frac{1}{\mu^T \Sigma^{-1} \mu}}$		$\sqrt{\frac{N}{\chi^2_{.99,k}}}$
Albert & Zhang (2010)	$MCV_{AZ} = \sqrt{\frac{\mu^T \Sigma \mu}{(\mu^T \mu)^2}}$	$\sqrt{\frac{MCV_{AZ}^2}{2N} (1 + 2MCV_{AZ}^2)}$	
Banerjee (2018)	$MCV_{B1} = \sqrt{\frac{1}{N \times k} \lambda(x^T x) - 1}$		
Auricchio, Giudici & Toscani (2024)	$MCV_{AGT} = \sqrt{\frac{N}{\mu^T \Sigma^{-1} \mu}}$		
Banerjee (2019)	$MCV_{B2} = \sqrt{\frac{\mu_m^T \Sigma_m \mu_m}{(\mu_m^T \mu_m)^2}}$		
Jayakumar, Kan, & Wilfred (2021)	$MCV_{JKW} = \sqrt{\frac{N-k}{k} F_{N-k,k,(\sqrt{N}/MCV_{VN})^2}}$		

註：N是樣本數，k是項目數， Σ 是變異數矩陣， μ 是平均數向量， λ 是最大特徵值， $\det()$ 係矩陣的行列式， $\text{trace}()$ 是矩陣的跡， x 是原始資料值矩陣的各項值/各項平均值所得資料矩陣， Σ_m 是以 x 資料矩陣分析所得變異數矩陣， μ_m 是以 x 資料矩陣分析所得平均數向量，MCV下標標示不同提出者的代號以利辨識。

又前述MCV係數多未提供標準誤（SE），或可參考Albert和Zhang（2010）做法，以bootstrap法模擬1000次資料分析所得的標準差替換為標準誤，以確認MCV值估算的精確性，或進一步做為進行信賴區間估計的參考依據。為此，針對三種MCV值估算法，使用某實證調查資料以bootstrap法進行1000次模擬產生資料進行MCV值分配型態檢視，分析結果見圖17。觀察各直方圖顯示MCV值分布頗類似常態分配，惟Reyment法和van Valen法所得MCV值分布的峰度略微高陡，Albert和Zhang法的MCV值分布略微平闊；進一步實施Kolmogorov-Smirnov和Shapiro-Wilk常態考驗，可知Reyment法及Albert和Zhang法的MCV值屬於非常態分配，van Valen法的MCV值則屬常態分配。大體看來，這三種估算MCV值方法適用bootstrap法，且模擬資料的MCV值分配接近常態分配型態，或能沿用基於z分配的95%信賴區間估計方法，分析MCV值的95%信賴區間，即 $LCI_{MCV} = \overline{MCV} - Z_{1-\alpha/2} \times \widehat{SD}_{MCV}$ ， $UCI_{MCV} = \overline{MCV} + Z_{1-\alpha/2} \times \widehat{SD}_{MCV}$ ，通常將 α 設定為.05，此處MCV是原資料所得MCV值，SDMCV是分析bootstrap法模擬資料所得MCV值的標準差。

圖17
以bootstrap法產生模擬資料之MCV值分配型態



註：使用實證資料(N=12)以bootstrap法進行1000次產生模擬資料所得MCV值。圖中顯示常態分配曲線。KS是Kolmogorov-Smirnov常態考驗統計量值及顯著性考驗結果，SW為Shapiro-Wilk常態考驗統計量值及顯著性考驗結果。* $p < .05$, ** $p < .01$

再者如何評斷 MCV 值優劣，不妨採取前述有關分類 CV、CVs、CDI 和 CDIs 值以 ± 1 個標準差為裁決依據的構想，即 $MCV < (M_{MCV} - SD_{MCV})$ 為極佳 (excellent)， $(M_{MCV} - SD_{MCV}) \leq MCV \leq (M_{MCV} + SD_{MCV})$ 為良好 (good)， $MCV > (M_{MCV} + SD_{MCV})$ 為不佳 (worst)，此處 M_{MCV} 和 SD_{MCV} 使用 bootstrap 法所得整體 MCV 值的平均數和標準差，較寬鬆標準係採用 ± 1.96 個標準差進行區分。

6. 研議 DM、SA、VCC 和 VCC_m 指數評估標準和顯著性考驗方法，增進應用效益

DM (DeMoivre index) 和 SA (strict agreement index) 兩指數分別屬於共識性指數 (consensus index) 和一致性指數 (agreement indices) (Meijering, Kampen, & Tobi, 2013)，並無評估標準和顯著性考驗方法不利於應用，對此可採取多次抽樣模擬分析做法，以確認其係數值的可能分配型態。表 39 和圖 18 顯示以隨機抽樣和 bootstrap 法抽樣所得 DM 和 SA 值分配情形，由圖 18 大致可看出 bootstrap 法抽樣資料分配型態近似常態分配，但略微偏於高峽峰型態，表 38 亦顯示類似資訊；但隨著增加抽樣次數如 2000，DM 值分配型態依然，SA 值分配型態會更近似標準化常態分配。另可得 DM 值多數偏小，SA 值略大，此與 Meijering、Kampen 和 Tobi (2013) 進行 1000 次模擬分析 60 個項目 20 位評者資料的模擬分析結果類似，而 bootstrap 法抽樣結果表現較優於隨機法抽樣，因此能採用 bootstrap 法抽樣協助考驗。簡言之，若要設定固定標準，DM 值 = 0 可接受，DM 值 ≥ 0.1 為佳，SA 值 ≥ 0.4 可接受，SA 值 ≥ 0.5 為佳；如果針對 bootstrap 法抽樣，則 SA 值落在 P5 和 P95 之間為可接受，低於 P5 為不可接受， \geq 中數為佳， $\geq P95$ 為優，DM 值落在 P5 和 P95 之間為可接受，低於 P5 為不可接受， \geq 中數為佳， $\geq P95$ 為優。又在多輪回調查時，可採用校正 SA 指數 (corrected strict agreement index, CSA)，其以第一輪回的 SA 值為比較基準，公式如下。

$$C_{irr} = \begin{cases} 1 & \text{if } x_i = x_j, X_i, X_j \text{ 是兩評估者對同一項目的評估值} \\ 0 & \text{otherwise} \end{cases}$$

$$C_{it} = \begin{cases} 1 & \text{if } \sum C_{irr} = \frac{m(m-1)}{2}, m \text{ 是評估者數} \\ 0 & \text{otherwise} \end{cases}$$

$$DM = \frac{\sum C_{it}}{k}, k \text{ 是項目數}$$

$$a_{it} = \frac{\sum_{i>j} C_{irr}}{m(m-1)/2}$$

$$SA = \frac{\sum a_{it}}{k}$$

$$CSA = SA_i - SA_1, i \text{ 是輪回序}, i > 1$$

表 39

DM、SA、VCC和VCCM指數值分配型態考驗結果之比較

	隨機抽樣		bootstrap法抽樣			
	DM	SA	DM	SA	VCC	VCC _M
最小值	0	0.15	0	0.280	0.484	0.402
最大值	0	0.30	0.250	0.814	0.871	0.602
P5			0	0.341	0.545	0.441
P95			0	0.629	0.743	0.538
平均數	0	0.2	0.011	0.471	0.634	0.484
		[0.199,0.203]	[0.008,0.014]	[0.466,0.477]	[0.630,0.638]	[0.482,0.486]
標準差	0	0.026	0.051	0.090	0.060	0.030
中數	0	0.197	0	0.462	0.628	0.481
偏態		0.603**	4.512**	0.537**	0.485**	0.496**
峰度		0.191*	18.399**	0.274**	0.165**	0.180**
常態考驗(K-S法)		0.082**	0.541**	0.057**	0.044**	0.044**
常態考驗(S-W法)		0.974**	0.203**	0.981**	0.985**	0.984**

註：進行1000次模擬分析4個項目10位評者資料。* $p < .05$ ** $p < .001$

其次，VCC (Ventana Coefficient of Consensus)和VCCM (modified VCC, VCC')兩指數 (Hsiao, Lin, & Chang, 2008; MacNeil, 1997) 可用於評估共識，分析公式如下 (即VCC_i和VCC_{Mi})，通常 $1 \geq VCC > 0$ ，當 $SD > \frac{range}{2}$ 時， $VCC < 0$ ，因此VCC=1表示完全共識， ≤ 0 是無共識，VCC正值時越大表示共識度越大，反之VCC負值時越大表示評者看法歧異越大。至於VCC_M部分，通常 $1 \geq VCC_M > 0$ ，VCC_M也可能 ≤ 0 和 > 1 ，VCC_M正值時越大表示共識度越大，VCC_M負值時越大表示評者看法歧異越大。這兩個係數係聚焦於單一項目，若針對整體觀之多個項目，對此提出VCC_T(Total Ventana Coefficient of Consensus)和VCC_{TM} (Total modified VCC)，分析公式如下，其值域出現情形如VCC和VCC_M。

$$VCC_i = 1 - \frac{2 \times \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}}{range} = 1 - \frac{2 \times SD_i}{range}, \text{ range} = \text{最大量尺值} - 1 = \text{限制最大值} - \text{限制最小值}$$

$$VCC_{Mi} = \frac{1}{2} \left(1 - \frac{2 \times SD_i}{range} \right) + \frac{1}{2} \left(\frac{Mode}{n} \right) = \frac{1}{2} VCC_i + \frac{1}{2} \left(\frac{Mode}{n} \right), \text{ n 是樣本數, Mode 是眾數}$$

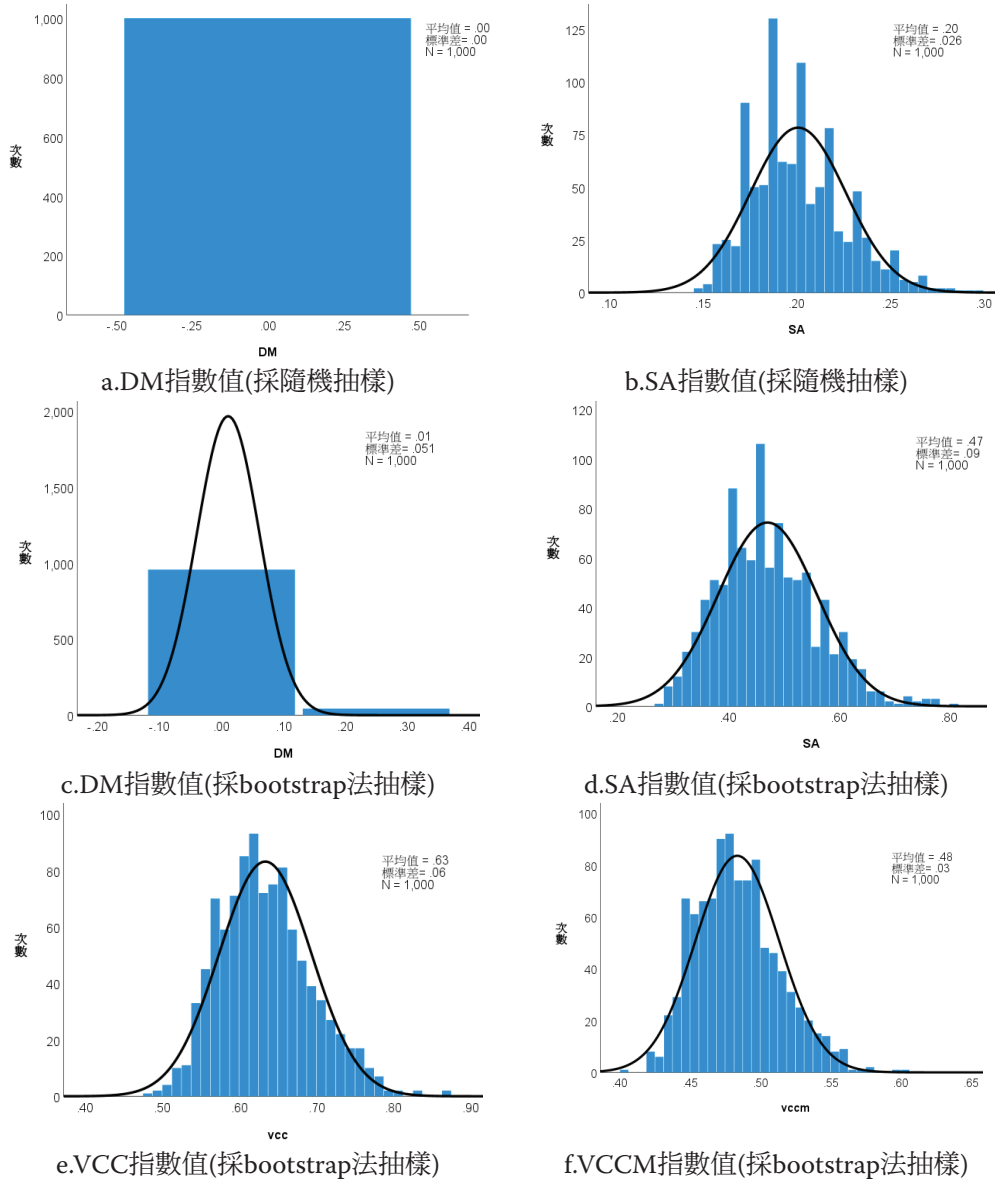
$$VCC_T = \frac{\sum VCC_i}{k}, \text{ k 是項目數}$$

$$VCC_{MT} = \frac{\sum VCC_{Mi}}{k}$$

而VCC和VCC_M欠缺評估標準和顯著性考驗方法，對此採取bootstrap抽樣模擬分析做法，以確認係數值的可能分配型態。表39和圖18顯示VCC_T和VCC_{TM}值分配情形，其顯示bootstrap法抽樣資料分配型態近似常態分配，但略微偏於右偏態型態，表39指出類似資訊；但隨著增加抽樣次數如5000，VCC_T和VCC_{TM}值分配型態更近似標準化常態分配。據此或能沿用基於z分

配的95%信賴區間估計方法，分析VCC值的95%信賴區間，將 α 設定為.05，即 $LCI_{VCC} = \widehat{VCC} - Z_{1-\alpha/2} \times \widehat{SD}_{VCC}$ ， $UCI_{VCC} = \widehat{VCC} + Z_{1-\alpha/2} \times \widehat{SD}_{VCC}$ ，而分析 VCC_M 值的95%信賴區間，亦可設定為 $LCI_{VCC_M} = \widehat{VCC_M} - Z_{1-\alpha/2} \times \widehat{SD}_{VCC_M}$ ， $UCI_{VCC_M} = \widehat{VCC_M} + Z_{1-\alpha/2} \times \widehat{SD}_{VCC_M}$ 。至於設定固定標準，參考圖18似乎可設定 VCC 值 ≤ 0 是無共識， < 0.5 低度共識， ≥ 0.5 中度共識， ≥ 0.7 為高度共識，為1是完全共識， VCC_M 值 ≤ 0 是無共識， < 0.4 是低度共識， ≥ 0.4 中度共識， ≥ 0.53 為高度共識， ≥ 1 是完全共識。

圖18
 DM、SA、VCC和VCC_M指數值模擬分析結果之分配型態



註：進行1000次模擬分析4個項目10位評者資料。

五、推廣應用Delphi分析已知多元分析方法和指數，驗證分析效益和解決應用難題

1. 檢驗應用多元統計分析方法所得共識性評估結果的一致性，注意採用統計方法的測量理念

根據論述 (Birko, Dove, & Özdemir, 2015; Cottam, Roe, & Challacombe, 2004; Hsiao, Lin, & Chang, 2008) 可用於 Delphi 分析評估共識性的係數包括 Fleiss Kappa (KF)、Conger Kappa (KC)、Clustered Mode (CM)、Extremity CPWA (XCPWA)、Clustered PWA (CPWA)、Mode (Md)、Pair-wise Agreement (PWA)、De Moivre index (DM)、Interquartile Range (IQR)、Group Conformity Index (GCI)、Average Percent of Majority Opinions (APMO)、Ventana Coefficient of Consensus (VCC)、VCC' (或稱 VCCM) 等，而論述 (Gwet, 2021; James, Demaree, & Wolf, 1984; van Oest, 2019) 指出 S-coefficient、Cohen's kappa、Scott's pi、Fleiss' kappa、Krippendorff's alpha、Gwet's AC1、Brennan-Prediger Coefficient、rwg 等係數能用於評量評分者間一致性 (inter-rater agreement)，當然也可適用於 Delphi 分析。但受限本文篇幅和探討焦點，僅能揀選部分進行討論和比較，所以這些未被討論的係數分析效益有待後續深究。

綜觀本文可用於評估 Delphi 分析共識性的統計方法，除了 CV 係數族外，其他均屬於單變量統計方法。以常見採用的 Cochran Q、Kappa 係數、Kendall w、Cronbach α 和主成分分析為例，分析結果實例見表 40，可見三個研究的 Cochran Q、Kappa 係數、Kendall w 分析結果皆指出多項目評估是達共識性，而 Cronbach α 用於考驗內部一致性信度，主成分分析考驗構念效度，表 40 顯示研究一和研究三指明內部一致性不佳，Cronbach α 值 < 0.52 ，甚至出現負值，主成分分析結果都指出抽取 2 個主成分，這些意味著該向度內的項目並非測量同一個構念，後續應該拆分項目成兩部分，此與前述 Cochran Q 等分析結果不同，這是因為這兩類資料分析方法背後的測量理念和測量對象不同所致，所以不能等同視之。

表 40

Delphi 分析使用多元統計方法實證檢視多項共識性結果之比較

統計量數	研究一(k=4, N=12)		研究二(k=4, N=10)		研究三(k=4, N=11)	
	分析值和95%CI	p	分析值和95%CI	p	分析值和95%CI	p
Cochran Q	3	.392	0.273	.965	4.675	.190
Best & Rayner, S	4	.261	0.364	.948	6.353	.096
Fleiss Kappa	0.151 [-0.152,0.454]	.328	-0.058 [-0.175,0.060]	.334	0.146[-0.206,0.498]	.416
Kendall w (異等級)	0.480	.631	0.028	.840	0.060	.580
Kendall w (同等級)	0.910	.350	0.030	.822	0.121	.261
Cronbach α	0.519 [-0.160,0.845]		-1.753 [-6.243,0.232]		0.315[-0.721,0.793]	
主成分分析 ^a	2成分(77.379%)		2成分(75.701%)		2成分(73.804%)	

註：原始資料取自期刊論文和學位論文集載資料。k是項目數，N是評估者數。^a採取K1法則選取主成分。

綜言之，基本上分析 Cochran Q、Kappa 係數、Kendall w 三者的結果應該相似，在 Delphi 分析中其測量作用是了解是評分者信度，或者亦能了解為內容效度，關心對象是評估者；反觀 Cronbach α 和主成分分析適用於了解內部一致性信度和構念效度，關注對象是項目。

2. 聚焦 CV 及衍生係數檢驗分析效益，推薦最優應用策略

前已指出 CV 族係數可自成一個體系，表 41 更進一步指出 CV、CDI、%CV 和 ICV 四個係數有關考驗的研究成果，可見除 CV 外，其他三個係數的有關研究仍屬不足。反之，是 CV 係數的研究成果甚多，但是有些則顯得紛雜，有時難以抉擇，如對於 CV、MCV 的 95%CI 和顯著性考驗探討就基於常態分配等多種假設，而衍生出多個考驗方法，並形成所得考驗結果值不盡相同的問題，且檢索文獻可知依然持續有論者提出新考驗做法，這讓使用者頓感莫衷一是的無奈，有待整合和比較，以取得更客觀可信賴的最優做法。

表 41

CV 及衍生係數有關考驗研究成果之比較

係數	伴生係數 ^a	母群體 正規化 係數值	係數值 判斷標準	信賴區間 估計考驗	顯著性 考驗	多值比 較	多值統合	去異 常值 影響
CV		CV _s	CV _{pop} Pearson 等	Miller 等	有	CV _{diff} , CV _{Rtrim}	CV _w , MCV, CV _{sd} , lnCVR, CV _{nat}	CV _{trim}
CDI	DC, DC _s	CDI _s , CDI _A	自訂					
%CV	GS							
ICV				有	有			

註：^a伴生係數指其伴隨原生係數而衍生，係數值意義與原生係數相反，如 DC=1-CDI，DC 是伴生係數，CDI 乃原生係數，兩者數值所表達意義相反。

綜上所述，可知 CV 族係數體系漸趨豐富和多元化，在評估 Delphi 分析的應用日趨重要，目前 CV 係數的研究焦點出現多元化面貌，在統合多值、比較多值兩方面已漸有研究成果可資採用，配合應用 Kendall w、Kappa、Cochran Q 等係數協助多值統合，已能提供判斷共識性達成與否的豐富資訊，此為最佳應用策略。而後續更進一步，可關注對於 CV 信賴區間考驗、多個 CV 值整合、多群體多項 CV 值整合的相關研究成果，以取得最佳分析效果的做法，增進應用效益。

3. 研議和應用校正評定量尺值的 IPRAS 指數和 DI 指數

再者，前已敘及宜思考使用 IPRAS 和 DI 兩指數，但是其計算公式是設定使用 9 點量尺，若採用其他值的評定量尺值如 Likert 5 點量尺，則宜根據使用的評定量尺值 (value of measurement scale, SV)，採取合適的調整數 (ws)，計算 IPR_w 和 IPRCP_w 指數，再分析其他的 IPRAS 和 DI 指數值，以取得合理的分析結果，計算公式臚列如下供參考。

$$\begin{cases} SV < 9, w_s = \frac{9}{SV} \\ SV = 9, w_s = 1 \\ SV > 9, w_s = \frac{SV}{9} \end{cases}$$

$$IPRAS_w = IPR_r + (AI \times CFA) \quad IPR_r=2.35$$

$$CFA=1.5$$

$$IPRCP_w = \left(\frac{IPR_L + IPR_U}{2} \right) w_s \quad IPR_L=P_{30} \quad IPR_U=P_{70}$$

$$IPR_w = (IPR_U - IPR_L) w_s$$

$$AI = |SV_{middle} - IPRCP_w| \quad SV_{middle} = \text{評定量尺值的中間值, 設定為 } 5$$

$$DI = \frac{IPR_w}{IPRAS_w}$$

六、發展專用於Delphi分析軟體和程式，增進分析效益和品質

過去有 Professional Delphi Scan、Real-Time Delphi (RTD)、Mesydel、Calibrium、Surveylet (www.calibrium.com)、Mesydel (www.mesydel.com)、RAND Expert Lens (www.rand.org/pubs/tools/expertlens) 等分析軟體面世 (Hasson, Keeney, & McKenna, 2025; Jorm, 2025)，又有 Welphi (<https://www.welphi.com>)、STAT95 (<https://www.stat59.com>)、edelphi (<https://www.edelphi.org>)、DelphiManager (<https://www.comet-initiative.org/delphimanager/>)、SmartDelphi (<https://www.smartdelphi.com/>) 等可進行線上 Delphi 調查或分析 Delphi 調查資料，其主要提供各輪回評估人員進行量化評定和簡易分析資料，其優勢包括結合線上諮詢作業，減少自行輸入評估資料的不便利，提供立即回饋評估資料的共識性檢視結果，但缺點是僅提供分析共識性的描述性統計資訊（如平均數、標準差），無法提供進階的推論統計資訊（如 Fleiss Kappa 係數），顯然有發展專用程式的必要。

其次，CV 是常見運用於 Delphi 分析的係數，SAS 提供 MEANS、UNIVARIATE、IML、TABULATE 等程序分析 CV，SPSS 提供 RATIO STATISTICS 程序分析 %CV（其顯示為 Mean Centered COV），亦有線上程式可分析 CV 值如 INCH CALCULATOR（見 <https://www.inchcalculator.com/coefficient-of-variation-calculator/>），比較兩個 CV 值差異如 MedCalc（見 https://www.medcalc.org/calc/comparison_of_coefficientsofvariation.php）。然而若要進行其他分析如 Kappa 係數、Kendall w 係數等，需要進行多次分析，顯然麻煩，若能整合則一次分析就畢其全功。

為補充既有分析軟體和程式提供分析功能的不足，從宏觀角度思考，針對量化Delphi分析流程（圖1和圖11），至少需要處理評估者專業性、單一輪回評估看法共識性、多輪回評估看法共識性差異和關聯性、評者特性對各輪回評估看法影響等部分的資料，據此需求，設計若干適用於SPSS分析平臺的程式見表42。這些程式均屬於巨集（marco）性質，需要設定一些分析參數後執行，各程式即可自動產出眾多有關資訊，並形成利於擷取編輯的分析報告，至於解讀其分析結果於附錄五至附錄十一中有簡要說明。

簡言之，分析單一輪回評估看法共識性宜用SIMPLEDELPHI7或ROMPERDELPHI程式，若有三輪回則應執行SIMPLEDELPHI7或ROMPERDELPHI程式三次，並使用CLUSTERNUMS程式確認單一輪回項目評估看法的合理分群數；若要比較兩輪回評估看法差異和變化，宜採用MULTIDELPHI3M程式，可知曉兩輪回評估看法共識性分析結果，及兩輪回看法關聯和差異情形；想知道評者特性對評估看法，宜使用COMPDELPHI4程式，可獲知不同特性子群體評者的評估看法差異或關聯情形；欲了解評者的專業性及評估做法，能使用DELPHICALC8M程式檢視專業性；想取得模擬資料，進行Delphi分析共識性評估分析，應運用SIMULATECV程式，產生模擬的評估資料（係Likert量尺值）；如有MULTIDELPHI3M程式或COMPDELPHI4程式產生的分析結果檔，宜考慮SHOWDELPHI3程式復原展示多輪回或多子群體的評估共識性結果。若要進行R型Delphi分析，應採用QTRANSDAT程式，將原本變項資料加以轉置，對調列和行形成新資料，再使用SIMPLEDELPHI7或ROMPERDELPHI程式進行分析。而ROMPERDELPHI程式可視為SIMPLEDELPHI7程式的簡要功能版本，執行速度較快，另新增評值%堆疊圖、異議評者偵測等分析功能。

表42

適用SPSS之Delphi分析巨集程式功能

程式	分析目的	分析功能
DELPHICALC8M	評估專家專業性	1. 描述統計分析專家判斷係數(ca),專家熟悉係數(cs),專家權威係數(cr),專家專業係數(DP)等 2. 辨識專性屬於偏離值性質的專家
SIMPLEDELPHI7	分析Delphi單次輪回調查資料的單項多項共識性	1. 分析單項描述統計(含集中量數,變異量數,相對地位量數),常態分配考驗及偏離值檢測,圖示分析 2. 分析CV及衍生係數(如CDI)及顯著性考驗,CV值分類,配對CV值差異比較 3. 分析一致性的單變項統計(如kappa係數),Cornbach α 及主成分分析,CVI及有關係數(如CVR)

程式	分析目的	分析功能
COMPDELPHI4	分析不同特性專家的評估差異影響	1. 分析不同特性群組專家的評估看法和共識性(含集中量數,變異量數,相對地位量數) 2. 彙總比較不同特性群體專家的評估看法和共識性評估差異
SHOWDELPHI3	顯示不同特性專家共識性評估結果	顯示不同特性群組專家的評估看法和共識性(含集中量數,變異量數,相對地位量數)
MULTIDELPHI3M	分析兩輪Delphi專家評估變化與差異	1. 描述統計分析兩輪回時單項和多項評估看法及共識性評估結果(含集中量數,變異量數,CV及衍生係數值,圖示分析) 2. 統計分析比較兩輪回時單項和多項評估看法及共識性變化(含差異考驗,關聯考驗,CV及衍生係數值)
ROMPERDELPHI	分析Delphi單次輪回調查資料的單項多項共識性	1. 分析單項描述統計(含集中量數,變異量數,相對地位量數), 圖示分析, ROMPER資料分析法, APMO值 2. 分析CV及顯著性考驗,CV值分類, CDI, DC 3. 分析一致性的單變項統計(如kappa係數),Cornbach α 及主成分分析,CVI及CVR, Aiken V和Aiken H
SIMULATECV	產生模擬值並分析CV值分配型態	1. 產生多樣本多變項特定量尺值的模擬資料 2. 產出CV值並分析其分配型態
CLUSTERNUMS	提供集群數評估建議	根據集群距離係數值,分析RSQ,SPRSQ,PSF係數,提供合理分群數建議
QTRANSDAT	轉置對調行與列資料	將原本變項資料加以轉置,對調列和行形成新資料,以進行R型Delphi分析

註：這些程式需要在IBM SPSS Statistics作業平台執行。

再者，為了比較兩輪回或多子群體評者評估結果，除了應用既有比較係數值差異的統計方法外，也提出一些新做法，簡述如下，而ROMPERDELPHI、COMPDELPHI和MULTIDELPHI三個程式就分別納入這些分析：

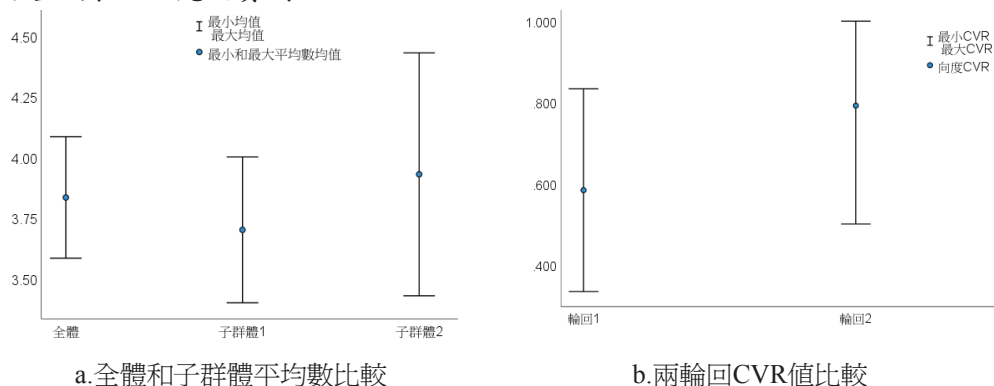
1. 約略區間展示 (rough interval display, RID)

針對單一項目，可比較兩輪回評估值的平均數、標準差、CV值、滿分比、達標比、CVI等，可進行差異考驗或關聯考驗，前者是進行兩個係數值差異或關聯的 t 考驗、 F 考驗或 z 考驗，以達 $p < .05$ 為判斷依據，後者是進行關聯性考驗(r)的 t 考驗。至於多項目時，可視為向度來處理，使用MCV、S-CVI等係數值，做綜合性了解；但是有些統計量數或係數如平均數、標準差、滿分比等，並未見可採用的對應係數或統計量數。對此可使用約略區間展示 (RID) 方法來協助，其是選擇多項目的最小和最大係數或統計量數做為區間的下限 (lower limit, LL) 和上限 (upper limit, UL)，再取最小值和最大值的平均值為中間代表值 (middle representative value, MRV)，或者選擇以向度係數值為代表值，由此三個數值構成區間值分布，其顯示形式如[最小值,中間代表值,最大值]，並繪製成圖，以利

直觀了解。圖 19 顯示應用實例，圖 19a 指出全體的平均值區間最小，表示全體評者評分相對接近，而子群體 2 評者的區間範圍很大，代表評分差異大，評者看法分歧性大；又此處三者區間有重疊，不能貿然說全體和兩子群體平均值沒有差異，需要進一步做個項目的平均數差異考驗來確定，反之若有一群體的區間不予其他群體區間重疊，也就明顯表示該子群體評者評分看法明顯異於其他子群體評者。圖 19b 是以向度係數為代表值，其解讀圖方法與圖 19a 相同，不再贅述。

圖 19

約略區間展示法應用實例



2. 跨樣本項目 Kappa 係數和 Kendall w 係數

Fleiss Kappa κ 係數和 Kendall w 係數常用於考驗同一批評者對同一批受評者的評分差異情形，因此能分析同一向度內多個項目評定看法的不一致情形。如果將兩輪回評者對同一批項目的評估看法，總合為一組資料，再採用 Kappa κ 係數和 Kendall w 係數亦能獲得分析值，此可稱為跨樣本項目 Kappa 係數 (cross samples of Kappa, kcs) 和 Kendall w 係數 (cross samples of Kendall w, wcs)，其係數值表示相依樣本時的 Kappa κ 係數和 Kendall w 係數，即不同情形時兩樣本數值的一致性程度，即計算公式與 Kappa κ 係數和 Kendall w 係數相同，但是對分析值的解釋做法則不同，分析值表示跨樣本或跨時間同樣本時的評分一致性。另一種做法是採用 Cohen Kappa，分析兩輪回評者對同一批項目的評估看法一致性。

3. 異議評者區辨係數和區辨方法

就量化 Delphi 分析取向而言，異議者是指評者的評分表現異於其他一般評者，其表現至少可從兩個角度觀察：一是異議者多評分偏低，Delphi 分析強調共識，評分多指涉對項目重要性、必要性、影響性、合宜性等性質的認定和評估，要想達成共識，應該呈現出評者都評定屬於較高程度，故評分應該呈現出偏向高評分的左偏態評值分配型態；二是異議者多採高分散性評分，即評分無特定偏好評值的情形，形成低闊峰分配，反之一般評者則採取高集中性評分，屬

於高狹峰分配，即多數評分偏向評定量尺值高分那一端。這兩者分別說明偏低評分（low rating）和離散評分（divergent rating）的評值分配樣態，對此能提出偏低評值程度（degree of low rating, DLR）和離散評值程度（degree of divergent rating, DDR）兩係數，DLR 值為 0~100，值越大表示選擇低值比率越高，DDR 值為 0~1，數值越大表示選擇越多不同評值，故這兩者都能顯示出評者評分是否與其他評者評分有共識性偏低的狀態。

據此可將 DLR 和 DDR 聯合用於區辨異議者，選取全體評者在兩係數值的平均為分界值，進而構成四分類的異議評者區分架構，見圖 20 是評者評分作為分類方格（classification grid of rater's rating behavior），簡稱評者分類方格。其分成四類：異議評者（dissensusor）、多值評者（divergent rating rater）、高值評者（higher rating rater）、低值評者（lower rating rater），異議評者是高 DLR 且高 DLR（ $DLR \geq MDLR, DDR \geq MDDR$ ），多值評者是低 DLR 且高 DDR（ $DLR < MDLR, DDR \geq MDDR$ ），高值評者是低 DLR 且低 DDR（ $DLR < MDLR, DDR < MDDR$ ），低值評者係高 DLR 且低 DDR（ $DLR \geq MDLR, DDR < MDDR$ ）。至於分類結果是否正確合理，可投入 DLR 和 DDR 係數值及分類結果，使用 MANOVA、集群分析和區別分析進行複核考驗，若 MANOVA 顯示 Pillai trace 等考驗值達 $p < .05$ 、集群分析指出明顯為 2 分群（即異議者和非異議者）、區別分析表明 4 分類正確率逾 80%，都能視為評者分類結果良好可接受。

$$DLR = \frac{\sum(R_i \leq R_{\text{median}})}{k} \times 100 \quad R_i: \text{評值} \quad R_{\text{median}}: \text{可評值的中間值} \quad k: \text{可用評值範圍}$$

$$DDR = \frac{\sum(R_i = R_{\text{scale}})}{k} \quad R_i: \text{評值} \quad R_{\text{scale}}: \text{可用所有合理評值} \quad k: \text{可用評值範圍}$$

圖 20

評者分類方格及分類標準

高 DDR 低	2 多值評者	1 異議評者
	3 高值評者	4 低值評者
	低	高
	DLR	

簡言之，上述程式提供非常豐富的分析資訊，有關解讀重點和如何選擇和陳述最適資訊於分析報告，表 43 列出一些呈現於 SIMPEDELPHI、ROMPERDELPHI、COMPDELPHI 和 MULTIDELPHI 程式分析結果報表的係數值及其判斷法則供參考，其中標示 * 者為建議在分析報告中應納入說明；至於常態分配考驗及偏離值（outlier）考驗可參看一般論述，此處未列出。基本上，評估項目共識性的係數分析結果反映的訊息（指達共識情形）應該相同或相仿，在極少數情形時可能發生相左，此時應以多數係數的判斷結果為主。

表43

SIMPLEDELPHI等程式進行Delphi分析所得分析係數值或結果之判斷參考法則

係數	係數值判斷	係數或方法	係數值或分析結果判斷
1.共識性考驗		2.信效度考驗	
*平均數	越大越好, 95%CI 不包括 0	*Cronbach α	越大越好, 應>0, 95%CI 不包括 0
*標準數	越小越好, 0 最好, 95%CI 不包括 0	*標準化 α	越大越好, 應<0
*單項 CV	越小越好, 95%CI 包括 0, >1 不佳	樣本校正 α	越大越好, 應>0
CV _s , CV ₂ ,	越小越好, >1 不佳, 0 最好	校正題項和總	越大越好, 應>0
CV _{pop} , CV ²		分相關 r_{it}	
CV _{trim}	越小越好, >1 不佳, 與 CV 值無顯著差異(p>.05)	刪項 α	越大越好, 應>0, 與 α 值達顯著差異
CV _{diff} (DCV)	越小越好, 0 最好, 考驗 p>.05	項目相關 r	越大越好, 應>0 且考驗 p<.05
*MCV	越小越好, 0 最好	Theta (θ)	越大越好
DC, DC _s	越大越好	*主成分數	為 1 最好, >1 不佳
CDI, CDIs,	越小越好	抽取成分解釋	宜>40%
CDIA		變異%	
*DI	越小越好, ≥ 1 不佳	*Kendall w	越大越好, 考驗 p<.05
lnCVR	越小越好, <0 為佳, <1 是輪回 1 的 CV 值優於輪回 2 的 CV 值	*Fliess Kappa	越大越好, 考驗 p<.05, 95%CI 不包括 0
IPRAS, IPR	IPRAS 值>IPR 值佳	Cohen Kappa	越大越好, 考驗 p<.05
CVCI 重疊%	越大越好, 100%最佳, 應>70%	Friedman χ^2	越大越好, 考驗 p<.05
RC	越大越好, 1 最好, 應>0.7	*Aiken V	越大越好, 考驗 p<.05
IQR, rIQR	越小越好	*Aiken H	越大越好, 考驗 p<.05
Q	越小越好	*Cochran Q	越小越好, 考驗 p>.05
*CQV, CR	越小越好	Best S	越小越好, 考驗 p>.05
滿分比	越大越好, 100%最好	*CVI	越大越好, 應達標準值, 為 1 最好
*達標比	越大越好, 100%最好, 應>80%	*CVR	越大越好, 應達標準值, 為 1 最好
DM	越大越好, 1 最好, 為 0 可接受	階層集群分析	1 集群為佳
SA, CSA	越大越好, 1 最好, 應 ≥ 0.4		
評值%堆疊圖	項目高值%越多越好		
評值森林圖	項目評值區間(P30 和 P70)越偏向高值端越好		
同意%,不同意	同意%>APMO 值,或不同意		
%,APMO	%>APMO 值,表示達共識		
VCC, VCC _M	越大越好		

註：這些係數值可見SIMPLEDELPHI、ROMPERDELPHI、COMPDELPHI和MULTIDELPHI程式分析結果報表。MULTIDELPHI程式將lnCVR設定為比較輪回2的CV值對輪回1的CV值。*為建議納入報告的係數值或分析結果。

而圖 21 揭示 Delphi 分析各階段可採用分析係數和分析方法，及能採用於分析的程式供參考，其中欲分析第二和三輪回評估看法變化或差異，可選擇 MULTIDELPHI3M 程式，輸入第二和三輪回評估值即可分析，想比較第一和三輪回評估值依此類推。若要進行 R 型 Delphi 分析，檢視評者的評估看法狀況，可採用 QTRANSDAT 程式。又 c. 項目信效度檢視部分亦可提早於輪回 1 階段之前進行，另邀請多位利害關係人先預試工具的信效度，及做必要的項目調整和敘述潤飾，此做法可邀請較多人員參加預試，滿足進行信效度考驗檢定力需求的人數，較佳於在第一輪回評估時，邀請較少數評者進行信效度考驗，可能衍生樣本數過少、代表性不足和檢定力偏低的疑慮。

圖 21

Delphi分析階段採用分析方法和可用程式對應

Delphi 分析階段	檢視工作	評估係數/分析方法
評者專業性評估 DELPHICALC8M	專業係數檢視	Cs, Ca, Cr, DP
輪回 1: a. 評估值特性檢視 SIMPLEDELPHI7, ROMPERDELPHI, SHOWDELPHI3 b. 共識性評估	評值偏離和分配分析	IQR, 盒狀圖, 偏態, 峰度, JB 考驗, 森林圖, ROMPER 資料分析表
	1. 單項共識性分析	平均數, 標準差, 中數, 眾數, 全距, 滿分比, 達標比, APMO, IQR, rIQR, CQV, CR, CV 及考驗和分類, CV ₂ , CV _s , CV _{pop} , CV _{trim} , CVIC, RC, CDI, CDI _s , DC, DC _s , IPRAS, DI
	2. 向度(多項)共識性分析	MCV, DM, SA, 集群分析, Fleiss Kappa, Kendall w
c. 項目信效度檢視	1. 評分者信度分析	Cochran Q, Best S, Aiken H
	2. 內容效度分析	CVI, CVR, Aiken V
	3. 內部一致性信度分析	相關(r), 題項和總分相關(r _{itc}), Cronbach α, 標準化 α, 樣本數校正 α, 刪項 α, θ 係數
d. 評者特性影響檢視 COMPDELPHI4	4. 構念效度分析 特性影響分析	主成分分析(主成分數, 解釋變異量%) 平均數, 標準差, CV, 滿分比, 達標比, CVI 等 差異考驗(各特性子群體和全體比較)
輪回 2: a.~d. 評估和檢視 e. 共識性差異/變化檢視 MULTIDELPHI3M	見輪回 1 前四階段 1. 共識性差異分析	同輪回 1 前四階段採用係數和分析方法 各項評值相同和不同數, 評值增減數, 關聯性考驗(r, C), 獨立性考驗, 各項平均數, 標準差, CV, 滿分比, 達標比, CVI 等差異考驗 (兩輪回比較), lnCVR, Cohen Kappa, Kendall w, Friedman χ ² 向度最大平均數, 標準差, CV, 滿分比, 達標 比, 向度 CVI, Cronbach α 等差異考驗(兩 輪回比較), 全部項目評值相關(r)考驗 Bowker 改變考驗
輪回 k: 評估值特性, 共識性, 評者 特性影響, 信效度檢視	2. 共識性變化分析	同上

伍、結論與建議

綜合前述討論，提出以下結論和建議：

一、結論

綜觀前述討論可知Delphi分析自RAND公司提出之後，不僅廣受採用，且與之有關的創新構想和做法不斷湧現，形成頗為龐大的方法體系，本文已介紹一二，但仍未窮盡說明全貌，有待日後更深入探析，儘管如此仍取得一些結論如下：

1. Delphi 分析體系儼然形成，但仍待續擴展和綜析

過去已累積龐雜的 Delphi 分析研究和討論成果，難以一窺全貌，本研究試圖構建一個廣義的研究體系（圖 12）及探討焦點架構（表 31）、應用流程（圖 1）、統計方法應用架構（圖 13）、適用統計方法和評估係數（表 43），足供研究者充分應用多元統計方法和評估係數以豐富和嚴謹 Delphi 分析結果及品質。

2. Delphi 分析遭遇涉及評者專業、評估工具有效、應用嚴謹、發展生態四類問題有待改善

總結和檢視應用文獻，本研究提出關係微觀和宏觀層面的八項問題，觸及評者專業、評估工具有效、應用嚴謹、發展生態四類課題，其彼此關聯和影響，影響 Delphi 分析合理性和效益甚大，必須重視。其中在微觀層面的七項問題討論，不僅提出諸多細項詳細討論，並輔以模擬分析和實際資料分析取檢視問題的核心事項，提供比單純論述更客觀的嚴謹分析數據做為佐證，進而破除一些迷思，提出頗多可採用的解決策略和適用的分析指數或係數，利於應用者更加有效運用 Delphi 分析，也提出研究者更多可思考進一步處理的待解問題。

3. 採用多元思維角度和多元方法及措施去有效處理遭遇問題

針對 Delphi 分析遭遇問題，提出從理論建構、架構和標準研提、流程調整、方法調整和創新、工具改良和創造等方向達 17 項的處理思維或改善做法，其內容具體且可應用性高，將有效降低既有問題的衝擊和影響，並大幅減少當前研究者做法與研究前沿成果的代際落差。其中創新提出諸多新評估分析品質的方法（如以 bootstrap 法分析 CV 等）和係數（如 MCV），補充既有可用分析工具的不足，而針對現有分析程式和軟體缺乏的困境，設計多個能提供眾多分析資訊的程式，經實際資料檢驗，發現確實有高實效，能大量降低分析成本（包括分析時間、分析應備知能）和提出具參考價值的分析品質評估資訊及證據。

二、研究貢獻

本研究有頗多突破過去研究和介紹侷限之處，說明如下：

1. 較完整彙總和提出系統化 Delphi 分析發展成果和運作架構、

過去關於 Delphi 分析的研討論述甚多，且涉及層面甚廣，包括分析取向、分析實務、遭遇問題、成果評論等，本文嘗試彙整和歸納出一些發展脈絡及系統性成果，初步提出分析體系和運作流程架構，以更完善反映既有的發展情形，並提出若干創思如 Q 型 Delphi 分析取向等，創新實務運作和理論。

2.實證檢視Delphi分析應用問題關鍵，以引介前人智慧、創新知和計量考驗可用改善做法

既有應用Delphi分析論述和應用存在與研究成果明顯落差、誤解、誤用和乏用分析係數和方法、分析成本偏高等問題，阻礙提升其運作品質和效益。本文配合資料實證和模擬考驗，逐一討論和釐清問題涉及的觀念和技術（如CVs、%CV、CDI），除引介既有可用做法（如MCV），亦新創不少思維與分析係數和方法，如辨識異議者及分類、採bootstrap法考驗共識性評估係數等，可大幅度改善現存問題。

3. 提出多個分析程式，降低分析成本和提升效益

共識性評估是Delphi分析的核心工作，結合眾多適用的統計分析是成功關鍵，本研究設計可於SPSS平臺運作的多個Delphi分析適用程式，廣納眾多逾42個共識性評估和考驗信效度方法及係數，適合多種情境需要，容易使用且分析快速，大幅降低分析成本和提升效益及確保品質。

三、建議

以下提出幾項建議供參考：

1. 綜析和提出 Delphi 分析取向類型理論，構建分析理論和應用體系及運作規範

隨著Delphi分析發展快速，前述討論可知已衍生紛雜的分析取向和方法，並不有利於理解和應用。為有效發揮使用效益，構建完善的分析理論和應用體系、及應遵行的運作規範和倫理實屬急迫和重要。在分析理論體系方面，包括依據學理、分析取向、測量方法、評估共識性方法和標準及等級分類、異議者區辨等，此部分的共識性評估通常依賴量化統計研究成果，但是有關係數和分析方法眾多且不斷提出，亟待檢驗分析成效和推介最適最優做法。在應用體系方面，涉及評選評者、選任諮詢者和協調者、選用評定量尺值、評項和評者信效度檢驗、偏離值取捨、選擇實施平臺（如郵寄、線上）、評估意見回饋與處理、異議評者汰留等諸多繁瑣課題，多數尚未取得最佳做法的共識，有待設計實驗進行考驗和比較，以提出嚴謹、客觀和合理的支持證據。至於運作規範和倫理方面，既有提出的運作規範看法亦屬紛雜，基本上訂定運作規範與分析流程所包含各環節工作項目有關，後者決定應遵行的運作規範、需注意的倫理議題和規範、及待檢核項目，這些均尚待統整和取得共識。

2. 善用合適程式進行 Delphi 分析，以最大化使用效益

欲完善Delphi分析，除採行嚴謹流程外，關鍵在於提供最適且最佳分析資訊，主要包括評者專業性、項目信效度、評項共識性、成果合宜性等四類資訊，各運作階段適用分析方法和係數可見前述提出的參考資料（圖18和表

41) 說明。基本上，進行此分析必須提供項目評分者信度、內容效度，最好加上跨樣本效度、內部一致性信度提供，若是在建構指標，則必須提供內部一致性信度、構念效度資訊。對此，屬於一般用途時宜使用 ROMPERDELPHI 程式，若採用 SIMPLEDELPHI7 程式能提供更豐富的考驗資訊，而同時應使用 DELPHICALC8M 程式說明評者專業性、以 COMPDELPHI4 程式說明跨樣本效度，如果要確認兩輪回評估看法的穩定性（可視為具有重測信度性質），應該使用 MULTIDELPHI3M 程式。

此外，隨著可考驗共識性的資料統計方法和係數不斷地推出，有關適用分析程式提供功能也應一併隨之提升和改進，而本文所提分析程式主要針對採取 Likert 量尺值的 Delphi 分析，而針對使用模糊集 (fuzzy set) 的模糊 Delphi 分析其適合分析程式不多 (Chan, Tan, & Mazlam, 2025)，後續有納入研發的必要。

3. 研議更多共識性評估方法和係數，考驗其效果，納入 Delphi 分析，並設計適用程式

前述已論及甚多共識性評估係數和方法，但仍有遺漏，如 Krippendorff alpha (α) (Krippendorff, 2019)、Gwet AC1 和 AC2 (Gwet, 2021)、G-indices (Bonett, 2022) 等係數可用於評估內容效度或評分者信度、基於 Bayesian 取向的 CV 係數、基於指數、t、Weibull 等分配型態的 CV 係數考驗方法、四分變異係數 (coefficient of quartile variation, CQV) (Bonett, 2006) 或稱四分離散係數 (quartile coefficient of dispersion, QCD) (Botta-Dukát, 2023)、全距係數 (coefficient of range, CR) 等，均益於提供更多共識性程度的資訊，但少見應用於 Delphi 分析，乃肇因於分析軟體不支援分析功能。因此，後續宜針對這些適用方法和係數，探討其效果和設計分析程式，以普及採用。

4. 推介具使用價值的 Delphi 分析方法

創新 Delphi 分析可謂日新月異，頗多新方法甚具應用價值，尤其 RAND 公司後續發展的 RAM 法和 ROMPER 法著重增進運作效率和分析嚴謹性，甚具採用效益，並有論者推出修正型構想，但是有關的推介論述卻不多，有礙參考使用及教學；另有其他線上分析系統亦有參考價值，有待推介。為豐富 Delphi 分析生態系統，積極做法是撰著論述以推介有應用潛力的分析方法，並評論該法的優劣和限制，以提供全面的了解。

參考文獻

陳青山、王聲湧、董曉梅、池桂波、荊春霞和朱麗 (2004)。在 Excel 中完成 Delphi 法評價指標的計算。《數理醫藥學雜誌》，17(1)，73-76。



- [Chen, Q., Wang, S., Dong, X., (2004). Calculation for the Delphi indices in EXCEL. *Journal of Mathematical Medicine*, 17(1), 73-76.]
- 葉連祺 (2018)。教育行政學位論文應用計量分析方法改善及相關量化分析發展。學
校行政，116，147-211。https://doi.org/10.6423/HHHC.201807_(116).0007
- [Yeh, L.-C. (2018). Improving on using quantitative analytic methods in educational
administration theses and the developments of related quantitative analysis. *School
Administrators*, 116, 147-211. https://doi.org/10.6423/HHHC.201807_(116).0007]
- 鄧振源 (2012)。多準則決策分析：方法與應用。鼎茂。
- [Teng, J.-Y. (2012). *Multi-criteria decision analysis: Methods and applications*.
Tingmao.]
- Abdi, H. (2010). Coefficient of variation. *Encyclopedia of Research Design*, 1(5),
169-171.
- Aerts, S., & Haesbroeck, G. (2017). Robust asymptotic tests for the equality of multi-
variate coefficients of variation. *TEST*, 26, 163-187. https://doi.org/10.1007/s11749-
016-0504-4
- Aerts, S., Haesbroeck, G., & Ruwet, C. (2015). Multivariate coefficients of variation:
Comparison and influence functions. *Journal of Multivariate Analysis*, 142,
183-198. https://doi.org/10.1016/j.jmva.2015.08.006
- Aerts, S., Haesbroeck, G., & Ruwet, C. (2018). Distribution under elliptical symmetry
of a distance-based multivariate coefficient of variation. *Statistical Papers*, 59(2),
545-579. https://doi.org/10.1007/s00362-016-0777-4
- Albatineh, A. N., Kibria, B. G., & Zogheib, B. (2014). Asymptotic sampling distribution of
inverse coefficient of variation and its applications: revisited. *International Journal
of Advanced Statistics and Probability*, 2, 15-20. https://doi.org/10.14419/ijasp.
v2i1.1475
- Albert, A., & Zhang, L. (2010). A novel definition of the multivariate coefficient of varia-
tion. *Biometrical Journal*, 52(5), 667-675. https://doi.org/10.1002/bimj.201000030
- Anderson, G. (2023). *A coefficient of variation for multivariate ordered categorical
outcomes*. University of Toronto, Department of Economics.
- Arachchige, C. N., Prendergast, L. A., & Staudte, R. G. (2022). Robust analogs to the
coefficient of variation. *Journal of Applied Statistics*, 49(2), 268-290. http://doi.org
/10.1080/02664763.2020.1808599
- Banerjee, A. K. (2018). Multidimensional indices with data-driven dimensional weights:

- A multidimensional coefficient of variation. *Arthaniti: Journal of Economic Theory and Practice*, 17(2), 140–156. <https://doi.org/10.1177/0976747918792644>
- Banerjee, A. K. (2019). *A class of multidimensional coefficients of variation*.
- Banik, S., Kibria, B. M., & Sharma, D. (2012). Testing the population coefficient of variation. *Journal of Modern Applied Statistical Methods*, 11(2), 325–335. <https://doi.org/10.22237/jmasm/1351742640>
- Barrios, M., Guilera, G., Nuño, L., & Gómez-Benito, J. (2021). Consensus in the delphi method: What makes a decision change?. *Technological Forecasting and Social Change*, 163, 120484. <https://doi.org/10.1016/j.techfore.2020.120484>
- Bedeian, A. G., & Mossholder, K. W. (2000). On the use of the coefficient of variation as a measure of diversity. *Organizational Research Methods*, 3(3), 285–297. <https://doi.org/10.1177/109442810033005>
- Beiderbeck, D., Frevel, N., von der Gracht, H. A., Schmidt, S. L., & Schweitzer, V. M. (2021). Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*, 8, 101401. <https://doi.org/10.1016/j.mex.2021.101401>
- Belton, I., MacDonald, A., Wright, G., & Hamlin, I. (2019). Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process. *Technological Forecasting and Social Change*, 147, 72–82. <https://doi.org/10.1016/j.techfore.2019.07.002>
- Birko, S., Dove, E. S., & Özdemir, V. (2015). Evaluation of nine consensus indices in Delphi foresight research and their dependency on Delphi survey characteristics: A simulation study and debate on Delphi design and interpretation. *PloS one*, 10(8), e0135162. <https://doi.org/10.1371/journal.pone.0135162>
- Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis*, 50(11), 2953–2957. <https://doi.org/10.1016/j.csda.2005.05.007>
- Bonett, D. G. (2022). Statistical inference for G-indices of agreement. *Journal of Educational and Behavioral Statistics*, 47(4), 438–458. <https://doi.org/10.3102/107699862210885>
- Botta-Dukát, Z. (2023). Quartile coefficient of variation is more robust than CV for traits calculated as a ratio. *Scientific Reports*, 13(1), 4671. <https://doi.org/10.1038/s41598-023-31711-8>



- Costa, N. H. D. A. D., Seraphin, J. C., & Zimmermann, F. J. P. (2002). Novo método de classificação de coeficientes de variação para a cultura do arroz de terras altas. *Pesquisa Agropecuária Brasileira*, 37, 243-249. <https://doi.org/10.1590/S0100-204X2002000300003>
- Cottam, H., Roe, M., & Challacombe, J. (2004). Outsourcing of trucking activities by relief organisations. *Journal of Humanitarian Assistance*, 1(1), 1-26. <http://www.jha.ac/articles/a130.pdf>
- Couto, M. F., Peternelli, L. A., & Barbosa, M. H. P. (2013). Classification of the coefficients of variation for sugarcane crops. *Ciência Rural*, 43, 957-961. <https://doi.org/10.1590/S0103-84782013000600003>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- Dalkey, N. C. (1967). *Delphi*. RAND Corporation. <https://www.rand.org/pubs/papers/P3704.html>
- Dalkey, N. C. (1969). *The Delphi method: An experimental study of group opinion*.
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458-467. <https://doi.org/10.1287/mnsc.9.3.458>
- Day, J., & Bobeva, M. (2005). A generic toolkit for the successful management of Delphi studies. *Electronic Journal of Business Research Methods*, 3(2), 103-116.
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67(4), 401-409. <https://doi.org/10.1016/j.jclinepi.2013.12.002>
- Dias, C., & Santos, C. (2021). Note on the coefficient of variation properties. *Brazilian Electronic Journal of Mathematics*, 2(4), 101-111. <https://doi.org/10.14393/BEJOM-v2-n4-2021-58062>
- Dodd, S. C. (1952). On percentage moments. *The Scientific Monthly*, 74(4), 220-222. <https://www.jstor.org/stable/20409>
- Dragostinov, Y., Harðardóttir, D., McKenna, P. E., Robb, D. A., Nessel, B., Ahmad, M. I., ... & Rajendran, G. (2022). Preliminary psychometric scale development using the mixed methods Delphi technique. *Methods in Psychology*, 7, 100103. <https://doi.org/10.1016/j.metip.2022.100103>

- Ferreira, A. A. S. N. de C., Dourado, L. R. B., Biagiotti, D., Santos, N. P. da S., Nascimento, D. C. N., & Sousa, K. R. S. (2018). Methods for classifying coefficients of variation in experimentation with poultrys. *Comunicata Scientiae*, *9*(4), 565-574.
- Ferreira, J. P., Schimldt, E. R., Schimldt, O., Cattaneo, L. F., Alexandre, R. S., & Cruz, C. D. (2016). Comparison of methods for classification of the coefficient of variation in papaya. *Revista Ceres*, *63*(2), 138-144. <https://doi.org/10.1590/0034-737X201663020004>
- Firth, A. M., O'Brien, S. M., Guo, P., Seymour, J., Richardson, H., Bridges, C., ... & Murtagh, F. E. (2019). Establishing key criteria to define and compare models of specialist palliative care: A mixed-methods study using qualitative interviews and Delphi survey. *Palliative Medicine*, *33*(8), 1114-1124. <https://doi.org/10.1177/0269216319858237>
- Fitch, K., Bernstein, S. J., Aguilar, M. D., Burnand, B., & LaCalle, J. R. (2001). *The RAND/UCLA appropriateness method user's manual* (No. RANDMR1269DGXIIRE).
- Fitch, K., Bernstein, S. J., Aguilar, M. D., Burnand, B., LaCalle, J. R., Lazaro, P., ... & Kahan, J. P. (2000). *The RAND/UCLA appropriateness method user's manual*. RAND Corporation.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). John Wiley & Sons.
- Forkman, F. J. (2005). *Coefficients of variation-An approximate F-test*. Licentiate Thesis. Swedish University of Agricultural Sciences.
- Forkman, J. (2009). Estimator and tests for common coefficients of variation in normal distributions. *Communications in Statistics: Theory and Methods*, *38*(2), 233-251. <http://dx.doi.org/10.1080/03610920802187448>
- Foth, T., Efstathiou, N., Vanderspank-Wright, B., Ufholz, L. A., Dütthorn, N., Zimansky, M., & Humphrey-Murto, S. (2016). The use of Delphi and Nominal Group Technique in nursing education: A review. *International Journal of Nursing Studies*, *60*, 112-120. <https://doi.org/10.1016/j.ijnurstu.2016.04.015>
- Fung, W. K., & Tsang, T. S. (1998). A simulation study comparing tests for the equality of coefficients of variation. *Statistics in Medicine*, *17*(17), 2003-2014. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980915\)17:17<2003::AID-SIM889>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0258(19980915)17:17<2003::AID-SIM889>3.0.CO;2-I)
- George, D., & Mallery, P. (2019). *IBM SPSS Statistics 25 step by step: A simple guide*



- and reference* (14th ed.). Routledge.
- Giannarou, L., & Zervas, E. (2014). Using Delphi technique to build consensus in practice. *International Journal of Business Science & Applied Management*, 9(2), 65–82. <https://hdl.handle.net/10419/190657>
- Gil, M. Á., Lubiano, M. A., Montenegro, M., & López, M. T. (2002). Least squares fitting of an affine function and strength of association for interval-valued data. *Metrika*, 56(2), 97–111. <https://doi.org/10.1007/s001840100160>
- Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9(3), 330–338. <https://doi.org/10.1016/j.sapharm.2012.04.004>
- Gluschenko, K. (2018). Measuring regional inequality: To weight or not to weight?. *Spatial Economic Analysis*, 13(1), 36–59. <https://doi.org/10.1080/17421772.2017.1343491>
- Gohres, H., & Kolip, P. (2023). Modified Delphi process to identify recommendations for action for the structural development of physical activity promotion in Germany. In M. Niederberger & O. Renn (Eds.), *Delphi methods in the social and health sciences concepts, applications and case studies* (pp.167–188). Springer.
- Grant, S., & Smart, R. (2024). ROMPER: The RAND/USC OPTIC method for policy expert ratings. *MethodsX*, 12, 102751. <https://doi.org/10.1016/j.mex.2024.102751>
- Gulhar, M., Kibria, G., Albatineh, A., & Ahmed, N. U. (2012). A comparison of some confidence intervals for estimating the population coefficient of variation: A simulation study. *Sort*, 36(1), 45–68.
- Gupta, R. C., & Ma, S. (1996). Testing the equality of coefficients of variation in k normal populations. *Communications in Statistics-Theory and Methods*, 25(1), 115–132. <https://doi.org/10.1080/03610929608831683>
- Guthrie, D. M., Williams, N., Beach, C., Buzath, E., Cohen, J., Declercq, A., ... & Martin, T. L. W. (2022). A multi-stage process to develop quality indicators for community-based palliative care using interRAI data. *PloS One*, 17(4), e0266569. <https://doi.org/10.1371/journal.pone.0266569>
- Gwet, K. L. (2021). *Handbook of inter-rater reliability* (5th ed.). STATAxis Publishing.
- Hallowell, M. R., & Gambatese, J. A. (2010). Qualitative research: Application of the Delphi method to CEM research. *Journal of construction engineering and man-*

- agement, 136(1), 99-107. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000137](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000137)
- Hanafin, S. (2004). Review of literature on the Delphi technique (2nd ed.). *National Children's Office*.
- Hart, J. (2012). Using basic statistics on the individual patient's own numeric data. *Journal of Chiropractic Medicine, 11*(4), 306-309. <https://doi.org/10.1016/j.jcm.2012.10.008>
- Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change, 78*(9), 1695-1704. <https://doi.org/10.1016/j.techfore.2011.04.005>
- Hasson., F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing 32*(4), 1008-1015. <https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x>
- Hasson, F., Keeney, S., & McKenna, H. (2025). Revisiting the Delphi technique-research thinking and practice: A discussion paper. *International Journal of Nursing Studies, 105*119. <https://doi.org/10.1016/j.ijnurstu.2025.105119>
- Holey, E. A., Feeley, J. L., Dixon, J., & Whittaker, V. J. (2007). An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Medical Research Methodology, 7*(52). <https://doi.org/10.1186/1471-2288-7-52>
- Hsiao, W. F., Lin, H. H., & Chang, T. M. (2008). Fuzzy consensus measure on verbal opinions. *Expert Systems with Applications, 35*(3), 836-842. <https://doi.org/10.1016/j.eswa.2007.07.040>
- Ishikawa, K. (1974). *Guide to quality control*. Asian Productivity Organization.
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*(1), 85-98. <https://doi.org/10.1037/0021-9010.69.1.85>
- Jayakumar, D. S., Kan, S. A., & Wilfred, S. (2021). On using the exact sampling distribution of multivariate coefficient of variation. *Thailand Statistician, 19*(3), 511-521. Retrieved from <https://ph02.tci-thaijo.org/index.php/thaistat/article/view/244514>
- Jorm, A. (2025). *Using the Delphi method to establish expert consensus: A practical guide*. Palgrave Macmillan.
- Jorm, A. F. (2015). Using the Delphi expert consensus method in mental health research. *Australian & New Zealand Journal of Psychiatry, 49*(10), 887-897. <https://doi.org/10.1177/0004867415600891>



- Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017). Guidance on Conducting and REporting DELphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine*, 31(8), 684–706. <https://doi.org/10.1177/02692163176906>
- Kalaian, S. A., & Kasim, R. M. (2012). Terminating sequential Delphi survey data collection. *Practical Assessment, Research & Evaluation*, 17(5), n5.
- Kalkur, T. A., & Rao, K. A. (2015). Pairwise comparison of coefficients of variation for correlated samples. *International Journal of Statistics and Applications*, 5(5), 231–236. <https://doi.org/10.5923/j.statistics.20150505.07>
- Kaufmann, P. R. (2016). Integrating factor analysis and the Delphi method in scenario development: A case study of Dalmatia, Croatia. *Applied Geography*, 71, 56–68. <https://doi.org/10.1016/j.apgeog.2016.04.007>
- Khodyakov, D., Grant, S., Kroger, J., & Bauman, M. (2023). *RAND methodological guidance for conducting and critically appraising Delphi panels*. RAND Corporation.
- Klenk, N. L., & Hickey, G. M. (2011). A virtual and anonymous, deliberative and analytic participation process for planning and evaluation: The concept mapping policy Delphi. *International Journal of Forecasting*, 27(1), 152–165. <https://doi.org/10.1016/j.ijforecast.2010.05.002>
- Kobus, J., & Westner, M. (2016, April). Ranking-type Delphi studies in IS research: Step-by-step guide and analytical extension. In *IADIS International Conference Information System* (pp. 28–38).
- Kraines, M. A., Uebelacker, L. A., Gaudiano, B. A., Jones, R. N., Beard, C., Loucks, E. B., & Brewer, J. A. (2020). An adapted Delphi approach: The use of an expert panel to operationally define non-judgment of internal experiences as it relates to mindfulness. *Complementary Therapies in Medicine*, 51, 102444. <https://doi.org/10.1016/j.ctim.2020.102444>
- Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE publications.
- Krishnamoorthy, K., & Lee, M. (2014). Improved tests for the equality of normal coefficients of variation. *Computational Statistics*, 29(1), 215–232. <https://doi.org/10.1007/s00180-013-0445-2>
- Kvålseth, T. O. (2017). Coefficient of variation: The second-order alternative. *Journal of Applied Statistics*, 44(3), 402–415. <https://doi.org/10.1080/02664763.2016.1174195>

- Landeta, J., & Lertxundi, A. (2024). Quality indicators for Delphi studies. *Futures & Foresight Science*, 6(1), e172. <https://doi.org/10.1002/ffo2.172>
- Linstone, H. A., & Turoff, M. (Eds.). (2002). *The Delphi method: Techniques and applications*. <http://is.njit.edu/pubs/delphibook/delphibook.pdf>.
- Lopes, B. G., Faria, G. A., Maltoni, K. L., Rocha, P. S., Peixoto, A. P. B., Oliveira, T. A. D., ... & Felizardo, L. M. (2021). Classification of the coefficient of variation for experiments with eucalyptus seedlings in greenhouse. *Revista Ciência Agronômica*, 52, e20207587. <https://doi.org/10.5935/1806-6690.20210050>
- Lorenzo, J. C., Yabor, L., Medina, N., Quintana, N., & Wells, V. (2015). Coefficient of variation can identify the most important effects of experimental treatments. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 43(1), 287-291. <https://doi.org/10.15835/nbha4319881>
- Lubiano, M.A., Montenegro, M., Pérez-Fernández, S., & Gil, M.Á. (2023). Analyzing the influence of the rating scale for items in a questionnaire on Cronbach coefficient alpha. In N. Balakrishnan, M. Á. Gil, N. Martín, D. Morales, & M. d. C. Pardo (Eds.), *Trends in mathematical, information and data sciences. studies in systems, decision and control* (Vol. 445, pp. 377-388). Springer. https://doi.org/10.1007/978-3-031-04137-2_32
- MacNeil, A. J. (1997). Case study method and electronic meeting system for principal preparation. In J. Willis, J. D. Price, S. McNeil, B. Robin, & D. A. Willis (Eds.), *Technology and teacher education annual* (pp. 247-251). (ED421921). ERIC. <https://files.eric.ed.gov/fulltext/ED421921.pdf#page=247>
- Martin, J. D., & Gray, L. N. (1971). Measurement of relative variation: Sociological examples. *American Sociological Review*, 496-502. <https://doi.org/10.2307/2093089>
- McKay, A. T. (1932). Distribution of the coefficient of variation and the extended "t" distribution. *Journal of the Royal Statistical Society*, 95(4), 695-698. <https://doi.org/10.2307/2342041>
- Meijering, J. V., Kampen, J. K., & Tobi, H. (2013). Quantifying the development of agreement among experts in Delphi studies. *Technological Forecasting and Social Change*, 80(8), 1607-1614. <https://doi.org/10.1016/j.techfore.2013.01.003>
- Miller, G. E. (1991). Use of the squared ranks test to test for the equality of the coefficients of variation. *Communications in Statistics-Simulation and Computation*, 20(2-3), 743-750. <https://doi.org/10.1080/03610919108812981>



- Naisola-Ruiter, V. (2022). The Delphi technique: A tutorial. *Research in Hospitality Management, 12*(1), 91–97. <https://www.ajol.info/index.php/rhm/article/view/230180>
- Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta-analysis of variation: Ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution, 6*(2), 143–152. <https://doi.org/10.1111/2041-210X.12309>
- Niederberger, M., & Renn, O. (2023). The group Delphi process in the social and health sciences . In M. Niederberger & O. Renn (Eds.), *Delphi methods in the social and health sciences concepts, applications and case studies* (pp.75–92). Springer.
- Niederberger, M., & Sonnberger, M. (2025). The participation of lifeworld experts in Delphi processes: A reflection on method and practice. *MethodsX, 14*, 103274. <https://doi.org/10.1016/j.mex.2025.103274>
- Niederberger, M., & Spranger, J. (2020). Delphi technique in health sciences: A map. *Frontiers in Public Health, 8*, 457. <https://doi.org/10.3389/fpubh.2020.00457>
- Niederberger, M., Schifano, J., Deckert, S., Hirt, J., Homberg, A., Köberich, S., ... & DEWISS network. (2024). Delphi studies in social and health sciences-Recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. *PLoS One, 19*(8), e0304651. <https://doi.org/10.1371/journal.pone.0304651>
- Päivärinta, T., Pekkola, S., & Moe, C. (2011). *Grounding theory from Delphi studies*. Thirty Second International Conference on Information Systems, Shanghai. <https://aisel.aisnet.org/icis2011/proceedings/researchmethods/4>
- Paré, G., Cameron, A. F., Poba-Nzaou, P., & Templier, M. (2013). A systematic assessment of rigor in information systems ranking-type Delphi studies. *Information & Management, 50*(5), 207–217. <https://doi.org/10.1016/j.im.2013.03.003>
- Pazilah, F. N., Hashim, H., & Yunus, M. M. (2024). Refining the TPACK framework: A fuzzy Delphi approach to 21st-century competency and self-efficacy constructs. *Cogent Education, 11*(1), 2428886. <https://doi.org/10.1080/2331186X.2024.2428886>
- Pélabon, C., Armbruster, W. S., & Hansen, T. F. (2011). Experimental evidence for the Berg hypothesis: Vegetative traits are more sensitive than pollination traits to environmental variation. *Functional Ecology, 25*(1), 247–257. <https://doi.org/10.1111/j.1365-2435.2010.01770.x>

- Potas, N., & Gamgam, H. (2019). Testing the population inverse-coefficients of variation and its application. *Eskişehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering*, *20*(4), 503–514. <https://doi.org/10.18038/estubtda.574511>
- Pourabdollah, A. (2020, July). Fuzzy number value or defuzzified value; Which one does it better?. In *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1–6). IEEE. <https://doi.org/10.1109/FUZZ48607.2020.9177533>
- Quan, H., & Shih, W. J. (1996). Assessing reproducibility by the within-subject coefficient of variation with random effects models. *Biometrics*, *52*(4), 1195–1203. <https://doi.org/10.2307/2532835>
- Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 73–79. <https://doi.org/10.1002/widm.2>
- Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375. [https://doi.org/10.1016/S0169-2070\(99\)00018-7](https://doi.org/10.1016/S0169-2070(99)00018-7)
- San-Jose, L., & Retolaza, J. L. (2016). Is the Delphi method valid for business ethics? A survey analysis. *European Journal of Futures Research*, *4*(1), 19. <https://doi.org/10.1007/s40309-016-0109-x>
- Schifano, J., & Niederberger, M. (2025). How Delphi studies in the health sciences find consensus: A scoping review. *Systematic Reviews*, *14*, 14. <https://doi.org/10.1186/s13643-024-02738-3>
- Schober, P., Mascha, E. J., & Vetter, T. R. (2021). Statistics from A (agreement) to Z (z score): A guide to interpreting common measures of association, agreement, diagnostic accuracy, effect size, heterogeneity, and reliability in medical research. *Anesthesia & Analgesia*, *133*(6), 1633–1641. <https://doi.org/10.1213/ANE.0000000000005773>
- Sekayi, D., & Kennedy, A. (2017). Qualitative Delphi method: A four round process with a worked example. *The Qualitative Report*, *22*(10), 2755–2763. Retrieved from <https://nsuworks.nova.edu/tqr/vol22/iss10/15>
- Seker, S. E. (2015). Computerized argument Delphi technique. *IEEE Access*, *3*, 368–380. <https://doi.org/10.1109/ACCESS.2015.2424703>
- Senior, A. M., Viechtbauer, W., & Nakagawa, S. (2020). Revisiting and expanding the



- meta-analysis of variation: The log coefficient of variation ratio. *Research Synthesis Methods*, 11(4), 553-567. <https://doi.org/10.1002/jrsm.1423>
- Shafer, N. J., & Sullivan, J. A. (1986). A simulation study of a test for the equality of the coefficients of variation. *Communications in Statistics-Simulation and Computation*, 15(3), 681-695. <https://doi.org/10.1080/03610918608812532>
- Sharma, K. K., & Krishna, H. (1994). Asymptotic sampling distribution of inverse coefficient-of-variation and its applications. *IEEE Transactions on Reliability*, 43(4), 630-633. <https://doi.org/10.1109/24.370217>
- Sheret, M. (1984). The coefficient of variation: Weighting considerations. *Social Indicators Research*, 15(3), 289-295. <https://doi.org/10.1007/BF00668675>
- Shoukri, M. M. (2011). *Measures of interobserver agreement and reliability* (2nd ed.). CRC press.
- Shoukri, M. M., Colak, D., Kaya, N., & Donner, A. (2008). Comparison of two dependent within subject coefficients of variation to evaluate the reproducibility of measurement devices. *BMC Medical Research Methodology*, 8(1), 24. <https://doi.org/10.1186/1471-2288-8-24>
- Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6(1), 1-21.
- Smithson, M. (1982). On relative dispersion: A new solution for some old problems. *Quality and Quantity*, 16(3), 261-271. <https://doi.org/10.1007/BF00144077>
- Sohst, R. R., Acostamadiedo, E., & Tjaden, J. (2023). Reducing uncertainty in Delphi surveys. *Demographic Research*, 49, 983-1020. <https://www.jstor.org/stable/48754931>
- Spranger, J., Homberg, A., Sonnberger, M., & Niederberger, M. (2022). Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 172, 1-11. <https://doi.org/10.1016/j.zefq.2022.04.025>
- Stine, R. (1989). An introduction to bootstrap methods: Examples and ideas. *Sociological Methods & Research*, 18(2-3), 243-291. <https://doi.org/10.1177/0049124189018002003>
- Strasser, A. (2017). Delphi method variants in information systems research: Taxonomy development and application. *Electronic Journal of Business Research Methods*, 15(2), 120-133.
- Suominen, A., Hajikhani, A., Ahola, A., Kurogi, Y., & Urashima, K. (2022). A quantitative

- and qualitative approach on the evaluation of technological pathways: A comparative national-scale Delphi study. *Futures*, *140*, 102967. <https://doi.org/10.1016/j.futures.2022.102967>
- Tapio, P. (2003). Disaggregative policy Delphi: Using cluster analysis as a tool for systematic scenario formation. *Technological Forecasting and Social Change*, *70*(1), 83-101. [https://doi.org/10.1016/S0040-1625\(01\)00177-9](https://doi.org/10.1016/S0040-1625(01)00177-9)
- Thangjai, W., Niwitpong, S. A., & Niwitpong, S. (2021). A Bayesian approach for estimation of coefficients of variation of normal distributions. *Sains Malaysiana*, *50*(1), 261-278. <http://dx.doi.org/10.17576/jsm-2021-5001-25>
- Themistocleous, C., Pagiaslis, A., Smith, A., & Wagner, C. (2019). A comparison of scale attributes between interval-valued and semantic differential scales. *International Journal of Market Research*, *61*(4), 394-407. <https://doi.org/10.1177/147078531983122>
- Tovohery, J. M., André, T., & Rajaonasy, F. D. (2022). The distribution law of Sharma's coefficient of variation. *International Journal of Science and Research*, *11*(1), 817-821. <http://dx.doi.org/10.21275/SR22114164739>
- van Grootven, B., McNicoll, L., Mendelson, D. A., Friedman, S. M., Fagard, K., Milisen, K., ... & Deschodt, M. (2018). Quality indicators for in-hospital geriatric co-management programmes: A systematic literature review and international Delphi study. *BMJ Open*, *8*(3), e020617. <https://doi.org/10.1136/bmjopen-2017-020617>
- van Oest, R. (2019). A new coefficient of interrater agreement: The challenge of highly unequal category proportions. *Psychological Methods*, *24*(4), 439-451. <https://doi.org/10.1037/met0000183>
- Verburg, K. (2022). *The proof of the pudding: Using outcome-based quality indicators in physical therapy*. <https://hdl.handle.net/2066/252888>
- von der Gracht, H. A. (2012). Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting and Social Change*, *79*(8), 1525-1536. <https://doi.org/10.1016/j.techfore.2012.04.013>
- Zartha Sossa, J. W., Halal, W., & Hernandez Zarta, R. (2019). Delphi method: Analysis of rounds, stakeholder and statistical indicators. *Foresight*, *21*(5), 525-544. <https://doi.org/10.1108/FS-11-2018-0095>

附錄一 Delphi分析可用評估題項示例

大抵上可知有多種調查看法、態度等適用的量尺測量形式（Lubiano, Montenegro, Pérez-Fernández, Gil, 2023; Themistocleous, Pagiaslis, Smith, & Wagner, 2019），包括：Likert量尺（Likert scale, LS）、區間值量尺（interval-valued scale, IVS）、語意差異量尺（semantic differential scale, SDS）、區間一致取向（interval agreement approach, IAA）量尺、視覺類比量尺（visual analogue scale, VAS）、模糊語意量尺（fuzzy linguistic scale, FLS）、模糊評定量尺（fuzzy rating scale, FRS）等，這些量尺部分亦適用於古典Delphi分析和模糊Delphi分析的調查題項形式。一般而言古典Delphi分析式要求評者針對題項提出重要性、必要性、適切性、可行性等看法，以多點評定量尺值或語意量尺表示程度，其中多點評定量尺值是指選擇1~k值，k=3~15，多數採用5、7或9點量尺值，亦見少數採取4、6量尺值，填選形式見圖22所示，分析時依據所選數值進行編碼和應用。

圖22

適用古典Delphi分析之調查題項形式

單一值1

單一值2

題項	小					大					
	1	2	3	4	5	題項	1很不重要	2不重要	3一般	4重要	5很重要
1.題項敘述			v			1.題項敘述			v		

至於適合模糊Delphi分析者有多種形式，一是單一值法，係選擇一個多點量尺值、或選擇一個語意詞再編碼為數值，此做法所得數值需要再根據特定規則轉化為模糊區間值，多數採取三角模糊數集，如(0.2,0.4,0.6)。二是直接設定區間值，以三角模糊數為例（形式為(a,b,c)），如填寫(0.2,0.4,0.6)表示針對項目看法的最低值、最佳值和最高值，見圖23的區間值2、4和6形式；或者是畫設區間加上標示最佳值，如圖23的區間值1、3和5形式；或是針對語意形式選項填上隸屬度值，全部選項的隸屬度值合計為1。

圖23
 適用模糊Delphi分析之調查題項形式

單一值1					單一值2						
題項	1	2	3	4	5	題項	1很不重要	2不重要	3一般	4重要	5很重要
1.題項敘述			v			1.題項敘述			v		
區間值1					區間值2						
題項	0.2	0.4	0.6	0.8	1.0	題項	最低值	最佳值	最高值		
1.題項敘述		v				1.題項敘述	0.2	0.4	0.8		
區間值3					區間值4						
題項	1	2	3	4	5	題項	最低值	最佳值	最高值		
1.題項敘述		v				1.題項敘述	1	2	4		
區間值5					區間值6						
題項	20	40	60	80	100	題項	最低值	最佳值	最高值		
1.題項敘述		v				1.題項敘述	25	50	90		
區間值7					區間值8						
題項	1很不重要	2不重要	3一般	4重要	5很重要	題項	最低值	最佳值	最高值		
1.題項敘述	0	0.2	0.3	0	0.5	1.題項敘述					

註：區間值1,3,5形式的v表示最佳值，線段兩端標示最低值和最高值設定。

附錄二 CV可能最大值和CVs之討論

以下依據公式 進行推導，為節省篇幅，省略列出由公式1推導到公式2的若干過程。

$$CV = \frac{SD}{M} = \frac{\sqrt{\frac{\sum(x-\bar{x})^2}{N}}}{\frac{\sum x}{N}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{\sum x^2}} \quad \text{公式 1}$$

$$CV = \sqrt{\frac{N\sum x^2}{(\sum x)^2} - 1} \quad \text{公式 2}$$

因為 $\sum x^2 < (\sum x)^2$ ，所以 $\frac{\sum x^2}{(\sum x)^2} < 1$

設 $C = \frac{\sum x^2}{(\sum x)^2}$ ， $0 < C < 1$ ，代入公式 2

則 $CV = \sqrt{NC - 1}$ ，因為 $\sqrt{NC - 1} < \sqrt{N - 1}$ ，所以 $CV < \sqrt{N - 1}$

可得知 CV 值最大不會超過 $\sqrt{N - 1}$ ，故 $\sqrt{N - 1}$ 可做為 CV 值的上限值，即

$$CV \in [0, \sqrt{N - 1}]$$

進一步可利用 $\sqrt{N - 1}$ 將 CV 值規範於 [0,1]，即建立正規化 CV 指數 (CV_s)， $CV_s =$

$\frac{CV}{\sqrt{n-1}}$ ， $CV_s \in [0,1]$ ，其比 CV 值更具應用價值，提供了具學理基礎的明確判斷值域，即

CV_s 值越接近 0 越好，越趨近於 1 則越差。再者，需要指出的是前述推導過程說明，與 Martin 和 Gray (1971, p.497)、Dias 和 Santos (2021, p.105) 在論文中所提出的推論過程略有不同。

附錄三 CDI可能最大值和CDI_s之討論

以下依據公式 $CV = \frac{SD}{M}$ 和 $CDI = \frac{M_i}{\max(M)} CV$ 進行推導，可得公式 3。

$$CDI = \frac{M}{\max(M)} CV = \frac{M}{\max(M)} \times \frac{SD}{M} = \frac{SD}{\max(M)} \quad \text{公式 3}$$

將 $CV = \sqrt{\frac{N \sum x^2}{(\sum x)^2} - 1}$ ，代入公式 3，得公式 4

$$CDI = \frac{M}{\max(M)} CV = \frac{M}{\max(M)} \times \sqrt{\frac{N \sum x^2}{(\sum x)^2} - 1} \quad \text{公式 4}$$

設 $\max(M) = VS$ ， VS (value of measurement scale) 是最大的採用評定量尺值， $VS > 0$ ，代入公式 3 和公式 4，得公式 5 和公式 6

$$CDI = \frac{M}{\max(M)} CV = \frac{SD}{\max(M)} = \frac{SD}{VS} \quad \text{公式 5}$$

$$CDI = \frac{M}{\max(M)} CV = \frac{M}{VS} \times \sqrt{\frac{N \sum x^2}{(\sum x)^2} - 1} \quad \text{公式 6}$$

因為 $\sqrt{\frac{N \sum x^2}{(\sum x)^2} - 1} < \sqrt{N - 1}$ ，根據公式 6，得知最大 CDI 值 $< \frac{M\sqrt{N-1}}{VS}$ ， $\frac{M\sqrt{N-1}}{VS}$ 可視為 CDI 的最大值

故 $\frac{M\sqrt{N-1}}{VS}$ 可做為 CDI 值的上限值，即 $CDI \in [0, \frac{M\sqrt{N-1}}{VS}]$

進一步可利用 $\frac{M\sqrt{N-1}}{VS}$ ，將 CDI 值規範於 $[0, 1]$ ，建立標準化 CDI 指數 (normalized CDI, CDI_s)， $CDI_s = \frac{CDI}{\frac{M\sqrt{N-1}}{VS}} = CDI \frac{VS}{M\sqrt{N-1}}$ (公式 7)， $CDI_s \in [0, 1]$ ，其判斷值域具學理基礎，比 CDI 值易於理解和比較，可謂更具應用價值，CDI_s 值越接近 0 越好，越趨近於 1 則越差。又能據此建立正規化共識指數 (normalized DC, DC_s)， $DC_s = 1 - CDI_s$ (公式 9)， $DC_s \in [0, 1]$ ，DC_s 值越大越好，越接近 1 為越好，DC_s=1 表示完全共識，DC_s=0 是毫無共識。而 CDI_s 可結合 CV 計算，簡化為公式 8，其顯示 $CDI_s = CV_s$ ，DC_s 亦可以精簡為公式 9。

$$CDI_s = CDI \frac{VS}{M\sqrt{N-1}} \quad \text{公式 7}$$

$$CDI_s = CDI \frac{VS}{M\sqrt{N-1}} = \left(CV \frac{M}{VS} \right) \times \frac{VS}{M\sqrt{N-1}} = CV \frac{1}{\sqrt{N-1}} = CV_s \quad \text{公式 8}$$

$$DC_s = 1 - CDI_s = 1 - CV_s \quad \text{公式 9}$$

附錄四 CV₂可能最大值和最小值之討論

以下依據公式 10 進行推導，可簡化取得公式 11，其便利於使用平均數和標準差進行計算。

$$CV_2 = \sqrt{\frac{CV^2}{1+CV^2}} \text{ 公式 10}$$

$$CV_2 = \sqrt{\frac{CV^2}{1+CV^2}} = \sqrt{\frac{\left(\frac{SD}{M}\right)^2}{1+\left(\frac{SD}{M}\right)^2}} = \sqrt{\frac{SD^2}{M^2+SD^2}} \text{ 公式 11}$$

因為 SD 值 ≥ 0 ，所以 CV₂ 值 ≥ 0 。又 VS $\geq M > 0$ ，VS 是最大的採用評定量尺值，故 VS² $\geq M^2 > 0$ ，設 M=VS，且 SD $\neq 0$ ，可知(SD² + M²) > SD²，即 $\frac{SD^2+M^2}{SD^2} > 1$ ，代入公式 11，可知 CV₂ 值 < 1 。據此可得 CV₂ 的值域為 0 \leq CV₂ < 1 。

通常使用 1~VS 為評定的數值值域，設 M=1 和 M=VS，且 SD $\neq 0$ ，代入公式 11。

因為 1² $<$ VS²，可知 $\sqrt{\frac{SD^2}{1^2+SD^2}} > \sqrt{\frac{SD^2}{VS^2+SD^2}}$ ，故可知 CV₂ 值的可能最大上限值為 $\sqrt{\frac{SD^2}{1^2+SD^2}}$ ，

和可能最小非 0 下限值為 $\sqrt{\frac{SD^2}{VS^2+SD^2}}$ ，即其可能出現值的值域為 $\left[\sqrt{\frac{SD^2}{VS^2+SD^2}}, \sqrt{\frac{SD^2}{1^2+SD^2}}\right]$ ；若

設 M=1 和 M=VS，SD ≥ 0 ，則 CV₂ 可能出現值的值域為 $\left[0, \sqrt{\frac{SD^2}{1^2+SD^2}}\right]$ 。

以使用 5 點量尺值和 SD=5 而言，則 $\sqrt{\frac{5^2}{1^2+5^2}} > CV_2 > \sqrt{\frac{5^2}{5^2+5^2}}$ 為

0.9806 $>$ CV₂ $>$ 0.7071。若使用 3 點量尺值和 SD=5，則 0.9806 $>$ CV₂ $>$ 0.8575。如果採用 7 點量尺值和 SD=5，則 0.9806 $>$ CV₂ $>$ 0.5812。改為採用 5 點量尺值和 SD=10，則

0.9950 $>$ CV₂ $>$ 0.8944。據此表明 CV₂ 值受到使用的評定量尺值和標準差值兩項變因的影響，而造成值域變動，標準差值越大，將增大 CV₂ 的最大上限值，量尺值越大，將可能減小 CV₂ 的最小下限值；也就是說，標準差值變動較明顯影響 CV₂ 的最大上限值，量尺值變動較明顯影響 CV₂ 的最小下限值，此特性值得注意。

附錄五~附錄十一

係說明自撰 SPSS 程式的參數設定、分析功能、分析結果解讀及程式碼，另有適用於 EXCEL 運作環境的程式，均限於篇幅，無法臚列分享。

