

啤酒與 學生氏 t 檢定

戈斯特對統計學最大的貢獻是他發明了學生氏 t-檢定 (Student t-test)，學生氏 t-檢定在科學和實務應用上早已成為不可缺少的理論與工具。

■ 劉仁沛

你知道金氏世界紀錄和健力士酒廠 (Guinness Brewery) 與統計學發展之間的關係嗎？健力士酒廠是位於愛爾蘭都柏林，具有 250 年歷史，生產生烈性黑啤酒的酒廠。1951 年 5 月 4 日，健力士酒廠的總經理休·比佛爵士在愛爾蘭打獵時，與同行的打獵伙伴爭論歐洲哪一種獵鳥飛得最快。為了解決這一問題，比佛爵士和在倫敦開實況調查公司的學生兄弟諾里斯及羅斯·麥克威特 (Norris and Ross McWhirter) 共同合作，於 1954 年 8 月發行第一版 174 頁的金氏世界紀錄。

但早在 1899 年，健力士家族了解到，在工業革命後，為了在新的 20 世紀的世界啤酒市場仍然保持領先的地位，必須採用科學的態度和實際的數據，改進啤酒釀製的過程和品質。為了達到這個目標，健力士酒廠僱用許多英國頂尖大學，如牛津或劍橋大學的畢業生，其中一位是剛自牛津大學畢業，擁有化學和數學雙學位的威廉·希力·戈斯特 (William Sealy Gosset)。



威廉·希力·戈斯特 (圖片來源：<http://www.causeweb.org/resources/fun/db.php?id=316>)



戈斯特提出更精確估計單位體積酵母菌濃度的方法，如此就能穩定地在生產過程中加入適當濃度的酵母菌，製造品質更穩定的啤酒。（圖片來源：種子發）

製造啤酒的重要步驟是先在瓶子內培養酵母菌，待瓶子內的酵母菌生長繁殖到適當數目後，再把酵母菌倒入大麥芽漿中發酵以製成啤酒。這時若倒入的酵母菌數目太少，會發酵不完全；太多則增加啤酒的苦味。為了確知瓶子中酵母菌的數目，當時的做法是先從瓶子中取部分液體樣本，在顯微鏡下計算樣本的酵母菌個數，再推估瓶子中酵母菌的總數。

因此如何精確地估計瓶中酵母菌數目，就變成一個重要課題。但是瓶子中的酵母菌會不斷地繁殖或死亡，瓶中酵母菌的總數並非一個固定值，而是一個會隨時變動的數目。

但是戈斯特了解，在單位體積（如 1cc）內酵母菌的數目雖然會變動，但變動的機率可以用統計學中的卜瓦松分布（Poisson distribution）描述。

卜瓦松分布是在單位體積或時間內，稀少事件發生次數的分布，如 1cc 血液內某種細菌個數的分布，或每天台灣高速公路發生死亡車禍次數的分布。描述卜瓦松分

布最重要的參數，是在 1cc 單位體積中平均的酵母菌量；卜瓦松分布的另一個特性是平均數和變異數相同。戈斯特提出更精確估計單位體積酵母菌濃度的方法，如此就能穩定地在生產過程中加入適當濃度的酵母菌，製造品質更穩定的啤酒。

戈斯特計劃發表他的酵母菌研究成果，但健力士酒廠為了保護公司的商業機密和智慧財產，明文禁止員工發表文章。戈斯特並沒有因為這項規定而放棄他的學術研究發表，他在《生物統計期刊》（*Biometrika*）主編卡爾·皮爾生（Karl Pearson）的鼓勵下，轉以「學生氏」（The Student）的筆名在 1907 年《生物統計期刊》上發表他的第一篇學術論文，並登出這項研究成果。

戈斯特對統計學最大的貢獻是他發明了學生氏 t-檢定（Student t-test），學生氏 t-檢定在科學和實務應用上早已成為不可缺少的理論與工具。

譬如想驗證某一品牌泡麵的防腐劑含量是否超出政府所訂的 3PPM（parts per million）上限，可以隨機抽取 10 包泡麵作

在 1937 年戈斯特因心臟病去世之前，他服務的健力士酒廠一直不知道他從事統計研究工作，並以學生氏筆名發表研究成果。

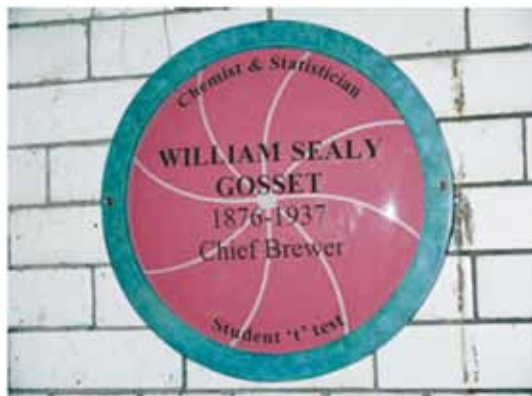
樣本，根據每包泡麵所測得的防腐劑含量，計算防腐劑含量的樣本平均數和標準差，以樣本平均數和 3PPM 的差來判斷該泡麵防腐劑平均含量是否高出 3PPM。

但樣本平均數有抽樣誤差，因此把樣本平均數和 3PPM 的差以標準差的倍數表示。在 20 世紀初若樣本數夠大（超過 30 以上），這一比值 t 的分布會趨近於標準常態分布。但在一般科學實驗中，如酒廠內的釀酒測試，樣本數很少會超過 20。

為了驗證「樣本及真正平均數的差」和標準差的比值 t 在小樣本下是否遵從標準常態分布，戈斯特利用英國犯人身體特徵資料做為母體，算出真正的平均值，然後每天下班後坐在桌前，在微弱的燈光下隨機取出一組樣本數是 4 的小樣本，計算平均值、標準差及比值 t （這是在 110 年前，計算機尚未萌芽的階段）。經過無數個夜晚，把數千個比值 t 的分布作圖後，戈斯特發現在小樣本下， t 不會服從標準常態分布。這種反覆抽樣的方法現在稱為蒙地卡羅模擬（Monte Carlo simulation），只是用電腦執行。

根據他所做出的分布，他導出平均數的差和標準差的比值 t 的理論分布，並在 1908 年再次以學生氏的筆名發表於《生物統計期刊》，這就是現在任何一本統計教科書都必須納入的學生氏 t - 分布和學生氏 t - 檢定。

戈斯特的統計研究都是在下班之後，在家中廚房的餐桌上完成的。他不但是一位傑出的化學家，更是一位有效率的行政專家。他在健力士酒廠一路晉升至首席釀



現在位於都柏林的健力士專賣店（Guinness Storehouse）中有一個戈斯特的紀念碑，上面寫著「化學家 & 統計學家 威廉·希力·戈斯特 首席釀酒師 學生氏 t - 檢定」（圖片來源：http://www.flickr.com/photos/tibs_ibc/2757806074/in/set-72157606688244632）

酒師，而且擔任許多重要的管理職務。

但是在 1937 年他因心臟病去世之前，健力士酒廠一直不知道戈斯特從事統計研究工作，並以學生氏筆名發表研究成果。許多統計研究者要和戈斯特見面，都必須像間諜電影般地祕密安排見面地點和時間，這大概是戈斯特在有生之年一直可把他的統計研究成果發表的原因吧。但是現在位於都柏林的健力士專賣店（Guinness Storehouse）中有一個戈斯特的紀念碑，上面寫著「化學家 & 統計學家 威廉·希力·戈斯特 首席釀酒師 學生氏 t - 檢定」。

劉仁沛

臺灣大學統計教學中心