

一、何謂邏輯回歸分析 (Logistic regression Analysis)？邏輯回歸分析 (Logistic regression Analysis) 有哪幾種做法？適用時機為何？請自選一個適當的研究主題，使用邏輯回歸分析來分析資料，請說明此研究的研究問題與目的、研究假設，並詳述如何進行資料分析、並使用虛擬資料呈現資料分析結果與解釋。

【擬答】

我們想要預測一件事情，最常用的統計工具就是「迴歸」(regression)，要被預測或被瞭解的變項叫做依變項 (Dependent variable)，它可以是名目變項 (nominal)、順序變項 (Ordinal)、等距變項 (interval) 以及比率變項 (ratio)。如果依變項是屬於後兩者，我們稱作連續變項 (Continuous)，那麼我們習慣用線性迴歸 (Linear regression) 去配適資料。



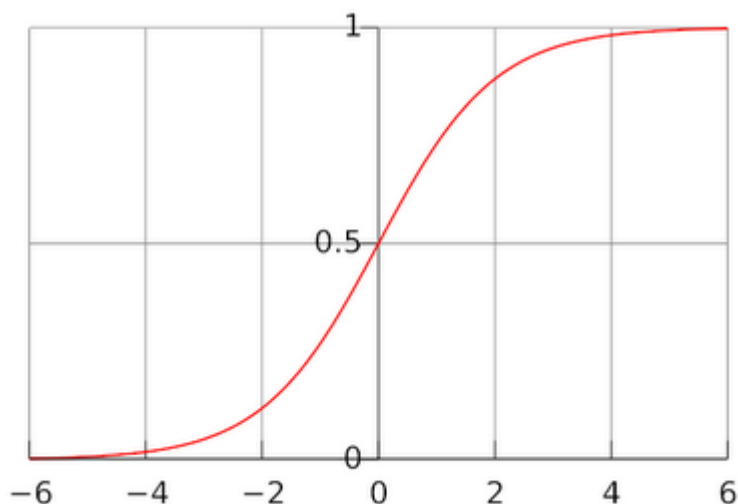
傳統線性迴歸的迴歸係數 (regression coefficient) 的解釋為「當自變項增加一個單位，依變項則會增加多少單位」，但是在邏輯回歸分析 (Logistic regression Analysis) 的迴歸係數解釋為「當自變項增加一個單位，依變項 1 相對依變項 0 的機率會增加幾倍」，也就是說「自變項增加一個單位，依變項有發生狀況 (稱為 Event) 相對於沒有發生狀況 (non-event) 的比值」，這個比值就是勝算比 (Odds ratio, OR)。勝算比值愈高，表示預測變項 (自變項) 與效標變項 (依變項) 關連程度愈強。

(一) 定義

邏輯回歸分析 (Logistic regression Analysis) 類似線性迴歸分析，主要在探討依變數與自變數間的關係。線性迴歸中的依變數 (Y) 通常為連續型變數，但邏輯回歸分析 (Logistic regression Analysis) 所探討的依變數 (Y) 主要為類別變數，特別是分成兩類的變數 (例如：是或否、有或無、同意或不同意……等)。

(二) 邏輯回歸分析和多元迴歸分析比較

- 1、多元迴歸分析效標變項和預測變項皆為連續變項，而邏輯回歸分析的效標變項為類別變項，預測變項為連續變項。
- 2、邏輯回歸分析的資料必須符合 S 型的機率分配，也稱為 Logit 分佈，常用的估計方法為最大概似法 (maximum likelihood estimate)，此方法可以用於更複雜的非線性估計。



(三) 作法

邏輯回歸分析 (Logistic regression Analysis) 的顯著性檢定包含：整體模式檢定與個別參數檢定二種，以下分述之：

1、整體模式檢定

SPSS 軟體提供兩種整體模式檢定資料：Pearson X^2 值與 Hosmer 以及 Lemeshow 檢定結果，當 Pearson X^2 值達顯著時，表示所投入的自變項中，至少有一個自變項能有效預測樣本在依變項之機率值；當 Hosmer 以及 Lemeshow 檢定值未達顯著時，表示模式愈佳。

2、個別參數檢定

SPSS 亦有兩種方法，分別為 Wald 檢定值與 Score 檢定值，當 Wald 檢定值達顯著時，表示自變項與依變項之間有顯著關連，但該值受到迴歸係數的影響，也就是迴歸係數很大時，Wald 的估計標準差就會膨脹，進一步導致 Wald 的估計值變得很小，使得犯型二錯誤機率就會提升，因此此時最好再參考 Score 檢定值是否達到顯著。

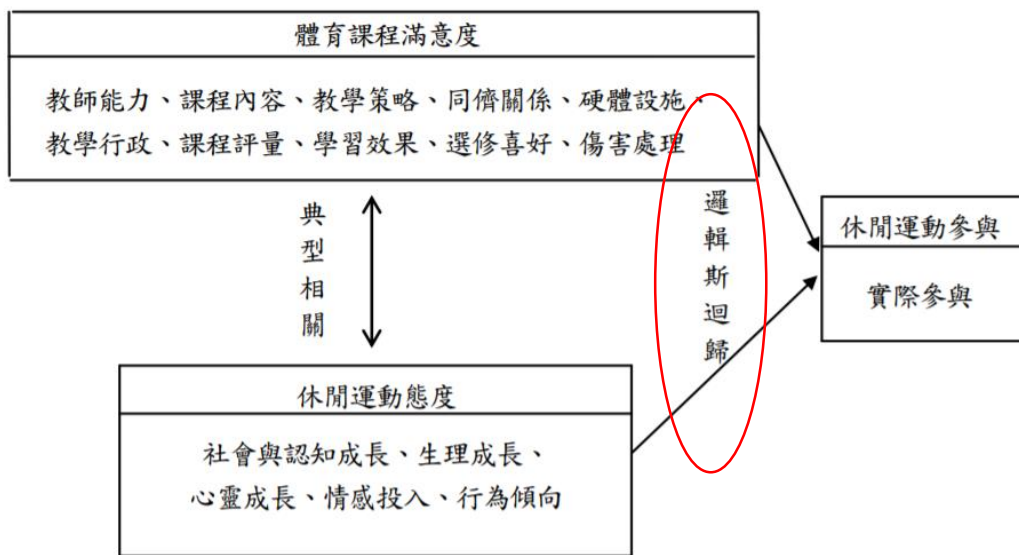
(三) 個案說明-1

【引用整理自：陳瑞辰、張川鈴 (2020)。臺灣師範校院學生體育課程滿意度、休閒運動態度與休閒運動參與之研究。《臺灣體育運動管理學報》，20(1)，59-86。】

1、研究假設：

- (1) 利用典型相關來瞭解體育課程滿意度與休閒運動態度間是否有相關情形。
- (2) 利用邏輯回歸分析 (Logistic regression Analysis) 探討體育課程滿意度與休閒運動態度對休閒運動參與的預測情形。

2、研究架構：



3、邏輯斯迴歸分析摘要表

體育課程滿意度、休閒運動態度對休閒運動參與之邏輯斯迴歸分析摘要表

預測變項	β係數	標準誤	勝數比
常數	-2.08*	.46	0.12
學習效果	0.77*	.14	2.16
硬體設施	-0.16	.09	0.85
教師能力	-0.16	.14	0.84
同儕關係	-0.19	.11	0.82
課程內容	0.21	.16	1.24
教學策略	0.15	.12	1.16
課程評量	-0.01	.13	0.98
選修喜好	0.04	.09	1.04
傷害處理	-0.16	.10	0.84

註：df=1；*p<.05；n.s.p>.05

體育課程滿意度、休閒運動態度對休閒運動參與之邏輯斯迴歸分析摘要表(續)

預測變項	β係數	標準誤	勝數比
教學行政	0.00	.09	1.00
社會與認知成長	0.16	.17	1.18
行為傾向	0.45*	.09	1.57
生理成長	0.04	.14	1.04
情感投入	-0.24	.14	0.78
心靈成長	-0.22	.14	0.80

整體模式適配度檢定 $\chi^2 = 123.56^*$

Hosmer-Lemeshow 檢定值 = 11.40 (n.s.)

註：df=1；*p<.05；n.s.p>.05

本研究透過邏輯迴歸分析 (Logistic regression Analysis) 進行分析，針對 15 個自變項所建立影響參與休閒運動之預測迴歸模型，其整體模式顯著性考驗 $\chi^2 = 123.56$ ($p < .05$)，達顯著水準；Hosmer-Lemeshow 檢定值為 11.40 ($p > .05$)，未達顯著水準，表示 15 個自變項所建立之迴歸模式適配度良好，亦即自變項可以有效預測依變項。

學生休閒運動參與預測因素中，學習效果、行為傾向等因素達顯著水準 ($p < .05$)。由於學習效果與行為傾向之勝數比 (odds ratio) 皆大於 1，故此兩項皆

為正向影響因素，亦即當學習效果或是行為傾向之影響因素得分提高，則學生參與休閒運動意願相對增加。

此外，將學習效果視為參與休閒運動之重要因素的學生，其參與運動的意願則是未將其作為考量因素之學生的 2.16 倍；將行為傾向視為參與休閒運動之重要因素的學生，其參與運動的意願則是未將其作為考量因素之學生的 1.57 倍。

(四) 個案說明-2

【引用整理自：王品方、鄭宇庭、郭訓志 (2018)。應用羅吉斯迴歸分析在個性化旅遊景點推薦模型之研究。 *Journal of Data Analysis* , 13(6) , 51-75 。】

研究方法：應用羅吉斯迴歸分析，將蒐集到的數據建模，找到每一種旅行方式較容易吸引到哪一些族群，並且對該族群做出該旅遊景點的推薦。

Y1	網去這麼好富	女性族群
		旅行時重玩&跟隨朋友類型的族群
		對於社群平台 Facebook 或 Instagram 上的旅遊資訊關
		注程度高的族群
Y2	塔上文青之旅	旅行時需要滿足嘗鮮類型的族群
		對於社群平台 Facebook 或 Instagram 上的旅遊資訊關
		注程度高的族群
Y3	飛簷走壁	旅行時最在意景點內容的族群
		旅行時重吃的族群
Y4	小當家	男性族群
Y5	越夜越美麗	女性族群
		能接受最多新台幣 500 元內旅行預算的族群
		喜愛秋季的族群
		對台北的美食景點熟悉程度高的族群
Y6	享受人生	旅行時傾向大自然秘境景點景點類型的族群
		能接受最多新台幣 1,000 以上旅行預算的族群
		只想要與自己朋友旅行的族群

在將資料匯入 clementine 進行分析前，先來做資料整理的步驟。先將從網路與紙本問卷蒐集到的答案合併過後，將問卷最後一題喜歡的旅遊景點作為此次的應變數 Y 來建模，由於 Y 當初設定的是複選題，因此將 Y 拆解成 Y1~Y6，表示

成 6 個變數，代表 6 種旅遊型態並建立 6 條旅遊景點推薦模型。接下來，將其他問卷題目當成自變數 X，由於 X 有連續型態與離散型態，因此將所有離散型態的 X 都設成虛擬變數來進行建模。

以下我們透過三個指標來評斷模型的效果，進而決定本文的最終模型，分別是 Omnibus Tests of Model Coefficients 來進行模型適合度檢定、Cox & Snell R Square 和 Nagelkerke R Square 來比較模型的解釋能力，最後用模型的 Classification Table 來看模型的預測能力。

表 1 Omnibus Tests of y1 Model Coefficients

Omnibus Tests of Model Coefficients (y1)			
Step	Chi-square	df	Sig.
enter	52.906	29	0.004
forwards	22.242	2	0
backwards	40.528	8	0

y1 的模式係數的 Omnibus 測試相當於線性迴歸裡的 ANOVA-F 檢定，探討羅吉斯迴歸模型中的 β 係數是否全部為 0，如下顯示：

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

由表 1 可以觀察到三種模型的顯著性 (Sig.) p 值 < 0.05，拒絕虛無假說。有足夠證據顯示此 y1 的羅吉斯迴歸模型是顯著的，具有預測能力。

表 2 Model Summary of y1

Model Summary of y1			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
enter	125.573(a)	0.291	0.424
forwards	156.237(a)	0.134	0.196
backwards	137.951(b)	0.231	0.337

接下來，根據表 2 顯示，取決於 Cox & Snell R Square 與 Nagelkerke R Square，enter 的模型之應變量可以被解釋的變異比例在 29.1%~42.4% 之間，為三種模型之中解釋能力最高者，因此 enter 模型是相較之下較好的選擇。

表3 Classification Table of y1

Classification Table(y1)					
	Observed	y5	Predicted		
			y5		Percentage
			0	1	Correct
Step 0	y5	0	113	0	100
		1	41	0	0
Overall Percentage					73.4
Step 1	y5	0	102	11	90.3
		1	21	20	48.8
Overall Percentage					79.2

再來我們根據enter的模型進一步探討預測效果，根據表3顯示，在不考慮自變數，假設所有觀察沒有選擇名為「網美這麼好當？」旅行景點(包含熱門打卡餐廳等，以下簡稱網美打卡景點)的人，模型能夠將在納入模型能夠將73.4%的觀測正確分類。當納入自變數時，模型能夠將79.2%的觀測正確分類（查看Overall Percentage一行），即納入這些自變數後能夠提高整體模型的預測能力。

表4 y1的羅吉斯迴歸模型表

variable	B	S.E.	Wald	df	Sig.	Exp(B)
sex_M(1)	-1.348	0.572	5.559	1	0.018	0.26
prefer_1(1)	0.241	0.66	0.133	1	0.715	1.272
prefer_2(1)	-2.723	1.275	4.563	1	0.033	0.066
prefer_3(1)	-0.033	0.615	0.003	1	0.958	0.968
playoreat_1(1)	0.889	0.694	1.64	1	0.2	2.433
playoreat_2(1)	2.731	1.144	5.697	1	0.017	15.353
satisfy_1(1)	18.595	40192.92	0	1	1	119099095.1
satisfy_2(1)	0.686	0.596	1.324	1	0.25	1.986
satisfy_3(1)	2.262	1.066	4.502	1	0.034	9.602
satisfy_4(1)	20.407	40192.96	0	1	1	728706642.6
fbig	0.379	0.132	8.217	1	0.004	1.46

確定模型後，如表4我們將數據顯示顯著(Sig.<0.05)的變數納入模型當中，

而當虛擬變數當中任何一個代表其變數的變數顯著時（如紅色顯示），我們會將整個變數都納入模型當中解釋，之後的模型將以此類推。

最後，我們可以對於選擇網美打卡景點的族群特性，分別用不同變數的 Odds Ratio(如黃色顯示)做一個歸納，其中表 4 的 $\text{Exp}(B)$ 即為 Odds Ratio。以下就相關結論說明之：

1、女性族群：

$\text{OR}=0.26 < 1$ ，這說明女性選擇網美打卡景點的勝算，比男性約多出 3.84 倍。

2、旅行時跟隨朋友類型的族群：

$\text{OR}=0.066 < 1$ ，這說明跟隨朋友類型的人選擇網美打卡景點的勝算，比起自己是傾向熱門打卡景點類型約多出 15.15 倍。

3、旅行時重玩的族群：

$\text{OR}=15.353 > 1$ ，這說明重玩類型的人選擇網美打卡景點的勝算，比吃玩並重的人約多出 15.353 倍。

4、旅行時需要滿足嘗鮮類型的族群：

$\text{OR}=9.602 > 1$ ，這說明旅行時需要滿足嘗鮮類型的人選擇網美打卡景點的勝算，比旅行時需要全方位滿足的人約多出 9.602 倍。

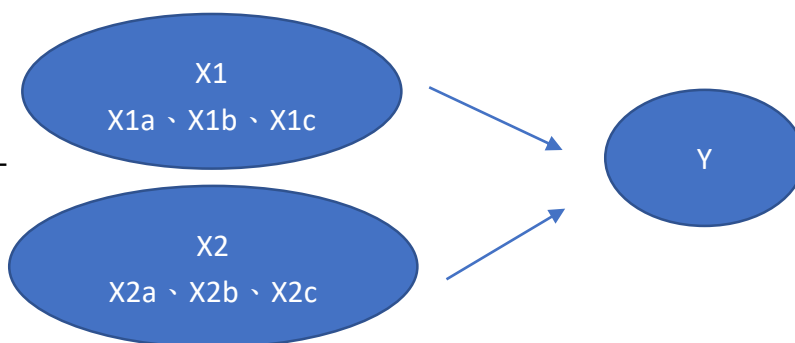
5、對於社群平台 Facebook 或 Instagram 上的旅遊資訊關注程度高的族群：

$\text{OR}=1.46 > 1$ ，這說明對於社群平台 Facebook 或 Instagram 上的旅遊資訊關注程度越高的人越容易選擇網美打卡景點，且關注程度每多一分，會選擇網美打卡景點的勝算約多出 1.46 倍。

（五）運用虛擬資料呈現資料分析結果與解釋

1、研究問題：運用邏輯回歸分析（Logistic regression Analysis）探究 X1 與 X2 對於 Y 的影響，探討 X1 對於 Y 以及 X2 對於 Y 有無正向影響。

2、研究變項架構



預測變項	β 係數	標準誤	勝算比
X1a	0.77*	.15	2.16
X1b	0.16	.09	0.85
X1c	0.19	.16	0.14
* $p < .05$ Pearson X^2 :123.56* Hosmer-Lemeshow:11.56*			

預測變項	β 係數	標準誤	勝算比
X2a	0.16	.16	0.18
X2b	0.45*	.12	1.57
X2c	0.04	.13	0.14
* $p < .05$ Pearson X^2 :123.56* Hosmer-Lemeshow:11.56*			

本研究透過邏輯回歸分析 (Logistic regression Analysis) 進行分析，針對 6 個自變項所建立影響參與休閒運動之預測迴歸模型，其整體模式顯著性考驗 $\chi^2 = 123.56$ ($p < .05$)，達顯著水準；Hosmer-Lemeshow 檢定值為 11.56 ($p > .05$)，未達顯著水準，表示 6 個自變項所建立之迴歸模式適配度良好，亦即自變項可以有效預測依變項。

X1 自變項中，X1a 因素達顯著水準 ($p < .05$)，X2 自變項中，X2b 因素達顯著水準 ($p < .05$)。由於 X1a 因素與 X2b 之勝數比 (odds ratio) 皆大於 1，故此兩項皆為正向影響因素，亦即當 X1a 或是 X2b 之影響因素得分提高，則對 y 有正相關。