جامعة محمد خيضر (بسكرة)
كلية العلوم الاقتصادية والتجارية والتسيير
قسم العلوم التجارية

السنة الجامعية: 2024/2023
المستوى: أولى ماستر تسويق
مقياس: الأساليب الكمية في التسويق 2

المحاضرة الثالثة: نموذج الاختيار المنفصل: التحليل اللوجستي Logistic regression

1. تمهيد:

يعتبر تحليل الانحدار من الطرق الاحصائية التي تستخدم لوصف، تفسير والتنبؤ بالعلاقة بين متغيرين أو أكثر، وقد تكون نماذج الانحدار خطية أو غير خطية، ورغم أن النماذج الخطية هي الأكثر استعمالا، إلا أن العديد من الظواهر ذات متغيرات غير كمية، وعلاقة غير خطية بين المتغيرات، ولذا تحتاج إلى نماذج احصائية ملائمة لطبيعتها وخصاصها، ومن أبرز هذه النماذج نجد الانحدار اللوجستي.

إن الاختلاف الرئيسي الذي يجعل الانحدارين اللخطي واللوجستي مختلفين عن بعضهما هو طبيعة المتغيرالتابع (متغير الاستجابة)، فعندما يكون فئويا (ثنائيا أو أومتعدد الفئات)، فعندئذ يتم إستخدام الانحدار اللوجستي، أما عندما تكون المتغير التابع كميا مستمرا، عندئذ يتم استخدام الانحدار الخطي.

عند محاولة استخدام تحليل الانحدار الخطي بطريقة المربعات الصغرى (OLS) لتوفيق البيانات ذات المتغيرات التابعة الثنائية أو متعددة الاستجابة، يبرز نوعان من رئيسيان من المشاكل، هما المبرارن لاستخدام الانحدار اللوجستي مكان الانحدار الخطي أو غيره من الأساليب الاحصائية الأخرى لتوفيق البيانات مع المتغير التابع الثنائي، وهذان النوعان من المشاكل هما:
تباين الخطأ - في حالة متغير تابع ثانئي أو متعدد الفئات- قد لا يتوزع توزيعا طبيعيا ، كما أن هذا التباين في الخطأ غير ثابت، كما هو مطلوب في تحليل الانحدار الخطي.
عدم إمكان تفسير القيم المتنبأ بها بوصفها احتمالات، حيت لا يمكن حصر هذه القيم بين الصفر والواحد، لذا لا يمكن الاستعانة بأساليب إحصائية أخرى ومنها نموذج الانحدار اللوجستي .

2. تعريف نموذج الانحدار اللوجستي:

نموذج الانحدار اللوجستي أو نموذج لوجيت logit هو نموذج إحصائي يسمح بدراسة العلاقة بين مجموعة من المتغيرات المستقلة x_i ومتغير نوعي تابع y. وبالتالي يسمح بالتنبؤ باحتمال وقوع حدث معين (نجاح/ فشل، شفاء/إصابة، شراء/ عدم شراء؛ منتج سليم/ معيب ...)

3. تطبيقات الانحدار اللوجستي:

- **الإحصاء الحيوي:** التنبؤ باحتمال حدوث نوبة قلبية لشخص ما خلال فترة معينة، حسب المعلومات الديمغرافية (العمر، الجنس) أو الطبية (مؤشرات بدنية، صحية، غذائية ..) أو السلوكية (التدخين، الكحول ...).
- **الطب والصيدلة:** التنبؤ باحتمالي إصابة فرد ما من عدم إصابته بمرض ما، من خلال معلومات عن الجانب الوراثي والمناعي؛ تقدير رد الفعل نحو الجرجات والمقارنة بين نجاعة الأدوية .
- **التأمينات:** فرز وتقسيم مجموعات العملاء في شركات التأمين حسب المخاطر، وتحديد مدى انجذابهم لمنتجات تأمينية معينة .
- **البنوك:** تنقيط العملاء أثناء دراسة ملفات القروض، لتحديد مدى قدرتهم على السداد من عدمه، وبالتالي اتخاذ قرار منح القرض أورفضه .
- **التسويق:** حساب توقعات ميل المستهلك إلى شراء منتج ما أو امتناعه عن الشراء، أو التنبؤ بالاعجاب بحملة إعلانية من عدمه.
- **سبر الآراء:** التنبؤ بقرار التصويت في الانتخابات اعتمادا على تنميط قبلي للمصوتين (مستوى اجتماعي، توجه سياسي، مستوى تعليمي ...).

4. أمثلة على الانحدار اللوجستي:

مثال (1):

تود شركة للتجارة الإلكترونية إرسال عروض ترويجية باهضة الضمن عبر البريد الإلكتروني إلى العملاء، وتريد معرفة ما إذا كان من المحتمل أن يستجيب عميل معين لذلك العرض أم لا، في التسويق يسمى هذا ميل الاستجابة للنمذجة. وقد صار الانحدار اللوجستي أمرا شائعا في الإعلان عبر الإنترنت، حيث يسمح للمسوقين بالتنبؤ بنسبة مئوية باحتمالية قيام مستخدم الويب بالنقر فوق إعلانات معينة.

مثال (2):

يريد بنك تجاري بناء نموذج تنبؤي لتقرير ما إذا كان سيصدر بطاقة إئتمان لعميل معين أم لا، ومعاولة التنبؤ بما إذا كان العميل سيتخلف عن السداد أم لا، بناء على خصائص معينة كالدخل السنوي، مدفوعات بطاقة الإئتمان الشهرية، عدد حالت التخلف عن السداد، في اللغة البنكية يسمى هذا نمذجة المخاطر البنكية.

5. أنواع الانحدار اللوجستي:

ينقسم الانحدار اللوجستي نوعين رئيسيين هما:

أ. التحليل اللوجستي الثنائي Binomial logistic regression :

يشترط فيه أن يكون المتغير التابع Y نوعيا، ويأخذ حالتين متنافيتين، مثل: نجاح أو فشل، ربح أو خسارة، مدخن أو غير مدخن، حامل لممرض أو غير حامل له، محقق لشرط ما أو غير محقق لو،... الخ)، وأن يأخذ مقابل الحالة المرغوبة الأولى القيمة العددية (1)، وأن يأخذ مقابل الحالة الثانية القيمة العددية (0)، أما المتغيرات المؤثرة في التصنيف X فيمكن أن تكون كمية أو نوعية أو مختلطة، وتأخذ قيمها ضمن مجالات أو فئات محددة، ولا يشترط عليها أن تحقق أية شروط مسبقة.

ب. التحليل اللوجستي المتعدد Multinomial logistic regression:

يعتبر امتداد بسيطا للانحدار اللوجستي الثنائي، حيث يشترط فيه أن يكون المتغير التابع Y نوعيا، ولكن يأخذ عدة حالات متنافية، مثل: مستوى التعليم، نوع المهنة، الحالة الاجتماعية .. الخ، وأن يأخذ مقابل إحدى الفئات القيمة (1)، ومقابل الفئات المتبقية القيمة (0).

6. الإطار الرياضي لتحليل الإنحدار اللوجستي:

يرتبط بناء وفهم وتطبيق نموذج الانحدار بمجموعة من المفاهيم الرياضية والاحصائية، من أهمها: الظواهر الثانئية أو توزيع برنولي؛ نسبة أو معدل الأرجحية؛ الاحتمالات الشرطية، الدالة اللوجستية، طريقة الإمكان الأعظم في تقدير المعلمات الاحصائية ... الخ، ولذا سنتناول بعضها فقط فيما يلي:

أ. الظواهر الثنائية وخواصها:

إن الظواهر الثنائية هي عبارة عن متغيرات ثنائية تأخذ حالتين A وĀ فقط (موافق أو غير موافق، نعم أو لا، نجاح أو فشل، ربح أو خسارة، قبول أو رفض،... الخ)، ولقد أصطلح على إعطاء المتغير التابع الثنائي Y قيمة (1) عندما تتحقق الحالة المرغوبة A، وقيمة الصفر (0) عندما تحقق الحالة غير المرغوبة Ā (عدم تحقق A)

مثال:

نفترض عند إجراء (n=100) تجربة على أية ظاهرة ثنائية، كانت نتائج تلك التجارب التي تخضع لتوزيع برنولي كما في الجدول التالي:

الجدول رقم (01): جدول توزيع برنولي

الحالة	$A= G_1$	$\bar{A}= G_2$	المجموع
قيمة التابع Y	1	0	1
احتمال التحقق	p	q	p+q=1
عدد التكرارات المطلقة	n_1	n_0	$n_1 + n_0= n$
توزع عدد التجارب	60	40	100

وعندما نكرر هذه التجربة n مرة سنحصل على عينة من قيم Y بحجم العينة n، وعناصرها تتوزع حسب الجدول السابق إلى مجموعتين هما:

- **المجموعة G_1:** هي مجموعة العناصر التي تقابل القيم (Y=1)(مجموعة الناجحين)، وتضم n_1 عنصرا، ويفترض أن يكون احتمال تحققها في كل تجربة ثابتا ويساوي p، وإن p يقدر بـ: 0.6 =60/100= p.
- **المجموعة G_2:** هي مجموعة العناصر التي تقابل القيم (Y=0)(مجموعة الراسبين)، وتضم n_0 عنصرا، ويفترض أن يكون احتمال تحققها في كل تجربة ثابتا ويساوي q، وإن q يقدر بـ: 0.4 =40/100= p.

وبناءً على نتائج هذه التجارب يمكننا تعريف عدة مؤشرات تستخدم في التحليل اللوجستي أهمها **نسبة الأرجحية (odds)**.

ب. مفهوم نسبة الأرجحية (odds):

إذا كانت امكانية فوز اللاعب A تساوي n_1، مقابل n_0 للاعب $\bar{A}$، وإذا قام اللاعبان بإجراء n= 100 تجربة، وفاز اللاعب A بـ n_1= 60 تجربة، وخسر n_0= 40 تجربة منها، فإن تعريف الأرجحية لحادث فوز اللاعب A على اللاعب $\bar{A}$ تعطى بالعلاقة التالية:

$$Odds(A)= \frac{\text{عدد مرات تحقق فوز A}}{\text{عدد مرات تحقق فوز } \bar{A}} = \frac{60}{40} = \frac{3}{2} = \frac{1.5}{1}$$

وعندها نقول أن إمكانية فوز A على $\bar{A}$ تساوي 60 مقابل 40، وهنا يفضل اختصار الكسر 60/40 إلى آخر عددين صحيحين مثل 3/2، ونقول أن امكانية فوز A على $\bar{A}$ تساوي 3 مقابل 2، وإنه من الأفضل تحويل الكسر الأخير إلى نسبة عدد إلى الواحد الصحيح، مثل 1.5/1، ونقول أن فوز A على $\bar{A}$ تساوي 1.5 مقابل 1، وتكتب على الشكل: (1 : 1.5)، وتكتب بالعكس بالنسبة للاعب $\bar{A}$ (1 :1.50).

وبطريقة مشابهة تعرف الأرجحية لفوز اللاعب $\bar{A}$ بالعلاقة:

$$Odds(\bar{A})= \frac{n_0}{n_1} = \frac{\text{عدد مرات تحقق فوز } \bar{A}}{\text{عدد مرات تحقق فوز A}} = \frac{4}{60} = \frac{2}{3} = \frac{1}{1.5}$$

ومن التعريفين السابقين نستنتد أن:

$odds(A)= 1/ odds(\bar{A})$ $\quad odds(A) \times odds(\bar{A}) = 1$

تعريف تحقق فوز اللاعب A: يعرف بالعلاقة التالية:

$P(A)= n_1/(n_1+n_0)= n_1/n= 60/100= 0.60= 1.5/(1+1.5)$

تعريف تحقق فوز اللاعب $\bar{A}$: يعرف بالعلاقة التالية:

$P(\bar{A})= n_0/ (n_1+n_0)= n_0/n= 40/100= 0.40= 1/(1+1.5)$.

من العلاقتين نستخلص أن: $P(A) + P(\bar{A}) = p+q= 1$

كما يمكن استخلاص العلاقة التي تربط بين الأرجحية واحتمال تحقق حالتها، حيث يمكننا كتابة:

$Odds(A)= n_1/n_0 = (n_1/n)/ (n_0/n)= P(A)/P(\bar{A})= p/q= p/(1-p)$. …….(1)

$Odds(\bar{A})= n_0/n_1 = (n_0/n)/ (n_1/n)= P(\bar{A})/P(A)= q/p= (1-p)/ p$ ……(2)

يمكن استخدام العلاقة (1) في استخراج نموج الانحدار اللوجستي.

ج. استخراج نموذج الانحدار اللوجستي:

لقد رأينا في المثال أن المتغير التابع Y هو متغير ثنائي، ويأخذ إحدى القيمتين (1) للنجاح أو (0) للرسوب، ومن شكل الانتشار، نلاحظ أن Y لا يصلح من وجهة نظر تحليل الانحدار، لأن يكون نتيجة لتركيب خطي أو غير خطي للمتغير المستقل X، لذلك يجب البحث عن بديل للمتغير التابع Y مرتبط به ويعبر عنه.

ومن جهة أخرى، نجد أن الاحتمال الشرطي لأن يأخذالمتغير التابع Y القيمة (Y=1) عند قيمة معطاة x يساوي احتمال أن ينتمي العنصر المعموم x إلى المجموعة G_1، ونكتب ذلك على الشكل التالي:

$P(G_1/x)= P(Y=1/x)= P_1(x)=$ Y الاحتمال اللاحق لـ

وهو احتمال النجاح عند أية قيمة معطاة x، وهو متغير تابع مستمر، ويأخذ قيمو في المجال [1 , 0] ، وهو يصلح لأن يكون بديلاً عن Y، لأنه أصبح من الممكن رياضياً دراسة علاقة $P_1(x)$ مع المتغير المستقل X .

واذا استطعنا أن نجد العلاقة بين هذا الاحتمال $P_1(x)$ والمتغير X، فإننا نكون قد تجاوزنا المشكلة، التي واجهتنا أثناء تمثيل Y عبر X .

وهكذا نجد أنه يجب علينا الآن أن نقوم بإيجاد قيم $P_1(x)$ المقابلة لجميع قيم x، حتى نستطيع أن نقابلها مع قيم x، ثم استخلاص علاقة الانحدار بينيما دون وضع شروط مسبقة على المتحول X .

لذلك نقوم بحساب قيم الاحتمالات $P_1(x)$ اللاحقة من علاقات بايز التي تأخذ الشكل التالي:

$$P_1(x)= P(G_1/x)= \frac{P * f(x/G_1)}{P * f(x/G_1) + q * f(x/G_0)}$$

حيث $f(x/G_1)$ و$f(x/G_0)$ هما التوزيعان التجريبيان لـ X ضمن المجموعتين G_1 وG_2 على الترتيب، وهما يحسبان (بعد تكرار التجربة n مرة) من التكرارات النسبية المقابلة لقيم X المختلفة كما يلي:

$f(x/G_1)= n_1(x)/n$ $\quad f(x/G_2)= n_0(x)/ n$ $\quad n_1 + n_2 = n$

حيث: $n_1(x)$ هو عدد تكرار مرات النجاح مقابل القيمة x؛ و$n_0(x)$ هو عدد تكرارات مرات الرسوب مقابل القيمة x.

وبعدها يمكننا أن نفترض أن العلاقة بين $P_1(X)$ وX هي علاقة انحدار خطي من الشكل: $P_1(x)= \alpha + \beta x$ ثم نقوم بحساب تقدير لـ α وβ بطريقة المربعات الصغرى، فنحصل على مستقيم يصل بين المجموعتين G_1 وG_2 كما هو مبين في الشكل السابق.

ومنه نحسب القيم النظرية للاحتمالات اللاحقة $P_1(x)$ الواقعة على ذلك المستقيم مقابل كل قيمة لـ x. ثم نقوم بحساب الاحتملات اللاحقة المتممة له: $P_0(x)$ من العلاقة: $P_0(x)= 1- P_1(x)$

وأخيرانقوم بمقارنة $P_1(x)$ مع $P_0(x)$، ونصنف أي عنصر جديد x وفق **القاعدة التالية:**

إذا كان: $P_1(x)\geq P_0(x)$ نصنف x في المجموعة G_1 (مجموعة الناجحين).

إذا كان: $P_1(x)< P_0(x)$ نصنف x في المجموعة G_0 (مجموعة الراسبين).

والخط المستقيم على الشكل السابق يوضح ذلك.

ولكن الشكل يظهر لنا أن جودة التمثيل لذلك المستقيم ضعيفة جدا (لأن قيمة صغيرة)، لذلك كان لابد من البحث عن حل آخر أو نموذج آخر لتمثيل العلاقة بين $P_1(x)$ وX، ومن أجل البحث عن تلك العلاقة، سنحاول الاستفادة من شكل العلاقة ونستبدل $P_1(x)$ بدالة مستمرة جدية ومناسبة، وهو متغير الأرجحية *(odds)*، والذي يعرف بدلالة الاحتمال $P_1(x)$ من خلال العلاقة:

$$Odds(A)= n_1/n_0 = (n_1/n)/ (n_0/n)= P(A)/P(\bar{A})= p/q= p/(1-p). \ldots\ldots(1)$$

والتي تأخذ الشكل التالي:

$$Odds(x)= \frac{P_1(x}{1-P_1(x)} = \frac{\text{احتمال تحقق Y}}{\text{احتمال عدم تحقق Y}} = \frac{n_1}{n_0}$$

حيث أن: $P_1(x)$ هو احتمال أن يأخذ المتغير التابع Y القيمة (1) عند القيمة x، أو احتمال أن ينتمي العنصر x إلى المجموعة G_1، ونكتب ذلك كما يلي:

$$P_1(x)= P(Y=1/x)= P(G_1/x) \ldots\ldots..(1)$$

ولإيجاد علاقة الانحدار بين هذه الأرجحية *odds* والمتغير المستقل X، نفترض أنيما يرتبطان بعلاقة خطية لوغاريتمية، والتي نكتبها كما يلي:

$$\ln (odds)= \left[\frac{P_1(x)}{1- P_1(x)} \right] = \alpha + \beta x$$

وتسمى الدالة اللوغاريتمية اليسرى باسم **$logit[P_1(x)]$**، ويكتب على الشكل التالي:

$$logit[P_1(x)]= \left[\frac{P_1(x)}{1- P_1(x)} \right] = \ln (odds) \quad ..... (2)$$

أي أن الدالة $logit[P_1(x)]$ هي تحويل الاحتمال $P_1(x)$ حسب المعادلة (2) إلى odds، ثم إلى الشكل اللوغاريتمي $\ln[P(x_1)/ (1-P_1(x))]$، وهو دالة مستمرة، تأخذ قيمها في المجال $[-\infty , +\infty]$، لأنه لدينا:

$$0\leq P_1(x)\leq 1$$

ومنه: $0\leq P_1(x)/ [(1-(P_1(x)] \leq +\infty$

وبالتالي فإن: $-\infty \leq \ln[P_1(x)/1-P_1(x)]\leq +\infty$

ومنه يمكننا افتراض أن العلاقة بين الدالة: $logit[P_1(x)]$ والمتغير X هي خطية وتأخذ الشكل:

$$logit[P_1(x)]= \alpha + \beta x= \ln(odds)$$

وبعد حساب القيم العددية لـ $logit[P_1(x)]$ من العلاقة (1) في الأعلى، يمكننا إيجاد التقديرات لـ α وβ بتطبيق طريقة المربعات الصغرى أو الإمكانية العظمى.

والآن نعود للعلاقة:

$$\ln (odds)= \frac{P_1(x)}{1- P_1(x)} = \alpha + \beta x$$

فنجد أنه يمكننا كتابته على الشكل التالي:

$$\frac{P_1(x)}{1- P_1(x)} = e^{(\alpha +\beta x)}$$

نقسم البسط والمقام في الطرف الأول على $P(x_1)$، فنجد أن:

$$\frac{1}{\frac{1}{P_1(x)} - 1} = e^{(\alpha +\beta x)}$$

ثم نقلب الطرف الأول، ونجعل أس العدد e بإشارة (-)، فنجد:

ثم تحويل المعادلة إلى:

$$\frac{1}{P_1(x)} = 1- e^{-(\alpha +\beta x)}$$

$$\frac{1}{P_1(x)} = 1 + e^{-(\alpha +\beta x)}$$

ثم نقلب الطرف الأول، ونجعل أس العدد e بإشارة (-) مرة ثانية، فنجد:

$$P_1(x)= \frac{1}{1 + e^{-(\alpha +\beta x)}}$$

ومنه احتمال أن ينتمي x إلى المجموعة G_1 يساوي:

$$P(G_1/x)= P_1(x)= \frac{1}{1 + e^{-(\alpha +\beta x)}}$$

ولحساب الاحتمال المتمم له، نقوم بحساب $P_0(x)$ من العلاقة:

$$P_0(x)= 1- P_1(x)= 1- \frac{1}{1 + e^{-(\alpha +\beta x)}} = \frac{e^{-(\alpha +\beta x)}}{1 + e^{-(\alpha +\beta x)}} = \frac{1}{1 + e^{-(\alpha +\beta x)}}$$

ومنه احتمال أن ينتمي x إلى المجموعة G_0 يساوي:

$$P(G_0/x)= P_0(x)= \frac{1}{1 + e^{(\alpha +\beta x)}}$$

ولاتخاذ قرار حول انتماء أي عنصر x لإحدى المجموعتين نطبق **القاعدة التالية**:

- إذا كان: $P_1(x)= P_0(x)$ نصنف x في المجموعة G_1 (مجموعة الناجحين).
- إذا كان: $P_1(x)\leq P_0(x)$ نصنف x ضمن المجموعة G_0 (مجموعة الراسبين)

والمنحنى الملتوي في الشكل (3) يوضح ذلك، مع ملاحظة أن G_1 حلت محل G_2، وG_0 حلت محل G_1 من الشكل (2)

ويمكننا تطوير هذه القاعدة بأخذ نسبة الاحتمالين: $P_1(x)/P_0(x)$ فنجد أن:

$$\frac{P_1(x)}{P_0(x)} = e^{\alpha +\beta x} \Longrightarrow \frac{P_1(x)}{1- P_0(x)} = e^{\alpha +\beta x}$$

قاعدة: نصنف العنصر x في المجموعة G_1 إذا كان: $e^{\alpha+\beta x}\geq 1$، ونصنفه في المجموعة G_0 إذا كان: $e^{\alpha + \beta x}\leq 1$.

ويمكن تحويل القاعدة إلى الشكل الخطي بأخذ اللوغاريتم النبيري للطرفين، فنحصل على القاعدة التالية:

قاعدة: نصنف العنصر x في المجموعة G_1 إذا كان: $\alpha x+\beta$ موجب أو معدوم، ونصنف x في المجموعة G_0 إذا كان $\alpha+ \beta x$ سالب تماما.

د. خواص الدالة اللوجستية:

يهدف التحليل اللوجستي في بالأساس إلى تصنيف عناصر المجتمع المدروس إلى مجموعتين أو أكثر (حسب عدد فئات متغير الاستجابة التابع: ثنائي أو متعدد الفئات)، وذلك باستخدام دالة التوزيع الاحتمالي اللوجستي التي وجدناه سابقا:

$$P(X)= \frac{1}{1 + e^{-(\alpha + \beta X)}} \quad -\infty < X<+\infty$$

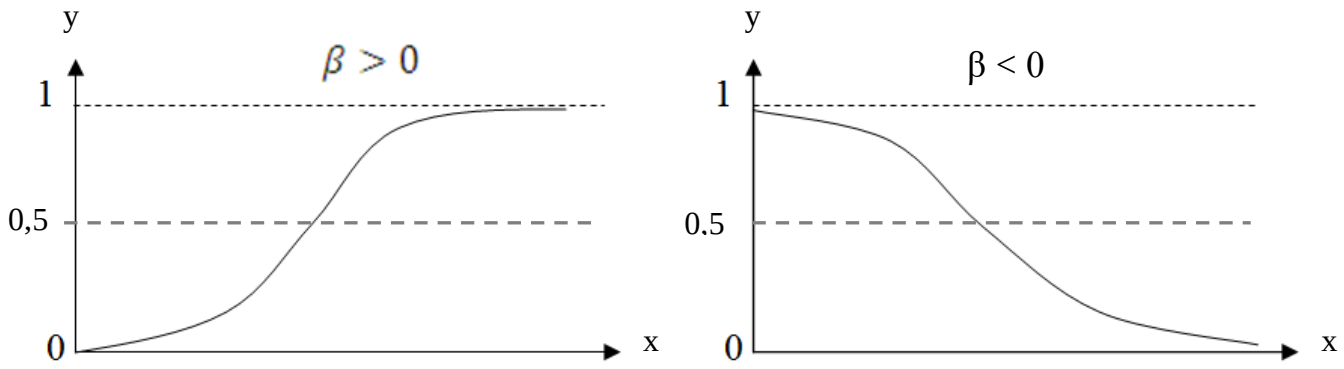
حيث $X= (x_1, x_2, \dots x_P)$ هو شعاع المتغيرات المستقلة؛ $P(x)$ هو الاحتمال المقابل له، ويأخذ الاحتمال قيمه في المجال [0 , 1]:

في حالة متغير مستقل وحيد: $-\infty < x<+\infty$ $P(X)= \frac{1}{1 + e^{-(\alpha + \beta x)}}$

$$P(X)= \frac{1}{1 + e^{-(\alpha+ \beta_1 x_1+ \beta_2 x_2+ .. \beta_P x_P}}$$ في حالة عدة متغيرات مستقلة: $-\infty < x_i < +\infty$

وهو يرسم في المستوي كما يلي (في حالة متغير مستقل واحد):

الشكل (01): منحنى الانحدار اللوجستي



نلاحظ أنه من أن أجل معامل الانحدال اللوجستي β موجبا تكون الدالة متزايدة، ومن أجل β سالبا، تكون الدالة متناقصة. ومن أجل: $x= -\alpha/\beta$، فإن بسط e سيساوي صفر، ومنه $e^0= 1$، ومنه التدالة تأخذ القيمة احتمال 0.50.

هـ. قاعدة التصنيف اللوجستي:

سؤال: كيف يمكن تصنيف مفردات المجتمع إلى مجموعتين (مجموعة المجاح: Y=1، مجموعة الفشل: Y=0)، انطلاقا من قيمة المتغير المستقل أو قيم المتغيرات المستقلة؟

نفرض أن P_1 احتمال تحقق الظاهرة، P_2 احتمال عدم تحققها، ويحسب P_1 بقسمة عدد مفردات المجتمع التي تحقق الظاهرة على العدد الاجمالي لمفردات المجتمع، ويكون: $P_2= 1 - P_1$

ومنه، يمكن تصنيف المجتمع G إلى مجموعتين G_1 وG_2، حيث:

$$Y= P(X_i)= 0,\ i \in G_1 \qquad Y= P(X_i)= 1,\ i \in G_2$$

تكون صيغة النموذج اللوجستي الذي يأخذ قيمه المستمرة في المجال [1 , 0] كما يلي:

$$P_1 > \frac{1}{1 + e^{-(a + b X)}} = P(X)$$

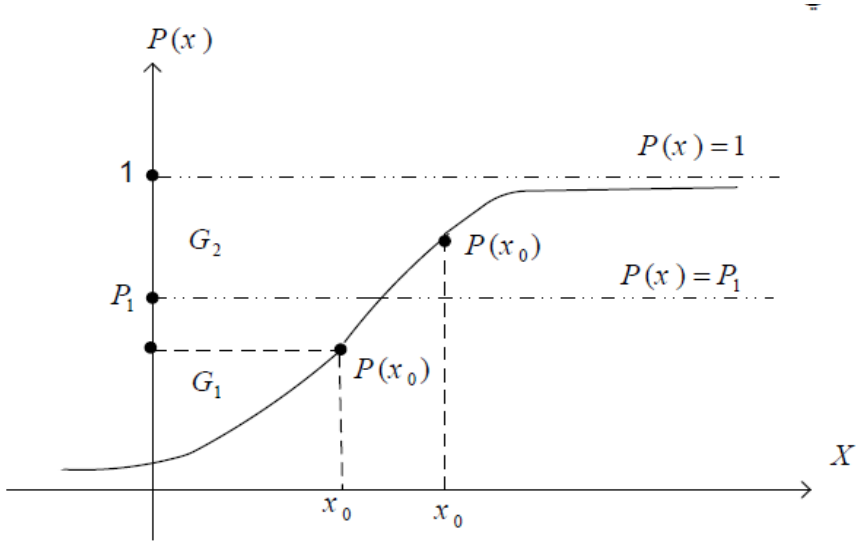
وهكذا تصبح قاعدة التصنيف إلى المجموعتين G_1 أو G_2 كما يلي:

إذا كان لدينا عنصرا جديدا x_0 من المجتمع، فإننا نصنفه في G_1 إذا كانت قيمة المتغير اللوجستي:

$$P(x_0)= \frac{1}{1 + e^{-(\alpha + \beta x)}}$$

أصغر من قيمة الاحتمال P_1، وهو معلوم لأنه عبارة عن نسبة المجموعة G_1 في المجتمع، أما إذا كانت قيمة $P(x_0)$ أكبر من الاحتمال P_1، فإننا نصنف x_0 في المجموعة G_2.

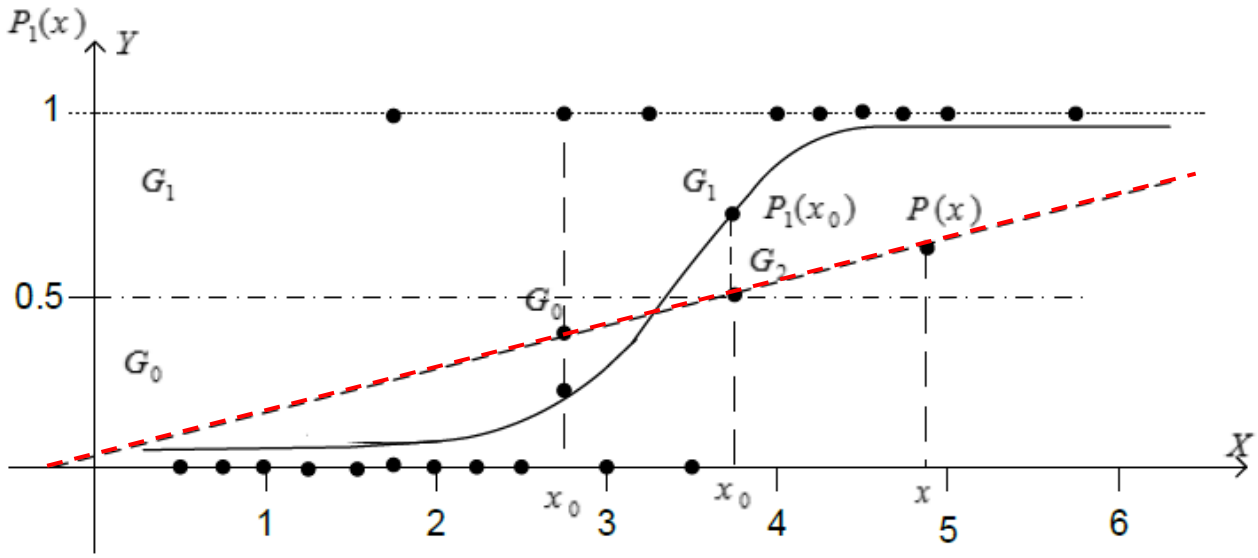
الشكل (2): قاعدة التصنيف للنموذج اللوجستي



و. تقدير معالم نموذج الانحدار اللوجستي:

لحساب معالم نموذج الانحدار اللوجستي α وβ رياضيا نلجأ إلى طريقة الإمكان الأعظم maximum likelihood *method*، وعمليا نلجأ إلى برنامج SPSS. وبالتالي يمكن رسم خط الانحدار اللوجستي، والاستفادة منه في تصنيف المفردات إلى مجموعة نجاح أو فشل كما يوضحه الشكل التالي لانتشار المفردات(البيانات):

الشكل رقم (3): شكل الانتشار للبيانات



نلاحظ من شكل انتشار البيانات أن المتغير المستقل X هو متغير مستمر كمي في هذا المثال، أما المتغير التابع Y فهو متغير منفصل (متقطع) وثنائي القيمة، فهو يأخذ القيمة (Y=1) في حالة النجاح، ويأخذ القيمة (Y=0) في حالة الرسوب، وإذا قمنا برسم النقاط (x_i , y_i)، حيث: P ... 1 =i.

كما نلاحظ (وهو الأهم) أن هذا المنحني يختلف جذرياً عن المستقيم المرسوم على نفس الشكل (خط الانحدار بالأحمر المنقط)، لأنه يقترب بطرفه الأيسر من نقاط المجموعة G_0، ويقترب بطرفه الأيمن من نقاط المجموعة G_1، هو يعطينا بدقة أفضل، احتمال أن ينتمي x إلى G_1 مقابل كل قيمة x من قيم X

ولحساب الاحتمال المتمم له، نقوم بحساب $P_0(x)$ من العلاقة:

$$P_0(x)= 1- P_1(x)= 1- \frac{1}{1 + e^{-(\alpha +\beta x)}} = \frac{e^{-(\alpha +\beta x)}}{1 + e^{-(\alpha +\beta x)}} = \frac{1}{1 + e^{-(\alpha +\beta x)}}$$

ومنه احتمال أن ينتمي x إلى المجموعة G_0 يساوي:

$$P(G_0/x)= P_0(x)= \frac{1}{1 + e^{(\alpha +\beta x)}}$$

ولاتخاذ قرار حول انتماء أي عنصر x لإحدى المجموعتين (مجموعة النجاح، مجموعة الفشل) نطبق **القاعدة التالية**:

- إذا كان: $P_1(x)= P_0(x)$ نصنف x في المجموعة G_1 (مجموعة النجاح).
- إذا كان: $P_1(x)\leq P_0(x)$ نصنف x ضمن المجموعة G_0 (مجموعة الفشل)

والمنحنى الملتوي في الشكل (3) يوضح ذلك، مع ملاحظة أن G_1 حلت محل G_2، وG_0 حلت محل G_1 من الشكل (2)

ويمكننا تطوير هذه القاعدة بأخذ نسبة الاحتمالين: $P_1(x)/P_0(x)$ فنجد أن:

$$\frac{P_1(x)}{P_0(x)} = e^{\alpha +\beta x}$$

قاعدة: نصنف عنصر x في المجموعة G_1 إذا كان: $e^{\alpha+\beta x}\geq 1$، ونصنفه في المجموعة G_0 إذا كان: $e^{\alpha + \beta x}\leq 1$.

ويمكن تحويل القاعدة إلى الشكل الخطي بأخذ اللوغاريتم النبيري للطرفين، فنحصل على القاعدة التالية:

قاعدة: نصنف العنصر x في المجموعة G_1 إذا كان: $\alpha x+\beta$ موجب أو معدوم، ونصنف x في المجموعة G_0 إذا كان $\alpha+ \beta x$ سالب تماما.