

統計與落點分析： 大學指考選填志願的輔助利器

任眉眉¹ 陳日昇² 詹嘉豪¹

¹國立成功大學統計學系

²國立成功大學計算機與網路中心

摘 要

大學指定考試分發入學前，考生需繳交志願卡，因應如何在一千多個校系中，選出自己喜歡而且上榜機會大的 100 個科系的問題，電腦落點分析遂成為考生選填志願的重要輔助工具。坊間的落點分析均建立在前提假設「前後兩年學生對校系基礎評價不變」上，本研究首次以統計理論探討該假設是否成立，並應用九十二和九十三學年度，聯合分發會公佈各校系最低錄取加權總分資料，進行實際驗證。在前提假設成立，文中也應用連續兩年度大考中心發佈的，各單一考科考生成績累計表和各校系之指考科目與加權方法，以及前一學年度各校系最低錄取加權總分等資料，應用統計理論與分析，建立一個落點預測的模式。並據此寫成一套有效的電腦落點分析軟體，置於網頁上，提供考生免費下載使用。

關鍵詞：最低錄取加權總分，最低錄取加權平均分數，落點分析，相關係數，簡單線性迴歸。

美國數學會分類索引：主要 62K05；次要 62J10。

1. 前言

以往大學聯考分成四類組，各校系通常只能選一類招生，且各類組採記相同科目，考五到六科，每年各校系最低錄取分數排行榜變動不大。為了導引考生在大學志願選填時能符合自己的性向，使各大學依據自己的特色單獨或聯合招收適才學生，遂於民國八十八年六月確定所謂的「考招分離，多元入學」方案，並於民國九十一學年首次實施。第一年實施的大學多元入學方案打破四類組限制，招生校系分成甲、乙、丙三案，甲、乙案要求先通過學測檢定，招生校系再自由採計零到三科指定考試（簡稱指考）成績，丙案類似舊的聯考制度，不檢定學測，採記五到六科指考成績。九十三學年大學多元入學方案是第三年，規定和前兩年差異甚大，甲、乙、丙三案併成一案，廢除採計零到三科的組合，只能採計三到六科，且部分校系要求考生需參加學測檢定。九十三學年招生簡章共計 270 頁，其中超過三分之一的校系採計考試科目和加權方法與前一年度不同，加上聯合分發會（簡稱聯分會）僅公佈最低加權錄取總分，不公佈最低原始錄取總分，考生如何在一千六百多個校系中，排出符合自己的志願，又要避免高分低就的順序呢？

每年指考成績寄發以後，坊間會推出很多收費的落點分析，收費金額由一百五十元至二百五十元不等，多數做過分析預測的考生都覺得效果不顯著。在大學入學考試中心（簡稱大考中心）的網頁 (1) 上針對落點分析，提供具體說明如下：落點分析（或錄取難易程度分析）的步驟是根據今年和去年兩個年度的累計人數表，首先把今年各科分數換算成去年各科分數，然後把換算後的分數按照各學系去年的採計和加權方式計算出加權總分，以如此得到的加權總分算出和去年最低錄取分數之間的「差距」，若「差距」 > 0 且離 0 愈遠，代表今年錄取機會愈大；若「差距」 < 0 且離 0 愈遠，代表今年錄取機會愈低；若「差距」在 0 上下，代表錄取機會屬於邊緣。相關資訊亦可參考中華電信 Hinet 學習網 (2) 及大同大學電機系網頁 (3)。大考中心亦進一步針對九十二和九十三學年度兩組資料進行準確度分析，並歸納造成預測不準的主要原因為前後兩個年度的採計科目與加權方式改變。事實上，大考中心是以去年權重作預測，忽略了今年權重的影響。當校系改變了採計的指考科目與加權方法時，去年的分數標準如何對應到今年來看？加權計分不同，導致錄取總分數有被灌水的可能性，每個校系新年度的錄取分數如何能被正確預估？

基本上,坊間推出的落點分析均建立在「前後兩年學生對校系基礎評價不變」的前提假設下,直接由電機專業人員負責撰寫程式,但是從未有任何文獻驗證前述基本假設是否成立。首先,本文首次提出如何驗證前述基本假設成立的統計理論,並根據九十二和九十三學年度,大考中心公佈的考試成績數據及聯分會公佈的校系最低錄取加權總分,應用統計方法驗證前述基本假設成立。其次,前後兩個年度,各校系會微調考科加權方法,且各考科考題難易度也不相同,針對各校系,我們定義四個指標,分別代表不同年度考科權重與不同年度考題難易度的指標,並探討其間的差異性。最後,在基本假設成立下,如何預測考生今年對某校系是否有上榜機會?又若兩個校系均有上榜機會時,何者機會大?文中針對新入學制度(考生可跨組選考並選填不同屬性之校系為志願),考量不同校系不同考科權重之意義,定義一套衡量錄取可能性高低的指標,提供考生選填志願時參考。實際應用上,為了免除考生與家長的負擔,我們根據所提出的落點預測方法,寫成一套有效的電腦落點分析軟體,提供考生利用網際網路免費下載使用,詳細操作說明亦刊載於網頁上。

在下一節中,我們先定義一些符號,進而陳述研究問題與統計方法。在第3節中,我們應用九十二和九十三學年度的資料,進行資料分析並整理結果。在第4節中,我們提出衡量錄取可能性高低的指標,建立一個落點預測的方法,據此寫成一套有效的電腦落點分析軟體。

2. 統計模式

聯分會所公佈的校系最低錄取加權總分是依據各校系採計科目和權重計算,採計科目愈多或加權權重愈大,均會導致最低錄取總分數有被灌水的可能性,故不宜直接應用該會公佈的數據,來探討前後兩年學生對校系的基礎評價是否改變。為了使各校系評價基準一致,本文建議使用加權平均作為比較的基礎,就是將最低錄取總分除以權數總和。如此一來,各校系的最低錄取加權平均分數就會介於0與100之間,我們就可以利用此同質性指標來表示學生對各校系的基礎評價。在本文中,我們不失通常性地假設各考科成績分佈為常態分佈(normal distribution)。由於不同年度各考科題目難易度不同,且同年度各考科題目難易度也不盡相同,本文進一步利用大考中心所公佈的,不同年度各考科成績的頂標、前標、均標、後標和底標(分別表示各考科第88%、75%、50%、

25% 和 12% 考生的成績), 來調整因題目難易度造成的差異, 增進預估各校系當年度最低錄取加權平均分數的準確性。在常態分佈假設下, 均標即代表該考科所有考生的平均成績, 因此在文中我們採用均標來調整指考成績; 成績屬於中上的考生, 可考慮僅與前 50% 考生競爭, 此時可採用前標成績來調整; 成績屬於中下的考生, 可考慮僅與後 50% 考生競爭, 此時可採用後標成績來調整。

2.1 符號定義

首先, 令 $k = 1, \dots, 9$ 分別代表各考試科目 (國文、英文、數甲、數乙、歷史、地理、物理、化學和生物); 針對某年度, 令 Y_k 代表考科 k 之成績, 則 $E(Y_k) = \mu_k$ 代表考科 k 成績之平均數 (mean), $Var(Y_k) = \sigma_k^2$ 代表考科 k 成績之變異數 (variance)。再令 $j = 1, 2$ 分別代表去年度 (九十二學年) 和今年度 (九十三學年), μ_{jk} 代表年度 j 考科 k 之平均成績, σ_{jk}^2 代表年度 j 考科 k 成績之變異數。由於大考中心沒有公佈各考科成績的變異數, 在變異數不可知的情況下, 為了簡化落點分析的過程, 我們假設所有考科成績具有常態分佈且不同年度各考科成績之變異數 (σ_{jk}^2) 均相等。

針對某固定校系, 令 X_{jk} 代表該校系年度 j 最低錄取總分考生考科 k 之成績, w_{jk} 代表該校系年度 j 考科 k 之權重, 則

$$\sum_{k=1}^9 w_{jk} X_{jk}, \quad j = 1, 2$$

分別代表去年和今年聯分會公佈的該校系最低加權錄取總分。針對該最低錄取總分考生, 其 k 考科成績應該與 Y_k , $k = 1, \dots, 9$ 具有相同分佈, 所以

$$E\left(\sum_{k=1}^9 w_{jk} X_{jk}\right) = \sum_{k=1}^9 w_{jk} \mu_{jk},$$

表示加權計分不同會導致錄取總分有被灌水的可能性。因此我們建議考量不同年度該校系的最低錄取加權平均分數

$$U_j = \frac{\sum_{k=1}^9 w_{jk} X_{jk}}{\sum_{k=1}^9 w_{jk}}, \quad j = 1, 2$$

作為比較的指標, 則

$$E(U_j) = \frac{\sum_{k=1}^9 w_{jk} \mu_{jk}}{\sum_{k=1}^9 w_{jk}}.$$

當不同年度各考科成績之平均數均相等時 ($\mu_{jk} \equiv \mu$), $E(U_1) = E(U_2) = \mu$, 此時不同年度的 U 值即可直接作為比較的指標。

事實上, 不同年度各考科題目難易度不同, 且同年度不同考科題目難易度也不盡相同, 透過標準化的過程, 我們進一步利用大考中心所公佈的均標 (在常態分佈的假設下, 均標即平均數), 來調整不同年度指考的成績, 定義

$$Z_j = \frac{\sum_{k=1}^9 w_{jk} (X_{jk} - \mu_{jk})}{\sum_{k=1}^9 w_{jk}}, \quad j = 1, 2$$

分別代表去年和今年該校系經調整考科題目難易度後的最低錄取加權平均分數, 則我們有 $E(Z_1) = E(Z_2) = 0$, 因此 Z 值更適合作為該校系調整後 (含考科難易度及權重) 的錄取指標。若再令

$$Q_{jl} = \frac{\sum_{k=1}^9 w_{jk} X_{lk}}{\sum_{k=1}^9 w_{jk}}, \quad j, l = 1, 2$$

則

$$q_{jl} = \frac{\sum_{k=1}^9 w_{jk} \mu_{lk}}{\sum_{k=1}^9 w_{jk}}, \quad j, l = 1, 2$$

分別代表該校系用年度 j 考科權重來計算年度 l 考科平均成績所計算出的指標, 同時我們有 Z_j 與 U_j 的關係式如下:

$$Z_j = U_j - q_{jj} = Q_{jj} - q_{jj}, \quad j = 1, 2.$$

2.2 問題與統計方法

2.2.1 基本假設的驗證及應用

為了提升落點分析的準確性, 在落點分析預測前, 須先檢驗其前提假設「前

後兩年學生對校系基礎評價不變」是否成立。在連續兩年放榜後, 根據某一學群或學類, 考量某校系 i 連續兩年的最低錄取加權平均分數 U_{1i} 及 U_{2i} , $i = 1, \dots, I$, 其中 I 為該學群或學類中校系總數。透過觀察 (U_{1i}, U_{2i}) , $i = 1, \dots, I$ 的散佈圖 (scatter plot), 我們可以初步瞭解該基本假設是否成立。亦可進一步應用 U_{1i} 與 U_{2i} 的樣本相關係數 (coefficient of correlation)

$$r_{U_1, U_2} = \frac{\sum_{i=1}^I (U_{1i} - \bar{U}_1)(U_{2i} - \bar{U}_2)}{\sqrt{\sum_{i=1}^I (U_{1i} - \bar{U}_1)^2 \sum_{i=1}^I (U_{2i} - \bar{U}_2)^2}},$$

來探討 U_{1i} 與 U_{2i} 的相關性, 其中 $\bar{U}_j = \sum_{i=1}^I U_{ji}/I$ 。若同時也考量不同年度各考科題目難易度所造成的影響, 則我們可用相同方法, 探討 2.1 節中經過調整的指標, Z_{1i} 與 Z_{2i} 的相關性。若前後兩年的指標 (U_{1i}, U_{2i}) 或 (Z_{1i}, Z_{2i}) 的線性相關程度高, 即代表基本假設「前後兩年學生對校系基礎評價不變」成立。此時, 我們可透過簡單線性迴歸模型 (simple linear regression)

$$E(U_{2i}) = \alpha + \beta U_{1i}, \quad i = 1, \dots, I$$

或

$$E(Z_{2i}) = \alpha + \beta Z_{1i}, \quad i = 1, \dots, I$$

透過對 α 和 β 的研究, 可以建立一個落點分析的預測模式; 有關簡單線性迴歸模型分析理論請參考 [4]。在第 3.1 節中, 我們會針對九十二和九十三學年度的資料, 進行簡單線性迴歸分析, 並於第 4 節中詳細說明有關落點預測的方法。

2.2.2 考科權重與考題難易度的影響

前後兩個年度, 各校系會微調考科加權方法, 且各考科考題難易度也不相同, 採計的考科權重與考題難易度是否會對各校系造成影響? 由於各校系僅有不同年度的兩個錄取指標 Q_{11} 及 Q_{22} , 這些指標僅能用來探討各校系不同年度間的差異性, 爲了進一步探討權重或考題難易度對各校系的影響, 針對各校系, 我們採用 2.1 節中已定義的四個指標 q_{jl} , $j, l = 1, 2$ 來探討考科權重與考題難易度是否會對各校系造成影響。在連續兩年放榜後, 根據某一學群或學類, 討論各校系 (q_{11}, q_{12}) 或 (q_{21}, q_{22}) 的關係, 可瞭解不同年度各考科題目難易度

對各校系造成的影響; 討論各校系 (q_{11}, q_{21}) 或 (q_{12}, q_{22}) 的關係, 可瞭解不同年度採計的考科權重對各校系造成的影響; 討論各校系 (q_{11}, q_{22}) 或 (q_{12}, q_{21}) 的關係, 可瞭解各考科題目難易度及採計的考科權重對各校系的交互作用效應。透過散佈圖的觀察及相關係數的計算, 我們可以初步了解各校系的四個數據 $q_{jl}, j, l = 1, 2$ 是否合乎統計獨立的假設, 若是則可應用二因子變異數分析法 (Two-way ANOVA) 做進一步的研究。

固定某學群或學類, 針對某校系 i , 考慮二因子變異數分析模型

$$q_{jli} = \mu + \alpha_j + \beta_l + (\alpha\beta)_{jl} + \epsilon_{jli}, \quad j, l = 1, 2, \quad i = 1, \dots, I,$$

其中 I 為該學群或學類中校系總數, α_j 表示年度 j 權重的因子效應, β_l 表示年度 l 考科難易度的因子效應, $(\alpha\beta)_{jl}$ 表示年度 j 權重與年度 l 考科難易度的交互作用效應, ϵ_{jli} 為獨立的常態誤差項。我們應用檢定虛無假設 $H_{01} : \alpha_1 = \alpha_2 = 0$ 是否成立, 來探討不同年度各考科題目難易度對各校系 q 值所造成的影響; 檢定虛無假設 $H_{02} : \beta_1 = \beta_2 = 0$ 是否成立, 來探討不同年度採計的考科權重對各校系 q 值所造成的影響; 檢定虛無假設 $H_{03} : (\alpha\beta)_{jl} = 0, j, l = 1, 2$ 是否成立, 來探討各考科題目難易度及採計的考科權重對各校系 q 值的交互作用效應; 有關二因子變異數分析理論請參考 [5]。在 3.2 節中, 我們針對九十二和九十三學年度的資料, 進行二因子變異數分析, 以便瞭解不同年度權重與考科難易度是否會對各校系的 q 值造成影響。

3. 資料分析

根據 2.1 節符號定義, 資料 (U_1, U_2) 可由聯分會九十二和九十三學年度公佈的資料, 經加權平均計算得到, 再利用大考中心九十二和九十三學年度公佈的各考科成績的均標, 調整而得資料 (Z_1, Z_2) 。為了資料完整性及一致性, 我們先剔除九十三學年度新增校系及加考術科的校系 (如國立台北師範學院音樂教育學系) 後, 研究資料集中共計有 1,332 個校系。再依招生簡章將這些校系分成 19 個學群, 其中某些校系會重複出現在不同學群中, 例如成功大學工業與資訊管理學系就重複出現在管理學群和工程學群中。表 1-1 列出所有校系中依 U_2 大小排序後, 前 20 個校系之 U_2 及 U_1 值; 表 1-2 列出工程學群中的對應資料供參考。我們將針對所有校系及各個學群分類, 進行資料分析並報告結果。

表1-1. 所有校系 ($I = 1,332$) 中依 U_2 排序之前20個校系名稱

學校名稱	科系	U_2	U_1
國立臺灣大學	醫學系 (自費)	90.05	87.07
國立臺灣大學	醫學系 (公費)	89.82	86.75
國立陽明大學	醫學系 (自費)	88.55	85.84
國立陽明大學	醫學系 (公費)	87.36	84.80
國立成功大學	醫學系	86.88	83.93
長庚大學	醫學系	86.23	83.15
臺北醫學大學	醫學系 (公費)	85.59	82.29
國立屏東師範學院	英語教育學系 (公費)-台東縣	85.00	76.13
國立臺灣大學	財務金融學系	84.58	83.91
高雄醫學大學	醫學系 (公費)	84.33	81.23
臺北醫學大學	醫學系 (自費)	84.18	81.28
國立臺灣大學	電機工程學系	84.15	82.77
慈濟大學	醫學系 (公費)	83.90	80.89
高雄醫學大學	醫學系 (自費)	83.29	80.46
慈濟大學	醫學系 (自費)	83.03	79.85
國立臺灣大學	牙醫學系	83.02	79.15
國立臺灣大學	物理學系	82.95	81.40
中國醫藥大學	醫學系 (公費)	82.62	79.97
輔仁大學	醫學系	82.57	79.60
國立交通大學	電機資訊學院學士班	82.37	80.90

3.1 基本假設的驗證及應用

爲了驗證「前後兩年學生對校系基礎評價不變」是否成立, 我們針對所有校系及各個學群分類, 採用 2.2.1 節所描述的方法, 製作 (U_1, U_2) 及 (Z_1, Z_2) 的散佈圖並計算其相關係數。圖1 爲部分分類群的 (U_1, U_2) 及 (Z_1, Z_2) 散佈圖, 圖中說明資料散佈大致均呈現線性關係; 表2 詳列各分類資料的相關係數, 其結果進一步說明前後兩年的資料, 明顯地高線性相關。據此, 我們首次定義 2.1 節中的 U 或 Z 爲代表學生對校系基礎評價的指標, 並應用此指標及實際資料, 驗證落點分析的前提假設「前後兩年學生對校系基礎評價不變」的合理性。

由表 2 觀察到在各分類中, (Z_1, Z_2) 的相關係數均比 (U_1, U_2) 的高 (其他學群

表1-2. 工程學群 ($I = 196$) 中依 U_2 排序之前 20 個校系名稱

學校名稱	科系	U_2	U_1
國立臺灣大學	電機工程學系	84.15	82.77
國立臺灣大學	材料科學與工程學系	81.42	78.30
國立清華大學	電機工程學系	79.30	76.79
國立交通大學	電子工程學系	77.72	75.92
國立交通大學	電信工程學系	75.83	74.10
國立臺灣大學	化學工程學系	74.98	70.94
國立清華大學	材料科學工程學系	74.98	71.72
國立交通大學	電機與控制工程學系	74.57	72.10
國立臺灣大學	機械工程學系	74.22	71.77
國立交通大學	材料科學與工程學系	73.54	70.81
國立清華大學	化學工程學系	72.66	69.41
國立成功大學	電機工程學系	72.59	70.63
國立清華大學	動力機械工程學系	71.33	68.38
國立臺灣大學	生物產業機電工程學系	71.19	68.26
國立清華大學	工程與系統科學系	70.16	67.48
國立臺灣大學	工程科學及海洋工程學系	69.95	65.29
國立成功大學	材料科學及工程學系	69.78	67.38
國立交通大學	機械工程學系	69.28	66.33
國立中央大學	電機工程學系	69.09	66.85
國立中山大學	材料與光電工程學系	68.85	63.14

除外), 且 (Z_1, Z_2) 相關係數值大致均高達 0.95 以上, 此現象符合我們建議將 U 值透過調整考科題目難易度調整轉換為 Z 值的理由。因此我們採用 (Z_1, Z_2) 的簡單線性迴歸模型

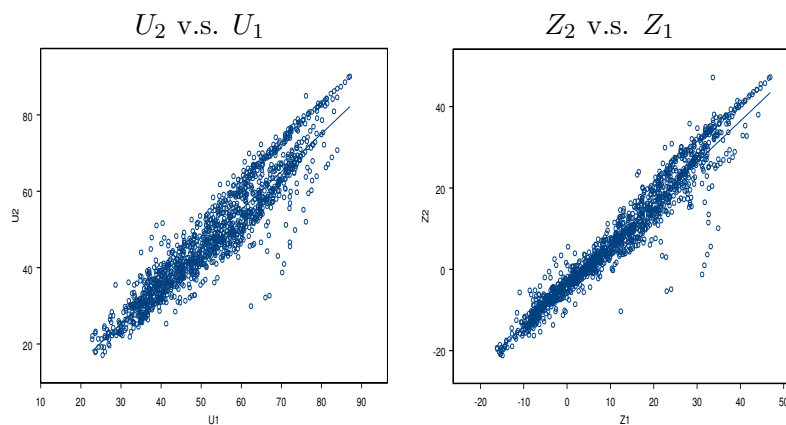
$$E(Z_{2i}) = \alpha + \beta Z_{1i}, \quad i = 1, \dots, I, \quad (1)$$

企圖應用 α, β 的訊息, 建立一個落點分析的預測模式。由於在選填志願時, 當年度該校系的最低錄取加權平均分數 U_2 值未知 (尚未放榜), 故 Z_2 值亦未知, 此時 α, β 的值, 無法於選填志願時預先估計出來輔助落點分析。換言之, α, β 的點估計值對落點分析無直接幫助, 因此我們專注於與 α, β 有關的假設檢定。

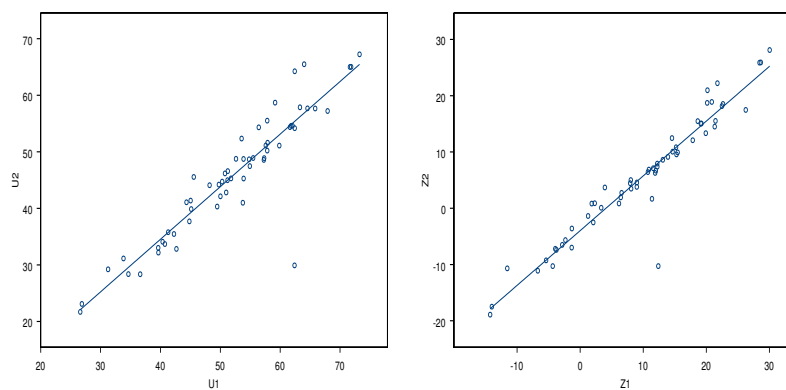
根據圖 1 顯示, 虛無假設 $H_0: \beta = 1$ 有可能成立, 且資料 (Z_1, Z_2) 顯示 α 的值似乎很小而且可以忽略不計。因此我們根據過去兩年資料, 在模型 (1) 下, 針對

圖1. 分類群的 (U_1, U_2) 及 (Z_1, Z_2) 散佈圖

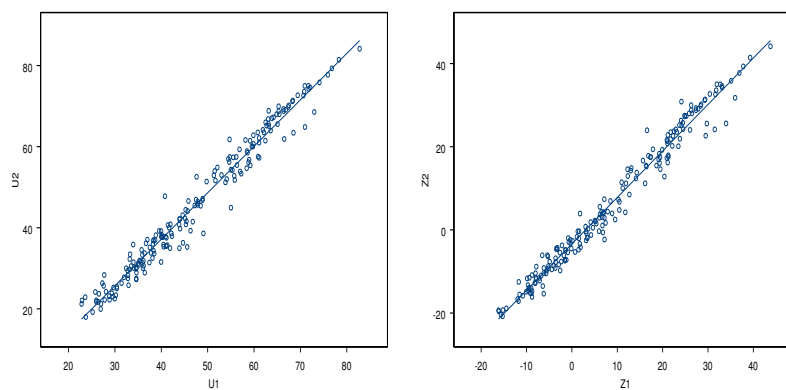
所有校系



大眾傳播學群



工程學群



文史哲學學群

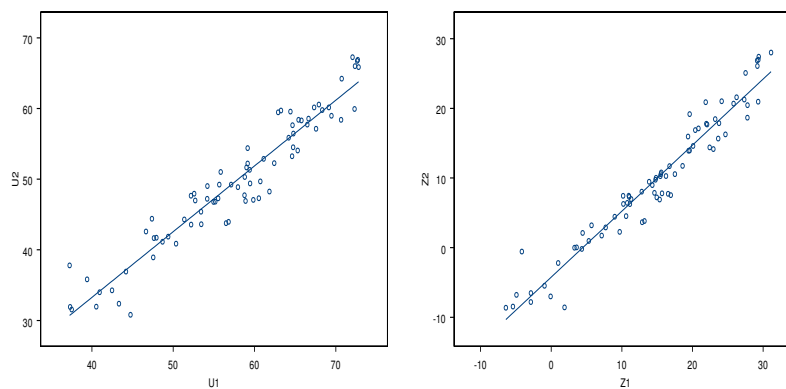


表2. 各分類中 (U_1, U_2) 及 (Z_1, Z_2) 的相關係數

分類	總數	r_{U_1, U_2}	r_{Z_1, Z_2}
所有校系	1332	0.9348	0.9663
大眾傳播學群	59	0.9166	0.9579
工程學群	196	0.9851	0.9880
文史哲學學群	77	0.9566	0.9682
外語學群	99	0.9415	0.9565
生命科學學群	49	0.9761	0.9803
地球與環境學群	36	0.9309	0.9815
法政學群	75	0.9777	0.9842
社會與心理學群	69	0.9413	0.9663
建築與設計學群	43	0.8646	0.9573
財經學群	142	0.9830	0.9867
教育學群	109	0.7851	0.7885
資訊學群	112	0.9811	0.9894
農林漁牧學群	41	0.9681	0.9730
管理學群	206	0.9637	0.9824
數理化學學群	116	0.8566	0.8683
醫藥衛生學群	142	0.9744	0.9875
藝術學群	14	0.9460	0.9518
體育學群	15	0.7979	0.8739
其他學群	4	0.9961	0.9815

各分類進行假設 $\beta = 1$ 及 $\alpha = 0$ 的檢定, 結果列於表 3, 其中估計值 $\hat{\beta}$ 均相當靠近 1, $\hat{\alpha}$ 相當靠近 -3 , t_1 -value 表示檢定 $H_0 : \alpha = 0$ 的統計量值, t_2 -value 表示檢定 $H_0 : \beta = 1$ 的統計量值, R^2 為決定係數 (coefficient of determination)。由於大多數學群的校系總數都大於 40, 在 0.05 顯著水準下, 若引用 $z_{0.025} = 1.96$ 做決策, 則假設 $H_0 : \alpha = 0$ 皆被拒絕; 在 9 個學群中, 假設 $H_0 : \beta = 1$ 被拒絕, 此結論與各學群的樣本數偏大有關。綜合言之, 仔細觀察表 3 中的 t -統計量值, 我們發現相對於 Z_1 與 Z_2 值的大小, α 的值可以忽略不計, 且 β 的值可以視為 1。因此我們採用模式 $\widehat{Z}_2 = Z_1$, 亦即應用

$$\widehat{U}_2 = U_1 - q_{11} + q_{22} \quad (2)$$

來估計當年度校系的最低錄取加權平均分數。

表3 迴歸模式(1) 下的相關統計量

分類	總數	$\hat{\alpha}$	t_1 -value	$\hat{\beta}$	t_2 -value	R^2
所有校系	1332	-3.9862	29.4372	1.0192	1.8462	0.9337
大眾傳播學群	59	-3.9819	-7.0522	0.9729	0.7021	0.9174
工程學群	196	-3.3632	-15.9112	1.1175	9.3253	0.9761
文史哲學學群	77	-4.2451	-8.5058	0.9484	-1.8233	0.9375
外語學群	99	-4.0186	-6.9287	0.9135	-3.0565	0.9148
生命科學學群	49	-3.7016	-6.0406	1.0637	2.0351	0.9609
地球與環境學群	36	-3.3448	-5.7890	1.0812	2.2431	0.9633
法政學群	75	-4.7518	-11.0712	1.0234	1.0833	0.9686
社會與心理學群	69	-4.1021	-8.2361	0.9731	-0.8486	0.9337
建築與設計學群	43	-2.9048	-6.1578	0.9081	-2.1472	0.9165
財經學群	142	-3.7175	-14.9224	1.0057	0.4071	0.9735
教育學群	109	-5.1607	-3.7588	0.8995	-1.6468	0.6217
資訊學群	112	-4.1867	-17.5067	1.0625	4.2230	0.9790
農林漁牧學群	41	-3.0235	-4.7402	1.1014	2.4200	0.9466
管理學群	206	-3.9630	-20.2009	1.0020	0.1493	0.9651
數理化學學群	116	-4.4162	-4.1385	0.9934	0.1241	0.7538
醫藥衛生學群	142	-2.8422	-8.6995	1.0653	4.5664	0.9752
藝術學群	14	-7.2748	-2.9695	1.2209	1.9463	0.9060
體育學群	15	-3.0621	-1.6151	0.5432	-5.4510	0.7637
其他學群	4	-6.9568	-1.8283	1.1116	0.7275	0.9633

3.2 考科權重與考題難易度的影響

接著討論不同年度, 採計的考科權重與考題難易度是否會對各校系造成影響? 表4列出 q_{jl} , $j, l = 1, 2$ 的相關係數, 資料顯示 q_{jl} 之間不具明顯線性關係。因此我們應用 2.2.2 節中二因子變異數模型

$$E(q_{jli}) = \mu + \alpha_j + \beta_l + (\alpha\beta)_{jl}, \quad j, l = 1, 2, i = 1, \dots, I,$$

針對各分類, 分別檢定假設

$$H_{01} : \alpha_1 = \alpha_2 = 0$$

$$H_{02} : \beta_1 = \beta_2 = 0$$

$$H_{03} : (\alpha\beta)_{jl} = 0, j, l = 1, 2$$

是否成立, 其對應之 p -value 整理於表 5 中。在 0.05 顯著水準下, 只有工程、外語、財經和管理等四個學群, 對不同年度採計的考科權重變化呈現顯著差異性; 除了工程、地球與環境和體育三個學群外, 均對不同年度採計考科考題難易度變化呈現顯著差異性; 考科權重變化與考題難易度變化均無顯著交互效應存在。綜合言之, 不同年度採計的考科權重變化沒有太大的影響, 但是不同年度考題難易度對 q 值有影響。

表 4. q_{ij} 之間的相關係數

分類	總數	$r_{q_{11}, q_{12}}$	$r_{q_{11}, q_{21}}$	$r_{q_{11}, q_{22}}$	$r_{q_{12}, q_{21}}$	$r_{q_{12}, q_{22}}$	$r_{q_{21}, q_{22}}$
所有校系	1332	0.3566	0.8200	0.6428	0.1732	0.4962	0.7707
大眾傳播學群	59	0.6949	0.6966	-0.0169	0.2157	0.3223	0.1301
工程學群	196	0.6071	0.7410	0.4938	0.3511	0.7091	0.6168
文史哲學學群	77	0.5610	0.7469	0.5332	0.3918	0.6933	0.6272
外語學群	99	0.6781	0.2609	0.1820	0.1062	0.2171	0.6225
生命科學學群	49	0.6298	0.6710	0.6092	0.4193	0.7613	0.8131
地球與環境學群	36	0.3384	0.8892	0.2773	0.4340	0.8975	0.5173
法政學群	75	0.5784	0.5939	0.0992	0.1418	0.3237	0.2620
社會與心理學群	69	0.1759	0.6778	-0.1887	-0.1316	0.5777	-0.0410
建築與設計學群	43	0.6070	0.4129	-0.2255	-0.0486	-0.0683	0.2254
財經學群	142	0.5573	0.5533	-0.0152	0.0957	0.3224	0.1177
教育學群	109	0.0498	0.9280	-0.0981	-0.0432	0.8116	-0.0348
資訊學群	112	0.6296	0.8362	0.4468	0.4478	0.6028	0.5829
農林漁牧學群	41	0.5799	0.6022	0.2912	0.2206	0.5618	0.3368
管理學群	206	0.3614	0.6813	0.1175	0.0765	0.6022	0.2150
數理化學學群	116	0.4788	0.7155	0.4062	0.4880	0.8727	0.6077
醫藥衛生學群	142	0.6707	0.6216	0.3636	0.3257	0.6371	0.6041
藝術學群	14	0.4333	0.8743	0.4174	0.2188	0.9472	0.2732
體育學群	15	0.2549	0.3650	0.0238	-0.0490	0.7275	0.2694
其他學群	4	0.9384	-0.1009	-0.0934	-0.3461	-0.2146	0.9046

表5. 檢定考科權重與考題難易度的影響之 p -value

分類	H_{01}	H_{02}	H_{03}
所有校系	0.0995	0.0000	0.9018
大眾傳播學群	0.7623	0.0000	0.8910
工程學群	0.0056	0.0767	0.7909
文史哲學學群	0.5416	0.0000	0.3595
外語學群	0.0014	0.0000	0.4831
生命科學學群	0.3814	0.0000	0.6851
地球與環境學群	0.2973	0.7589	0.8443
法政學群	0.9837	0.0000	0.5990
社會與心理學群	0.9904	0.0000	0.9910
建築與設計學群	0.4483	0.0001	0.4157
財經學群	0.0193	0.0000	0.2716
教育學群	0.8045	0.0000	0.8253
資訊學群	0.2879	0.0000	0.6800
農林漁牧學群	0.5426	0.0000	0.9173
管理學群	0.0198	0.0000	0.7079
數理化學學群	0.4438	0.0068	0.9346
醫藥衛生學群	0.9843	0.0000	0.8894
藝術學群	0.9693	0.0000	0.5481
體育學群	0.2393	0.3581	0.9885
其他學群	0.1981	0.8193	0.7918

4. 電腦落點分析

根據 3.1 節資料分析結果, 在基本假設成立下, 我們採用某校系去年的最低錄取加權平均分數 U_1 , 配合該系去年採計的考科權重及考科平均成績所計算出的指標 q_{11} 及該系今年採計的考科權重及考科平均成績所計算出的指標 q_{22} , 經由 (2) 式

$$\widehat{U}_2 = U_1 - q_{11} + q_{22}$$

估計今年度校系的最低加權平均錄取分數, 由於 q_{11} 與 q_{22} 的同質性, 相對於 U_1 值的大小, $q_{11} - q_{22}$ 的值不會影響 \widehat{U}_2 大於 0 的基本要求。有別於大考中心或一般坊間在落點分析時, 應用差距大小代表錄取機會, 我們將介紹一個實用

的指標, 提供考生衡量 (預測) 其成績相對於某一校系會被錄取的機會。

針對某一考生而言, 令 Y_k 表示其考科 k 之成績, 固定某校系今年之考科權重 w_{2k} , 定義

$$Y = \frac{\sum_{k=1}^9 w_{2k} Y_k}{\sum_{k=1}^9 w_{2k}}$$

代表此考生相對於該校系的加權平均分數。基本上, Y 值愈大表示此考生成績較好, 被該校系錄取的機會愈大; 同時 \widehat{U}_2 代表今年度校系的最低錄取加權平均分數, 根據 3.1 節資料分析顯示, 在基本假設成立下, \widehat{U}_2 可以代表學生對校系基礎評價的指標, 其值愈小表示校系基礎評價較差, 此考生被該校系錄取的機會愈大。因此, 我們定義一個實用的指標

$$T = \frac{Y}{\widehat{U}_2} = \frac{Y}{U_1 - q_{11} + q_{22}}$$

作為該考生預測其成績相對於某一校系會被錄取的機會。值得注意的是, Y 和 \widehat{U}_2 的值均介於 0 和 100 之間, 所以 T 值的範圍為 $[0, \infty)$; 當 T 值小於 1 時, 表示考生成績低於估計的最低錄取加權平均分數; 當 T 值大於 1 時, 表示考生成績高於估計的最低錄取加權平均分數; T 值愈大表示被該校系錄取的機會愈大。再者, 在 3.1 節中, $\hat{\alpha}$ 值均為負數, 表示 (2) 中 \widehat{U}_2 值稍有高估, T 值稍有低估, 亦即 T 值保守地預測考生會被錄取的機會。針對所有校系研究校系基礎評價時, 3.1 節資料分析亦顯示相當好的結果, 因此指標 T 亦適用於不同學群或學類的比較, 這突破了傳統上不同學群或學類不知如何評比的問題。

透過文中統計理論分析, 輔以實際資料分析作驗證, 我們依此建立一套電腦落點分析軟體, 置放於「統計網路學習館」網頁上 (<http://estat.ncku.edu.tw>), 提供免費下載。該軟體套用資料庫系統, 考生只需輸入指考各科成績及學測成績, 再依學校、學群、學類、公私立和地區別等屬性, 可大量選取有興趣的校系 (最多為全部校系總數), 快速計算考生成績相對應於各校系的 T 值, 並自動進行排序, 方便考生於 T 值大於 1 且較靠近 1 的 100 個 (九十四年度) 校系中, 再依自己的興趣選填志願。如此一來, 一般高中升學輔導室建議考生「採用 1/3 低分高就, 1/3 合理落點, 1/3 安全保障的穩當選填志願法」就不值得再

鼓勵了。

5. 結論

考招分離, 多元入學方案, 本意是導引考生在大學志願選填時能符合自己的性向, 使各大學依據自己的特色單獨或聯合招收適才學生。但是在傳統校系評價基礎下, 大學指定考試分發入學前, 選填志願遂成爲每年十萬考生及其家人的困擾問題。基於推廣統計應用的理念及曾爲考生家長的困擾經驗, 遂完成此研究。本文首次以統計理論定義代表學生對校系基礎評價的指標, 並應用統計方法及實際資料驗證落點分析的前提假設成立。進一步應用大考中心發佈的各單一考科考生成績累計表, 各校系之指考科目與加權方法, 及前一年度校系的最低錄取加權總分, 寫成一個有效的電腦落點分析軟體, 提供考生及家長一個免費且效率高的選填志願參考。但是填寫志願最重要的還是, 依自己的意願和性向, 搭配自己的成績優勢, 避免高分落榜和高分低填, 找自己喜好的校系填入。

致謝詞: 本研究承蒙成功大學資工所 94 級碩士張維舜先生幫忙撰寫部份電腦落點分析軟體, 特此誌謝; 感謝評審的寶貴意見。

參考文獻

大考中心的網頁。 <http://test.x01.org/2003/data/027.htm>

中華電信 *Hinet* 學習網。 <http://analyse.eleam.hinet.net>

大同大學電機系。 <http://www2.tisnet.net.tw>

Neter, J., Kutner, M. H. and Wasserman, W. (1996) *Applied Linear Regression Models*. 3rd edition. McGraw-Hill College.

Montgomery, D. C. (1991) *Design and Analysis of Experiments*. 3rd edition. John Wiley & Sons, New York.

[民國 94 年 3 月收稿, 民國 94 年 5 月接受。]

STATISTICAL FORECASTING AFTER THE COLLEGE ENTRANCE EXAMINATION

Mei-Mei Zen¹, Zue-Seng Chen² and Chia-Hao Jhan¹

¹Department of Statistics, National Cheng-Kung University and

²Computer and Network Center, National Cheng-Kung University

ABSTRACT

After the important College Entrance Examination, a senior high school student has to select at most 100 of 1600 departments for advance study. Since the weight of counting the scores of nine individual examinations is determined by each department, which varies year by year, the problem of selection and priority of certain departments becomes a big trouble to students and parents yearly. In this paper, following statistical concepts and analysis, we build up an index for each student to represent the possibility of entering any interested department. A real data are applied to check the efficiency of our index. Based on this index, a software is written and offered on website for free download.

Key words and phrases: Weighted score, simple linear regression.

AMS 2000 subject classifications: Primary 62K05; secondary 62J10.