الانحدار اللوجستي

1- المقدمة:

الهدف من الانحداربصفة عامة هو التنبؤ بالمتغير التابع بدلالة متغير او اكثر مستقل ( غير تابع) والانحدار اللوجستي

له نفس الهدف. فان السؤال الذي يمكن طرحه هو: **- ما الفرق بين الانحدار اللوجستي والانحدار الخطي؟**

- الفكرة الأساسية هي شكل المتغير التابع.

**2- أهم الأسس النظرية:**

اسمي نوعي ( أي يأخذ قيمتين فقط Dichotomique - يستخدم الانحدار اللوجستي عندما يكون المتغير التابع ثنائي

) مثل نعم أو لا . ولا يوجد شروط بالنسبة للمتغيرات المستقلة يمكن أن تكون متغيرات فئوية

 .يمكن أن تكون خليط من الاثنين كما او متصلة Catégorielle

- فإذا كان المتغير التابع ثنائي فإننا نستخدم الانحدار اللوجستي الثنائي.

أكثر من فئتين فإننا نستخدم الانحدار اللوجستي متعدد الفئات Y - أما إذا كان

 Régression logistique multinomiale

 مثل مستوى التعليم فاننا نستعمل الانحدار اللوجستي الرتبي ordinal رتبي Y اذا كان المتغير التابع - اما

 Régression logistique ordinale

 ثنائي مثال:Y يستعمل الانحدار اللوجستي الثنائي عندما يكون المتغير التابع

- القبول او الرفض في الالتحاق بالماستر

- البقاء في الجامعة حتى الحصول على الدرجة مقابل عدم الاكتمال

- الموت مقابل الحياة في حالة الاطفال حديثي الولادة

- الاصابة بمرض ما مقابل عدم الاصابة

فاذا كما نلاحظ المتغير يحمل قيمتين فقط يمكن التعبير عنهما بواحد والصفر.

صفر يعني الفشل والواحد يعني النجاح.

فمثلا نريد التنبؤ بقبول الطالب في الجامعة أو الإصابة بمرض ما.

 قبول الطلب يرمز له بالواحد و عدم قبوله يرمز له بالصفر

 عند استعمال الانحدار اللوجستي يجب التأكد من أن البرنامج يقوم بنمذجة النجاح (1)

\* السؤال : لماذا لا نستخدم الانحدار الخطي؟

لان هدا الأخير يستخدم بيانات مختلفة من الشكل :

 تأخذ قيم متصلةY كما هو ملاحظ

|  |  |
| --- | --- |
| التحصيلy | الدافعيةX |
| 2527504340 | 300320500422420 |

فاذا قمنا بثمثيل معطيات هذا الجدول بيانيا

فيمكن ان نرسم خط ينبأ بدقة لقيمة التحصيل

Y=ax+b

من الشكل :

|  |  |
| --- | --- |
| التحصيلy | الدافعيةX |
| 25275043 | 300320500422 |

فاذا قمنا بالتمثيل البياني

فيمكن ان نرسم خط ينبأ بدقة لقيمة التحصيل

 في الانحدار الخطي Y=ax+b

 تمثل مقدار التغير الذي سيحدث في قيمة المتغير التابع عندما يزيد المتغير المستقل درجة واحدة  =a

- في الانحدار اللوجستي قيم المتغير التابع تكون بيانيا نحصل على الشكل

|  |  |
| --- | --- |
| التحصيلy | الدافعيةX |
| 300320500422344 | 001010 |

نلاحظ ان النقاط اما موجودة عند الصفر او موجودة عند الواحد

فلا يمكن تمثيل هذه النقاط بخط مستقيم وهذا يعني ان المعادلة الخطية لا تصلح ويمكن تمثيلهم على شكل

فاذا نقوم بتحويل معادلة الانحدار الخطية الى معادلة لوجستية

Y=ax+b يجب ان تكون دوما موجبة

 دوما موجب ax+b باستعمال الدالة الاسية سوف سيكون

وهي معادلة الانحدار اللوجستي

 هي قيمة الزيادة التي a وهنا لا يمكن القول أن

قيمة الزيادة التي تحدث في قيمته =a

 بدرجة واحدة xاللوغاريتم نسبة الترجيح عندما يزيد

 نسبة الترجيح

- اذا كانت نسبة الترجيح لحدث ما تساوي الواحد فهذا معناه أن احتمال وقوعه يساوي احتمال عدم وقوعه

خصائص هامة لوغاريتم نسبة الترجيح

اذا كان نسبة الترجيح

أي أن احتمالية وقوع الحدث اقل من احتمالية عدم وقوعه فان لوغاريتم نسبة الترجيح تكون سالبة

- اذا كانت نسبة الترجيح اكبر من الواحد أي أن احتمالية وقوع الحدث اكبر من احتمالية عدم وقوعه فان لوغاريتم نسبة الترجيح تكون موجبة

- اذا كان نسبة الترجيح تساوي 1 أي أن احتمالية وقوع الحدث تساوي احتمالية عدم وقوعه فان

 Y وx لوغاريتم نسبة الترجيح تساوي الصفر 0 وهذا يعني لا يوجد علاقة مابين

Spssتطبيق الانحدار اللوجيستي الثنائي في

:SPSS لدينا قاعدة بيانات في برنامج

**المتغيرات هي :**

 Admis = القبول = y المتغير الذي نريد تفسيره وهو

أي القبول في الجامعة الذي يعتمد في بعض البلدان على بعض المتغيرات وهي:

gre=اختبار الذكاء

gpa= المعدل التركيبي

rang =رتبة الجامعة في هذا المثال

 أربع مستويات من 1 إلى 4 وهو متغير رتبي حيث 4 أحسن جامعة و 1 اقل جامعة Rang يأخذ

 **الهدف الأساسي من الانحدار التنبؤ وسوف نهتم إذا بالقبول في الجامعة**

**نضع بالنسبة للمتغير Y = القبول، و يرمز للقبول ب 1 و للرفض ب 0 .**

 القيمة الأعلى أي 1 هي التي Spssوهذا أساسي لأنه عند القيام بالنموذج اعرف باني اتنبا بماذا ففي

سوف يعمل لها النمذجة فاذا عكسنا ذلك بالخطأ فسوف نعمل نموذج للرفض وبذلك التفسير سوف يختلف و بالتالي من الضروري معرفة نحن نعمل نموذج للقبول او للرفض.

**مراحل الاختبار:**

 **Fréquences Descriptives Analyse في البداية و عن طريق**

 rangالذي سوف اتنبا به والمتغير المستقل الرتبي Admis و هذا بهدف التعرف على معطياتنا فندخل المتغير التابع وبذلك اتعرف على عدد الحالات : Ok وبعد ذلك

400 حالة و 00 حالة مرفوضة والرفض 273 حالة (% 68.3) و 127 حالة قبول (% 31.8)

- ثم نقوم بالبحث عن الترابط ما بين المتغيرات التابعة والمستقلة وذلك بهدف البحث عن التعددية الخطية الارتباطية اوما .la multicolinéarité يسمى بالارتباطية الخطية المتعددة اي مشكل

Analyse - corrélation - bivariée

وندخل 4 متغيرات حتى ابحث عن ارتباطهم بالمتغير التابع وارتباطهم فيما بينهم لاتحصل على مصفوفة الارتباط.

\* (نجمة) للمتغيرات المرتبطة عن مستوى الدلالة 0.05 و \*\* عند مستوى دلالة 0.01Spssفيضع

 فنلاحظ ان بين نريد ارتباطاتAdmisفنتحصل على الجدول الاتي بحيث في

 هناك ارتباط 0.184 gre و Admis

 (0.178 ) gpa ونفس الشيئ مع

 ولو كان ضعيف الا انه دال بصغة عامة هناك ارتباط ولكن (-0.243) rang وارتباط سالب مع

 احصائيا . فالدلالة الاحصائية مرتبطة بحجم العينة فاذا انتقلنا الى فنلاحظ ان هناك ارتباط مع وارتباط سالب مع

ولكن ضعيف ليس في مرحلة وفي الأخير ما بين ارتباط سالب ضعيف جدا 0.057 اذا ليس هناك مشكل وبذلك نتوقع ان النموذج لن يفسر بقدر كثير في التباين في وهو مابين في الانحدار الخطي وهو يوجد ايضا في الانحدار اللوجستي

- دون ان ننسى انه ثم رفض 68.3

وكل هذا سوف يظهر في فيكون هناك عمود سوف يحسب وعمود سوف يحسب ( بناء على درجته في هل يكون الطالب في الرفض او القبول)

فمثلا الطالب الذي تحصل على 0 ثم رفضه وفقا لنموذجي اي درجته

يعطى منحنى بياني يوضح العلاقة مابين

وننفذ الانحدار اللوجستي واهم شيئ الجملة الموجودة في الاخير

وهذا يعني ان احتمالية التنبا مرتبطة بعضوية فئة القبول اي ان نعمل برمجة للقبول

ونرجع الى النتائج :

الاول (0) يبدا بنموذج اول وهذا يعني ليس لدينا ولا شئ في النموذج وبما انه ليس لدينا معلومة ثانية فسوف اقوم برفض الكل 273 مرفوض في الواقع و 127 التي هي مقبول في لكن اقوم برفضها في لانه ليس لدينا معلومة ثانية وهذا تامين لي فاذا كانت هناك معلومة اخرى كنت من الموضوع . ويعني هذا النمودج انه قد تم التنبؤ بالمتغير التابع دون وجود اي متغير مستقل ولهدا لا يوجد في هذا النموذج الا المقدار الثابث وهو دال ، وهذا معناه انه يوجد متغيرات اخرى لم ندخلها في النموذج يمكن ان تفيذ في التنبؤ

اي الثابت هنا يمثل الخطا.

فالمتغير الذي ادخلناه هو دال نعم اصغر 0.05 وهذا يعني ان نحول سوف يحسن من النموذج لانه هو دال

ولهذا في الاول وهو واثبت ان قيمته دالة وهو يسمى وهو يشبه اختبار تحليل التباين الذي كان يجري في حالة الانحدار الخطي درجة الحرية هنا تساوي 1 لانه ليس لدينا الا متغير واحد .اذا اضافة يحسن النموذج

ففي الحد

نعلم ان الانحدار اللوجستي هو لهذا لدينا

الدلالة هنا تعني ان وجود يسهم في التنبؤ بالقول بصورة احصائية

لدينا في ما يسمى ب

وهي

في حالة وجود عدة نماذج في المقارنة فهو احسن نموذج فيمكن استعمالها للحصول على احسن نموذج

فالنموذج الافضل هو الذي فيه اقل قيمة

فقيمة تعتمد على قيمته وهو مايسمى ب لوغاريتم الترجيح ولدينا وهو طريقة لحساب في الانحدار اللوجستي

0.032 اي 32 من التباين الموجود في القبول يمكن تفسيره بمعلومتية

 اي جدول التصنيف نلاحظ اننا مازلنا عند 68.3 رغم اننا ادخلنا لم يحدث تحسن في نسبة التنبا الصحيحة مازال النموذج يضع كل الاشخاص في فئة الرفض لان اسهامه ليس واضح

فهل عند ادخال يكون هناك تحسن

الرفض تصنف جدا صحيح لكن القبول صنف خطا فنجد 100 رفض و 00 للقبول

ففي الجدول الموالي الاساسي المتغيرات المعادلة : نلاحظ ان معامل الانحدار هو موجب اي توجد علاقته موجبة بين القبول في الجامعة والمعدل يمكن القول اذا انه اذا ازدادت درجة درجة واحدة فان مقدار لوغاريتم نسبة الترجيح سوف يزيد بهذا المقدار والذي هو 1.051

- لن افسرما هو غير دال

- هي دالة لان اصغر 00.5 فيجب اذا تفسيره عن طريق اخذ

ماذا تعني اذا زادت درجة الفرد عن المتغير المستقل درجة واحد فان نسبة الترجيح هذا الشخص سوف تزداد لتصبح 2.861 اي سوف تتضاعفمرتين و اي يقبل بالكلية ب 2.861 مقابل رفضه

- فاذا ادخلنا بقية المتغيرات ولكن نحدد ان هو متغير وليس متصل ويجب التذكير ان يتعامل مع

على ان

فاذا كان متغير متكون في فئات فهو ياخذ 3 والرابع يعتبرها فئة مرجعية ، والفئة المرجعية هي الفئة الاخيرة

واضع ثم

نلاحظ الجدول

إن 4 صفات 1.2.3.4 فقد تم تحويل 4 صفات إلى 3 حيث 1 يمثل الأول 2 يمثل المتغير الثاني والثالث يمثل المتغير 3 أما الفئة 4 فيها أصفار وهي تعبر عن الفئة المرجعية

كما ألاحظ أن ارتفع شيئا ما إلى 0.099

وفي الجدول نلاحظ ارتفاع 71 بدلا من 68.3 كما نلاحظ 93

ففي فئة القبول 97 صنفوا خطا و 30 صنفوا صح بعد إدخال 3 متغيرات

في جدول

لدينا بالنسبة وهي دالة

وهذا يعني اذا زادت درجة واحدة فان نسبة الترجيح سوف تزيد ب 2.235 اي فرصة الفرد سوف تزيد مرتين وربع تقريبا بوجود متغيرات أخرى

- بالنسبة فرصة الفرد سوف تزيد ب 1.002 فيمكن القول ان فرصة الفرد سوف تزيد 2 عشرة من المائة

- بالنسبة يفسر لدينا المجموعة 4 فنقول اذا الطالب قدم بالجامعة في المرتبة 1 فهذا معناه ان فرصته في القبول سوف تكون 4 أصناف و 7 من عشرة فرصته لو قدم في جامعته في المرتبة 4فرصته او نسبة ترجيحه

نفس الشيء

- فاذا عكست اي اخذت 1 سوف تختلف عوض اخذ

ملاحظة هي رمز التصنيف وليس جامعة 1

- فرصة الفرد سوف تقل ب 49 ( 1 – 0.509 ) اذا قدم في الجامعة في المرتبة 2 عن الجامعة من المرتبة 1 بما ان سالبة في حالة فئة مرجعية الجانعة 1.

النتيجة واحدة ولكن الصيغة تختلف.

ي