壹●前言

在高中我們學會分析二維數據資料，來研究兩筆資料間的關連性。若只用二維分析的方法，則變數數量不敷使用，所以我們請老師指導我們如何從二維迴歸直線，延伸到三維的迴歸平面。我們首先探討國中基測的數學及自然科平均值和大學學測的數學成績的關係，再更進一步將數學科和自然科的成績當作兩個變數，個別考慮，藉由最小平方法與配方法求得迴歸平面公式，最後推廣 n 維空間的迴歸超平面公式，但由於變數較多，所以我們利用函數求極值點微分為零的想法，求得推導出迴歸超平面的係數，進而得到我們要的結果。

研究目的

我們大都認為數學與自然科的成績表現關係密切，也同意國中基測成績愈高，則大學學測考好的〝機會〞應該愈大，因此才會以成績篩選入學學生。針對這兩點直觀的猜測，我們以本校畢業生作為研究對象，探討國中基測的數學科和自然科分數的平均值，與大學學測的數學成績，是否呈現正相關？更進一步，將國中基測的數學科和自然科成績個別考慮，探討它們與大學學測數學成績的相關性，因此必須將二維的迴歸直線，推廣到三維的迴歸平面，應該如何定義與計算。

貳●正文

一、研究過程

問題的緣由：

在生活中，時常會碰到資料數據的解讀。特別是我們關心的數據中，若存在某種相關性， 我們會想了解某一資料數據變化是否會影響到另一資料數據變化？本作品就是想探討一個科目的成績高低，是否會影響到相關學科的成績，例如，我們大都認為，自然學科與數學科成績的相關性，會比自然學科與社會學科或是語文學科的相關性來得大，在高中數學機率與統計這單元裡，我們學到相關係數這個定義：給定 n 組資料$(x_1,y_1),(x_2,y_2),(x_3,y_3),\cdots\cdots,(x_n,y_n)$其相關係數$r_{xy}$定義為

$$r_{xy}=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}\qquad \text{其中}\ \overline{x}=\frac{x_1+x_2+\cdots+x_n}{n}\ ,\ \overline{y}=\frac{y_1+y_2+\cdots+y_n}{n}$$

我們可以利用相關係數去判別我們所關心的資料是否有高度的相關性，例如以本校去年 337 個畢業生為例，如圖 1 為去年高三畢業生國中基測數學成績與大學學測數學成績的資料散佈圖，我們知道可以利用最小平方法求出數據的迴歸直線 $y=ax+b$

，其中$a=\dfrac{r_{xy}s_y}{s_x}$，$b=\overline{y}-\dfrac{r_{xy}s_y}{s_x}\overline{x}$，$s_x=\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}}$ 圖中黑色直線爲此資料的迴歸直線，我們可以利用此直線去做分數的預測。不過我們高中數學所學資料分析爲二維資料，而我們想要分析的是國中基測的數學科和自然科成績兩科，與大學學測數學科成績的相關性，因此必須將二維的迴歸直線，推廣到三維的迴歸平面。

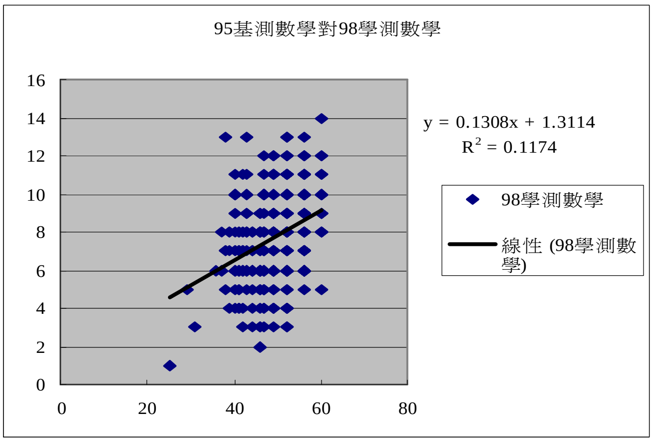


圖 1. 國中基測數學成績與大學學測數學成績的資料散佈圖

二、資料迴歸分析的探討

(一)、三維資料迴歸分析討論：

假設我們從母體中隨機抽樣n筆三維資料，(x_1,y_1,z_1)、(x_2,y_2,z_2)、$\cdots(x_n,y_n,z_n)$，由於資料點是在三維空間，我們很自然地會想利用最小平方法求迴歸直線的作法，去求空間中的迴歸平面，即考慮平面方程式 z=ax+by+c ，其中a,b,c爲待定的係數，定義殘差爲$e_i=z_i-\hat{z}_i$ ，其中$\hat{z}_i=ax_i+by_i+c$ ，我們想求a,b,c爲多少時，殘差平方和最小。在推導我們的結果前，先來引述一個引理。

引理 1：

若函數 $f(x,y)=ax^2+2bxy+cy^2+dx+ey+f$ ，其中 $a>0$ ，$b^2-ac<0$則 $f(x,y)$會在$(x,y)=(h,k)$發生最小值，此時$h=\dfrac{\begin{vmatrix}-d & 2b\\ -e & 2c\end{vmatrix}}{\begin{vmatrix}2a & 2b\\ 2b & 2c\end{vmatrix}}$ ，$k=\dfrac{\begin{vmatrix}2a & -d\\ 2b & -e\end{vmatrix}}{\begin{vmatrix}2a & 2b\\ 2b & 2c\end{vmatrix}}$

pf ：首先令$x=x'+h$ ，$y=y'+k$代入 $f(x,y)$

$$f(x'+h,y'+k)=a\left(x'+h\right)^2+2b\left(x'+h\right)\left(y'+k\right)+c\left(y'+k\right)^2+d\left(x'+h\right)+e\left(y'+k\right)+f$$

$$f(x'+h,y'+k)=ax'^2+2bx'y'+cy'^2+(2ah+2bk+d)x'+(2bh+2ck+e)y'+f(h,k)$$

令 x' 項係數與 y' 項係數爲零即

$\begin{cases}2ah+2bk+d=0\\2bh+2ck+e=0\end{cases}$ ，利用克拉瑪公式求得

$h=\dfrac{\Delta_h}{\Delta}=\dfrac{\begin{vmatrix}-d & 2b\\-e & 2c\end{vmatrix}}{\begin{vmatrix}2a & 2b\\2b & 2c\end{vmatrix}}$ ， $k=\dfrac{\Delta_k}{\Delta}=\dfrac{\begin{vmatrix}2a & -e\\2b & -e\end{vmatrix}}{\begin{vmatrix}2a & 2b\\2b & 2c\end{vmatrix}}$ ，注意 $\Delta=\begin{vmatrix}2a & 2b\\2b & 2c\end{vmatrix}=4\left(ac-b^2\right)>0$

此時 $f\left(x'+h,y'+k\right)=ax'^2+2bx'y'+cy'^2+f'$ 其中 $f'=f\left(h,k\right)$ ，利用配方法得到

$$f\left(x'+h,y'+k\right)=a\left[x'^2+\frac{2b}{a}x'y'+\left(\frac{b}{a}y'\right)^2\right]-\frac{b^2}{a}y'^2+cy'^2+f'$$

$$=a\left(x'+\frac{b}{a}y'\right)^2+\left(\frac{ac-b^2}{a}\right)y'^2+f'\geq f'$$

當〝=〞成立時 $\Leftrightarrow\begin{cases}x'+\dfrac{b}{a}y'=0\\y'=0\end{cases}\Leftrightarrow\begin{cases}x'=0\\y'=0\end{cases}$

所以當 $(x,y)=(h,k)$ 時， $f\left(x,y\right)$ 有最小值爲 $f\left(h,k\right)$ ，故得證

下面我們開始推導迴歸平面該如何求得：

首先若從母體中隨機抽樣 n 筆三維資料， $\left(x_1,y_1,z_1\right)$ 、 $\left(x_2,y_2,z_2\right)$ 、 $\cdots\left(x_n,y_n,z_n\right)$ ，將此 n 筆資料標準化後得到 $\left(x_1',y_1',z_1'\right)$ 、 $\left(x_2',y_2',z_2'\right)$ 、 $\cdots\left(x_n',y_n',z_n'\right)$ 其中 $x_k'=\dfrac{x_k-\bar{x}}{s_x}$ 、 $y_k'=\dfrac{y_k-\bar{y}}{s_y}$ 、 $z_k'=\dfrac{z_k-\bar{z}}{s_z}$ ，先考慮標準化後的平面方程式爲 $z'=ax'+by'+c$ ，殘差爲 $e_i=z_i'-\hat{z}_i'$ ，其中 $\hat{z}_i'=ax_i'+by_i'+c$ ，我們希望求何時 a,b,c 使得 $\sum_{i=1}^{n}e_i^2$ 最小，以下是我們的推導：

$$\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}\left(ax_i'+by_i'+c-z_i'\right)^2=\sum_{i=1}^{n}\left[\left(ax_i'+by_i'-z_i'\right)+c\right]^2$$

$$=\sum_{i=1}^{n}\left(ax_i'+by_i'-z_i'\right)^2+2c\sum_{i=1}^{n}\left(ax_i'+by_i'-z_i'\right)+\sum_{i=1}^{n}c^2$$

$$=\sum_{i=1}^{n}\left(ax_i'+by_i'-z_i'\right)^2+2ac\sum_{i=1}^{n}x_i'+2bc\sum_{i=1}^{n}y_i'-2c\sum_{i=1}^{n}z_i'+nc^2$$

$$=\sum_{i=1}^{n}\left(ax_i'+by_i'-z_i'\right)^2+nc^2\left(\text{由於}\overline{x'}=0\Leftrightarrow\sum_{i=1}^{n}x_i'=0,\overline{y'}=0\Leftrightarrow\sum_{i=1}^{n}y_i'=0,\overline{z'}=0\Leftrightarrow\sum_{i=1}^{n}z_i'=0\right)$$

因爲我們希望上式和最小，令$c=0$代入上式得到

$$\sum_{i=1}^{n}\left(ax_i'+by_i'-z_i'\right)^2=\left(\sum_{i=1}^{n}x_i'^2\right)a^2+\left(2\sum_{i=1}^{n}x_i'y_i'\right)ab+\left(\sum_{i=1}^{n}y_i'^2\right)b^2-2\left(\sum_{i=1}^{n}x_i'z_i'\right)a-2\left(\sum_{i=1}^{n}y_i'z_i'\right)b+\sum_{i=1}^{n}z_i'^2$$

若考慮函數

$$f(a,b)=\left(\sum_{i=1}^{n}x_i'^2\right)a^2+\left(2\sum_{i=1}^{n}x_i'y_i'\right)ab+\left(\sum_{i=1}^{n}y_i'^2\right)b^2-2\left(\sum_{i=1}^{n}x_i'z_i'\right)a-2\left(\sum_{i=1}^{n}y_i'z_i'\right)b+\sum_{i=1}^{n}z_i'^2$$

$$@\ Aa^2+2Bab+Cb^2+Da+Eb+F$$

由柯西不等式得：$\left(\sum_{i=1}^{n}x_i'^2\right)\left(\sum_{i=1}^{n}y_i'^2\right)\geq\left(\sum_{i=1}^{n}x_iy_i\right)^2\Rightarrow B^2-AC\leq 0$

一般而言我們考慮$B^2-AC\neq 0$

接著我們利用引理 1，$a=\dfrac{\begin{vmatrix}-D & 2B\\ -E & 2C\end{vmatrix}}{\begin{vmatrix}2A & 2B\\ 2B & 2C\end{vmatrix}}=\dfrac{-DC+DE}{2\left(AC-B^2\right)}$ ，$b=\dfrac{\begin{vmatrix}2A & -D\\ 2B & -E\end{vmatrix}}{\begin{vmatrix}2A & 2B\\ 2B & 2C\end{vmatrix}}=\dfrac{-AE+BD}{2\left(AC-B^2\right)}$

其中$A=\sum_{i=1}^{n}x_i'^2,B=\sum_{i=1}^{n}x_i'y_i',C=\sum_{i=1}^{n}y_i'^2,D=\sum_{i=1}^{n}x_i'z_i',E=\sum_{i=1}^{n}y_i'z_i',F=\sum_{i=1}^{n}z_i'^2$

注意上式的a,b可化爲下列的式子：

$$a=\frac{-DC+DE}{2\left(AC-B^2\right)}=\frac{-(\sum_{i=1}^{n}x_i'z_i')(\sum_{i=1}^{n}y_i'^2)+(\sum_{i=1}^{n}x_i'z_i')(\sum_{i=1}^{n}y_i'z_i')}{2[(\sum_{i=1}^{n}x_i'^2)(\sum_{i=1}^{n}y_i'^2)-(\sum_{i=1}^{n}x_i'y_i')^2]}=\frac{-r_{xz}r_{yy}+r_{xz}r_{yz}}{2(r_{xx}r_{yy}-r_{xy}^2)}$$

$$b=\frac{-AE+BD}{2\left(AC-B^2\right)}=\frac{-(\sum_{i=1}^{n}x_i'^2)(\sum_{i=1}^{n}y_i'z_i')+(\sum_{i=1}^{n}x_i'y_i')(\sum_{i=1}^{n}x_i'z_i')}{2[(\sum_{i=1}^{n}x_i'^2)(\sum_{i=1}^{n}y_i'^2)-(\sum_{i=1}^{n}x_i'y_i')^2]}=\frac{-r_{xx}r_{yz}+r_{xy}r_{xz}}{2(r_{xx}r_{yy}-r_{xy}{}^2)}$$

這樣，我們利用配方法，求得迴歸平面。

三維的迴歸平面爲$\dfrac{z-\overline{z}}{s_z}=a(\dfrac{x-\overline{x}}{s_x})+b(\dfrac{y-\overline{y}}{s_y})$ ，其中a,b如上所述。

如果繼續推廣，我們可以考慮 n 筆 m+1 維資料,$(x_{11},\mathbf{K},x_{m1},z_1),(x_{12},\mathbf{K},x_{m2},z_2)\mathbf{K}(x_{1n},\mathbf{K},x_{mn},z_n)$

我們一樣可以定義 m+1 維空間中的迴歸超平面，型如$z=a_1x_1+a_2x_2+\mathbf{K}+a_mx_m+a_{m+1}$

然後考慮殘差平方和最小求得係數，不過由於變數過多，所以若用前面的配方法，可能不太恰當，所以我們必須尋求更有效率的求法，來求得迴歸超平面的係數。事實上，我們知道微積分中，求極值問題我們都可以利用微分等於零，即所謂的臨界點，再來

判斷臨界點是否真爲極值。在多變數函數中，要判別一個臨界點是否爲極值，是個非常難的問題，但是如果我們知道所求函數必有極值的話，那問題就變得比較容易。

(二)、m+1 維資料迴歸分析討論：

考慮 n 筆 m+1 維資料,$(x_{11},\cdots,x_{m1},z_1),(x_{12},\cdots,x_{m2},z_2)\cdots(x_{1n},\cdots,x_{mn},z_n)$，定義 m+1 維空間中的超平面，型如 $z=a_1x_1+a_2x_2+\cdots+a_mx_m+a_{m+1}$，我們想要求殘差平方和 $\sum_{i=1}^{n}e_i^2$ 最小，其中 $e_i=z_i-\tilde{z}_i$ ， $\tilde{z}_i=a_1x_{1i}+a_2x_{2i}+\cdots+a_mx_{mi}+a_{m+1}$ ，首先我們先證明此平方和必有最小值，在此先定義符號。

定義：1.假如 $x=(x_1,\ldots,x_m,x_{m+1})\in\mathbb{R}^{m+1}$ ，則 x 的長度定義爲 $\|x\|=\sqrt{x_1^{\,2}+\ldots+x_{m+1}^{\,2}}$

2.函數 $f(x)$ 對變數 x_i 偏微分定義爲

$$\frac{\partial f(x)}{\partial x_i}=\lim_{t\to 0}\frac{f(x+te_i)-f(x)}{t},\quad e_i=(0,..1..,0)\text{其中第}i\text{分量爲1其它爲0}$$

3.函數 $f(x)$ 的梯度定義爲 $\nabla f(a)=\left(\frac{\partial f(a)}{\partial a_1},\frac{\partial f(a)}{\partial a_2},\cdots,\frac{\partial f(a)}{\partial a_m},\frac{\partial f(a)}{\partial a_{m+1}}\right)$

引理 2：

若實函數 $f(x)$ 爲定義在 $\mathbb{R}^{m+1}$ 上的連續函數，且 $\lim_{\|x\|\to\infty}f(x)=\infty$ 則函數 $f(x)$ 必有最小值。

pf ：因爲 $\lim_{\|x\|\to\infty}f(x)=\infty$ 所以存在 $K>0$ 使得 $f(x)>f(0)+1$ 對所有 $\|x\|\geq K$ ，

但我們知道連續函數 $f(x)$ 在 $\|x\|\leq K$ 必有極值，所以 $f(x)$ 有最大最小值在 $\|x\|\leq K$ ，設

$M_1=\underset{\|x\|\leq K}{Max}\,f(x)$ ， $m_1=\underset{\|x\|\leq K}{Min}\,f(x)$ $\Rightarrow m_1\leq f(x)\leq M_1$, 對所有的 $\|x\|\leq K$

又因爲 $m_1\leq f(0)<f(0)+1<f(x)$,對所有 $\|x\|\geq K$

所以 $m_1\leq f(x)$, 對所有的 $x\in\mathbb{R}^{m+1}$ ， m_1 爲 $f(x)$ 在 $\mathbb{R}^{m+1}$ 的最小值 ，故得證。

由於我們考慮殘差函數爲 $f(a)=\sum_{i=1}^{n}(a_1x_{1i}+a_2x_{2i}+\cdots+a_mx_{mi}+a_{m+1}-z_i)^2$ 爲一個連續函數且 $\lim_{\|a\|\to\infty}f(a)=\infty$ ，所以利用引理 2 知道殘差函數必有最小值。

但是我們知道極值點 a 必須滿足 $\nabla f(a)=0$ ，換句話說即所求的點 a 必須滿足下

面的方程組：

$$\begin{cases}\dfrac{\partial f(a)}{\partial a_1}=\sum_{i=1}^{n}2x_{1i}(a_1x_{1i}+a_2x_{2i}+\dots+a_mx_{mi}+a_{m+1}-z_i)=0\\ \dfrac{\partial f(a)}{\partial a_2}=\sum_{i=1}^{n}2x_{2i}(a_1x_{1i}+a_2x_{2i}+\dots+a_mx_{mi}+a_{m+1}-z_i)=0\\ \vdots\\ \vdots\\ \dfrac{\partial f(a)}{\partial a_{m-1}}=\sum_{i=1}^{n}2x_{m-1i}(a_1x_{1i}+a_2x_{2i}+\dots+a_mx_{mi}+a_{m+1}-z_i)=0\\ \dfrac{\partial f(a)}{\partial a_m}=\sum_{i=1}^{n}2x_{mi}(a_1x_{1i}+a_2x_{2i}+\dots+a_mx_{mi}+a_{m+1}-z_i)=0\\ \dfrac{\partial f(a)}{\partial a_{m+1}}=\sum_{i=1}^{n}2(a_1x_{1i}+a_2x_{2i}+\dots+a_mx_{mi}+a_{m+1}-z_i)=0\end{cases}$$

上式爲一個 m+1 元一次聯立方程組，因此可以利用克拉瑪公式或是高斯消去法求得方程組的解。注意到上式方程組的最後一式爲

$$\begin{aligned}&\sum_{i=1}^{n}(a_1x_{1i}+a_2x_{2i}+\dots+a_mx_{mi}+a_{m+1}-z_i)=0\\ &\Rightarrow a_1\sum_{i=1}^{n}x_{1i}+a_2\sum_{i=1}^{n}x_{2i}+\dots+a_m\sum_{i=1}^{n}x_{mi}+\sum_{i=1}^{n}a_{m+1}=\sum_{i=1}^{n}z_i\\ &\Rightarrow a_1\sum_{i=1}^{n}x_{1i}+a_2\sum_{i=1}^{n}x_{2i}+\dots+a_m\sum_{i=1}^{n}x_{mi}+na_{m+1}=\sum_{i=1}^{n}z_i\\ &\Rightarrow a_1\overline{x_1}+a_2\overline{x_2}+\dots+a_m\overline{x_m}+a_{m+1}=\overline{z}\\ &\Rightarrow a_{m+1}=\overline{z}-(a_1\overline{x_1}+a_2\overline{x_2}+\dots+a_m\overline{x_m})\end{aligned}$$

所以我們解得迴歸超平面爲 $z-\overline{z}=a_1(x_1-\overline{x_1})+a_2(x_2-\overline{x_2})+\ldots+a_m(x_m-\overline{x_m})$ 其中 $a_{1,}a_2,\ldots a_m$ 爲上式方程組的解。

三、研究結果

研究一開始從課本所學到的二維資料分析的迴歸直線出發，來分析我們關心的資料是否有關係存在，雖然可以得知資料是否有關係，不過，由於納入考慮的因素太少而不敷使用，漸漸地二維的回歸直線失去了它的實用，我們想要得知是否還有其他更精確的預測，所以我們利用引理 1 經由配方法求得三維空間中的迴歸平面，不過，這些方法終究只能應付較少變因的問題，所以我們必須找尋更有效率的求法，來求得高維空間的迴歸超平面。事實上我們知道微積分中，求極值的問題我們都可以轉化成函數微分等於零，接著再來判斷所求點是否真爲極值，要在多變數函數中，判別一個臨界點是否真爲極值，是個非常難的問題，所以我們利用引理 2 得知，若函數 $f(a)=\sum_{i=1}^{n}(a_1x_{1i}+a_2x_{2i}+\dots+a_mx_{mi}+a_{m+1}-z_i)^2$ 且 $\lim_{\|a\|\to\infty}f(a)=\infty$，則 $f(a)$ 在微分等於零時的點必爲極值點。，同時我們知道發生的極值點 a 必須滿足之前所提到的 m+1 元一次聯立方程組，利用克拉瑪公式或是高斯消去法可解出 a，因此我們解得迴歸超平面爲 $z-\overline{z}=a_1(x_1-\overline{x_1})+a_2(x_2-\overline{x_2})+\ldots+a_m(x_m-\overline{x_m})$，其中 $a_{1,}a_2,\ldots a_m$ 爲上式方程組的解。

四、討論

在此我們想分析本校去年高三畢業生的國中基測數學與自然成績，和大學學測數學成績的相關性，首先利用二維資料分析，先求基測數學與自然的平均值，再來計算與大學學測數學的相關性，圖 2 為資料點畫在座標平面上，x 軸方向為基測數學自然成績平均(滿分 60)，y 軸方向為大學學測數學成績(滿級分 15)，在此定義一個符號：相關係數平方為 R^2

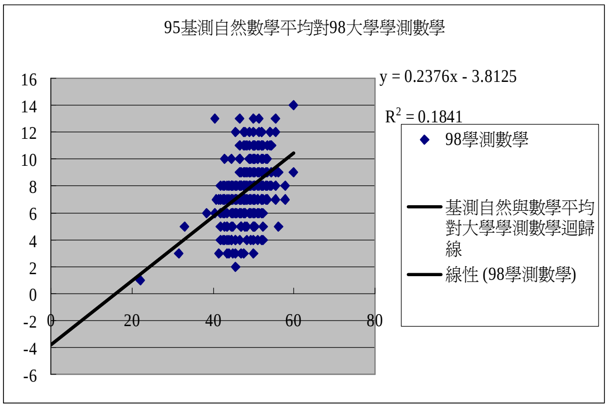


圖 2. 國中基測數學與自然成績平均對大學學測數學成績二維迴歸

我們可以看得出來，資料點比前面只考慮基測數學對學測數學分布的情況更為集中，R^2 值　較前面大，可以看出數學科和自然科一起納入考慮，似乎可以提高相關性，不過 R^2 值還是很低。另一方面將基測數學自然成績平均，比較看不出基測數學還是基測自然，和學測數學的相關性哪個比較大，若影響大小可以量化的話，將幫助我們解讀三者之間的關係，所以我們將前面推導的迴歸平面拿來套用在此資料上，圖 3 為基測數學和自然分別當做 x，y 變數，然後考慮對學測數學的迴歸平面圖，x 軸方向為基測數學成績(滿分 60)，y 軸方向為基測自然成績(滿分 60)，z 軸方向為大學學測數學成績(滿級分 15)，

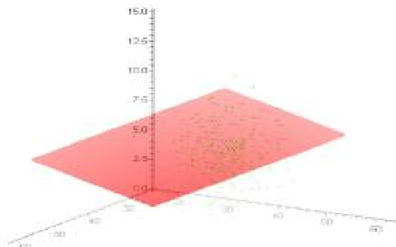


圖 3. 國中基測數學和自然成績對大學學測數學成績三維超平面迴歸

迴歸平面方程式為 $z = 0.1108x + 0.1311y - 4.0157$ ，在此我們得到平面方程式中 x 係數 0.1108，y 係數 0.1311，x 係數小於 y 係數，似乎可以解釋，此畢業生基測自然的影響程度大於基測數學的影響程度，這結果令我們意外，因為直觀上我們會認為基測數學與學測數學的相關程度是比較大的，但是此資料發現事實不然，所以之前我們把基測數學和自然平均求迴歸直線，會看不出兩者影響學測數學的程度大小。所以我們發覺

推廣到迴歸超平面時，變數的係數，可以看成各變數對所對應資料的相關程度，事實上，我們有這樣的猜想，對相同一筆資料做最小平方法得到的迴歸直線和迴歸平面，甚至迴歸超平面，所得到的殘差平方和最小值，是否有某種關係，感覺上若考慮的變因增加，迴歸函數應該越精準，換言之，就是平方和最小值，是否會隨者維度增加而變小，這將是我們未來想要研究證明的問題。

參●結論

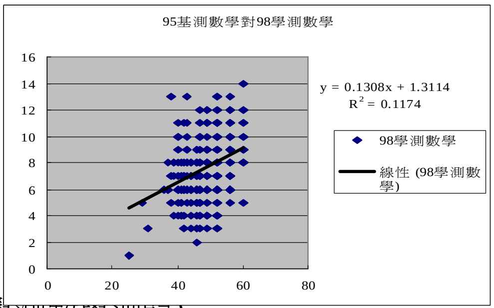
在生活中，時常會碰到資料數據的解讀，特別是我們關心的數據，若存在某種相關性，我們會想從中了解其中一資料數據變化是否會影響到另一資料數據變化，像是在校成績、製作過程解析或是實驗數據……，常會遇到要分析兩種以上的數據之間的關係，我們知道數據產生時必有各種原因影響而變化，所以如果只靠直覺判斷，常會有判斷錯誤的情形發生。若要很客觀地分析之間的關係就是相關係數，若想進一步去調查數據間的預測函數，就可以應用回歸分析來處理，本作品主要利用配方的手法完成三維的迴歸平面，公式如下：

$$\frac{z-\overline{z}}{s_z}=a\left(\frac{x-\overline{x}}{s_x}\right)+b\left(\frac{y-\overline{y}}{s_y}\right)\quad \text{其中 } a=\frac{-r_{xz}r_{yy}+r_{xz}r_{yz}}{2(r_{xx}r_{yy}-r_{xy}^2)}\text{ ， } b=\frac{-r_{xx}r_{yz}+r_{xy}r_{xz}}{2(r_{xx}r_{yy}-r_{xy}^{\ 2})}$$

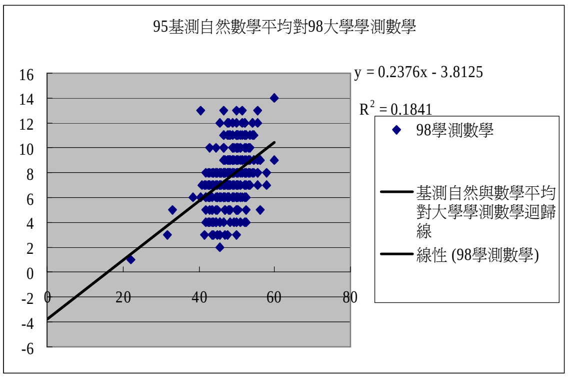
最後我們將本校去年高三畢業生，做了一個實際數據的分析：

一、二維資料分析(國中基測數學對大學學測數學迴歸)

如右圖資料相關係數為 0.3426，為中度正相關。

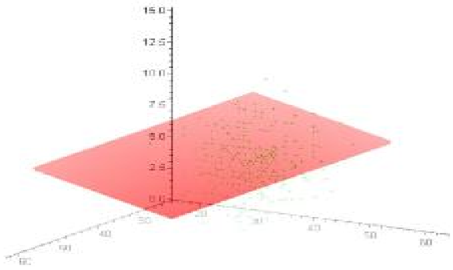


二、二維資料分析(國中基測數學與自然平均對大學學測數學迴歸)



如上圖資料相關係數為 0.429，為中度正相關，不過此相關係數大於上面的相關係數，可見如將國中自然基測成績考慮進來，顯然相關係數變大。

三、三維資料分析(國中基測數學與自然對大學學測數學迴歸)



如上圖，此資料 x 係數 0.1108，y 係數 0.1311，可以看出基測自然比對基測數學影響學測數學的影響程度來得大。發現到不管是二維或是三維做出的結果，基測自然和數學成績，與學測數學的散佈圖，資料點都相當分散，相關係數並不高，可見高中三年的努力與否，仍會使三年後的學測成績產生很大的變化。

肆●引注資料

1、普通高級中學教科用書，高中選修數學(二)，龍騰文化
2、墨爾(民 87)。統計，讓數字說話！臺北市：天下文化
3、洪維恩(民 90)。數學魔法師 Maple 6。臺北市：碁峰資訊