



稽核轉型與加值－資料分析 案例

掌握數位化及資料化之變革環境，同時借力新興科技及資料驅動分析方法，進而優化內部稽核規劃、證據蒐集及報告撰擬等過程，有效促成稽核轉型及提升附加價值，並促進整體政府更良好的政府資料交換與共享環境，發揮內部稽核在良善治理之積極角色。

周靜幸（審計部審計業務研究委員會組長）

壹、前言

數位化是當今世界各國公私部門最熱門的話題，資料無所不在，且每個人都在產生資料。「資料化（datafication）」對社會產生的好處不勝枚舉：醫生使用複雜的演算法協助安排更符合病人需要之治療方法、智慧家居系統幫助人們管制能源使用、系統舞弊及貪腐可能因分散式帳本技術（一般稱為區塊鏈）而被根除。雖然巨量資料、人工智慧、資料分

析、網路安全及智慧系統廣受全球創新者讚許，但也伴隨惡意個人設法濫用不斷成長之資料，以製造假新聞、或竊取他人身分等侵犯隱私權問題。

就政府內部稽核人員而言，為提供內部稽核服務及增加稽核價值，往往會蒐集大量資料，但實際運用所蒐集資料之規模，恐未及十億位元組（gigabytes,GB）、兆位元組（terabytes,TB）或千兆位元組（petabytes,PB）。本文除簡要說明內部稽核資料分析內

涵、運用領域及測試類型、技術外，並進一步說明班佛定律（Benford's Law）觀念及運用案例等，期與政府稽核人員分享與持續成長。

貳、認識資料分析－變與不變

2019 年 OECD 出版「誠正分析：強化貪腐及舞弊風險評估之資料驅動方法」報告指出，資料價值鏈包括資料蒐集/產製、資料處理、資料分享、資料運用（再運用）－資訊、

資料運用（再運用）－決策制定等一系列過程，其中資料運用而進行資料分析所蒐集之資訊及知識，能有效解決監督、洞察及前瞻等關鍵問題（表 1）。

資料分析是分析原始資料及從資料流程中作出結論之科學，目的是在發現及檢查各種模式。它不僅可幫助稽核人員管理及解釋持續增加的資料數量，而且可以透過分析大量資

料而不是小型非代表性樣本，讓他們能蒐集進一步的稽核證據。稽核人員可透過辨識資料之各種模式、集群及異常值，獲得更深入的見解，使他們能夠在資料分析期間偵測及判斷因果關係之攸關性。

內部稽核資料分析係稽核規劃、蒐集證據及撰擬報告等過程之一種程序，無論是採人工或自動化，其運用時機、目的、範圍及專業判斷是不變的。

資料分析之運用已經存在好幾十年，例如數位化未普及的時代，稽核人員運用人工手抄資料、進行橫縱向與跨期間比較分析、比率分析等，獲得分析性證據。但隨著數位化不斷革新、計算能力的增強、以及外來資料的興盛，正徹底翻轉、改變內部稽核確認性服務之架構－即由被動事後稽核轉為主動事發當下之稽核，此一轉變除可提升稽核過程之效率及效果，更可提高交易本身的準確性。

資料分析按其目的不同，一般可分為敘述性、診斷性、預測性及規範性等 4 類型，彙整如表 2。

表 1 資料分析可解決之關鍵問題

	監督	洞察	前瞻
資訊	過去已發生哪些事情？ (報導)	現在正在發生哪些事情？ (警惕)	未來將發生哪些事情？ (推斷)
知識	這些事情如何及為何發生？ (建立模型、實驗設計)	接下來的最佳行動是什麼？ (建議)	可能發生最好 / 最糟的情況是什麼？ (預測、最佳化、模擬)

資料來源：作者自行整理。

表 2 資料分析類型

類型	內涵
1. 敘述性分析	敘述過去的資訊，這些資訊可能有助於確定未來會發生什麼事。
2. 診斷性分析	使用資料以瞭解事情為何以那種方式發生。這有助於進行根本原因分析及決定各種模式。
3. 預測性分析	結合敘述性分析及診斷性分析之發現，依據過去資訊就未來可能發生的事情，預測或做出有根據的猜測。
4. 規範性分析	資料分析之更高階目的，這可能包括為降低風險之建模選項、確認有助機構之趨勢、及概述或定義未來行動。

資料來源：作者自行整理。

參、資料分析運用領域及測試類型

依據國際內部稽核協會（The Institute of Internal Auditor, IIA）研究基金會 2016 年調查發現，內部稽核最常運用資料分析之職能領域及示例如下頁表 3。

資料分析測試類型概分臨時測試（Ad Hoc）及重複 /

專題

連續測試 (Repetitive/Continuous) 兩種，茲說明如下：

一、臨時測試之目的在取得特定問題之答案，透過瀏覽及調查資料以檢視是否存在任何異常跡象。例如將供應商主檔與員工主檔之地址欄位進行比較配對，如配對紀錄發現

異樣，可能代表存有幽靈供應商。透過執行臨時測試，有利發現可能發生錯誤或舞弊之情事。但是，這仍然是非常耗費人工及時間，而且，當這類異常情事是相對普遍的，進行下列重複 / 連續性測試是較有用的。

二、重複 / 連續測試之目的在對大量資料執行測試，以利辨識異常情況，一般利用晚上執行程式，以測試所有交易、及時收到趨勢與模式通知及將任何異常通報管理階層。例如，採購卡濫用是一個普遍的問題，因為特大型機構通常執行大量採購卡採購交易，透過執行自動化測試程式就所有採購卡交易發生當下進行測試，除可以確保遵循相關控制，並可使稽核團隊騰出時間調查可能出錯的其他領域或專注於那些需要大量時間及人工的任務及專案。

表 3 資料分析之職能領域及示例

內部稽核職能	資料分析運用之示例
遵循稽核	<ul style="list-style-type: none"> ● 評估所有交易之費用報告及採購卡使用情形。 ● 透過使用細項帳戶費用資料以辨識需要進行調查之異常及趨勢，以執行供應商稽核。 ● 評估監管機構之要求。 ● 辨識各種資料系統之不佳資料品質及不完整性，其為未遵循風險之關鍵驅動因素。
舞弊、風險評估、偵測及調查	<ul style="list-style-type: none"> ● 辨識高舞弊風險領域及評估控制。 ● 辨識幽靈員工、潛在虛假供應商及關係人交易或員工－供應商關係。 ● 強調顯示對機構構成最大財務及 / 或聲譽風險之異常資料。 ● 調查資產挪用詭計之癥狀，以回答「是誰、做什麼、在何時、在何地」等問題。
營運績效	<ul style="list-style-type: none"> ● 找出有關支出分析之關鍵指標，例如付款時間、放棄提前付款折扣及付款效率。 ● 執行重複付款分析及收回。 ● 執行收入保障分析。 ● 執行滯銷庫存分析。 ● 辨識跨產業及業務項目之關鍵績效與關鍵風險指標。
內部控制	<ul style="list-style-type: none"> ● 預期性，例如業務持續性計畫。 ● 偵測及改正，例如控制帳戶調節。 ● 指示，例如行為規範。 ● 預防性，例如密碼、存取控制。 ● 執行職責分工分析。 ● 執行使用者存取分析。 ● 評估控制績效。 ● 異常報告，例如，辨識指出控制無效或弱點之潛在異常值。

資料來源：作者自行整理。

肆、資料分析技術

就內部稽核目的之資料分析而言，一般認為非常有效的幾種資料分析技術如下：

- 一、計算統計參數（例如平均數、中位數、眾數、標準差、最大及最小值），以辨識出異常的交易。
- 二、透過分類，以找出各個資料元素群組的態樣及關聯

性。

三、將數值分層，以辨識出不尋常（特別高或特別低）的數值。

四、利用班佛定律進行位數分析，以辨識自然發生資料組合中不太可能出現的特定數字。

五、連結不同的資料來源，以辨識出在不同的系統中不當配對的內容，例如姓名、地址及帳戶號碼。

六、重複性測試，以辨識出單純或複雜的機構交易，例如付款、薪資、請款或費用報表細項。

七、間隔測試，以找出序號資料中遺漏的號碼。

八、加總數值，以檢查可能有誤的控制總數。

九、驗證資料輸入日期，以辨識不恰當或可疑的過帳或資料輸入時間。

十、趨勢分析及時間序列分析，計算隨時間變化的資料趨勢及依據趨勢模式就繼續發生之假設進行預測。

十一、比率分析，常用的三個比率為最高值對最低值

（極大值/極小值）、最高值對次高值、當年度對上年度。

十二、偶數金額，偶數（四捨五入至元）金額不會經常發生。因此，四捨五入到十位數、百位數及千位數的數字可能被視為異常且應該更仔細地檢查。

為「2」的頻率次之，比率約為 17.6%，依序遞減，當首位數值遞增至「8」及「9」時，比率則降到只剩 5.1% 及 4.6%（圖 1）。若資料庫中首位數字出現頻率未符前述規律即屬異常情事，內部稽核人員通常會再進行深入調查。

一、5 種分析類型

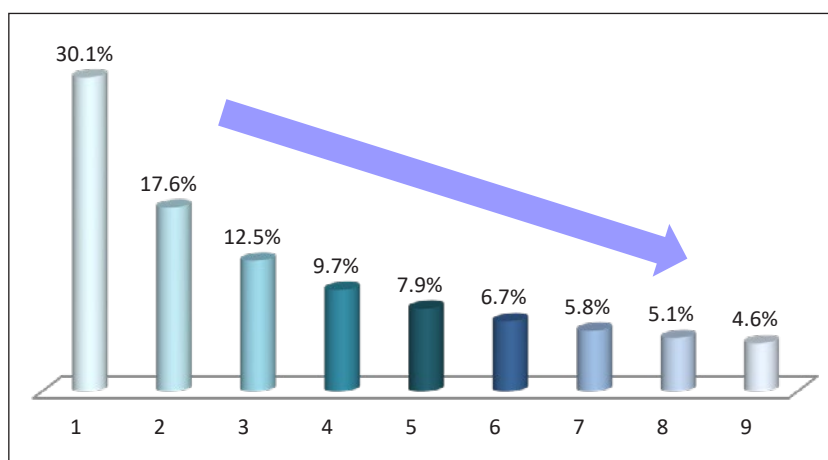
班佛定律之基本測試包括第一個位數分析、第二個位數分析、前兩個位數分析、前三個位數分析及最後兩個位數分析，茲說明如下：

（一）第一個位數分析（First-digit Analysis）

伍、班佛定律基本觀念及運用案例

班佛定律係指任何自然產生的龐大資料庫，數字首位數值出現「1」的頻率最高，比率約為 30.1%、首位數值

圖 1 數字首位數值出現 1 至 9 之次數比率



資料來源：作者自行繪製。

專題

稽核人員可以以圖表示「1」到「9」每個位數之預期及實際發生情況。他們可以進一步深入探討異常差異，以利進行分析及行動。

(二) 第二個位數分析 (Second-digit Analysis)

與第一個位數分析一樣，第二個位數分析是合理性的測驗。例如一家醫療保健公司針對 21,000 多筆薪資紀錄之第二個位數分析，結果顯示第二個位數出現「0」的次數是預期值的兩倍。數字「5」顯示出現頻率比預期高出 60%。根據這些發現，這些紀錄被認為是存有舞弊的。

(三) 前兩個位數分析 (First Two-digit Analysis, F2D)

一個數字中的前兩個位數有 90 種可能的組合 (10 到 99)。例如 110,364 的前兩個位數是「11」。在 F2D 測試中，班佛定律指出「11」是前兩個位數的可能性為 3.8%。

(四) 前三個位數分析 (First Three-digit Analysis, F3D)

在 F3D 測試中，具有

900 種可能的組合 (100 到 999)，有利深入分析大型資料集。有關在 10,000 個或更多交易集中找出異常重複交易，此方法提供更大的精確度。

(五) 最後兩個位數分析 (Last Two-digit Analysis)

一個數字的最後兩個位數存有 100 種可能的組合 (00 到 99)。每種組合的預期比例為 1%。任何小數點均四捨五入或被列入已知數字中。

二、使用時機

班佛定律最適合用於超過 1,000 筆紀錄之資料集，且這些紀錄包含至少四個位數的數字。隨著資料集規模的增加，其與預期頻率之接近程度也會隨之增加。然而，並非所有資料都適合進行這類測試，班佛定律之有效及無效應用如表 4。

三、較小資料集運用範例—世界人口資料 (261 個國家)

首先針對 2014 至 2018 年世界人口資料 (B-F 欄位)，重複運用 EXCEL 的函數公式

表 4 班佛定律之有效及無效應用

有效應用	非有效應用 (人為控制所產生數字而非自然產生)
<ul style="list-style-type: none"> ➢ 由數字 (例如，數量 * 價格) 的數學組合所產生的數字 ➢ 交易資料 (例如，付款、銷貨) 	<ul style="list-style-type: none"> ➢ 數字是按指定順序編號 (例如，支票號、採購訂單、發票編號、身分證字號等) ➢ 數字是人工分配的結果 (例如，透過行銷或一定門檻所設定的價格) ➢ 包含特定數位的帳戶 (例如，為記錄 25 美元退款所開立之帳戶) ➢ 內含最低或最大金額之帳戶 (零用金支出規定最低 10 美元至最高 40 美元) ➢ 航空公司每架飛機乘客人數 ➢ 電話號碼 ➢ 1,000 筆或較少交易的資料集 ➢ 公式所產生的資料 (例如，YYMM#### 是保險單或貸款號碼) ➢ 受最大或最小數字限制的資料 (例如，差旅費、出席費及審查費、應公開招標採購金額、每小時工資率等) ➢ 沒有紀錄之任何交易，諸如偷竊、回扣及圍標

資料來源：作者自行整理。

=LEFT (B2,1)，即可得到 261 個國家各年度人口數字第 1 個數值 (G-K 欄位)，其結果如圖 2。

接著利用函數公式 =COUNTIF (\$G\$2:\$K\$262,1)，分別計算上述 261 個國家 5 個

年度人口數字第 1 個數值為 1 至 9 之出現次數，結果如圖 3。

圖 3 中，數值 1 出現 382 次；數值 2 出現 204 次；數值 3 出現 171 次…等，將各個首位數值出現的次數繪製如下頁圖 4。

在這個相對較小的資料集中 (261 乘以 5 年等於 1305 筆資料集)，其結果也大致 (但不是完全，以 9 為首的數字筆數為 75 筆，大於 7 及 8 為首筆數 71 及 53 筆) 遵循班佛定律。通常較大資料集產生的曲線會比相對較小的資料集更緊密地吻合班佛定律曲線。因此，可以作出結論－班佛定律分析趨勢可以驗證人口是真實數字，未存在潛在虛假資料之情形。

當班佛定律分析之曲線未符合預期時，稽核人員應再努力驗證資料的真實性，建議作為包括：(1) 重新考量原始資料是否適合班佛定律分析：在進一步調查前，稽核人員應重新檢查資料集，以發現對某些數字存在內部偏差的可能性，如果發現偏差，則應忽略班佛定律分析結果。(2) 應用分析性複核程序：如果班佛定律預測不適用於特定資料集，則應懷疑異常及尋求資料集確實有效之其他確認；稽核方法論的「分析程序」為進行分析性複核提供指引。(3) 重新思考內部控制：考量是否以建立可靠的控

圖 2 261 個國家 5 個年度人口數字第 1 個數值之分布情形

Country Name	2014	2015	2016	2017	2018	2014	2015	2016	2017	2018
Aruba	103,774.00	104,341.00	104,872.00	105,366.00	105,845.00	1	1	1	1	1
Afghanistan	33,376,904.00	34,413,603.00	35,383,128.00	36,296,400.00	37,172,386.00	3	3	3	3	3
Angola	26,941,779.00	27,884,381.00	28,842,484.00	29,816,748.00	30,809,762.00	2	2	2	2	2
Albania	2,889,104.00	2,880,703.00	2,876,101.00	2,873,457.00	2,866,376.00	2	2	2	2	2
Andorra	79,213.00	78,011.00	77,297.00	77,001.00	76,006.00	7	7	7	7	7
Arab World	387,907,748.00	396,028,278.00	404,024,433.00	411,898,965.00	419,790,588.00	3	3	4	4	4
United Arab Emi	9,214,175.00	9,262,900.00	9,360,980.00	9,487,203.00	9,630,959.00	9	9	9	9	9
Argentina	42,669,500.00	43,131,966.00	43,590,368.00	44,044,811.00	44,494,502.00	4	4	4	4	4
Armenia	2,912,403.00	2,925,553.00	2,936,146.00	2,944,809.00	2,951,776.00	2	2	2	2	2
American Samoa	55,791.00	55,812.00	55,741.00	55,620.00	55,465.00	5	5	5	5	5
Antigua and Bart	92,562.00	93,566.00	94,527.00	95,426.00	96,286.00	9	9	9	9	9
Australia	23,475,686.00	23,815,995.00	24,190,907.00	24,601,860.00	24,982,688.00	2	2	2	2	2
Austria	8,546,356.00	8,642,699.00	8,736,668.00	8,797,566.00	8,840,521.00	8	8	8	8	8
Azerbaijan	9,535,079.00	9,649,341.00	9,757,812.00	9,854,033.00	9,939,800.00	9	9	9	9	9
Burundi	9,844,297.00	10,160,030.00	10,487,998.00	10,827,024.00	11,175,378.00	9	1	1	1	1
Belgium	11,209,057.00	11,274,196.00	11,331,422.00	11,375,158.00	11,433,256.00	1	1	1	1	1
Benin	10,286,842.00	10,575,952.00	10,872,067.00	11,175,204.00	11,485,048.00	1	1	1	1	1
Burkina Faso	17,586,017.00	18,110,624.00	18,646,378.00	19,193,284.00	19,751,535.00	1	1	1	1	1
Bangladesh	154,520,167.00	156,256,276.00	157,970,840.00	159,670,593.00	161,356,039.00	1	1	1	1	1
Bulgaria	7,223,938.00	7,177,991.00	7,127,822.00	7,075,947.00	7,025,037.00	7	7	7	7	7
Bahrain	1,336,075.00	1,371,851.00	1,425,791.00	1,494,074.00	1,569,439.00	1	1	1	1	1
Bahamas, The	370,633.00	374,206.00	377,931.00	381,761.00	385,640.00	3	3	3	3	3
Bosnia and Herz	3,482,104.00	3,429,361.00	3,386,267.00	3,351,527.00	3,323,929.00	3	3	3	3	3
Belarus	9,474,511.00	9,489,616.00	9,501,534.00	9,498,264.00	9,483,499.00	9	9	9	9	9

資料來源：作者自行繪製。

圖 3 261 個國家 5 個年度人口數字第 1 個數值為 1 至 9 之出現次數

	2014	2015	2016	2017	2018	2014	2015	2016	2017	2018	Digits	Count	percent
1	103,774.00	104,341.00	104,872.00	105,366.00	105,845.00	1	1	1	1	1	1	382	0.29272
2	33,376,904.00	34,413,603.00	35,383,128.00	36,296,400.00	37,172,386.00	2	2	2	2	2	2	204	0.15632
3	26,941,779.00	27,884,381.00	28,842,484.00	29,816,748.00	30,809,762.00	3	3	3	3	3	3	171	0.13103
4	2,889,104.00	2,880,703.00	2,876,101.00	2,873,457.00	2,866,376.00	4	4	4	4	4	4	133	0.10192
5	79,213.00	78,011.00	77,297.00	77,001.00	76,006.00	5	5	5	5	5	5	110	0.08429
6	387,907,748.00	396,028,278.00	404,024,433.00	411,898,965.00	419,790,588.00	6	6	6	6	6	6	106	0.08123
7	9,214,175.00	9,262,900.00	9,360,980.00	9,487,203.00	9,630,959.00	7	7	7	7	7	7	71	0.05441
8	42,669,500.00	43,131,966.00	43,590,368.00	44,044,811.00	44,494,502.00	8	8	8	8	8	8	53	0.04061
9	2,912,403.00	2,925,553.00	2,936,146.00	2,944,809.00	2,951,776.00	9	9	9	9	9	9	75	0.05747
10	55,791.00	55,812.00	55,741.00	55,620.00	55,465.00	1305							

資料來源：作者自行繪製。

專題

制措施以檢測或防止不當行為。

(4) 考量來源：重新考量所獲資料的來源，它們是內部生產的還是從外部來源獲得的？如是外部來源，應驗證資料來源之可靠及完整性。(5) 針對異常資料進行全部查驗，例如，上述以 9 為首的 75 筆人口數字。

陸、結語

當今，稽核人員因數位化轉型或稽核自動化、資料分析而產生新的角色，其可運用諸如 Microsoft Excel、CaseWare IDEA 等工具，製作稽核工作底稿及執行審核程序，但人工作業仍普遍存在，包括重複性

輸入、客戶資料清理、資料轉置及規則導向資料分析等方面。唯有當稽核人員不再花費大部分時間執行重複且基本工作時，始能將資源集中在更具挑戰性及關鍵性的工作上，尤其是評估重大風險，而這種「人機合作 (man-machine cooperation)」模式將是未來稽核的核心。

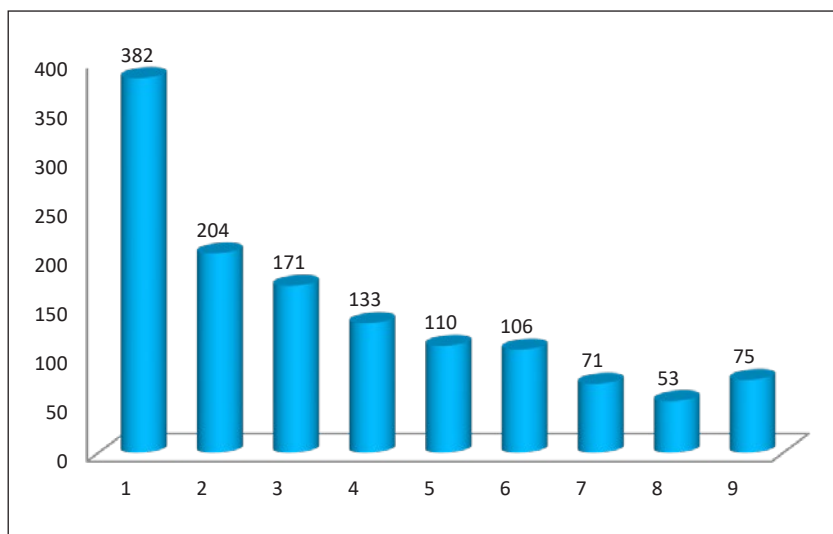
亞歷山大大帝 (Alexander the Great) 名言「沒有知識，就無法關注技能。沒有技能，就無法發揮力量。沒有力量的知識可能無法被應用。」成功資料分析繫乎基本知識、技能及有力支持等三要素，即使像

微軟這樣大型軟體公司的內部稽核人員，也必須獲得正確基礎知識及技能。行政院主計總處為因應數位化稽核時代，已製作電腦稽核模組及範例並置於官網供各機關參考，積極推動政府稽核人員運用資料分析於稽核工作，各機關稽核人員在此有力支持下，可以讓稽核工作因資料蒐集、使用、分析而增加價值，並促進整體政府更良好的政府資料交換與共享環境，有效發揮內部稽核在公部門良善治理之積極角色。

參考文獻

1. The Organisation for Economic Co-operation and Development (OECD) (2019), Analytics for Integrity : Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessments.
2. Galvanize (2019), Detecting and Preventing Fraud with Data Analysis.
3. The Institute of Internal Auditor (April 2020), Benford's Law in a Big Data World.
4. The Institute of Internal Auditor (August 2011), Global Technology Audit Guide (GTAG) 16 : Data Analysis Technologies. ❖

圖 4 世界人口數字首位數值出現 1 至 9 之次數



資料來源：作者自行繪製。