

# 應作迴歸觀

## ——資料、不確定性與人工智慧法學

何漢歲\*

*“Data is the sword of the 21st century, those who wield it well, the Samurai”.*

Jonathan Rosenberg, former SVP of Products at Google

藉由這篇為了慶賀本所建所十周年而作的文章，我想寫下在起手法實證研究，乃至於討論人工智慧法學如何可能前，應該要知道的，有關資料與方法的二三事。英諺有云：「魔鬼藏在細節裡」，當下載巨量原始資料可以透過電腦程式在彈指間完成的現在，為了得到後續用以分析之資料所做的處理，研究者是否本於良知，對資料的正確性與完整性（integrity）勉力以求，並詳實揭露所做之完整結果，以免誤導<sup>1</sup>；接收者是否在選擇相信研究者之餘也對資料的可靠性有合理確信，這些似乎已逐漸在大數據的汪洋及推陳出新的運算方法浪頭中幾被遺忘。

個人堅持「先確認資料可靠，再講求分析效果」的理由很難用有邏輯的文字表達。雖說現在的自己是由過去的選擇與足跡所形塑，我的求學過程與生命經驗實在沒有什麼特別之處，我個人除了體重外也沒什麼份量到可以分享「我的學思歷程」，只是素人一枚，有幸成為中央研究院法律學研究所的一員，讓我的學術生涯路徑遠不同於其他統計博士；也因為出身非法律背景，也許我在處理和數字有關的問題時的想法與作法會跟其他人不同。關於我以下所提出對資料與方法的所思所想內容，將充滿著我會計與統計科班出身的色彩：摻雜著回憶殺、一些統計科

---

\* 中央研究院法律學研究所研究助技師。

<sup>1</sup> 科技部對研究人員學術倫理規範第3點，<https://www.most.gov.tw/most/attachments/a8ff2bb9-84ae-41ec-b539-bc54d9085811>（最後瀏覽日：2021年6月20日）。  
穩定網址：<http://publication.iias.sinica.edu.tw/40518012.pdf>。



普，有些地方可能是讀者沒有注意過的、有些地方可能會被認為過於理想、有些地方可能不會被認同，但無論如何很感謝有這個榮幸，可以讓我為這些「寫了不知何處適合發表，不寫又覺得有點可惜」的內容安排一個歸宿。這篇文章也將會是我一直在緩慢醞釀的，一系列關於統計科學哲學、人工智慧法學，以及法律文本探勘相關論述的先導。

在進入正題之前，不可免俗地，我想我得先交代一下我是怎麼半路踏進法律所的。在民國107年12月開始任職研究助技師之前，我其實早在博士論文口試通過後，102年8月起，就在本所法實證中心全職工作了。算一算，到今年7月底我就在本所全職工作滿八年。至於為何會來到法律所工作，則是一段必須感謝黃丞儀老師助我結下的緣分，而這要從我99年夏天結束為期一年的千里馬計畫回到臺灣說起。當時的我應該是博士班六年級了，修業年限倒數中而且每個學期都在為錢煩惱。大約是在100年的春天，我在網路上（ptt的兼職板）看到了一則誠徵兼任助理啟事的刊登，是來自中研院法律所的求才，詳細內容已不可考，我只記得當時我不太確定這份工作是要幫忙分析資料，還是比較像要當一個統計學家教，「總之我都可以勝任吧」我當時想，所以就投了履歷。

那則徵才啟事並沒有寫出是法律所的哪位老師在徵求，後來有同事知道後開玩笑道「妳都不知道是誰就投履歷了喔」，說真的當時我沒有想太多，去面試的感覺也很愉快，這份兼職工作一直維持到102年7月我博士論文口試通過，靠著這份兼職一個月八千元的收入與家父每月一萬元的奧援，讓我撐過博士班的最後一哩路。幸運的是，大概也是在我博士論文口試後，法實證中心開了助理缺，我就順勢投了履歷。我想，當時的中心執行長黃國昌老師錄取我的關鍵原因，極有可能是履歷上寫著我曾經擔任過法律所兼任助理，而我也懷著報恩的心情迎接可以經濟獨立的日子。

## 統計、資料

這篇文章的寫作當下正值強化新冠肺炎第三級疫情警戒的時刻，每日下午2時準時關注指揮中心報告幾乎成為一種習慣。一天下來新增了多少確診個案，是境外移入或是本土？性別、居住地與年齡的（次數）

分布範圍如何？又新增多少死亡案例？這是大眾理解的統計——計數再匯總，有時會輔以性別、年齡、居住地等人口統計變數加以區隔。而從5月22日起的確診個案補登（即「校正回歸」）可以讓大眾體會到，數據產出伴隨的細節很多，得出的數字本身有時也是需要定義、說明的，如果能早些揭露檢驗塞車的狀況，在解讀確診數字時就能將量能不足之處納入考慮，使數字更能允當地表達目前以天為基礎的疫情狀況報導。就個案資料合計至以天為單位的資訊而言，難免會有程度不一的時間差，理想上是要以某一時間點為界，將進入流程（如抗原快篩為陽性正在等候PCR核酸檢測結果者）到結案（如PCR檢測結果為陰性、死亡等）各狀態的人數分布加以揭露才是完整，而只要做法（定義、認列條件及衡量觀念等）一致並加以揭露，就能增加每天數字的可比較性與溝通價值。

至於個案層級的資料，如檢驗結果，在醫事檢驗是以標準作業程序操作的前提下，應不需要考慮檢驗結果會因不同的操作人員或機台而不同，這意味著可合理假設醫檢師或儀器不會成為誤差來源。至於常聽到的「偽陽性」、「偽陰性」這類的錯誤，由於跟統計推論中的假設檢定及人工智慧中預測的評估指標有關，將在後面的篇幅繼續說明。綜上所述，在解讀與分析資料前，關於資料品質應該注意到的四個重點：一、轉換現象或抽象概念於有限資料欄位時，如以統計數字表示疫情現況，是否充分完整；二、即便是計數所得，但凡對使用者操作上重要有用的資訊，都應該充分揭露，更不用說將文字等非結構化資料類型化時，所採用的分類標準、特殊情況如何處理等，均應在不失重要性的原則下，以適當的方式提供之。還有另外兩點：三、觀察或衡量的層次為何？四、可能的誤差來源來自何處？後兩點主要是影響如何以統計模型描述人的行為，例如慰撫金的審酌、量刑等。此部分將待稍後我為大家簡單介紹統計學（術）後敘之。

## 知易行難的判決文本結構化

以上對資料的描述，其實預設了所處理的是結構化的資料——資料已依事先定義的資料型態逐欄整齊排列，最常見的是資料以二維表格或

試算表方式儲存，每一列是一個觀察單位，每一欄是一個變數（或稱變項、屬性），例如「每日新增死亡新冠肺炎確診個案表」中，每列代表一名病例，欄位則有案號、性別、年齡、慢性病史、活動接觸史、發病日、症狀、採檢日、住院／隔離日、確診日、死亡日等。年齡及日期相關的欄位概念上沒什麼模糊性，性別也有自然定義好的類型，至於症狀的內容，選項如咳嗽、呼吸困難、發燒等，應是早已類型化，只是可以複選，所以看起來不如接觸活動史是單選來得工整；慢性病史在早期的公布亦是可複選，如高血壓、糖尿病等病症，但目前僅公布有／無。上述的欄位資料都是已經過整理過的，例如活動史必定是經過辛苦的疫調後統整歸納出可能的選項。然而，儘管這樣的資料讀進電腦十分容易且直觀，離「乾淨」可供分析的狀態卻還有一點距離，這是由於電腦讀資料的方式是視一格的内容只會有一個值／選項，因此可能的改進之處包括：將複選題中所有的選項做逐項（從多選項可置於一欄轉為一欄僅代表一項）、二元的編碼，如一欄是有／無高血壓，再一欄是有／無糖尿病，依此類推；以及就住院／隔離日欄位中，非因確診住院的病例，如因其他原因住院、到院前死亡等，另起新欄備註之。

所以，欲以法院判決做為法實證研究素材<sup>2</sup>，可以如何進行呢？民事訴訟法第226條與刑事訴訟法第51條第1項、第308條、第309條、及第314條之1等分別規範了民事判決書與刑事判決書上的記載事項，個人認為判決文本是帶標籤的非結構化資料，而非半結構化資料。判決中很明顯有些資訊如法條、當事人姓名、律師姓名等，因其帶有規律的格式，不難以程式從中擷取再存成結構化資料。而隨著自然語言處理領域中命名實體識別（Named Entity Recognition, NER）任務技術的進步，提出如金額、百分比等數量詞辨識，及部分法律領域專有名詞的偵測，目前都有現成的套件<sup>3</sup>可供線上或線下操作，只是經驗上，判決書裡事實與理

---

<sup>2</sup> 個人目前歸納了五類已用於法實證研究的素材：一、法院裁判（含大法官解釋）；二、法律條文及公司章程；三、心理學實驗；四、（抽樣）調查；及五、檔案（archives）。

<sup>3</sup> 例如本院資訊所中文詞知識庫小組（CKIP Lab）專案研發之專有名詞辨識，完整辨識清單可見<https://github.com/ckiplab/ckiptagger/wiki/Entity-Types>（最後瀏覽日：2021年6月20日）。

由的提煉，單靠專有名詞的堆砌是很難充分描述的，特別是讓抽象的法律概念能（盡量）透過類型化後的文字完整表達，實屬不易。即使研究者有敏銳的嗅覺，想的到判決文字中可將抽象概念顯化的代理變數，法官使用的工具與表達方式也非一成不變，比方說欲探討「資力」，也就是要在判決中尋找描述「資力」的相關文字並加以編碼，有的法官可能用查詢稅務部門得到的年收入，有些可能是列出其土地、房車、存款、股票投資及貴重物品等財產，當然，有時候也會付之闕如。抽象的概念在法學研究中所在多有，如何設計並維持資訊品質，研究者除了要能在有限的欄位中對已知文字內容有效評估剪裁，使其類型化後「簡潔、可靠又不失其攸關性」外，也必須確保可驗證性——對於同一筆文字若由不同編碼者操作，也要能得到相同的結果。

以上刻劃文字資料結構化的不易之處，並非是想嚇跑欲投入法實證研究的人，相反地，也許富有法律專業訓練的讀者更有能力接受這樣的挑戰。這並不是鼓勵大家投入勞力密集的「工人智慧」，而是從製造資料的源頭出發，在考慮司法制度的運作、法律人所受將具體個案事實如何置於特定構成要件下的訓練，以及職業養成過程所受之思維邏輯薰陶後，就問題意識與規範上可能的特徵，選擇與了解自然語言處理的人協力設計合適的資訊檢索與擷取流程。我認為法律人做實證研究的困難並不在於沒有學過統計：從統計課程存活下來的努力，會讓你在形式上知道如何與統計軟體溝通，但對資料實質內容的判斷與掌握，才是有用研究的關鍵。正因為法官行為／判決研究面對的是人，應該更要正面面對法官的文字，想靠一知半解的華麗炫技，我認為是改變不了「垃圾進，垃圾出」的。

由於巧婦難為無米之炊，追求有效率、充分允當地對法律文本結構化亦是我個人短期學術目標中最重要的一項。在結束關於資料這個回合的論述前，我想有必要提出目前有部分法實證研究（輔）以單字／詞做為分析判決文本的基礎——將文本關鍵詞彙轉換為帶有語意訊息的詞向量（word embeddings/vector/representation，又稱詞嵌入），再將該詞向量做為一個特徵納入分類預測模型。轉換的目的在於，使相似的字詞所對應的向量表示也會類似，也就是說，語意相似（以餘弦相似度衡量

之)的字詞有較近的距離，且一個字詞可與其他字詞互為表意、類比。由於類神經網絡演算法的一日千里，目前的詞向量技術，例如常聽到的BERT (Bidirectional Encoder Representations from Transformers, 基於變換器的雙向編碼器表示技術)，可以訓練出讓不同語境 (語境即上下文訊息) 下的同一個字詞有不同的詞向量表示，也就是有更精確的語意理解，使進一步的自然語言處理任務，如文本分類，有更好的表現。至於可以如何想像向量呢？請打開手機的數字鍵盤，在綠色通話鍵上有著每3個數字就換一列的配置，橫著看的(1,2,3)至(\*,0,#)就是列向量，直著看的(1,4,7,\*)至(3,6,9,#)就是行向量，而這些列向量或行向量堆疊成了一個4x3的矩陣。假設我們目前考慮的是列向量，即可將詞1至詞4表示為詞1 = (1,2,3)，詞2 = (4,5,6)，依此類推。

## 將判決丟進電腦中

人工標記資料的高成本，使得許多資工勇者們前仆後繼發展像BERT這樣的預訓練語言模型，口語中說的「將判決丟進電腦中」，就我個人的理解是以程式進行以下步驟：一、將判決中事實與理由的部分讀進電腦；二、資料清理或前處理 (如遺漏值的填補、錯誤的改正、亂碼或特殊符號的難字處理，使資料型態一致的調整等一切使資料合於分析指令輸入的格式)；三、自動分／斷詞 (segmentation) 後移除停用詞 (stopwords)；四、辨認出重要 (關鍵) 字詞並加以向量化。當然後續自然語言處理任務的進行也是以程式自動化進行，唯此處對比的是「將文本資訊結構化後輸入電腦」再進行所需分析，而幫助文本結構化的自然語言處理工具除了分／斷詞及移除停用詞之外，還有句法分析 (syntactic analysis, 即parsing) 及語義分析 (semantic analysis) 等。至於詞向量是否適合用在法實證研究乃至於人工智慧法學，個人目前不成熟的想法是認為，法實證研究與文本分析在處理分類問題與分群問題上，目標一致，但猜測以詞向量為基礎直接進行分類或分群或許聊備一格；運用詞向量技術在其他自然語言處理任務，如文本情感分析與意見探勘的表現，是否會勝過結構化資料，則可以期待。至於結構化文本資料與詞向量在人工智慧法學將扮演何種角色，因為還牽涉到推論與可解

釋性等學理上的討論，我指的是具體而言，應該以什麼樣的標準對資料輸入後得到的規則進行評價，是必須先針對方法進行釐清的。

## 統計學，學什麼？

統計學是一門為了做出好的判斷，對數值化資料進行蒐集、組織、呈現、分析，以及詮釋的科學，其中又再區分為敘述統計與推論統計。敘述統計是以傳遞資訊的方式對觀察到的資料進行組織、摘要與呈現的方法。組織資料一般來說就是分組劃記，通常會製作成次數分布或表格的形式。至於摘要或呈現的部分，則可分成以數值方式，或是以作圖方式表之，前者如大家耳熟能詳的平均數、中位數、變異數等，後者如直方圖、盒鬚圖、散布圖等。資料的摘要與呈現方式與本身的測量尺度有關，例如對性別等名目資料計算平均數或中位數是沒有意義的，僅能以眾數描述資料的中心位置，讀者可以自行從任何一本統計學教科書查閱測量尺度的分類，及各種適用的數值量測與統計圖。在此僅稍微解釋兩個在數位時代常聽到，但其實有年紀的兩個名詞，第一是探索式資料分析（Exploratory Data Analysis, EDA）：利用簡單的統計值與易繪之統計圖迅速摘要資料的組成結構、外型特徵與可能形式（pattern）的一種手法，亦可協助研究者在資料檢誤階段發現可能的錯誤。第二是資料視覺化（data visualization）：運用資訊圖表將繁雜的數據簡化成為易於吸收的內容。可以是進行複雜或嚴謹的分析之前對資料的認識，也可以是分析結果的呈現。

提到視覺化，不得不說相對於疫情指揮中心習以數字和文字排列在一起的隨文數字傳遞統計資訊，雙北市政府對視覺化疫情數據的努力，充分體現了一張圖勝過千言萬語的精神。臺北市政府甚至建立了所謂的「疫情數據儀表板」<sup>4</sup>供內部使用，所謂的數據儀表板，是一種在一個頁面同時配置多個相關變數視覺化的作法，例如臺北市公開的疫情儀表板

---

<sup>4</sup> 臺北市政府資訊局新聞稿，臺北市科技防疫再升級「台北通實聯制、紓困線上申辦、疫情數據儀表板」，2021年6月1日，[https://doit.gov.taipei/News\\_Content.aspx?n=4B2B1AB4B23E7EA8&sms=72544237BBE4C5F6&s=F351486D6A6AEEB0](https://doit.gov.taipei/News_Content.aspx?n=4B2B1AB4B23E7EA8&sms=72544237BBE4C5F6&s=F351486D6A6AEEB0)（最後瀏覽日：2021年6月20日）。

截圖，畫面上同時有各國累積確診數、各國累積死亡數、全球死亡率、每日臺北市、臺灣、世界、美國確診數、統計逐日疫情控管指數及每日桃機入境人數。除了按時更新資料外，儀表板最重要的是可產生互動式的效果——設計友善的介面讓使用者依需求調整圖的呈現，例如將確診數長條圖從預設的由多至少排序改為由少至多，或是設定數據呈現的期間等，使決策者能更快、更全面掌握所需要的數字。新北市政府推出的新北災訊E點通<sup>5</sup>——災訊地圖則是結合了地理資訊系統，以不同顏色標示各行政區內疫情的風險程度，更透過疊加不同圖層，整合並即時呈現各種資訊如火警、救護車出勤動態，篩檢站、熱區防疫中心位置及其他即時災害相關通知。

推論統計做的則是透過樣本對未知的母體參數進行估計（*estimation*）與假設檢定（*hypothesis testing*），部分學者會把區間估計（*interval estimation*）納入推論統計的範圍。如果說敘述統計是入世的，學習推論統計——將隨機性或不確定性納入資料分析中，可說是有入世與出世一體兩面的思維——在大一學統計學時忙著把實際資料帶入公式般的檢定統計量計算出檢定統計值，大二學數理統計學時則要以符號（*notations*）推導檢定統計量的由來，始知其然更知其所以然。這邊要強調統計推論是一個專有名詞，不等於「統計」加上「推論」，更不是只要是使用統計出的數字來進行敘述，都是統計推論。統計推論背後有著統計哲學的蘊含及嚴謹數學證明的支持，是統計人的核心競爭力：知道公式由來的推導過程，能為資料選擇合適的現有方法並正確解讀，如果已知方法不敷使用而能自行發展，則是統計學術研究的層次。

未知的世界很大，能觀察到的現象很少，學習估計與假設檢定即是學習如何以有限的資料合理地見微知著而非盲人說象，已如前述。舉例而言，自從社區感染出現後，普篩聲浪再起，由於政府選擇不普篩，究竟還有多少染疫黑數不得而知，但是透過目前得到的確診情況，例如將各行政區的（樣本）確診率依人口加權後得到之整體（樣本）確診率乘以全臺灣人口，或許是一個可供參考的確診數（上限）。這個對估計的

---

<sup>5</sup> 新北災訊E點通網站，<https://e.tpf.gov.tw/>（最後瀏覽日：2021年6月20日）。

舉例略為複雜，一般我們會用以樣本平均數估計母體平均數（即期望值），或是以樣本變異數估計母體變異數等聽起來十分直觀的做法。「需要什麼數學證明嗎？」也許有人會有這樣的疑問，事實上統計學家期待一個好的估計量要有像是不偏（unbiased）、一致（consistent）、有效（efficient）等好的性質，這些性質的操作都源自數學上的定義，是否成立也都是以數學證明出來的。常用的估計也不見得會具備所有好的性質，例如以最大概似估計法得到的樣本變異數估計量，其實是有偏的（biased），只是當樣本數足夠大時，漸進不偏。通常統計上提到估計指的就是以上說明的點估計，但是母體參數就是樣本估出的值的機率有多高？答案是不是0就是1，而且除非普篩，否則永遠不會知道答案。因此統計學家希望能有一組合理的上下界，使得在考慮隨機性下，從同一個母體抽出的樣本即使不盡相同（得到的點估計值也因此不盡相同），但有相當高的比例，如通常給定的95%，未知的真正母體參數會落在此區間中。此即近日食藥署公告國產新冠肺炎疫苗緊急使用授權審查標準中，免疫橋接研究方法提到之95%信賴區間（個人預設是較為嚴格的雙邊區間估計）的本質含意，雖然區間的推導，必與該藥理學指標定義有關，閾值的決定也是專業考量，但萬一臨床資料算出的下界一翻兩瞪眼沒有落在要求之上，則可能不是效果不足，就是受試者接種後反應不太一致。

除了可用樣本估計母體參數外，我們也會想對母體參數提出假設，並比較此假設值與樣本值是否有顯著差異，做為進一步下決定的依據（是否有足夠的證據拒絕有關該母體的敘述）。母體參數的假設可以是經驗值、預設或是現況，上述提及之閾值數字也是一例。又如高端疫苗解盲公布的臨床試驗數據中，0.7%的疫苗組受試者及0.4%的安慰劑組受試者有發燒症狀，單靠點估計 $0.7\% > 0.4\%$ 可以直接宣稱疫苗組比安慰劑組容易發燒嗎？假設檢定就是想回答這樣的事，或者換另外一個想法是，0.3的差距夠大嗎？想像如果用刻度為1的尺來量，肯定不算大，但如果以刻度為0.1的尺來量，應該就有差了。量尺的長相即公式乃根據數學推導而得，至於刻度單位的粗細則是與原本母體的變異（或以樣本估計之）及樣本數大小有關，整體而言刻度越粗表示越不容易估得精確

(precision)，因此0.3的差距極有可能是因為抽樣誤差或偶然所致，不足掛齒。在變異程度固定為常數時，樣本數越大刻度會越細。然而，在隨機的世界，不論量尺刻度的大小，統計假設檢定所得的結論都有犯錯的可能，就好比檢測工具皆有偽陰性、偽陽性之可能一樣，只是我們最終可以用PCR檢測認定是否確診，而在真實的世界裡事實是無法得知的。

如果快篩（根據觀察資料做成的結論）與PCR檢測（未知的實際情況）結果一致，表示快篩的結果正確，沒有犯錯，故錯誤有兩種。以判斷是否染疫為例，統計上會將受檢者沒有帶原的情況設為虛無假設（null hypothesis），有帶原的情況設為對立假設（alternative hypothesis）。第一種錯誤，也是統計檢定想法中認為比較嚴重的錯誤，稱型一錯誤（type-I error），是虛無假設為真，但資料顯示拒絕此虛無假設（可類比偽陽性，即快篩呈現陽性但PCR檢測為陰性）；第二種錯誤，統計專有名詞稱型二錯誤（type-II error），是虛無假設為偽，但資料顯示不拒絕<sup>6</sup>此虛無假設（可類比偽陰性，即快篩呈現陰性但如以PCR檢測會是陽性）。型二錯誤的補集是當虛無假設為偽且資料也拒絕此虛無假設，表示了此檢定的檢定力，可類比檢驗醫學中的敏感性（sensitivity）或是分類問題在評估預測好壞時會用到的召回率（recall）；至於型一錯誤的補集，就我個人理解在統計上尚未有術語指稱之，但檢驗醫學中的特異性（specificity）則表達了此概念。另外統計上有所謂的偽發現率（False Discovery Rate, FDR），衡量的是在多重檢定問題中，拒絕虛無假設時犯型一錯誤的程度，與機器學習中分類問題在評估預測好壞時常用到的精確率概念上互補。

<sup>6</sup> 我們可以證明數學命題中的真偽，但對於假設要決定的是接受（不拒絕）或不接受（拒絕）。拒絕虛無假設表示有足夠（自樣本得出）的證據否定虛無假設對母體參數的敘述，但仍有可能犯型一錯誤。由於型一錯誤較為嚴重，通常會給它設定上限稱之為顯著水準，如常用的0.05。顯著水準一旦決定，對應的拒絕域——計算出的檢定統計量之值落入拒絕域中即代表要拒絕虛無假設，也會同時被一併決定。當計算出的檢定統計量之值沒有落在拒絕域時，顯示沒有足夠的證據可以拒絕虛無假設，此時比較保守的講法是「不拒絕」虛無假設，也有人直接說「接受」虛無假設，並不為過，畢竟接受不代表該虛無假設為真，只是還沒有資料足以做為否定的證據。

## 虛無假設通常不是研究假設，對立假設才是

虛無假設的寫法決定了是否犯錯，與犯了那一種錯誤，那麼，它的寫法是恣意的嗎？確實是可以視問題而定，但不可以隨意。一般來說會先寫下的其實是對立假設，它通常是結論會引起興趣的，例如喝醋可預防AZ疫苗引發的血栓疑慮；改變現況的，例如國產疫苗的效力優於已上市疫苗；萬一檢定出錯較難彌補者，例如判被告有罪；或是研究者期待成立的猜想，例如確診死亡與慢性病史有關。根據以上凡例推敲與之相對的虛無假設，不難歸納出虛無假設通常指涉的是現況或傾向否定者（如無關）。「『虛無假設是雖然想推翻它，卻又儘量特別保護它，不讓它輕易被推翻，除非有顯著的差異。如此一旦推翻，才有說服力』<sup>7</sup>，『新的推論也才不會很快又被推翻』」<sup>8</sup>。不過，實證研究中也不盡然都是想拒絕虛無假設，最明顯的例子是抽樣調查在收案後，需要以卡方適合度檢定確認樣本與母體兩者組成結構特性（通常是人口統計變數）沒有顯著差異，此即樣本代表性。要補充的一點是，如果是帶有數字的假設，虛無假設一定是含有等號的敘述，對立假設不需要與之互補：可以是不等於、大於、小於，或是等於另一個值。

## 蒐集資料是統計人的守備範圍嗎？

前面定義統計學時有提到統計學的科學範圍包括蒐集資料。可能是因為民意調查是最常聽到統計數字的時候，一般人可能會以為統計人會參與第一線的原始資料蒐集，其實不然。舉例來說，選舉調查或是市場調查通常是基於政治學與行銷學的各種興趣，社會學家也擅長使用調查訪問的方式對樣本進行長期追蹤，負責調查的訪員不會需要有統計訓練的背景做為門檻。又像目前採取「熱區圍堵、冷區殲滅」策略的新冠肺炎檢測，從檢體採集、處理到結果判讀、上傳，也必須由醫事人員為之，與統計專業無涉，精準疫調也非統計人的專長。統計學對於數據收

<sup>7</sup> 黃文璋，談統計誤差——假設檢定篇，黃家小館，2016年8月28日，<http://www.stat.nuk.edu.tw/huangwj/stat-article/stat58.pdf>，頁22（最後瀏覽日：2021年6月20日）。

<sup>8</sup> 黃文璋，統計顯著性，數學傳播，29卷4期，頁37（2005年）。

集的幫助依專業程度多寡或可分為三類，第一類是將研究領域的成果應用於實際的樣本蒐集規劃問題上，例如在調查準備階段如何決定合適的抽樣設計，包含抽樣方法的選擇、最適樣本數的配置等，又如最近媒體經常報導的平行（parallel）、雙盲、隨機分配、具對照組的設計，是生物統計中的臨床試驗（clinical trial）設計的一種。實（試）驗設計的目的在比較不同處理（treatments）的效果（effects）是否有差異，例如在第二期臨床試驗中，為找到目標族群的最適劑量，建議應採多劑量組別<sup>9</sup>的設計。統計人的第二類貢獻是資料蒐集完畢後，整理與檢誤資料<sup>10</sup>。這類工作的技術層次並不高，一般來說修過統計學的人員就可以執行。只是由於「垃圾進、垃圾出」，做為將資料以合於（分析方法之）格式「讀」入軟體中，並以程式指令操作電腦產出分析結果的技術擔當，統計人無法與資料正確性切割，矇著眼睛對錯誤視而不見。美國統計學會曾寫下，統計人要能產出值得信賴的、可靠的資料<sup>11</sup>，我認為指的就是這類分內之事。最後一種其實算是純粹靠累積經驗與旁徵博引得到的，關於事前或事後的編碼建議，以及對應過錄編碼簿格式的編纂：因為知道如何配置資料才容易讀取與分析，以及那些問題選項是可以透過程式邏輯性處理，使問卷設計與編碼過錄儘量友善且防呆。

## 統計學術研究

科技部統計學門將統計學術研究分為五個次領域，分別為數理統計與機率、統計方法、生物統計、工業統計以及其他應用統計，各次領域項下又列舉了數個學術專長<sup>12</sup>，例如生物統計項下其中一門學術專長是

<sup>9</sup> 王亞蕾，疫苗研發的臨床參考要點，當代醫藥法規月刊，98期，頁2（2018年）。

<sup>10</sup> 關於調查研究資料的管理與保護、問卷編碼設計與過錄，及資料的整理與檢核等相關推動，本院人社中心調研中心深耕許久，讀者可自行搜尋相關積累，如<https://survey.sinica.edu.tw/?p=4837>。

<sup>11</sup> 現在在美國統計學會網站上的「What Do Statisticians Do?」已然不是當年的文字（果然統計也是持續在進化的）。我將貼在我辦公室的部分紙本內容寫下。Statisticians are experts in the following: Producing trustworthy data; Analyzing data to make their meaning clear; Drawing practical conclusions from data.

<sup>12</sup> 自然科學及永續研究發展司，[https://www.most.gov.tw/nat/ch/detail?article\\_uid=3c16b238-4d7d-4611-a86a-294ba628b67f&menu\\_id=dabd2410-af58-4314-8392-c7070c8e8373&](https://www.most.gov.tw/nat/ch/detail?article_uid=3c16b238-4d7d-4611-a86a-294ba628b67f&menu_id=dabd2410-af58-4314-8392-c7070c8e8373&)

臨床試驗與製藥統計，迴歸分析、類別資料分析（如邏輯式迴歸）則被歸類為統計方法底下的模型分析學術專長中，機器學習則是其他應用統計的一員。這邊想先嘗試說明的是，雖然在名稱上僅有一門「數理統計與機率」看起來跟數學有關，是高深的理論統計研究，事實上其他的次領域都是在機率與數理統計的基礎之上，使用統計模型為（各種領域）資料的推論問題提出新的分析工具，通常雖不會特別說成「○○統計方法」，但其實都是統計方法。以定義、假設（assumption）與符號證下的新定理或新模型，才是統計學術論文發表的重點，而且統計學術論文寫法要求精鍊，看起來簡潔的定理，其證明就算會放在附錄供參，證明過程也多有省略，讀者可能得花各把月才能自己證過一遍。再者應用統計（方法）非統計應用，通篇不見實際的數據是家常便飯。如果有用新方法操作一兩個實際例子，也是在跟現有方法做比較，跟在實證研究中讓數字說故事的情況截然不同。我認為所有的統計博士都是被期待要以高階數理能力發展統計理論、創新統計方法<sup>13</sup>經營學術職涯。如果說法學讀到最後是在讀哲學，那麼統計讀到最後應該就是在讀數學吧。

我碩士畢業後曾在本院統計所擔任過一年研究助理，博士論文口試通過後在本所工作的這幾年，在院內遇過幾次現任統計所所長陳君厚老師，有幸能被他記得。他在知道我成為研究助技師後曾問我博士論文寫的是什麼？是不是跟實證或模型分析有關的題目所以才會在法律所協助實證研究。當我說我寫的是實驗設計（更精確來說是最適設計）時，他似乎有點驚訝，因為觀察性研究大概用不到實驗設計，心理學實驗的進行通常也不會採用理論的結果。前面提到實驗設計是實驗進行前所做的安排，更具體來說，是決定各種處理要配置多少受試者，至於如何隨機指派誰接種疫苗或誰注射安慰劑，都是額度決定之後的事。例如有網路媒體報導，國產疫苗二期臨床試驗中，實驗組與對照組兩者人數比例是

---

l=ch，附加檔案「學門代碼.xls」（最後瀏覽日：2021年6月20日）。

<sup>13</sup> 政治大學統計系網站，<https://stat.nccu.edu.tw/uploads/asset/data/59cf02a81d41c843540036f/統計學系博士班核心能力對應表.pdf>（最後瀏覽日：2021年6月20日）。政大統計系的師資背景大學為數學主修的比例應該是國內統計系所為低者，政大尚且如此期待，其他學校乃至於統計學界對數學證明能力的要求，不難想像。

6:1，而國外三期臨床試驗的比重卻是1:1，這個比例有何邏輯性<sup>14</sup>？理論上的最佳比例就是最適（連續）設計在推導的，答案在此先賣個關子<sup>15</sup>。如果要能正確比較出實驗組與對照組受試結果的差異，最理想的情況，就是每位受試者如複製人般一模一樣，且不存在其他干擾因子，但即使如此，實驗組的結果也可能不盡相同（有變異），對照組的結果也可能不盡相同，這是由於未知且不可控制的干擾因子所導致，統計上提出的解決辦法就是透過隨機化讓此干擾因子的效果平均掉。理想上實驗單位越同質越好，退而求其次來說，如果能找到一個區隔變數，可能研究者對此變數本身的效果並不感興趣，也不覺得此變數與處理（即實驗組與對照組）會相互影響，但是覺得此變數是實驗結果變異的來源之一，例如欲以食藥署近期核准的三款居家快篩商品進行快篩試驗，試劑本身的效果不重要，也不會有試劑對疫苗或安慰劑（舉例）特別敏感，但已知此三款試劑準確度或有不同，必須考慮其對實驗結果的影響。這種已知且可控制的變數稱為集區（blocking）因子，透過把相似環境放在一起——集區內實驗單位同質，集區間實驗單位異質，可以在分析時排除集區因子造成的變異，更正確地比較實驗組與對照組的實驗結果是否有差異。有集區因子時的隨機化，會是在集區內隨機分派誰打疫苗誰注射安慰劑，而非直接隨機分派實驗組或對照組受試者至集區。至於每個集區要配置多少實驗組或對照組，我的博士論文剛好有處理到這個問題，且我的發現是基於假定（assume）實驗結果來自廣義線性模型生成的分配時所得出，較傳統線性模型更為廣泛適用。

隨機化與集區化是實驗設計三項基本原則中的二項。第三項是重複性，指的是同一種實驗狀況設定下至少要做二次實驗，例如三種居家快篩試劑應至少各有兩名實驗組與對照組受試者，如此才能計算處理與集區因子的變異。除了已知且可控制與未知且不可控制的干擾因子外，統

<sup>14</sup> 謝柏宏，高端疫苗受試人數邏輯解析，經濟日報，2021年6月10日，<https://health.udn.com/health/story/121833/5521953>（最後瀏覽日：2021年6月20日）。

<sup>15</sup> 如果把實驗組與對照組都當成處理來看，兩者樣本配置在假定反應變數是常態分配時，應該要越接近越好。如果是其他模型，則最適樣本配置要看實驗組與對照組的變異程度，但如果是廣義線性模型下，即使變異程度差非常多，均分樣本數也不失為一個有效的配置。

計上還有一種干擾因子是已知但不可控制的，這種干擾因子通常是無法分組的連續變數，最常舉的例子就是假設大一微積分是統一會考，欲比較會考成績是否因不同老師教法不同而有所差異，由於學生數學能力參差不齊，且一定對學習微積分有所影響，應該要以移除自身數學能力後的表現評價老師的教法是否有差，故可以考慮使用共變異數分析，將微積分成績根據大學指考的數學成績做調整後，再對不同老師進行比較，方為合理。

### 不確定的知識 + 了解其中的不確定程度 = 可用的知識<sup>16</sup>

前面提及分析資料前要注意觀察或衡量的層次如何，以及可能的誤差來源。階層概念的重要性在於，在觀察人的行為時，預設每個人的屬性或本質相近是不合理的，比方說同一位法官的判決可能有其偏好與慣性（所以跟其他法官的決策模式不同）、同一件判決的數名原告因共享相同的案件事實故同質性高，不同判決的原告同質性低，此處的法官與案件就是一個集區因子，在計量模型的術語中稱為一個群聚（cluster），其層級在判決與原告之上。因此，在以統計模型描述法官行為時，需要將同一法官、同一案件的相似性納入考慮<sup>17</sup>，才不會直接配適傳統假定觀察值皆獨立的模型，而產生估計上的偏誤，導致後續不正確的假設檢定結論。換一個角度來說歧異，思考一個數字，如原告期待的慰撫金數額，只想到平均數或中位數等中心量數固然有感，但缺乏對數字分散程度的掌握，極有可能會因實際不如預期而感到失望，所以更有價值的發掘應該要針對慰撫金審酌的機制，也就是資料是如何生成的。追著這個概念學習法律資料（learning from legal data），或許會對法學研究發展有一些啟發，微觀上可以衡量判決實務的不確定性並反饋於規範面：確定那些事情在原則、概念上可理解，但具體操作時構成要件缺乏定義、

<sup>16</sup> See C. RADHAKRISHNA RAO, *STATISTICS AND TRUTH: PUTTING CHANCE TO WORK* (2d ed. 1997). 原文為Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge.

<sup>17</sup> 可參考拙作：何漢歲、張永健，法實證研究方法進階導論：固定效果、隨機效果、群聚標準誤，月旦法學雜誌，259期（2016年）。

標準不一，或因案情等各種不同有不同的實現；又或者本來就有裁量空間的決定，是否或又會是以何種外界未知的規律在運作，又該如何詮釋運作背後的選擇進而發展新的學理或改變現有的作法。在宏觀上我認為對誤差或變異的肯認可以是法釋義學之外的新知識框架的建立，如果能不止從文義、體系、解釋、比較法、學說、價值、例外等形塑對法律概念的想像與面向，而是抱持敬畏面對相同法律概念或制度的實踐對不同人造成的不確定結果，透過仔細思辨並以統計模型表達那些是不確定中的確定、其餘的不確定性從何而來，有什麼可能原因與之有關，以及如何量度這些不確定。儘管由觀察數據歸納出（特例推向通例）的規則仍是不確定的知識，一旦能量化其中的不確定性，就成了另一種確定的知識<sup>18</sup>。

### 統計學是探求真理必備的工具<sup>19</sup>

說真的，在人的世界講可以得到真理也許是有點輕佻了，就算數學已被認為是最高真理，做為數學基礎的演繹推理也不是沒有邏輯缺陷的<sup>20</sup>。我個人不太喜歡為統計方法歌功頌德，只是沒有辦法接受會有人提到統計想到的就是謊言，提到機器、深度學習就五星吹捧。雖然沒有特別考古，但現今機器學習模型的標配如線性迴歸、決策樹、邏輯式迴歸、 $k$ -近鄰演算法（ $k$ -Nearest-Neighbors,  $k$ NN）、區別分析（線性分類）、群集分析（簡單如 $K$ -means）與主成分分析（用以降維）等都是早在我大學時就學過的，表示這些方法在更早之前就被數學家（或兼統計學家）提出了（提出的意思是背後的定理都已被嚴格證明）。單純貝氏分類器說真的大學時應該沒有特別學過，只熟悉背後的貝氏定理（其實也就是條件機率），後來研究所時學的貝氏分類器或基於機率的分類器數學結構上都比較複雜（假定較為放鬆）。至於支持向量機（Support

---

<sup>18</sup> RAO, *supra* note 16, at 54. 原文為It was realized that although the knowledge created by any rule of generalizing from the particular is uncertain, it becomes certain knowledge, although different kind, once we can quantify the amount of uncertainty in it.

<sup>19</sup> RAO, *supra* note 16, at 121. 原文為Statistics: An Inevitable Instrument in Search of Truth.

<sup>20</sup> RAO, *supra* note 16, at 50.

Vector Machine, SVM) 等非線性分類器演算法的提出因為年代較近，演進的過程較為有感，我的理解是數學家（或兼統計學家）與計算機科學家都對此仍在持續發展中的方法有所貢獻。其他像是如一兩層的類神經網絡模型做為非線性迴歸的一種，在統計系大學部應該也會上到，但是更複雜的架構，或許要修到研究所程度的「統計學習」課程才會學到。另外像是正規化迴歸中的脊迴歸（ridge regression）雖然也是大學部程度的迴歸課程會教，但至少從前的脈絡不是在解決變數選擇或過度擬和的問題，不過函數中加入懲罰項早就是模型選擇指標常用的一種手法。至於另一個正規化迴歸中的最小絕對緊縮與選擇算子（Least Absolute Shrinkage and Selection Operator, LASSO），現在應該至少是碩士班課程的標配，畢竟變數個數遠超過觀察值的寬資料（wide data）也是90年代末才開始出現的<sup>21</sup>。以上的敘述大部分是我的修課經驗，歡迎感到懷疑的讀者去翻翻看稍微有點年紀的統計系大二以上各必修教科書確認。此處想表達的是，從資料中學習並不是電機資訊科學／工程獨有的專長，他們的數學能力也很好，證出許多解決機器學習問題中需要的凸最佳化（convex optimization）演算法的數學證明，程式也寫得很厲害，但是不知道為什麼部分人對於人工智慧的見解可以接受「我們不知道為什麼，但是預測得很好」。人腦或許也是黑盒子，但就法官判斷而言，我們希望人工智慧學的應該是法官的腦而不是可以神預測的半仙的腦——機器不會從資料中自動學習法律制度與規範，也不是從過去判決中得出相當精準的規則就等同模仿了法官的腦，或是一定程度的保證未來能產出正確的決定。是故我個人認為重點在於學習判決資料要反推的是法官的大腦如何運作（只是我們習慣將之簡化成對顯著變數的詮釋），而不是為預測而預測。反過來說，硬要對預測高的模型解釋，真的能得到真實的近似嗎？我沒有答案，只覺得就判決量化分析而言，監督式學習（supervised learning）之所以缺乏解釋與說明預測結果與證據資料的連結方式，就是因為它追求的僅是預測值與實際值間的相近，沒有專注在建立描述資料生成機制、可以把確定與不確定分開的統計模型。而且，

---

<sup>21</sup> BRADLEY EFRON & TREVOR HASTIE, COMPUTER AGE STATISTICAL INFERENCE: ALGORITHMS, EVIDENCE, AND DATA SCIENCE 298 (2016).

即使用了預測力不高但可解釋的迴歸模型，坦白說我個人也從未看過電腦科學家如計量經濟學者般將資料層次結構或誤差來源納入模型中考量。

## 一切有爲法（演算法），如機器學習，如分類問題，應作迴歸觀

前面提到許多堪稱古老的統計方法迄今仍被普遍使用，其中原因除了統計方法是以堅實的數學證明做為後盾，一旦被證明為真，那便是「定理恆久遠，方法永流傳」。數學證明通常始於特例，待特例證成再逐步放鬆假定，一步步朝通則邁進，這是數學證明的特殊化原則<sup>22</sup>。所以統計方法也不是停留在過去，會一直有後浪嘗試證明更一般化的結果，或是根據資料特徵改變已知模型的假定，例如線性混合模型（Linear Mixed Model, LMM）在傳統迴歸之上加入隨機效應，藉由假定「同一群聚的觀察值有相同的變異（程度），不同群聚觀察值有不同變異（程度）」，推導當資料結構有層次時，能更正確得到統計顯著性的估計與檢定。資料與方法的關係，總體來說是資料狀況多而分析方法少，或許也可以說是一種方法養百樣資料，比方欲探討決定慰撫金酌定與量刑的因素有那些，都可以使用迴歸分析。此處「迴歸」是一個概括性的統稱，泛指以統計模型建立資料中反應變數與解釋變數間的關係，舉凡傳統線性迴歸、廣義線性模型（Generalized Linear Model, GLM）下的邏輯式迴歸（反應變數為二元）、卜瓦松迴歸（Poisson regression，反應變數是計數值，如快篩陽性數）、反應變數為順序（ordinal）尺度的迴歸等。粗略地說，大部分的時候根據反應變數的測量尺度就可以決定使用上開那一種迴歸，當模型生成不只有隨機誤差時，研究者得依自身對複雜模型的可理解或接受程度，更進一步選擇如（廣義）線性混和模型等較為細緻的工具。這邊要提出的一個看法是，我個人當然會希望模型能越符合資料生成機制越好，再複雜也沒在怕，但是我認為研究者應該要選擇（但不需堅持）自己能駕馭的方法，也學習別人使用的方法

---

<sup>22</sup> 參見翁立衛，數學解題過程中的特殊化原則，科學教育月刊，304期（2007年）。

並內化，但不要做方法的奴隸，或為配適而配適。例如還是很多研究會把反應變數為計數或比例（如原告與有過失）的資料以傳統線性迴歸配適，我不會這樣做，但方法上我會給過。這樣看起來似乎靠迴歸就可以一招半式闖天下？我只能說任何模型「沒有最適合，只有更適合」、所有模型都是虛構小說，但有些故事比其他的好<sup>23</sup>。統計方法的數學證明會需要有一些假定做為前提，不是所有猜想都能被順利證出，也不是每一筆的資料都可以找到夠合適的方法，而且有現成軟體可以進行分析，有一些方法上的小洞還有待遇到問題的人（例如在下）再想方設法處理之。

以上談到的迴歸似乎還是落在老套的建立模型範疇中，事實上迴歸分析是一個兼容推論式與演算法式（如機器學習程序）統計分析的統一框架。例如統計學習<sup>24</sup>可看作是無母數迴歸的一種形式<sup>25</sup>；神經網絡是專門針對預測的複雜迴歸方法<sup>26</sup>等。我想至少於法實證研究中（或者扣除非我專業的規範面討論，姑且暫稱為法律資料分析），在探討反應變數與解釋變數間關係時（或稱監督式學習），所使用的不管是統計方法或機器學習演算法，在數學理論架構上都不出迴歸的守備範圍，絕對不是機器學習領域分為迴歸問題與分類問題<sup>27</sup>，這是其一。其二是所有統計推論的結果都需要透過演算法落地。換句話說，演算法是統計學家執行問題解決的步驟，而推論是統計學家為什麼這麼做的理由，所以即使有人認為演算法就只是把實際數字帶入公式自動得到結論，也不足為奇。統計學的中心在過去的70年間已經從傳統（50年代以降）的數學與邏輯

<sup>23</sup> RICHARD A. BERK, *STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE* 22 (3d ed. 2020). 原文為All models are fiction, but some stories are better than others.

<sup>24</sup> 此處不區分機器學習與統計學習。

<sup>25</sup> BERK, *supra* note 23, at 33. 原文為As a first approximation, statistical learning can be seen as a form of nonparametric regression in which the search for an effective mean function is especially data intensive.

<sup>26</sup> EFRON & HASTIE, *supra* note 21, at 371. 原文為Neural nets are elaborate regression methods aimed solely at prediction.

<sup>27</sup> 分類就是對二元（或多元）結果的預測，個人認為真正的問題在於用什麼規則得到預測。

轉而朝向側重計算<sup>28</sup>。在此要先澄清的是這裡的計算指的不是數數或是計算檢定統計量的值，而是任何透過數值、非公式化方式得到統計推論結果的計算，例如在小樣本時，用無母數靴拔重抽法（bootstrap）估計母體變異數的方式，是將樣本視為母體，從中以抽後放回的方式重複抽樣，再以這些樣本變異數的變異數做為母體變異數的估計。這個看起來神奇的方式背後不意外也是有數學證明支撐的；常用來估計預測誤差的交叉確認（cross-validation）也是所謂「計算」的一種。又如前面提及在我大學時沒什麼學到貝氏方法（Bayesian methods），當時還會聽到頻率學派與貝氏學派之爭——大多數的統計學家是頻率學派，少數是貝氏學派，兩者對參數的想法有著哲學上根本的不同，故後續的推論方式與使用術語也有所差異。貝氏統計要對許多參數計算數值積分，即使當時還沒有所謂的「大數據」，運算量仍十分龐大（computer-intensive），過去硬體也不似現在進步，所以我個人認為貝氏統計在過去不算主流（但現在覺得貝氏方法會是實證法學中分析法官行為模型的主力之一）。頻率與貝氏學派的界線早已逐漸模糊，計算速度的飛奔也使得缺乏解析解（analytic solution）的統計推論問題得到良好的近似。這邊的點在於，演算法式的或程序性的計算動作是只要有把任何方法運用到實際資料都會進行的，至今仍然有很多統計領域是像上面敘述那樣理解並實作演算法。然而，隨著大數據時代的來臨，演算法式的處理大量資料將演算法與統計推論慢慢連接了起來，而預測是一個演算法發展遠遠超前其推論證成（inferential justification）的領域<sup>29</sup>，不代表預測沒有推論問題需要處理，反而應該要有更多統計學家投入其中。

## 我們要有怎樣的人工智慧法學？

坦白說我還不知道，但想先提出兩件需要意識到的事。首先，透過

---

<sup>28</sup> EFRON & HASTIE, *supra* note 21, at xvii. 原文為The center of the field has in fact moved in the past sixty years, from its traditional home in mathematics and logic toward a more computational focus.

<sup>29</sup> EFRON & HASTIE, *supra* note 21, at 209. 原文為Prediction, perhaps because of its model-free nature, is an area where algorithmic developments have run far ahead of their inferential justification.

大規模的資料學習，確實可以得到理論上對群體來說最好的結果，但是未必對個體來說是最好的，就如同疫苗整體來說保護力很高副作用很低，所以還是會讓人產生想接受施打的誘因，但是不能排除個體接種後產生嚴重不良反應或還是確診並轉為重症的可能，所以人工智慧法學的定位為何？是否可依照各別案情不同有所調整？換句話說，相似程度要多高可以成立所謂的「同案同判」？不同之處是否需要額外審酌，都是在運用預測結果前應該要思考的事。第二件事情是統計上顯著不必然意味實務上顯著（practical significance），白話而言就是統計上顯著的變數不必然代表該變數真的是重要變數。這是因為只要觀察值夠多，即使是微小的差異也會使得該變數統計上顯著。在目前大數據或是崇尚觀察值越多越好的研究風氣下，這個現象是數字解讀者應該要戒慎恐懼的。

最後我想先以個人經驗對法實證研究做一個可能不太成熟的分類，我不知道這樣的層次（或是怎樣的層次）是否可以投影到人工智慧法學上，但直覺既然機器學習技術還在持續發展中，我想至少技術所對應能處理的內容應該會有層次之分吧。除了對象是大法官的研究之外，法實證研究目前處理的核心問題，我想主要可說是在一、檢驗「書本中的法律」與「事實運作中的法律」兩者差異和對規範面的反饋；二、裁量決定模型／機制的發現，三、以及／或者對法官、律師（辯護人、代理人）行為的探討。如若再加上四、現況的觀察，就構成了分析的階層。後三者是從觀察事實運作中的法律開始，提出可能的行為科學理論並加以驗證，最後透過模型的建立展現顯著相關因素的效果，從而以推論、預測或解釋與已知規範對話。

## 沒有走在傳統統計學術研究的軌道上，我選擇了一條沒有人走過的路，而這讓一切變得如此不同<sup>30</sup>

我很感謝我的博士論文指導教授說服口試委員讓我畢業，恩師是位自律、井井有條的人，如果讓我重新選擇，我仍希望能成為她的學生。可是我的博士班過得真的有點辛苦，所以我一定要先感謝法律所給我很

---

<sup>30</sup> 改寫自美國詩人羅伯特·佛洛斯特（Robert Frost）的著名詩作〈未行之路〉（The Road Not Taken）的末兩句。

多的肯定，讓我能真正重拾信心。

之所以會說過得辛苦，主要是因為缺錢，有非常長的時間我都必須另外找校內外兼職，甚至在去千里馬前（五年級）我找了一份一個月待遇二萬但幾乎快要是全職的資料分析工作，這當然會影響到我的研究進度，而且多少會羨慕別的博士生只需要專注在課業與研究，不需要為錢煩惱（不過老闆是自己選的怨不得別人）。即使到現在我都不能保證如果又被戳到關於讀博士班時的痛處能一笑置之。比方說，畢業後有一個過去算要好的博士班學妹曾跟我說「研究就是錢堆出來的，妳就是最好的例子」。現在想想其實她說的沒錯，但當時聽了很刺耳，可能也覺得「不要拿我跟妳們比」。也有政大的老師到現在還會三不五時碎唸我當年為什麼沒有專注在研究上而是去賺錢（啊就缺錢啊），我只能回他說「老師你的意思是有錢人才能讀博士班嗎」（讀者不用擔心我跟這位老師講話就是這麼直接）。然後我還記得幾年前有一次到人社中心聽演講，有一位老師在報告結束的提問脫口而出說「博士班唸比較久就是比較笨啊」，呃，我想我聽的是一場量化研究的演講。

很多心中的情緒，無論理性或不理性，確實都在成為研究技術人員後一掃而空了，但我不是因為現在的位置所以覺得沒有走另一條路是對的，因為口試通過那時我就想過去業界工作，還為此裸考多益。老天與許多貴人的眷顧讓我能法律所安身立命，每天通體舒暢，我想我在這邊的所見所聞，接收到的刺激與成長還有許多美好的人事物，是即使我被評價為優秀的統計博士都不可能得到的，我很感恩法律所讓我能充分做自己（即使還不是研究助技師的時候），所以我能用更超然的角度看統計。每一天我都覺得自己才是能讓別人尊重、肯定統計專業的人。每一天我都覺得因為在法律所，我會有各種的可能。最後感謝每一位都像大樹一樣可以carry我的老師與同事們，即使有些人年紀比我小。感謝大家對我的愛護、包容、提點與幫助，我在本所深深感到聰明的人特別幽默也特別會表達。也感謝前後任所長林子儀老師與李建良老師，來了這麼多年我終於很自然地用「老師」來稱呼所長了。特別感謝蘇彥圖老師與張永健老師，他們為我做的肯定比我能想到的要多更多，尤其是張永健老師，我想我今後還是會繼續麻煩他。

## 參考文獻

### 一、中文部分

- 王亞蕾（2018），疫苗研發的臨床參考要點，當代醫藥法規月刊，98期，頁1-10。
- 何漢葳、張永健（2016），法實證研究方法進階導論：固定效果、隨機效果、群聚標準誤，月旦法學雜誌，259期，頁167-181。
- 翁立衛（2007），數學解題過程中的特殊化原則，科學教育月刊，304期，頁29-37。
- 黃文璋（2005），統計顯著性，數學傳播，29卷4期，頁29-38。
- （2016），談統計誤差——假設檢定篇，<http://www.stat.nuk.edu.tw/huangwj/stat-article/stat58.pdf>。

### 二、外文部分

- Berk, Richard A. 2020. *Statistical Learning from A Regression Perspective*. 3d ed. New York, NY: Springer International Publishing.
- Efron, Bradley, and Trevor Hastie. 2016. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. New York, NY: Cambridge University Press.
- Rao, C. Radhakrishna. 1997. *Statistics and Truth: Putting Chance to Work*. 2d ed. Singapore: World Scientific Publishing.

