

# 雲林縣雙語成績分析

## ETA/ETF 雙語教育對英文成績的分別

研究目的:利用變異數分析說明不同教學方法(eta,etf,else)是否對英文成績有顯著差異

資料來源:由雲林縣的各小學四年級的各個英語成績作為分析

資料標籤:由(ETA 美籍英語教學助理)，ETF (美籍英語教學助理)，ELSE(其他教學助理)做為不同的類別標籤當成不同英語教學方法，每個標籤內的不同學生都有多筆英文成績

資料分布:eta 教學方法(11 筆)，etf 教學方法(3 筆)，ELSE(141 筆)，由此可以看出資料分布極不均勻，沒有資料缺失

實驗方法:

將 data 匯入 python 後首先我們移除用不到的學校代碼，以及英文總體答對率(因為我們直接對各科作分析，因此總體答對率並不再我們的考慮範圍)，再根據雙語教育的名冊將不同教學方法的學校將其分為 eta/etf/else，以統一類別變數，其後我們直接對原始數據做 one-way-manova 以粗略分析不同的教學方法(eta/etf/else)，而結果如下

Multivariate linear model						
=====						
Intercept	Value	Num DF	Den DF	F Value	Pr > F	
-----						
Wilks' lambda	0.1718	8.0000	145.0000	87.3884	0.0000	
Pillai's trace	0.8282	8.0000	145.0000	87.3884	0.0000	
Hotelling-Lawley trace	4.8214	8.0000	145.0000	87.3884	0.0000	
Roy's greatest root	4.8214	8.0000	145.0000	87.3884	0.0000	
-----						
學校名稱	Value	Num DF	Den DF	F Value	Pr > F	
-----						
Wilks' lambda	0.9382	16.0000	290.0000	0.5879	0.8924	
Pillai's trace	0.0625	16.0000	292.0000	0.5882	0.8922	
Hotelling-Lawley trace	0.0653	16.0000	233.7121	0.5884	0.8913	
Roy's greatest root	0.0530	8.0000	146.0000	0.9672	0.4640	
=====						

我們由圖可以獲得以下三種結論:

- Wilks' lambda = 0.9382,  $p = 0.8924$  : 表示學校名稱變量在模型中不顯著。
- Pillai's Trace、Hotelling-Lawley 和 Roy 的統計量也給出一致的結果，F 值都很小， $p$  值遠大於 0.05。
- 結論：教學方法(eta/etf/else)對多變量模型的影響不顯著，這表示學校名稱與響應變量之間並沒有統計學上的關聯。

由於這只能代表總體的差異，但我們仍然不知道不同教學方法對各科的差異，因為教學方法可能對某科有顯著差異，但是他在整體差異不高導致差異不顯著，因此我們還要對各個變因作變異數分析也就是 anova。

但是由於各個教學方法的數據量差異極大，因此我們對較小的資料(etf,eta)做 over sampling 取到 major(else)，並且因為一組成績是同個考生考的，這表彼此間具有相關性，於是為了生成的虛擬資料高分的都高分，低分的都低分，因此我們把 eta,etf 的資料合併，並依據總分分出四群(low,high,medium1,medium2)讓不同分群的學生只在他對應的總分區間抽樣取出虛擬成績，以避免 over sample 對分析結果產生太大的誤差，而在 over sample 之後，為了瞭解要使用哪種變異數分析的方式，我們對我們要做的單科成績做出#Levene's test，以聽力\_語音聽辨為範例，結果如下

statistic:76.96  
p\_value:0.00

由於 p-value 很小

結論：

組間的變異數不相等，即方差不齊性。因此我們使用#welth anova 來計算聽力\_語音聽辨對不同教學方法是否具有統計上的差異，結果如下

Source	ddof1	ddof2	F	p-unc	np2
0 group	1	167.897033	0.353844	0.552746	0.001262

- $p\text{-unc} = 0.5527$  :
  - $p > 0.05$  : 我們未能拒絕虛無假設。
  - 這意味著不同組別之間的均值沒有統計學上的顯著差異。
- 部分  $\eta^2$  ( $np2$ ) = 0.0013 :
  - 部分  $\eta^2$  表示組間變異占總變異的比例。
  - 這個數值非常小，說明組別變異對總變異的貢獻極其有限。

因此我們可以得出結論:不同的教學方法對於聽力\_語音聽辨沒有統計上的明顯差異，而我們也能照做得知不同教學方法對其他考科

結論:由於其他人的#Levene's test p 都極小，因此我們都使用#welth anova 而在其中，

### 結果整理

項次	指標	ddof1	ddof2	F 值	p-unc	np2	結果解釋
1	group	1	235.242844	4.952611	0.027	0.01738	顯著差異 ( $p < 0.05$ )
2	group	1	230.021245	2.678346	0.103089	0.009475	不顯著 ( $p > 0.05$ )
3	group	1	243.155086	21.99719	0.000005	0.072839	顯著差異 ( $p < 0.05$ )
4	group	1	164.504186	8.886551	0.003309	0.030761	顯著差異 ( $p < 0.05$ )
5	group	1	212.899618	0.002314	0.961678	0.000008	不顯著 ( $p > 0.05$ )
6	group	1	164.504186	8.886551	0.003309	0.030761	顯著差異 ( $p < 0.05$ )
7	group	1	217.319999	7.307295	0.007411	0.025434	顯著差異 ( $p < 0.05$ )

(np2)：

- 顯著結果的效果量 np2 大多在 0.017~0.072 之間，表示組別對變異的解釋力有限（中等偏小）。
- 例如第 3 項("聽力\_文化節慶理解")的效果量 0.072839 是最高的，說明組別對變異的解釋能力相對較強

#因此由結果得知不同教學方法只對("聽力\_文化節慶理解")有較為顯著差異

附錄:python 執行相關性檢定

```
聽力_辭彙聽辨 statistic:17.90
聽力_辭彙聽辨 p_value:0.00
聽力_教室生活辭句理解 statistic:8.98
聽力_教室生活辭句理解 p_value:0.00
聽力_文化節慶理解 statistic:15.74
聽力_文化節慶理解 p_value:0.00
閱讀能力_字詞辨識 statistic:65.26
閱讀能力_字詞辨識 p_value:0.00
閱讀能力_句子理解 statistic:33.12
閱讀能力_句子理解 p_value:0.00
閱讀能力_字詞辨識 statistic:65.26
閱讀能力_字詞辨識 p_value:0.00
閱讀能力_文化節慶理解 statistic:27.63
閱讀能力_文化節慶理解 p_value:0.00
```

