

حل سلسلة تمارين رقم (02) حول التحليل اللوجستي

حل التمرين الأول:

1. حساب معدل الشراء عدم الشراء:

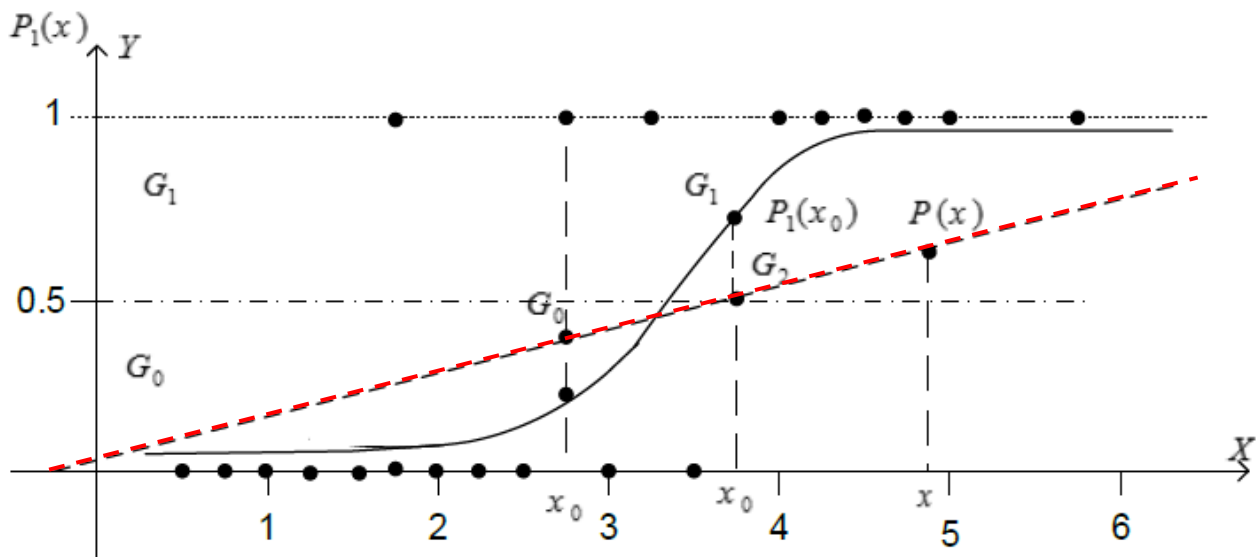
من الجدول (1) في التمرين، نلاحظ أن عدد العملاء الذي قاموا بالشراء من المتجر الإلكتروني: $n_1 = 10$ ، وعدد العملاء الذي تصفو الموقع خلال الأسبوع من دون شراء: $n_0 = 10$ ، وذلك من بين عينة العملاء ($n=20$)، أي أن المعدل العام للشراء:

$$p = n_1 / (n_1 + n_2) = n_1 / n = 10 / 20 = 0.50$$

وبالمقابل فإن المعدل العام لعدم الشراء:

$$q = n_2 / (n_1 + n_2) = n_2 / n = 0.50$$

2. رسم شكل انتشار البيانات:



- المتغير التابع Y فهو متغير منفصل (متقطع) وثنائي القيمة، فهو يأخذ القيمة ($Y=1$) في حالة الشراء، ويأخذ القيمة ($Y=0$) في حالة عدم الشراء؛

- إذا قمنا برسم النقاط (x_i, y_i)، حيث: $i = 1 \dots 20$ (عيل)، نجد أن قيم Y الصفرية (حالات عدم الشراء) تتموضع على محور الفواصل (OX)، وتقع أغلبها نحو اليسار (بالقرب من نقطة المبدأ O)، أما قيم Y المساوية لـ 1 (حالات الشراء)، فتقع على المستقيم ذو المعادلة $Y=1$ ، وأغلبها تتموضع على يمين هذا الخط، كما توجد بعض النقاط المتقابلة في المنطقة الوسطى.

3. تحليل جدول المخرجات:

أ. اختبار Wald:

هو وسيلة لاختبار أهمية متغيرات تفسيرية معينة في نموذج إحصائي. وفي الانحدار اللوجستي لدينا متغير تابع ثنائي ومتغير مستقل مفسر واحد أو أكثر، ولكل متغير تفسيري في النموذج سيكون هناك معلمة مرتبطة به، وهنا يعتبر اختبار Wald أحد الطرق لاختبار ما إذا كانت المعلمات المرتبطة بمجموعة من المتغيرات المستقلة التفسيرية تساوي أصفار أم لا، فإذا كان اختبار Wald ذا دلالة معنوية ($\text{Sig} < 0.05$) بالنسبة لمتغير تفسيري معين أو مجموعة من المتغيرات التفسيرية، فإننا نستنتج أن المعلمات المرتبطة بهذه المتغيرات ليست صفراً، وبالتالي يجب تضمين المتغيرات في النموذج، وإذا لم يكن اختبار والد ذا دلالة، فيمكن حذف هذه المتغيرات التفسيرية من النموذج.

لبيان أهمية معامل الانحدار في نموذج الانحدار اللوجستي نستخدم Wald، والذي له توزيع مربع كاي، نقارن مستوى الدلالة المعنوي لـ test wald (في الجدول) مع معيار المعنوية المعتمدة 0.05، ونلاحظ من الجدول الثاني

للمخرجات، أن عدد ساعات التصفح X ترتبط معنوياً مع احتمال الشراء من المتجر الإلكتروني ($Y=1$)، وذلك لأن قيمة P حسب اختبار Wald تساوي 0.0167، وهي أصغر من قيمة مستوى الدلالة 0.05.

ب. معادلة نموذج الانحدار اللوجستي:

تمثل معادلة نموذج الانحدار اللوجستي أو معادلة لوجيت بالعلاقة بين عدد ساعات التصفح الإلكتروني x ولوعار يتم احتمال الشراء $\text{logit}[P_1(x)]$ ، وهي:

$$\text{logit}[P_1(x)] = \left[\frac{P_1(x)}{1 - P_1(x)} \right] = \alpha + \beta x ; \quad x \in [0, +\infty[\text{ أو } [0, 6]$$

استخر $\text{logit}[P_1(x)] = \left[\frac{P_1(x)}{1 - P_1(x)} \right] = 1.5046x - 4.0777 = 1.5046[x - 2.71]$ هو ما يفيدنا في الإجابة على السؤال الموالي:

ج. معادلة الأرجحية:

ومنه نجد معادلة الأرجحية (النسبة بين احتمال الشراء واحتمال عدم الشراء) بدلالة x ، وهي:

$$\text{odds}(Y) = \frac{P_1(x)}{1 - P_1(x)} = e^{1.546(x - 2.71)}$$

4. استنتاج فترة التصفح التي يكون يتعادل عندها احتمال الشراء وعدم الشراء: من الصيغة:

$$\text{odds}(Y) = \frac{P_1(x)}{1 - P_1(x)} = e^{1.546(x - 2.71)}$$

ومنها نجد أنه إذا كان عدد الساعات $x = 2.71$ فإن: $\text{odds}(Y) = e^0 = 1$ ، وفي هذه الحالة نستنتج أن إمكانية الشراء (odds) تساوي إمكانية عدم الشراء، وأن احتمال كل منهما يساوي 0.50، وفي هذه الحالة (عندما $x=2.71$)، القول: أرجحية الشراء مقابل عدم الشراء هي كما يلي: 1 مقابل 1، وكتب كما يلي (1:1).

5. مناقشة قرارات الشراء من أجل زمن تصفح $x=1$ و $x=6$ بالساعات:

لحساب احتمال الشراء من أجل فترة تصفح x معينة: $P_1(x)$ من النموذج، نجد أن:

$$P_1(x) = \frac{1}{1 + e^{-(-4.0777 + 1.5046x)}} = \frac{1}{1 + e^{4.0777 - 1.5046x}}$$

أ. عندما يكون $x = 1$ نجد أن احتمال الشراء وعدم الشراء يساويان.

$$P_1(1) = \frac{1}{1 + e^{4.0777 - 1.5046(1)}} = 0.0709 \quad P_0(1) = 1 - 0.071 = 0.93$$

وهذا يعني أن احتمال الشراء لمن يتصفح الموقع لمدة ساعة واحدة أسبوعياً ($x=1$)، فإن احتمال الشراء ضعيف جداً (حوالي 7% فقط)، وهو أقل بكثير من احتمال عدم الشراء (حوالي 93%)، لذلك نصنف هذا العميل في المجموعة G_0 (فئة غير المشتريين أو العملاء ضعيفي المردودية للمؤسسة). وتكون نسبة الأرجحية لهذا العميل (الذي يتصفح الموقع ساعة واحدة):

$$\text{Odds}(1) = P_1(1)/P_0(1) = 0.0709/0.93 = 0.076$$

ويمكن كتابتها بالشكل التالي: (1 : 13.11) ≈ 0.076 ؛ احتمال الشراء تقريبا 13 مرة احتمال عدم الشراء.

ب. عندما يكون $x = 6$ نجد أن احتمال الشراء $P_1(6)$ وعدم الشراء $P_0(6)$ يساويان:

$$P_1(6) = \frac{1}{1 + e^{4.0777 - 1.5046(6)}} \approx 0.99 \quad P_0(6) \approx 0.01$$

وهذا يعني أن احتمال شراء العميل الذي يتصفح الموقع مدة ($x=4$) ساعات يحاد يصل إلى 100%، ولا مجال لمقارنته باحتمال عدم الشراء (0.01)، لذلك نصنف ذلك هذا العميل في المجموعة G_1 (مجموعة المشتريين). وتكون نسبة الأرجحية لهذا العميل (الذي يتصفح الموقع 6 سا):

$$\text{Odds}(6) = P_1(6)/P_0(6) = 0.99/0.01 = 99$$

ويمكن كتابتها بالشكل التالي: (1 : 99) ≈ 99.00 ؛ احتمال الشراء تقريبا 99 مرة احتمال عدم الشراء. أي أن العميل بعد تصفحه للموقع لمدة 6 ساعات في الأسبوع فإن أكيد تقريبا سيشتري.

ملاحظة:

بنفس الطريقة يمكن حساب احتمال الشراء والأرجحية بالنسبة لقيم x: 2؛ 3؛ 4 و 5 كما في الجدول التالي:

الجدول رقم (3): قيم التحليل اللوجستي

عدد ساعات الدراسة X	مؤشرات النجاح في الامتحان		
	قيمة ln (odds)	قيمة odds	احتمال النجاح $P_1(x)$
1	-2.57	$0.076 \approx 1 : 13.3$	0.07
2	-1.07	$0.35 \approx 1 : 12.91$	0.26
2.71	0.00	$1 \approx 1 : 1$	0.50
3	0.44	$1.56 : 1$	0.61
4	1.94	$6.69 : 1$	0.87
5	3.45	$32.33 : 1$	0.97
6	4.95	$99.00 : 1$	0.99

6. اقتراحات تحسين مبيعات الموقع الإلكتروني للمؤسسة:

بما أن مشتريات العملاء تتزايد بتزايد فترة التصفح للموقع الإلكتروني (علاقة طردية اثبتتها تحليل الانحدار اللوجستي)، فإن زيادة جاذبية الموقع، تؤدي إلى بقاء العملاء مدة أطول في تصفح منتجات وخدمات المؤسسة، وبالتالي ارتفاع احتمالات الشراء، ويمكن زيادة جاذبية الموقع من خلال:

- جعل تصميم الموقع منظم، من خلال عدد مناسب من العناصر المرئية والمعلومات، وإلا كانت هناك صعوبة في جذب انتبه الزوار، مما يدفعهم للمغادرة بسرعة؛

- جعل المحتوى سهل القراءة، بأن يمزج بين الكتابة والرسومات، الصور والفيديوهات، الأصوات والألوان بشكل مناسب ومتوازن؛

- العناية بطرق التنقل في موقع الويب، حتى يسهل على الزوار العثور على محتوى معين، وذلك باستخدام أسلوب الترتيب الهرمي في عرض النوافذ والروابط المتعلقة بالموقع الإلكتروني؛ وعرض المحتويات الأكثر صلةً أولاً، ثم الأقل صلةً فالأقل.

- جعل الموقع يستجيب ويتكيف مع الأجهزة المختلفة، بحيث يمكن للعملاء الوصول إلى المتجر الإلكتروني عبر أي جهاز مثل الهواتف المحمولة والأجهزة اللوحية وأجهزة الكمبيوتر، كما يبدو الموقع جيداً لزوار الهواتف المحمولة، سواء كان أندرويد Android أو آيفون iPhone أو أي نوع آخر من الأجهزة.

- جعل المحتوى أطول، حتى يشجع على البقاء على الموقع لفترة أطول، من خلال تقديم معلومات وافية عن منتجات وخدمات الموقع، شرط أن تكون مهمة ومفيدة للزوار.

- تقديم معلومات عالية الجودة في الصفحة الأولى للموقع (واجهة المتجر)، حيث تحتوي العروض الجديدة، الإعلانات، التخفيضات.

- التواصل مع العملاء، من خلال نافذة الاقتراحات والتعليقات، الدردشات الحية، وهي قناة اتصال مباشرة وفورية يمكن من خلالها تقديم خدمة عملاء جيدة، كما أنها تقدم بيانات قيمة لتحليل تجربة الزوار.

- توجيه الزوار إلى صفحات أخرى ذات صلة في الموقع، حيث كلما زاد عدد الصفحات التي يصل إليها الزائر، زادت مدة بقائهم في الموقع الإلكتروني، والاعتدال في الروابط الخارجية لأنها ترسل الزوار بعيداً عن الموقع.

حل التمرين الثاني:

يتم تنظيم المعطيات في جدول الطوارئ contingency table، وهو ما يوضحه الجدول التالي، علماً أن: Diseased مصابين؛ Healthy أصحاء؛ Exposed تعرض، Not exposed عدم تعرض للإشعاع

مجموع	أصحاء (H) Healthy	مصابين (D) Diseased	
E= 400	HE= 380	DE= 20	التعرض (E) Exposed
N= 600	HN= 594	DN= 06	عدم التعرض (N) Not exposed

حساب الخطر النسبي:

الخطر المطلق للإصابة بالفيروس في حالة التعرض للإشعاع هو: $DE / VE = 20/400 = 0.05$
والخطر المطلق للإصابة بالفيروس في حالة عدم التعرض للإشعاع هو: $DN / VN = 6/600 = 0.01$
يصعب الحكم على أثر التعرض للإشعاع على الإصابة بالفيروس من عدمه، لأن هناك من تعرضوا للإشعاع ولم يصابوا بالفيروس، وهناك من لم يتعرضوا للإشعاع وأصيبوا بالفيروس، لذا فإحدى الطرق الواضحة لمقارنة المخاطر هي استخدام النسبة بين الاثنين، أي الخطر النسبي:

$$\text{Relative risk} = \frac{DE/E}{DN/N} = \frac{DE/(DE+HE)}{DN/(HE+HN)} = \frac{20/400}{6/600} = 5$$

حساب نسبة الأرجحية:

أما نسبة الأرجحية فهي مختلفة، احتمال الإصابة بالفيروس إذا تعرض الشخص للإشعاع هي:
 $DE/HE = 20/380 \approx 0.053 \approx 5.3\%$
وا احتمال الإصابة بالفيروس في حالة عدم التعرض للإشعاع هي:
 $DN/HN = 6/594 \approx 0.010 \approx 10\%$
ومنه نسبة الأرجحية هي النسبة بين الاثنين:

$$\text{odds ratio} = \frac{DE/HE}{DN/HN} = \frac{20/380}{6/594} \approx 0.053 / 0.010 = 5.3$$

كما هو موضح في هذا المثال، في حالة فيروس نادر مثل هذه، تكون نسبة الخطر النسبي ونسبة الأرجحية هي نفسها تقريباً بحكم التعريف، الفيروس النادر يعني أن: عدد المتعرضين $\approx$ عدد الأصحاء المتعرضين ($E \approx HE$)، و: (عدد غير المتعرضين $\approx$ عدد الأصحاء غير المصابين بالفيروس ($N \approx HN$))، وبالتالي، فإن القواسم في نسبة المخاطرة والأرجحية النسبية هي نفسها تقريباً ($400 \approx 380$ و $600 \approx 594$).