

逢甲大學學生報告 ePaper

探討糖尿病的高危險因子  
high-risk factor of diabetic mellitus

組別：第一組 生物醫學類

作者：葉玫欣、林怡君、許芷暄、陳昀妮、吳紘慶、黃紹璋

學號：D9860503、D9860399、D9860339、D9860415、D9817519、D9860270

系級：統計學系三年乙班

開課老師：陳婉淑 教授

課程名稱：迴歸分析

開課系所：統計學系

開課學年：100 學年度 第 1 學期

## 目次

摘要.....	2
Abstract.....	3
前言.....	4
<b>第一章 何為糖尿病</b>	
一、糖尿病的定义.....	5
二、糖尿病的分類.....	5
<b>第二章 資料描述</b>	
一、資料來源與採用.....	7
二、篩選後變數解釋.....	7
<b>第三章 研究方法</b>	
一、診斷多重共線性.....	9
二、選取重要變數前的模型建立.....	9
三、選取重要變數的方法與過程.....	10
四、選模結果小結.....	11
<b>第四章 分析最佳模型中的殘差</b>	
一、殘差假設.....	13
二、診斷異常點.....	14
三、解決殘差不符常態及常數.....	15
<b>結論</b> .....	16
<b>參考文獻</b> .....	17
表附錄.....	18
圖附錄.....	22

## 摘要

本文為對中國醫藥大學2001至2003年期間參加糖尿病檢測的民眾，對檢測的資料做分析與了解，其中欲探討的是在多元的檢測項目下，可能會導致糖尿病的高危險因子。醫學證實糖尿病的發生與血糖變化有密切相關，因此控制血糖是預防糖尿病的不二法則。而空腹血漿血糖（又稱飯前血糖值）是判斷罹患糖尿病與否的重要準則，本文將會探討舒張壓、糖化血色素、低密度脂蛋白.....等八項測量值與飯前血糖值之間的關係，並採取五種迴歸重要變數的篩選方式，「向前選取法」、「向後消去法」、「逐步選取法」、「Cp選取法」、「校正後的複判斷係數法」，以此挑選出最終的模型。並且嚴格地審查最終模型，須檢查有無異常點以及影響點、是否符合殘差的四項基本假設，「殘差平均數是否為零」、「殘差是否來自常態」、「殘差是否獨立」、「殘差變異數是否為常數」，並對異常點、影響點以及違反殘差假設的部分另做修正，以確保最終模型能對可能導致糖尿病的高危險因子這宗旨提供有效並且可信的解釋力，並以此作為糖尿病預防以及防範的重要典範與參考。

關鍵字：糖尿病、空腹血漿血糖、迴歸變數篩選

## Abstract

This report is to investigate 500 patients which were collected by China Medical University from 2001 to 2003. These patients participated in the detection of diabetes. We extract partial data set which may relate in a high risk factor for diabetes mellitus. The response variable is fasting blood glucose. The explanatory variables include family medical history, smoking, drinking, diastolic blood pressure, glycated hemoglobin, low-density lipoprotein, total cholesterol, and triglyceride. Forward Selection, Backward Elimination, Stepwise Selection, Cp selection method, and all subset selection are used to choose the best subset model. We conclude that diastolic blood pressure, glycated hemoglobin and triglyceride are the most important explanatory variables to fasting blood glucose. Diagnostic checking and residual analysis are performed to check whether the final model is adequate.

**Keywords:** diabetes, fasting plasma glucose regression variable screening

## 前言

根據衛生署公告九十一年國人十大死因排行中，糖尿病已超越事故性傷害，於排名中位居第四名。到了九十二年，糖尿病依然為國人十大死因中的第四，且死亡率大幅增高一成以上。糖尿病與國人死亡率間有著密切的關係。

近年來國內罹患糖尿病的人口不斷增加，一方面是由於人口結構老化，根據研究顯示，男性 40~64 歲因罹患糖尿病的死亡率是一般人的 2.5 倍，女性在 40~64 歲間則是一般人的 3.1 倍。另一方面是國內的飲食習慣日漸西化，速食業發達導致許多兒童在很小的年紀裡就養成喜歡吃速食餐的習慣，因此兒童肥胖的程度比例也日漸攀升，罹患糖尿病不再是老年人的專利，且年齡層有降低的趨勢。

國人也逐漸意識到糖尿病的危險性，在糖尿病的預防上不再只是著重於年齡較大的人，對於糖尿病的發生、如何預防糖尿病、降低疾病發生的機率，成為國人關心的議題。因此我們想要探討糖尿病的高危險因子有哪些，平常可以透過哪些數值的檢測了解自己是否為糖尿病的高危險群，即早發現即早接受治療與控制。

引發糖尿病的發生是許多原因綜合之下的結果，可能跟飲食習慣有關，也可能與家族基因有關。但最重要的是要知道如何預防以及控制病情繼續惡化，因為糖尿病是不會痊癒的，且病情控制不當會引發許多糖尿病的併發症，例如腦中風、心臟病、高血壓等。因此希望藉由我們的研究報告能找出哪些數值是可以當為參考依據，讓大家平常可以注意自己身體的各項測量指數，一發現有某些數值偏高或異常，就應該立即就醫診斷做更進一步的確認，發現的早對病情的控制才更於輕鬆容易。我們更希望透過這個報告讓大家對糖尿病有初步的認識，而不是對疾病充滿恐懼，也曉得糖尿病已經不是老年人才會罹患的疾病。倘若不幸罹患糖尿病，也不需以懼怕的心情去面對，而是能勇敢的戰勝糖尿病。

## 第一章 何為糖尿病

一、 糖尿病的定義：美國國家糖尿病資料小組(NDDG)(1979) 對糖尿病的診斷標準和分類做出了最新的三大修正，只要符合其中一項即為罹患糖尿病，以下是三種判斷的依據。

(1) 空腹血漿血糖(又稱飯前血糖值) $\geq 126\text{mg/dl}$ ：

由於飯前血糖值檢查容易執行，且患者接受度高。空腹的定義是至少 8 小時內未攝取任何卡路里。

(2) 口服葡萄糖耐受試驗第 2 小時血漿血糖 $\geq 200 \text{ mg/dl}$ ：

進行口服葡萄糖耐受試驗前必須禁食至少 8 小時。進行口服葡萄糖耐受試驗前先立即測量空腹血糖值，然後再喝下含有 75 克葡萄糖的糖水，喝完 2 小時後再測量一次血糖，此方法是為了驗證空腹血漿血糖判別法的準確性。

(3) 典型的高血糖症狀或高血糖危象且隨機血漿血糖 $\geq 200\text{mg/dl}$ ：

若隨機血糖測試結果在 200 mg/dL 或以上，並且出現典型糖尿病的症狀：尿多、喝多、不明原因的體重減輕，其他症狀還包括疲勞、視力模糊、容易飢餓、傷口不易癒合，且隨機測量血漿葡萄糖濃度二次都 $\geq 200\text{mg/dl}$ 。

二、 糖尿病的分類：美國糖尿病學會專家委員會提出的糖尿病最新分類主要是以「致病原因」為分類依據，分為以下三大類，

(1) 第一型糖尿病 (**Type I diabetes**):

第一型糖尿病是一種自體免疫，免疫疾病是由於身體的免疫系統對自身作出攻擊而成的，糖尿病患者的免疫系統對自身分泌胰島素的胰臟β細胞作出攻擊並殺死他們，結果胰臟並不能分泌足夠的胰島素，有一些第一型糖尿病患者，需要每天服用胰島素來維持生命。在現今的科學界，科學家還未能找到身體攻擊胰臟β細胞的真正原因，但他們相信這個

## 探討糖尿病的高危險因子

疾病跟遺傳因素和環境因素例如病毒，是有關連的。

### (2) 第二型糖尿病 (**Type II diabetes**)

第二型的患者通常能夠分泌足夠的胰島素，但因為未知的原因，身體不能有效地使用胰島素，這個情況叫作胰島素阻抗性，而且過了一段時間後胰島素的分泌亦會減少。而且是最普遍的糖尿病，差不多 90~95% 的糖尿病人是二型的，通常在 40 歲時病發，55 歲為病發的高峰期，八成第二型患者是過胖的，第二型患者通常會有肥胖、高血壓、高血脂等疾病。

### (3) 妊娠期糖尿病 (**Gtational diabetes**)

妊娠期糖尿病是過去沒有診斷出糖尿病的婦女，在懷孕的過程中出現高血糖症的現象，根據統計的研究結果，有 3~10% 的婦女會罹患妊娠期糖尿病。目前並沒有發現特定的原因，但是通常於生產後就會回復正常，但生產後罹患糖尿病的機率會比一般同年齡婦女高。此種類型屬於特例，因此本研究並不將此類型糖尿病列入探討範圍內。

## 第二章 資料描述

### 一、 資料來源與採用：

本研究的資料是由中國醫藥大學生統所李采娟教授提供。資料為 2001~2003 年期間，參加糖尿病檢測的民眾，樣本數為 500 人。原始資料中是對參與檢測的民眾做一份測量，其中共包含多筆不同的測量項目，經過整理、分析之後，我們找出一些非引發糖尿病的因素，例如：病人編號、就診原因、糖尿病的併發症……等，將其排除在探討糖尿病的高危險因子之外。經過嚴格謹慎地篩選過後，我們發現有八個測量項目極有可能為引發糖尿病的高危險因子，因此採用這八個項目作為討論的變數。《表一》是對其中五項數值變數做基本統計量，以便觀察其範圍與平均值。其中發現在總膽固醇項目中最小值為 0，但醫學上人類總膽固醇不可能為 0，因此推斷可能是紀錄數值上的人為疏失，本研究將此觀察值刪去不列入考慮。其餘三項變數為虛擬變數，在此不做基本統計量。

最後根據糖尿病的三種判斷準則，本研究採用第一種飯前血糖值判斷法作為判斷準則，將飯前血糖值作為迴歸模型中的應變數來探討這八項自變數與其之間的線性關係。

### 二、 篩選後變數解釋：

- (1) x1 家族病史：家族病史是一個簡單但強大的基因組工具，可用於確定有疾病危險的個體和人口，例如心臟病、癌症和糖尿病。
- (2) x2 抽菸：根據世界衛生組織定義，抽菸的定義是一生中抽的煙數超過 100 根，且過去 30 天內曾經使用菸品。本研究以虛擬變數來表示抽菸的有無，1 代表有抽菸，0 代表沒有抽菸。
- (3) x3 飲酒：根據家庭醫學專業醫師表示，一天喝 30 克的酒精量，大約是 600c.c 的台灣啤酒，為正常飲酒，但超過 30 克即為過量。本研究以虛擬變數來表示飲酒的有無，1 代表有抽菸，0 代表沒有抽菸。
- (4) x4 舒張壓：心臟舒張時，測量血管壁所承受之壓力稱舒張壓。如果所測



## 探討糖尿病的高危險因子

得的舒張壓高於 95 則為高血壓。

- (5) x5 糖化血色素：血紅素是紅血球裡面的蛋白質，用來將氧運送給周邊組織利用，同時將周邊組織的二氧化碳運回肺部以排出體外。而紅血球在血液中運行時，血中的葡萄糖可進入紅血球中而依附在血紅素上，形成糖化血紅素。血紅素一旦被葡萄糖依附後，就會持續保持這種狀態，直到此紅血球死亡為止。血糖越高，糖化血紅素的比值就越高，而紅血球的壽命是 120 天，所以，糖化血紅素可用來反應最近三個月來的血糖控制狀況。正常人的糖化血紅素值約在 5% 左右，在醫學上糖尿病患的糖化血紅素希望能控制在 7% 以下，減少病人發生併發症的機率。
- (6) x6 低密度脂蛋白：人體重要的脂蛋白，可運送膽固醇到全身供細胞使用，當濃度過高時，容易在血管壁造成堆積，形成血管粥狀硬化，阻塞血管，而當低密度脂蛋白超過 130 mg/dl 時是發生心血管疾病的高危險群。
- (7) x7 總膽固醇：膽固醇是體內的一種脂質，主要在血管的強化和維持，由脂肪酸結合的酯型及各別分離的游離型兩種，合稱為總膽固醇。總膽固醇的檢查，是動脈硬化及心臟病等不可或缺的判斷準則。正常值為 130~200mg/dl，超過這範圍時，可能會引發是甲狀腺機能低弱症、末端肥大症、腎病變、肥胖等。
- (8) x8 三酸甘油酯：是血液中的一種脂質，體內會將尚未使用的熱量轉為三酸甘油酯，儲存在脂肪細胞內，而在未進食期間，荷爾蒙會釋出三酸甘油酯當能源。當三酸甘油酯過高時，則身體產生心臟疾病、中風、肥胖、糖尿病等的機率也會隨著增大。

## 第三章 研究方法

### 一、 診斷多重共線性：

多重共線性表示獨立變數之間存在著高度相關，也就是獨立變數若存在多重共線性，代表獨立變數間對應變數有相同的貢獻關係，因此必須踢除或重令一新獨立變數來表示對應變數之間的關係。例如身高與體重存在多重共線性，可以身體質量指數來表示兩者之間的關係。診斷多重共線性必須由 VIF 來判斷，若  $VIF > 10$ ，則該變數存在多重共線性，所以由《表二》可以發現，各變數的 VIF 皆小於 10，表示變數之間沒有多重共線性的問題，所以不需要重令新變數或踢除變數。

### 二、 選取重要變數前的模型建立：

如上述，飯前血糖值為本研究的應變數，並與整理資料後最終取得的八個可能會引發糖尿病的高危險因子自變數做模型配適。以下為其完整的全模型迴歸式：

$$\hat{y} = 4.92651 + 0.19072x_1 + 0.24024x_2 - 0.0461x_3 + 0.02311x_4 + 0.63373x_5 \\ - 0.00237x_6 + 0.00363x_7 + 0.00123x_8$$

由全模型迴歸式中可以看出各自變數與應變數之間的關係， $x_1$  (家族病史)，此變數與應變數間呈現正相關，也就是說家族裡若曾有人罹患過糖尿病，則民眾罹患糖尿病的機率會增加，因此飯前血糖值也會隨著上升； $x_2$  (抽菸)，若是每增加一單位的抽菸量，則其飯前血糖值就會增加 0.24024 單位； $x_3$  (飲酒)，每增加一單位的飲酒量，飯前血糖值卻減少了 0.0461 單位，透露適量的飲酒反而有益健康。根據全模型迴歸式， $x_4$  (舒張壓)、 $x_5$  (糖化血色素)、 $x_7$  (總膽固醇)與  $x_8$  (三酸甘油脂)資料顯示，參數的係數值皆為正號，表示四個自變數與飯前血糖值間都成正相關，但  $x_6$  (低密度脂蛋白)而言卻是相反，其與飯前血糖值間是負相關，也就是說每增加一單位的低密度脂蛋白，飯前血糖值就會減少 0.00237 單位。

## 探討糖尿病的高危險因子

### 三、 選取重要變數的方法與過程：

本研究用了 5 個選取重要變數的判別方法，來決定 8 個解釋變數中有哪些是真正導致糖尿病的高危險因子。《表三》是整理出各項選取法所選出的重要變數，可以發現這五種選取法所選出的變數皆為 x4 舒張壓、x5 糖化血色素、x8 三酸甘油酯，以下詳細的介紹各變數的選模方法。

#### (1)向前選取法(Forward Selection)：

首先將所有解釋變數置於模型外，接著遵循判斷準則，將模型外 p-value<0.1 的解釋變數選進迴歸模型中。此時模型中的變數即為該選取法所選出的重要變數。此方法得到最佳預測飯前血糖值的解釋變數是 x4 (舒張壓)、x5 (糖化血色素)、x8 (三酸甘油酯)，最佳的模型組合式為

$$\hat{y} = -37.88821 + 0.66132x_4 + 17.56625x_5 + 0.04696x_8$$

#### (2)向後刪去法(Backward Elimination)：

先將所有的變數均納入迴歸模型中，之後再逐一將對模式貢獻最小 p-value>0.15 的解釋變數踢除，直到所有變數均達到標準為止。踢除貢獻最小的解釋變數之後，模型內剩餘的解釋變數即為重要變數為 x4 (舒張壓)、x5 (糖化血色素)、x8 (三酸甘油酯)，這三個重要變數所組成的最佳模型配適如下：

$$\hat{y} = -37.88821 + 0.66132x_4 + 17.56625x_5 + 0.04696x_8$$

#### (3)逐步分析選取法 (Stepwise Selection)：

逐步分析的作法首先是迴歸模型內不放入任何解釋變數，然後採用向前選取法，作為第一步驟的判斷，判斷的準則是如果 p-value < 0.1 則可進入模型內，接著第二步是將已選進模型內的解釋變數使用向後刪去法將模型內不符合的解釋變數予以剔除，判斷準則是 p-value > 0.15 的變數會被排除於模型之外。經由以上步驟得到最佳預測飯前血糖值的解釋變數是 x4 (舒張壓)、x5 (糖化血色素)、x8 (三酸甘油酯)，最佳的模型組合為

$$\hat{y} = -37.88821 + 0.66132x_4 + 17.56625x_5 + 0.04696x_8$$

## 探討糖尿病的高危險因子

### (4)Cp 準則：

Cp 準則為選取 Cp 值最接近參數個數 p 的模型組合。由《表四》及《圖一》得知，選模結果 x1(家族病史)、x4 (舒張壓)、x5 (糖化血色素)、x8 (三酸甘油酯)為重要變數，其 Cp 值與參數個數最為接近，因此選擇此模型配適。

### (5)校正後的複判斷係數法(Adjust Rsquare Selection)：

校正後的複判斷係數法是針對判定係數做調整，判斷準則是選取調整後判定係數最高的模型。因為當解釋變數愈多時判定係數的解釋能力必定愈強，因此必須加以調整使其在相同的基準之下做判斷。我們利用調整判定係數作為比較準則，以達到選模準確性。由《表四》的結果可以得知，在調整後判定係數欄選出最高的值即為此選取法所選出的重要變數為 x4 (舒張壓)、x5 (糖化血色素)、x8 (三酸甘油酯)。

## 四、選模結果小結：

綜合上述五種選模方法，可以觀察到 x4 (舒張壓)、x5 (糖化血色素)、x8 (三酸甘油酯)由這三個解釋變數所組成的模型為最終的選模結果，因此本研究最後採用這三個變數作為探討引發糖尿病的高危險因子之重要變數。最佳迴歸模型為：

$$\hat{y} = -37.88821 + 0.66132x_4 + 17.56625x_5 + 0.04696x_8$$

觀察此迴歸模型可以發現 x4(舒張壓)的係數約為 0.66，即每增加一單位的舒張壓，飯前血糖值亦會隨著增加，其散佈圖如《圖二》所示，大多數的人其舒張壓皆介於正常範圍 50~90mmHg 之間，圖中的觀察值與 y 呈現微弱正相關。x5(糖化血色素)對 y 的貢獻為每增加一單位，會增加約 17.57 單位的飯前血糖值，相較於變數 x4(舒張壓)、x8(三酸甘油酯)而言，增加的幅度較大，可推論與飯前血糖值的關係較為密切，且藉由觀察其散佈圖《圖三》，圖中很明顯的可以看出兩者存在顯著的正相關。而 x8(三酸甘油酯)對 y 亦是呈現增加的趨勢，每增加一單位的 x8(三酸甘油酯)，y 隨著增加約 0.05，此貢獻

## 探討糖尿病的高危險因子

不多，且由其散佈圖《圖四》中，可以看到兩者之間並沒有很明顯的正相關，但圖中似乎又有微弱的相關性，因此可以考慮將  $x_8$ (三酸甘油酯)做資料上的轉換，使其對  $y$  的貢獻可以呈現出較高且明顯的正相關。



## 第四章 分析最佳模型中的殘差

### 一、殘差假設：

產生完整的線性迴歸模型須通過諸多條件限制及嚴厲的考驗，因此即便是經由選模機制所選出的最佳配式模型也不一定是合格的，還須對被挑選的配適模型做殘差的四項基本假設，殘差平均數是否為零、殘差是否來自常態、殘差間是否獨立、殘差變異數是否為常數，以驗證此模型是否合格。

#### (1)檢查殘差平均是否為零

首先須設立虛無假設( $H_0: \mu = 0$ )與對立假設( $H_1: \mu \neq 0$ )，若是不拒絕虛無假設，則殘差平均數為零。並用三種檢定方式 Student's test、Sign test signed 以及 Rank test，判斷平均數是否為零。由《表五》可以得知，這三種檢定方式的 p-value 值皆大於顯著水準 0.05，所以不拒絕虛無假設，也就是符合殘差平均為零的假設。

#### (2)檢查殘差是否來自常態

四項檢定中，以這點最為重要，如果殘差服從常態，則使用最小平方法(OLS)時具有不偏性，估計也較準確，但若是殘差不服從常態，雖然依舊可以使用 OLS 做估計，但所估計出來的未必是最有效的。首先須設立虛無假設( $H_0: \epsilon_i \sim \text{Normal}$ )與對立假設( $H_1$ :反對  $H_0$ )，若是不拒絕虛無假設，則殘差來自常態。利用以下四種檢定方式，Shapiro-Wilk test、Kolmogorov-Smirnov、Cramer-von Mises 以及 Anderson-Darling，來判別殘差是否來自常態。由《表六》可以得知，p-value 值皆小於顯著水準 0.05，代表拒絕虛無假設，且從《圖五》的 Normal probability plot 也可以發現並非一 45 度直線，即殘差並非來自常態分配，所以對此須另做探討並解決。

#### (3)檢查殘差是否獨立

首先須設立虛無假設( $H_0: \rho = 0$ )與對立假設( $H_1: \rho > 0$ )，若是不拒絕虛無假設，則殘差間獨立。經由最佳選模後，誤差項必須是要互相獨立的，如果不相互獨立表示此模式有自我相關的情形。利用 Durbin-Watson D test 檢定方法來判別，

## 探討糖尿病的高危險因子

當檢定統計量的值很接近零，代表誤差項間有自我相關，表殘差間不相互獨立；反之當檢定統計量的值很接近 2 時，則誤差項間沒有自我相關，即符合殘差間獨立的假設。由《表八》可知 Durbin-Watson D 檢定統計量的值為 1.92，相當接近 2，不能拒絕虛無假設，因此符合殘差間獨立的假設。

### (4)檢查殘差變異數是否為常數

由《圖六》、《圖七》中我們可以觀察到，自變數 x4 與 x5 的殘差圖中，其散佈的點主要都均勻的落在以零為中心的區域內，並沒有隨著自變數的值增高亦跟著上升，所以符合殘差變異數為常數的假設。但《圖八》自變數 x8 的殘差圖其散佈點大多集中在 0~300 之間，並非均勻的落在以零為中心的區域內，因此不符合殘差變異數為常數的假設，所以對自變數 x8 而言還須在做討論。

## 二、診斷異常點：

(1)離群值(outlier)：所謂離群值是指數據中有一個或多個數值與其他數值相比差異較大。從散佈圖可以很明顯看出離群值的位置較偏遠於其他值，但圖中無法明確告知第幾個觀察值是離群值，因此必須透過 SAS 程式的結果，利用標準化殘差(Student Residual)或標準化去值殘差(Rstudent)，取絕對值大於 3 者即為離群值，《表九》是整理出所有資料中的離群值，可以發現兩種方法所找出的離群值並不完全相同，因此皆須列入考慮。

(2)影響點：影響點對迴歸線的估計與配適具有影響力，因此影響點的存在會影響到配適出的模型是否為最佳迴歸式，所以必須偵測並踢除影響點，才能得到最佳配適模型。診斷影響點的方法有五種，分別是 DFFITS、Hat matrix element、Cook's distance statistic、DFBETAS、COVRATIO，經由整理後得到《表十》，可以發現每種方法所找出的影響點都有數個，每個觀察值皆至少重複二、三次。

異常點的存在會影響選模後所配適出的模型是否為最佳迴歸式，因此需考慮異常點存在的影響性，本研究將異常點逐一踢除後，以求配適後即為最佳模型。

## 探討糖尿病的高危險因子

### 三、解決殘差不符常態及常數：

先前探討的殘差四項基本假設中，由於不符合常態與常數的這部分，所以須做探討與解決，使最終選模過後的配適模型合格，並且能夠有效地解釋自變數與應變數間的關係。首先，我們觀察《圖二》、《圖三》、《圖四》的散佈圖，發現《圖四》y 與 x8 變數的散佈圖雖然沒有特別明顯的相關性，但似乎有存在著微弱的正相關，且在殘差的第四項檢定項目裡，也發現 x8 變數不符合常數。因此決定將 y 與 x8 個別取對數做轉換，並且刪去可能會影響資料分析結果的異常點與影響點，重新配適模型後，由《表七》的結果可以看到，四種檢定方式其顯著水準皆大於 0.05，符合殘差為常態的假設。此外，觀察《圖九》做轉換後 y 與 x8 的散佈圖呈現明顯的正相關，且從《圖十》中也可以看到其散佈的點主要都均勻落在以零為中心的區域內，符合殘差為常數的假設。將變數做轉換後殘差的四項基本假設都符合了，驗正了由 x4(舒張壓)、x5(糖化血色素)、x8(三酸甘油脂)與 y(飯前血糖值)所配適的模型是合格的，可以有效地解釋自變數與應變數間的關係。



## 結論

文章一開始曾提到本研究的宗旨為可能會導致糖尿病的高危險因子，並利用飯前血糖值來判斷是否為罹患糖尿病的準則，從篩選重要變數、殘差四大檢定、異常點與影響點的偵測，到解決殘差不符常態及常數的假設，最後由  $x_4$ (舒張壓)、 $x_5$ (糖化血色素)、 $x_8$ (三酸甘油脂)與  $y$ (飯前血糖值)組成了最佳的配適模型，其估計迴歸式為  $\hat{y} = -37.88821 + 0.66132x_4 + 17.56625x_5 + 0.04696x_8$ 。由於應變數的改變會跟隨自變數做變換，從估計迴歸式中可以看到三個自變數的參數係數皆為正號(正相關)，表示當舒張壓、糖化血色素、三酸甘油脂其所測量的值越高時，飯前血糖值也會高，當飯前血糖值增高時罹患糖尿病的機率也會跟著上升。其中又以糖化血色素的參數係數值為最高，指每增加一單位的糖化血色素，飯前血糖值會增加 17.56625，是舒張壓與三酸甘油脂的好幾倍。查閱醫學文章發現，糖尿病的診斷目前雖仍是以血糖為主，但美國病理學會(1997)認為糖化血色素超過標準可能會罹患糖尿病，於是提出將糖化血色素做為治療評估的計劃，台北榮民總醫院也參與此計劃，由此可見糖化血色素過高對人體所帶來的危害，另外醫學文獻也記載三酸甘油脂以及舒張壓較高的人，罹患腦中風、糖尿病、心臟病、動脈硬化等疾病的機率比正常人高出二到三倍，由其對當今糖尿病被列為國人十大死因的社會來說，人民更應該重視糖尿病的議題，預防勝於治療。

## 參考文獻

- a. Michael H. Kutner , Christopher J. Nachtsheim , John Neter and William Li.2005.Applied Linear Statistical Models.5<sup>th</sup> ed.,361-368. Boston :McGraw-Hill Irwin
- b.林志娟、林志鴻、張慶暉。2009。迴歸分析入門：SAS 程式報表精析解讀與個案建模範例。初版，8-1~8-21。臺北市：臺灣東華。
- c.王濟川、郭志剛。2004。Logistic 迴歸模型：方法及應用。初版：236-253。臺北市：五南。
- d.蔡樹濤。2006。糖尿病。初版：39-41。臺北市：新潮社。
- e.馬心婷、羅麗萍、竹幼婷、吳家璇、羅欣怡。2005。名醫的糖尿病聖經。初版：111-115。臺北市：杏一醫療。
- f.高醫醫訊。2001。糖尿病專刊。網址：<http://www.kmu.edu.tw/~kmcj/>。上網日期：2001-05-01

## 表附錄

《表一》

	舒張壓	糖化血色素	低密度脂蛋白	總膽固醇	三酸甘油脂
最小值	44	4.4	50	0	13
第一百分位數	70	6.522	96	167.5	100
中位數	80	7.6	117.7	192	138
平均值	78.54	8.008	118.7	192.6	169.3
第三百分位數	86	8.975	140.6	222.5	200.8
最大值	120	15.7	199.6	368	894

《表二》

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr>  t	Variance Inflation
Intercept	1	4.92654	1.00357	4.91	<.0001	0
X1	1	0.19072	0.21738	0.88	0.3810	1.03638
X2	1	0.24024	0.29005	0.83	0.4082	1.31822
X3	1	-0.04610	0.36826	-0.13	0.9005	1.28926
X4	1	0.02311	0.01039	2.22	0.0269	1.04618
X5	1	0.63373	0.05499	11.52	<.0001	1.05060
X6	1	-0.00237	0.00549	-0.43	0.6659	2.56160
X7	1	0.00363	0.00447	0.81	0.4180	2.76096
X8	1	0.00123	0.00109	1.12	0.2636	1.45311

探討糖尿病的高危險因子

《表三》

向前選取法	x4、x5、x8
向後刪去法	x4、x5、x8
逐步分析選取法	x4、x5、x8
Cp 準則	x 1、x4、x5、x8
調整後判定係數選取法	x4、x5、x8

《表四》

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
3	0.3612	0.3675	1.2511	x4 x5 x8
4	0.3610	0.3694	2.3544	x1 x4 x5 x8
4	0.3609	0.3692	2.4057	x2 x4 x5 x8
5	0.3607	0.3711	3.5271	x1 x2 x4 x5 x8
4	0.3596	0.3679	3.0275	x3 x4 x5 x8

《表五》

Tests for Location : Mu= 0			
Test	Statistic	p Value	
Student's t	t -0.00085	Pr >   t	0.9993
Sign	M -19.5	Pr >=   M	0.0856
Signed Rank	S -3899.5	Pr >=   S	0.2126

探討糖尿病的高危險因子

《表六》

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.950053	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.080971	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.851551	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.807331	Pr > A-Sq	<0.0050

《表七》

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.995381	Pr < W	0.2731
Kolmogorov-Smirnov	D	0.022225	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.026155	Pr > W-Sq	>0.02500
Anderson-Darling	A-Sq	0.258894	Pr > A-Sq	>0.02500

《表八》

The REG Procedure	
Model : MODEL1	
Dependent Variable : y	
Durbin-Watson D	1.920
Number of Observations	489
1 <sup>st</sup> Order Autocorrelation	0.039

探討糖尿病的高危險因子

《表九》

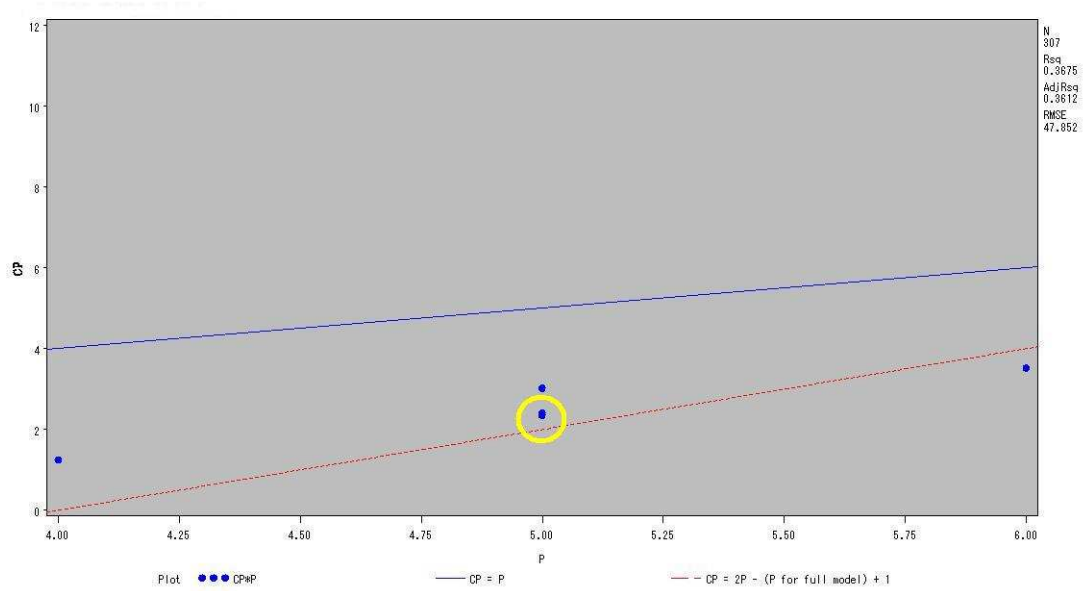
Student Residual	101,137,376,409,469,471
Rstudent	99,327,351,397

《表十》

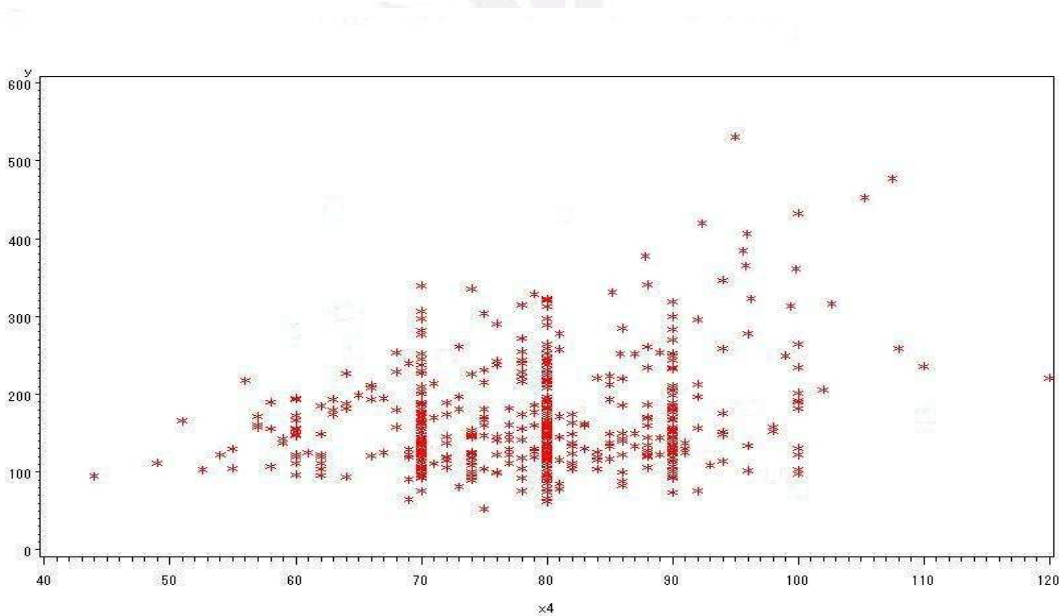
Hat Diag	2,11,12,16,28,38,89,99,163,187,191,204,208,224,227,269,272,278,281,289,290,313,320,327,350,352,355,356,357,360,367,388,392,397,401	
CovRatio	2,12,16,38,27,75,99,107,132,133,163,167,187,191,194,200,204,208,222,224,227,249,254,261,274,278,281,290,313,327,344,346,351,352,355,360,363,388,390,392,397,401	
DFFITs	11,25,27,28,73,75,99,107,206,222,224,269,274,281,320,327,350,351,357,372,405	
DFBETAS	I	7,25,27,50,70,73,75,99,103,106,194,200,201,224,231,272,274,280,281,287,311,351,357,372,397,405
	x4	23,25,50,73,75,99,103,146,194,200,222,224,231,272,274,280,287,299,311,322,327,351,357,372,397
	x5	2,11,13,75,99,106,107,122,140,206,222,230,248,261,263,264,269,289,298,314,320,327,351,354,357,380,397,405
	x8	25,27,28,71,75,99,200,222,224,264,281,327,346,350,355,356,367,397,405

## 圖附錄

《圖一》 Cp 圖

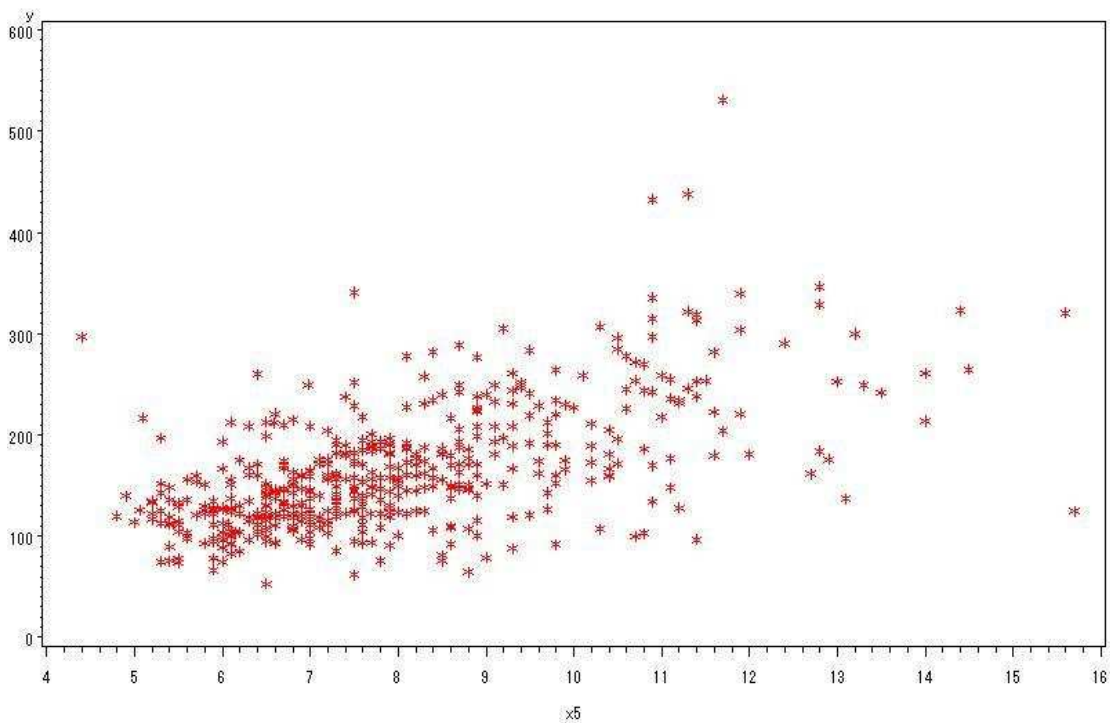


《圖二》變數 x4 散佈圖

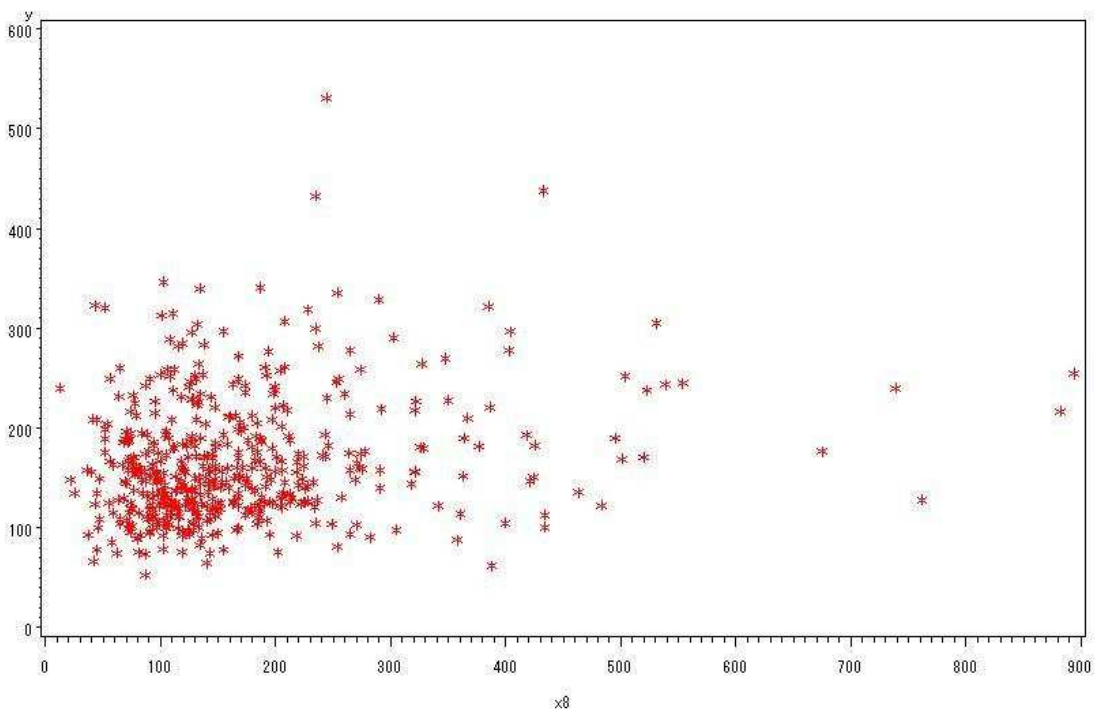


探討糖尿病的高危險因子

《圖三》變數 x5 散佈圖



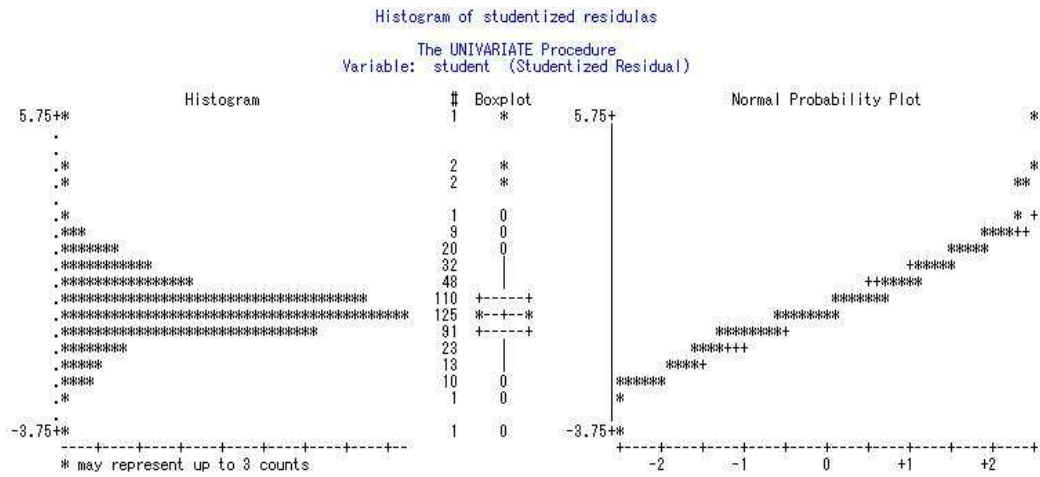
《圖四》變數 x8 散佈圖



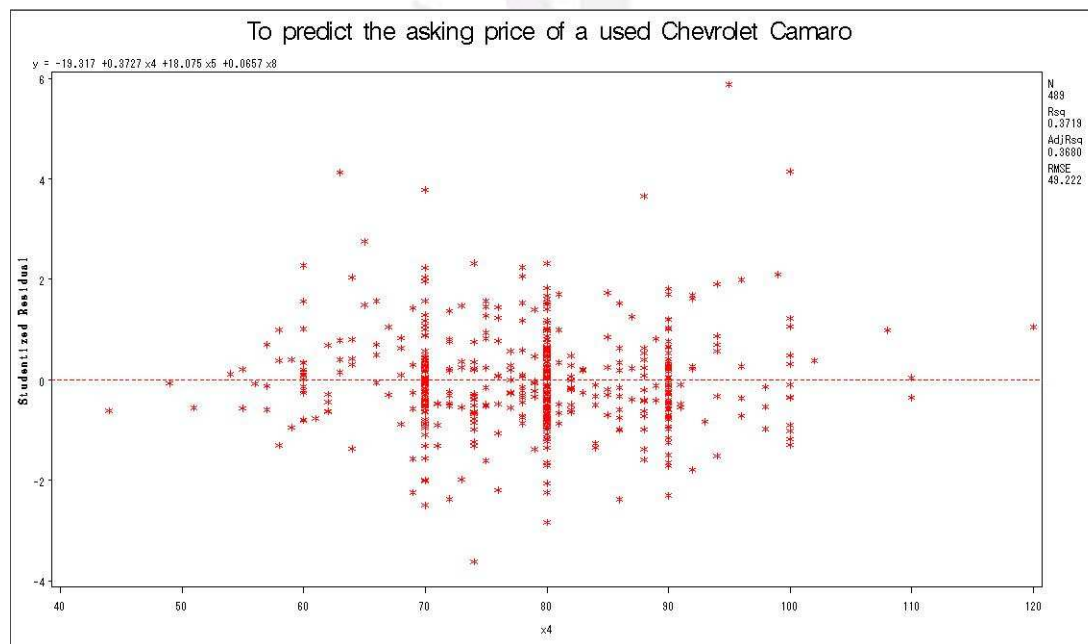


探討糖尿病的高危險因子

《圖五》

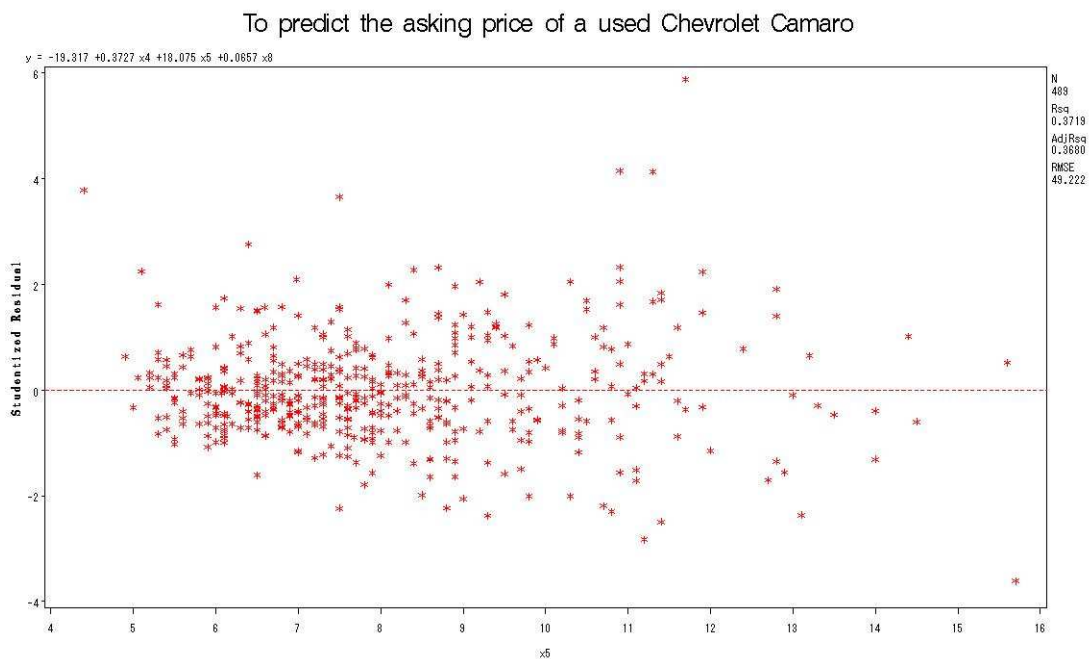


《圖六》變數 x4 標準化殘差圖

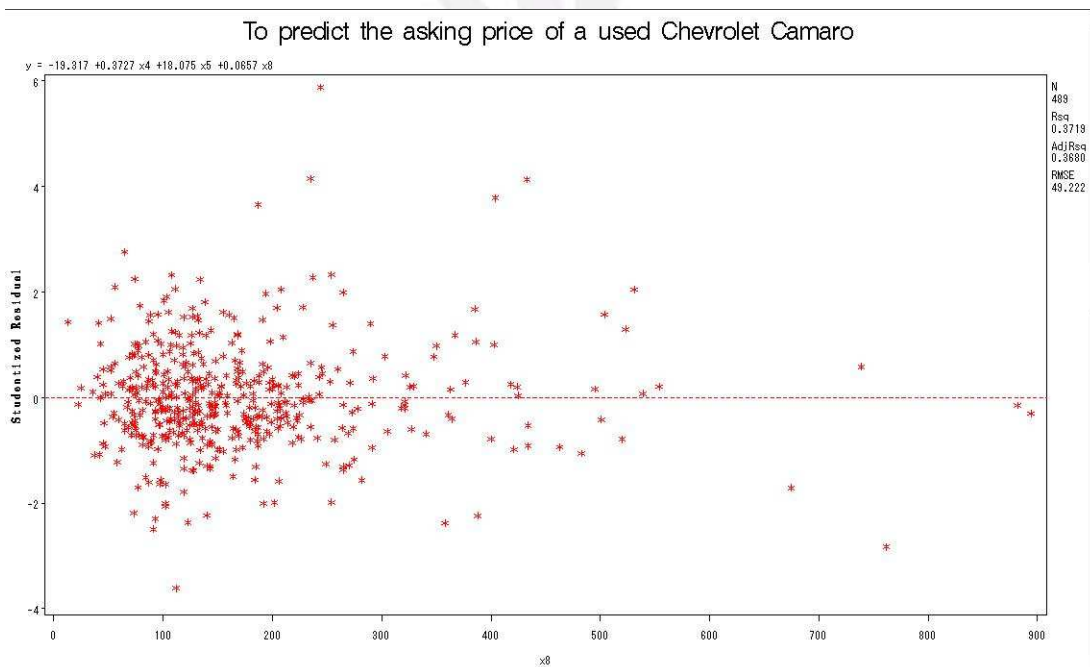


# 探討糖尿病的高危險因子

## 《圖七》變數 x5 標準化殘差圖

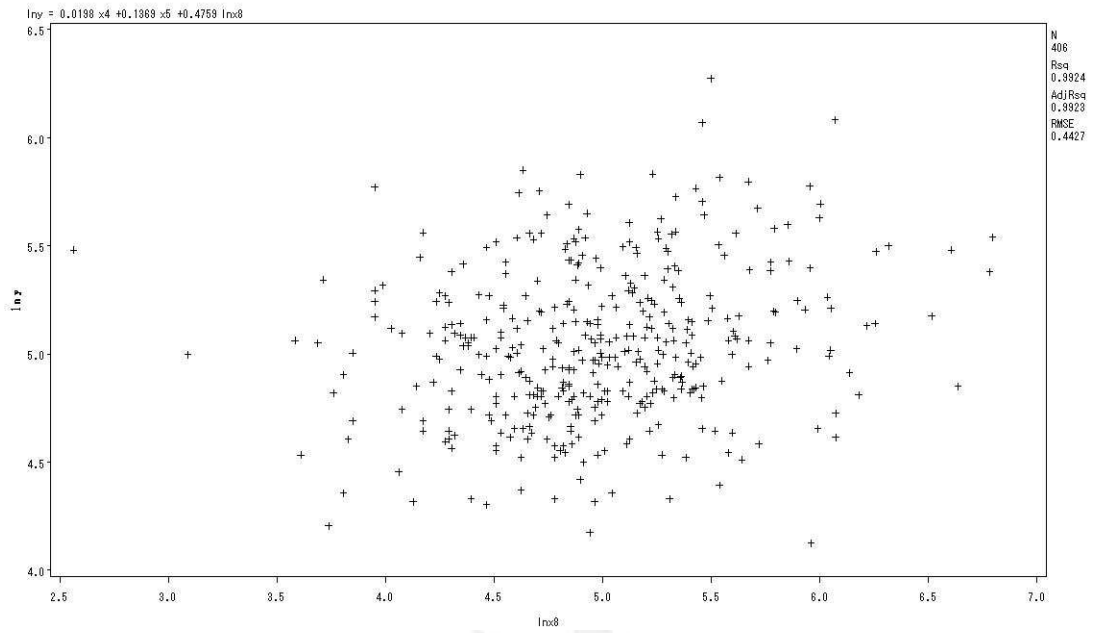


## 《圖八》變數 x8 標準化殘差圖



# 探討糖尿病的高危險因子

## 《圖九》轉換後 x8 散佈圖



## 《圖十》轉換後 x8 殘插圖

