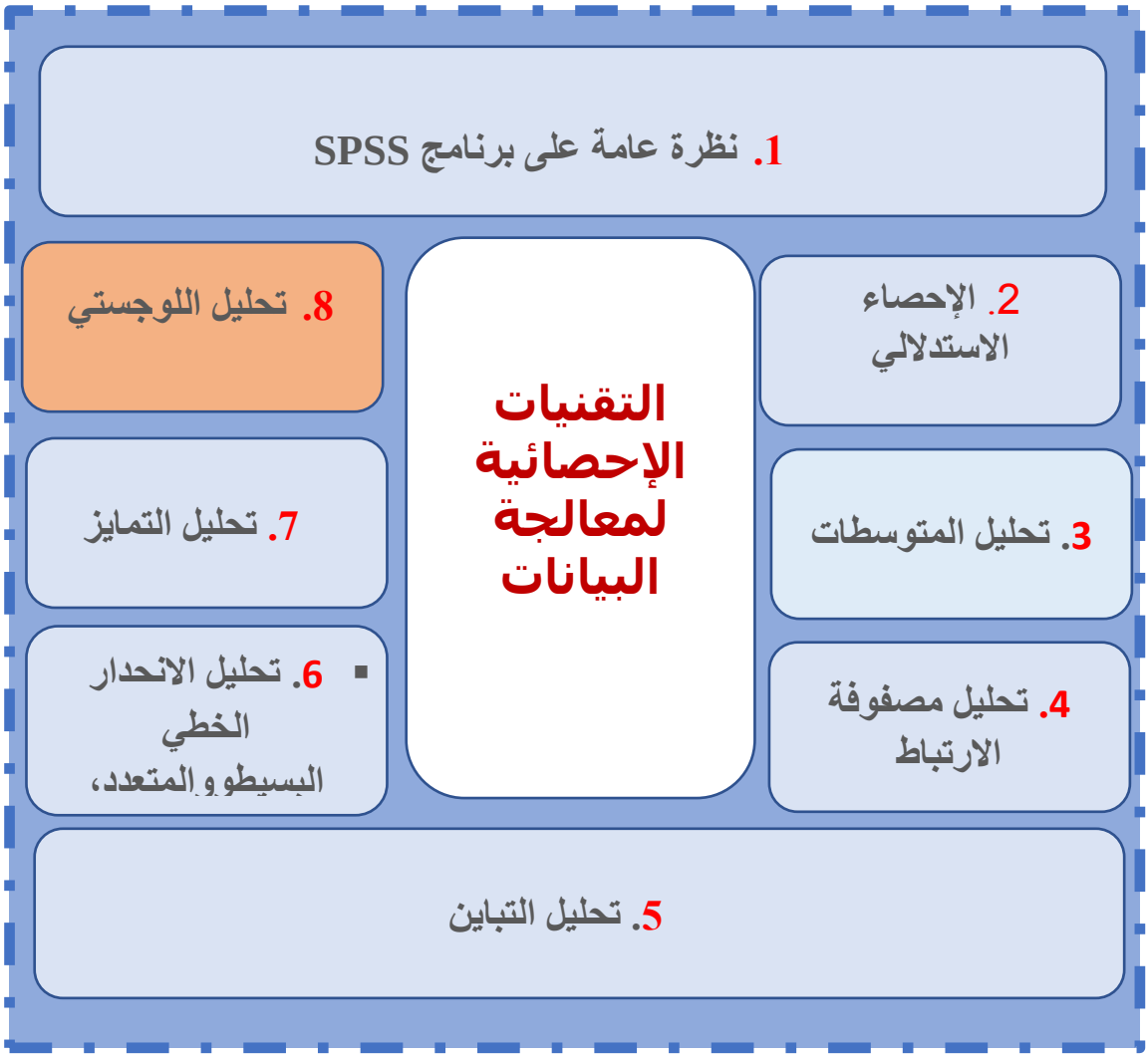
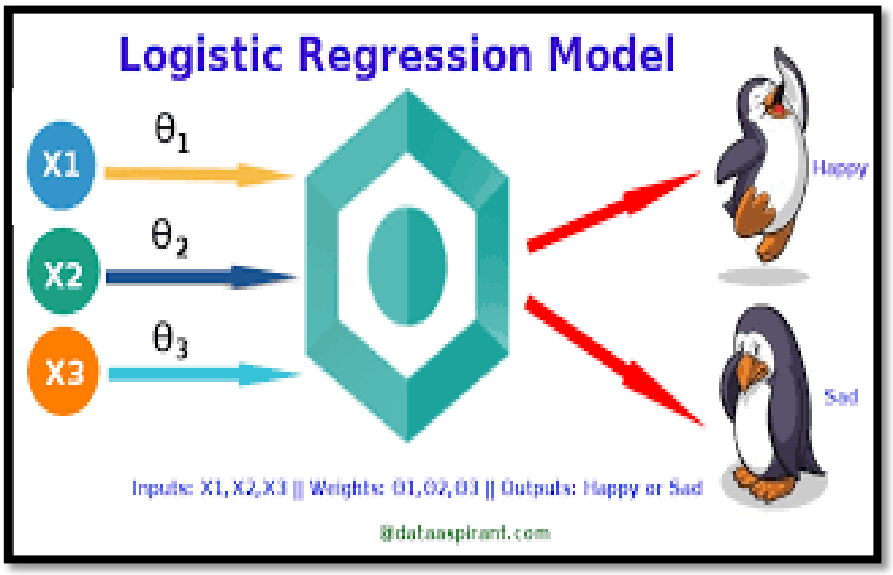

1. نظرة عامة على برنامج SPSS
2. الإحصاء الاستدلالي
3. تحليل المتوسطات
4. تحليل مصفوفة الارتباط
5. تحليل التباين
6. تحليل الانحدار الخطي البسيط و المتعدد،
7. تحليل التمايز
8. تحليل اللوجستي
التقنيات الإحصائية لمعالجة البيانات

الفصل الثامن:
الانحدار اللوجستي



المصدر: google image 8

خطة الفصل

المقدمة
1. الانحدار اللوجستي مقابل الانحدار الخطي
2. الفرضية الصفرية
3. أمثلة عن الانحدار اللوجستي
4. مقدمات منطقية
5. التطبيق العملي
5 النص التطبيق
2.5 وصف قاعدة البيانات
1.5 الاجابة
1.1.5 السؤال رقم 1

خطة الفصل

1.1.5 السؤال رقم 1
1.1.1.5 المنهجية والهدف
2.1.1.5 تفسير البيانات
3.1.1.5 كتلة 0: بداية بلوك
4.1.1.5 الكتلة 1: الطريقة = أدخل
5.1.1.5 اختبار HOSMER و LEMSHOW
6.1.1.5 أهمية المتغيرات المساهمة في تحقيق الظاهرة
7.1.1.5 متغير في المعادلة
8.1.1.5 جدول التصنيف
2.1.5 السؤال 2
1.2.1.5إشارة
2.5 الاستنتاج
6. كيفية عرض نتائج اختبار تحليل الارتباط في المقالات وفي الأطروحات

الأهداف التعليمية للفصل

1. معرفة كيفية بناء الانحدار اللوجستي والشروط والحيثيات المصاحبة
2. معرفة كيفية قراءة نتائج الانحدار اللوجستي وتفسيرها
3. الإحاطة بمختلف المفاهيم المؤسسة للانحدار اللوجستي
4. القيام بتطبيقات على الانحدار اللوجستي
5. التعرف على كيفية عرض نتائج اختبار الانحدار اللوجستي في المقالات العلمية وفي الأطروحات

قاموس المصطلحات

المصطلحات باللغة الإنجليزية	المصطلحات باللغة الفرنسية	المصطلحات باللغة العربية
Linear regression	Régression linéaire	الانحدار الخطي
Logistic regression	Régression logistique	الانحدار اللوجيستي
Interpretation of data	Interprétation des données	تفسير البيانات
Statistical technique	Technique statistique	تقنية إحصائية
Predictive technique	Technique prédictive	تقنية تنبئية
Algorithms	Algorithmes	خوارزميات
Threshold	Seuil	العتبة
Null hypothesis	Hypothèse nulle	الفرضية الصفرية
Qualitative variable	Variable qualitative	متغير كيفي
Quantitative variable	Variable quantitative	المتغير الكمي
Statistical model	Modèle statistique	نموذج إحصائي
Regression model	Modèle de régression	نموذج انحدار

قاموس المصطلحات (تابع)

المصطلحات باللغة الإنجليزية	المصطلحات باللغة الفرنسية	المصطلحات باللغة العربية
Hosmer and Lemeshow test	**Test de Hosmer et Lemeshow**	**اختبار Hosmer و Lemeshow**
Omnibus tests of model coefficients	Tests omnibus des coefficients du modèle	الاختبارات الشاملة لمعاملات النموذج
summaries, processing case summaries et depending variable encoding	Résumés, traitement des résumés de cas et codage des variables dépendantes	ألقاء نظرة على ملخصات الحالة ومعالجة ملخصات الحالة تبعًا لتشفير المتغير
Classification table	Tableau de classification	جدول التصنيف
Contingency and table of Hosmer and Lemeshow test	Contingence et tableau du test de Hosmer et Lemeshow	الطوارئ وجدول اختبار Hosmer و Lemeshow
Variable in the equation	Variable dans l'équation	متغير في المعادلة
Variable not in the equation	Variable absente de l'équation	متغير ليس في المعادلة
Model summary	Résumé du modèle	ملخص النموذج

()

المقدمة

الانحدار اللوجيستي هو تقنية إحصائية تنبئية انطلاقا من بيانات لمجموعة من الملاحظات الحقيقية لظاهرة علمية المراد دراستها. الانحدار اللوجيستي هو تقنية تنبئية، وهو يهدف إلى تكوين نموذج يسمح بتنبؤ أو بتفسير القيم المأخوذة من القيمة الكيفية أو النوعية.

الانحدار اللوجيستي يختبر نموذج انحدار مع أن المتغير التابع ثنائي (يرمز ب 0-1)، أما المتغيرات المستقلة ترتيبية أو مستمرة أو نسبية. ومنه، الانحدار اللوجستي هو نموذج إحصائي يسمح بدراسة العلاقات بين مجموعة من المتغيرات الكمية Xi ومتغير كيفي Y. فهو نموذج خطي عام باستعمال دالة لوجستية كدالة ربط.

الانحدار اللوجستي يسمح بتوقع احتمال أن حدث يحصل على (قيمة 1) أو على قيمة (قيمة 0) بعد تقدير عوامل الانحدار. هذه النتيجة تتغير بين 0 أو 1. لما القيمة المتوقعة هي أكبر من العتبة، الحدث من المحتمل أن يقع، بيد أن القيمة إن كانت أقل من نفس العتبة، الحدث لا يمكن أن يحصل.

يعد الانحدار اللوجستي أحد أكثر خوارزميات التعلم الآلي شيوعًا في التعامل مع مشاكل التصنيف الثنائي، والتي تحتوي على قيمتين لكل فئة، بما في ذلك تنبؤات مثل "هذا أو ذاك" و "نعم أو لا" و "أ أو ب ". الغرض من الانحدار اللوجستي هو تقدير احتمالات الأحداث وتحديد العلاقة بين الخصائص واحتمالات نتائج معينة.

في تحضير وترجمة هذا الفصل تم الاعتماد على المراجع التالية:

Amroune Boudjemaa (2008).
Baillargeon et Rainville (1978)
Darren et Mallery (2008)
Gavard et al. (2008)
Maurer (1999)
Plaisent et al. (2004)
Spiegel (1993)
Zuccaro (2008)

1. الانحدار اللوجستي مقابل الانحدار الخطي[1]

" الانحدار اللوجستي الثنائي يشبه للانحدار الخطي. الفرق الرئيسي بين الانحدار اللوجستي والانحدار الخطي هو أن الانحدار اللوجستي يوفر نتيجة ثابتة، بينما يوفر الانحدار الخطي نتيجة مستمرة.

في الانحدار اللوجستي، النتيجة، مثل المتغير التابع، لديها فقط عدد محدود من القيم الممكنة. ومع ذلك، في الانحدار الخطي، تكون النتيجة مستمرة، مما يعني أنه يمكن أن تحتوي على أي عدد لا حصر له من

[1] https://whatis.techtarget.com/fr/definition/Regression-logistique#:~:text=R%C3%A9gression%20logistique.%20La%20r%C3%A9gression%20logistique%20est%20une%20m%C3%A9thode,important%20dans%20la%20discipline%20de%20l%27%20apprentissage%20automatique.

القيم الممكنة. يُستخدم الانحدار اللوجستي عندما يكون متغير الاستجابة اسميا، مثل نعم / لا، وصواب / خطأ، ونجاح / فشل.

يُستخدم الانحدار الخطي عندما يكون متغير الاستجابة مستمرًا، مثل عدد الساعات والطول والوزن. على سبيل المثال، بالنظر إلى البيانات المتعلقة بالوقت الذي يقضيه الطالب في الدراسة ونتائج امتحاناته، يمكن أن يتنبأ الانحدار اللوجستي والانحدار الخطي بأشياء مختلفة. مع تنبؤات الانحدار اللوجستي، يُسمح فقط بقيم أو فئات محددة. لذلك، يمكن أن يتنبأ الانحدار اللوجستي بما إذا كان الطالب قد نجح أو فشل. نظرًا لاستمرار تنبؤات الانحدار الخطي، مثل الأرقام في النطاق، يمكنهم التنبؤ بنتيجة الطالب على مقياس من 0 إلى 100.

أسفله بعض الأمثلة على أسئلة البحث التي يمكن أن يجيب عليها الانحدار اللوجستي:

- هل عدد ساعات الدراسة ومستوى القلق والجنس تتنبأ بالنجاح أو الفشل؟
- ما هو احتمال تجاوز وزن صحي باتباع أسلوب حياة سيئ؟

2. الفرضية الصفرية

الفرضية الصفرية العامة هي أن مجموعة المتغيرات المستقلة (النموذج) تفشل في تفسير وجود / غياب المتغير التابع بشكل أفضل من نموذج بدون متنبئ "

3. أمثلة عن الانحدار اللوجستي[2]

" مثال 1:

تود شركة التجارة الإلكترونية التي ترسل عروضًا ترويجية باهظة الثمن عبر البريد إلى عملائها معرفة ما إذا كان من المحتمل أن يستجيب عميل معين لتلك العروض أم لا. على سبيل المثال، سيرغبون في معرفة ما إذا كان هذا المستهلك سيكون "مستجيبًا" أم "غير مستجيب". في التسويق، يسمى هذا ميل للاستجابة للنمذجة.

أصبح الانحدار اللوجستي شائعًا بشكل خاص في الإعلان عبر الإنترنت، مما يسمح للمسوقين بالتنبؤ، بنسب مئوية بنعم/لا ، باحتمالية قيام مستخدم الويب بالنقر فوق إعلانات معينة.

مثال 2:

تبني شركة بطاقات الائتمان نموذجًا لتقرير ما إذا كانت ستصدر بطاقة ائتمان للعميل أم لا ومحاولة التنبؤ بما إذا كان العميل سيتخلف عن السداد أم لا بناءً على خصائص مثل الدخل السنوي ومدفوعات بطاقات الائتمان الشهرية وعدد حالات التخلف عن السداد. في اللغة المصرفية، تسمى هذا نمذجة المخاطر الافتراضية.

[2] https://whatis.techtarget.com/fr/definition/Regression-logistique#:~:text=R%C3%A9gression%20logistique.%20La%20r%C3%A9gression%20logistique%20est%20une%20m%C3%A9thode,important%20dans%20la%20discipline%20de%20l%27%20apprentissage%20automatique.

مثال 3:

الرعاية الصحية لتحديد عوامل الخطر للمرض وتخطيط التدابير الوقائية.

مثال 4:

تطبيقات التنبؤ بالطقس للتنبؤ بتساقط الثلوج والظروف الجوية.

مثال 5:

تطبيقات التصويت لتحديد ما إذا كان الناخبون سيصوتون لمرشح معين.

مثال 6:

التأمين، للتنبؤ بفرص وفاة حامل الوثيقة قبل انتهاء السياسة بناءً على معايير مثل الجنس والعمر والفحص البدني.

مثال 7:

المصرفية للتنبؤ بفرص تعثر طالب القرض من عدمه، بناءً على الدخل السنوي، والتخلف عن السداد في الماضي والديون السابقة."

4. مقدمات منطقية[3]

أنواع المتغيرات التي يجب استخدامها:

1) المتغير التابع (التنبؤي): متغير اسمي ثنائي التفرع: يجب أن يكون الأخير متغيرًا ثنائي التفرع اسميا حقيقيًا وليس متغيرًا مستمرًا معاد ترميزه في مجموعتين؛

2) قم بتضمن المتغيرات ذات الصلة: يجب تضمين جميع المتغيرات ذات الصلة في النموذج وتلك التي لا يتم حذفها؛

3) استقلالية الملاحظات والبقايا: لا يمكن للفرد أن يكون جزءًا من مجموعتين على سبيل المثال مع قياسات ما قبل الاختبار اللاحق؛

4) العلاقة خطية بين التحويل اللوجستي بين المتغيرات المستقلة والمتغير التابع.

5) عدم وجود علاقة خطية متعددة كاملة أو عالية: يجب ألا تكون هناك علاقة خطية كاملة، أو علاقة خطية عالية جدًا بين اثنين أو أكثر من المتنبئين. لذلك، لا ينبغي أن تكون العلاقات قوية جدًا بينهم؛

[3] http://spss.espaceweb.usherbrooke.ca/pages/stat-inferentielles/regression-logistique.php

6) لا توجد قيم قصوى للمخلفات أو البقايا: كما هو الحال في الانحدار المتعدد، تؤثر القيم المتبقية المعيارية التي تزيد عن 2.58 أو أقل من -2.58 على معاملات النموذج وتحد من نوعية ملاءمة النموذج؛

7) حجم العينة: يجب أن تكون العينة كافية للسماح بالتحليل. تم اقتراح 10 ملاحظات على الأقل لكل متغير مستقل؛

8) عينة مناسبة للتنبؤات الفئوية: عندما يتم تفسير متغيرة مستقلة مع متغيرة تابعة، يجب ألا تحتوي أي خلية على أقل من ملاحظة واحدة ويمكن أن تحتوي كحد أقصى 20٪ من الخلايا على 5 ملاحظات على الأقل

5. التطبيق العملي

1.5 النص التطبيق

يتم تقديم بيانات درجة الائتمان لـ 323 عميلًا من عملاء البنوك في ملف درجة الائتمان (**credit. sav**). باستخدام التحليل التمييزي، صغ نموذج الفروق بين درجتي الائتمان.

الأسئلة

1) ما هو المتغير الأكثر تمييزًا؟
2) تفسير الأوزان التمييزية؟
3) تقييم ترتيب النموذج؟
4) رتب هذه الحالات الخمس الجديدة؟

بطاقة AMEX الخاصة بالعمر والدفع

المستخدم	الدفع	العمر	بطاقة AMEX
ذوي الياقات البيضاء	أسبوعي	38	نعم
محترف	أسبوعي	27	لا
مدير	شهري	44	نعم
مدير	أسبوعي	33	نعم
فني	شهري	26	لا

2.5 وصف قاعدة البيانات

في هذه المشكلة الثانية، لدينا قاعدة بيانات لـ 323 عميل بنك. هناك مجموعتان حسب نوع البعد:

المجموعة 1: سوء التصنيف الائتماني
المجموعة 2: تصنيف ائتماني جيد.

نريد أن نفهم بشكل أفضل الفرق بين المجموعتين من خلال 4 متغيرات مستقلة:

1. الوظيفة التي يؤديها العميل: 5 فئات (إدارية، مهنية، ذوي الياقات البيضاء، فني، عامل بسيط).
2- الدفع (نوع الراتب): بفئتين (الراتب الأسبوعي والراتب الشهري)
3. عمر العميل: مصنّف إلى 3 فئات (أقل من 25 و 25 إلى 35 وما فوق 35)
4- حيازة بطاقة ائتمان AMEX: من فئتين (العملاء الذين ليس لديهم بطاقة AMEX وأولئك الذين لديهم بطاقة)

باستخدام نفس البيانات من ملف credit card.sav ، قم بعمل انحدار لوجستي.

1. فسر النتائج
2. قارن نتائج التحليل المميز والانحدار اللوجستي

1.5 الاجابة

1.1.5 السؤال رقم 1

فسر النتائج؟

الاجابة:

1.1.1.5 المنهجية والهدف

هناك عدة خطوات ضرورية لتفسير نتيجة الانحدار اللوجستي

أ- : ألقاء نظرة على ملخصات الحالة ومعالجة ملخصات الحالة تبعًا لتشفير المتغير (summaries, processing case summaries et depending variable encoding)

ب- الكتلة 0: بداية بلوك (Beginning Block)

- جدول التصنيف (Classification table)
- متغير في المعادلة (Variable in the equation)
- متغير ليس في المعادلة (Variable not in the equation)

ت- الكتلة 1: الطريقة = أدخل (Method = Enter)

- الاختبارات الشاملة لمعاملات النموذج (Omnibus tests of model coefficients)
- ملخص النموذج (Model summary)
- اختبار Hosmer و Lemeshow (Hosmer and Lemeshow test)
- الطوارئ وجدول اختبار Hosmer و Lemeshow (Contingency and table of Hosmer and Lemeshow test)

- جدول التصنيف (Classification Table)
- متغير في المعادلة (Variable in the equation)

سنجري انحدارًا لوجستيًا لمقارنته بتحليلنا المميز السابق للمشكلة 2. وسنقارن، بعد ذلك، النموذجين وفقًا للطريقتين ونقارن بين الوظائف المتأصلة وقوة تصنيف الوظائف وفقًا للنموذجين.

2.1.1.5 تفسير البيانات

في حالتنا لدينا:

1. متغير ثنائي تابع: سوء الائتمان و الائتمان الجيد

2. 4 متغيرات فئوية مستقلة وهي:

1) أولا. نوع الوظيفة (5 فئات)
2) ثانيا. نوع الدفع (2 فئات)
3) ثالثا. عمر العميل (3 فئات)
4) رابعا. لديه بطاقة ائتمان AMEX (2 فئات)

في ملخص معالجة الحالة لدينا، لدينا 323 موضوعًا للدراسة "SBU" "Subject Business Unit"، منها 51 وحدة SBU بياناتها غير مكتملة، 272 وحدة SBU بها بيانات كاملة، وبالتالي تمثل نسبة 84.2٪ موضوع تحليلنا من الانحدار اللوجستي، انظر الجدول الموجز أدناه.

Récapitulatif de traitement des observations

		N	Pourcentage
Observations non pondérées[a]			
Observations sélectionnées	Incluses dans l'analyse	272	84,2
	Observations manquantes	51	15,8
	Total	323	100,0
Observations non sélectionnées		0	,0
Total		323	100,0

a. Si la pondération est active, consultez la table de classification pour connaître le nombre total d'observations.

مجموعة ترميز متغير تعتمد

Codage de variable dépendante

Valeur d'origine	Valeur interne
Mauvaise	0

Bonne	1

حالتنا المرجعية هي مجموعة الاحتمالات السيئة = 0
ومجموعة نقاط الائتمان الجيدة هي = 1

3.1.1.5 كتلة 0: بداية بلوك

في "جدول التصنيف" لدينا 166 حالة من العملاء الذين لديهم ائتمان سيئ و106 لديهم ائتمان جيد يبلغ إجمالي عددهم 272 عميل.

جدول تصنيف الجدول

Table de classification[a,b]

			Prévisions		
	Observé		Cote de crédit		Pourcentage correct
			Mauvaise	Bonne	
Pas 0	Cote de crédit	Mauvaise	166	0	100,0
		Bonne	106	0	,0
	Pourcentage global				61,0

a. La constante est incluse dans le modèle.

b. La valeur de coupe est ,500

منحدر: اختبر ما إذا كان المنحدر كبيرًا

الفرضية الصفرية H0: βj= Ø "مجموعة فارغة"، وهذا يعني أنها ليست مهمة:
لا يساعد المتغير في تقدير الفروق بين المجموعات بمعنى آخر: لا يساهم في تفسير الاحتمالات والاحتمالات

في هذا السياق ووفقًا للجدول التالي، نقوم باختبار اختبار Walds: اختبار لمعرفة ما إذا كانت المعلمات مهمة

Wj $\mathbf{Wj = (\beta j / E.S\ de\ \beta j)^2}$
Wj = (ميل / خطأ معياري)^2
هذه المنحدرات لها توزيعات لذلك تحتوي على أخطاء متعلقة بالتقدير

الجدول: متغير في المعادلة

Variables de l'équation							
		B	E.S	Wald	ddl	Sig.	Exp(B)
Pas 0	Constante	-,449	,124	13,016	1	,000	,639

توزيع عادي Z ويسمح بتحديد ما إذا كان المتغير مهمًا. يمكنك استخدام اختبار Z ولكن أسهل طريقة هي استخدام نفس GIS

H0: المتغير مهم إذا كان مستوى الأهمية أكبر > 0.05

في هذه الحالة (Sig = 0.000 < α = 0.005) ، نرفض الفرضية الصفرية أن المتغير ليس مهمًا.

4.1.1.5 الكتلة 1: الطريقة = أدخل

الاختبارات الشاملة لمعاملات النموذج

للتأكد من أن تكامل المتغيرات المستقلة الأربعة أدى إلى تحسين نموذجنا مقارنةً بالمتغيرات المستقلة، فإننا نجري الاختبار التالي (**block1**):

الجدول: الاختبارات الشاملة لمعاملات النموذج

Tests composites des coefficients du modèle				
		Khi-carré	ddl	Sig.
Pas 1	Pas	244,791	7	,000
	Bloc	244,791	7	,000
	Modèle	244,791	7	,000

يقارن هذا الاختبار النموذج بدون متغيرات مستقلة والنموذج بمتغيراتنا الأربعة المستقلة.

الفرضية الصفرية التي تتوافق مع هذا الاختبار هي كما يلي:

H0 : $\chi^2=Ø$: المتغيرات المتكاملة لا تضيف شيئًا إلى النموذج ، أي أن النموذج الذي يحتوي على أربعة متغيرات مستقلة ليس مهمًا. نحن هنا نحاول رفض H_0.

وفقًا لاختبار **Block**، نرى أن "Sig = 0.000" ، أي أن فرصنا في أن نكون مخطئين برفض H_0 هي صفر. ثم نرفض H0 وبالتالي يمكننا أن نؤكد أن نموذجنا مهم وأن متغيراتنا التوضيحية الأربعة التي أضفناها في هذا الاختبار قد حسنت نموذجنا.

الآن بعد أن تأكدنا من أهمية نموذجنا، سيكون من المثير للاهتمام تقييم الاختلاف اللوجستي بين متغيراتنا الملحوظة ومتغيراتنا المقدرة.

5.1.1.5 اختبار HOSMER و LEMSHOW

الفرضية الصفرية التي تتوافق مع هذا الاختبار هي كما يلي:

H0 : χ^2=Ø: لا يوجد فرق لوجستي بين المتغيرات الملحوظة والمتغيرات المقدرة ، بمعنى آخر ، ملاءمة نموذجنا ممتازة. يتم قياس ذلك من خلال اختبار Hosmer و Lemshow. في هذا الاختبار نسعى لقبول H_0.

الجدول: اختبار Hosmer و Lemshow

Test de Hosmer et Lemeshow			
Pas	Khi-carré	ddl	Sig.
1	53,094	8	,000

وفقًا للاختبار الذي أجريناه، فإن " « Sig = 0.50 > α=0.05 »" (يكون احتمال قبول H_0 مرتفعًا جدًا). مما يتيح لنا أن نقول إن التوافق بين البيانات الملحوظة والبيانات المستنسخة بواسطة النموذج ليس جيدًا (نحن نقبل فرضيتنا الصفرية) وبالتالي، في حالتنا لا يوجد فرق لوجستي كبير بين المتغيرات الملحوظة والمتغيرات المقدرة. يظهر هذا في جدول الطوارئ لاختبار Hosmerو Lemeshow المعروض أدناه، انظر القيم الملحوظة والقيم المقدرة.

الجدول: جدول الطوارئ لاختبار Hosmer و Lemeshow

Tableau de contingence pour le test de Hosmer et Lemeshow						
		Cote de crédit = Mauvaise		Cote de crédit = Bonne		Total
		Observé	Attendu	Observé	Attendu	
Pas 1	1	20	20,000	0	,000	20
	2	25	26,923	2	,077	27
	3	20	20,507	1	,493	21
	4	21	20,198	0	,802	21

	5	20	20,192	1	,808	21
	6	25	23,796	1	2,204	26
	7	23	20,434	7	9,566	30
	8	11	11,112	11	10,888	22
	9	0	1,466	20	18,534	20
	10	1	1,372	63	62,628	64

6.1.1.5 أهمية المتغيرات المساهمة في تحقيق الظاهرة

Variables de l'équa

		B	E.S	Wald	ddl	Sig.	Exp(B)	Intervalle de confiance 95% pour EXP(B)	
								Inférieur	Supérieur
Pas 1[a]	Type d'emploie			10,611	4	,031			
	Type d'emploie(1)	3,419	1,613	4,491	1	,034	30,542	1,293	721,518
	Type d'emploie(2)	-1,231	1,076	1,310	1	,252	,292	,035	2,404
	Type d'emploie(3)	,044	1,108	,002	1	,968	1,045	,119	9,173
	Type d'emploie(4)	-17,958	6352,810	,000	1	,998	,000	,000	.
	Type de paiement(1)	-4,019	,658	37,261	1	,000	,018	,005	,065
	Age du client	3,019	,551	30,010	1	,000	20,474	6,952	60,302
	Possède une carte de crédit AMEX(1)	,051	,481	,011	1	,915	1,053	,410	2,704
	Constante	-2,270	1,147	3,919	1	,048	,103		

a. Introduction des variables au pas 1 : Type d'emploie, Type de paiement, Age du client, Possède une carte de crédit AMEX.

7.1.1.5 متغير في المعادلة

دعنا نواصل تحليلنا على أي حال نرى الآن مساهمة كل متغير مستقل:

بادئ ذي بدء، سنحاول تقييم أهمية المعلمات المختلفة للانحدار اللوجستي. في الواقع، من المثير للاهتمام معرفة ما إذا كانت متغيراتنا تساهم في شرح فرص واحتمالات وقوع حدثنا، أي: منح تصنيف ائتماني جيد أو تصنيف ائتماني سيئ. بعد ذلك، سنتمكن من الوصول إلى الخطوة الأكثر إثارة للاهتمام في الانحدار اللوجستي، وهي تفسير معاملات الاحتمالات.

الجدول: متغير في المعادلة

- نلاحظ أنه بالنسبة للمتغير المستخدم، فإن **0.05** =α < **0.632** = Sig. ومع ذلك، فإن الفرضية الصفرية التي تتوافق مع هذا الاختبار هي:

H0 : χ^2=Ø أي أن المتغير المستقل لا يساهم في تقدير الفروق بين مجموعتنا.

هنا وفي هذا المتغير، نقبل H_0 ويمكننا، بعد ذلك، أن نأخذ في الاعتبار أن متغير التوظيف لدينا ليس له تأثير كبير على تحقيق حدثنا ولا يساعد في تقدير الاختلافات بين مجموعاتنا.

- نوع متغير الدفع: احتمال رفض H_0 عندما يكون صحيحًا هو صفر (Sig = 0.000). نحن نرفض، إذن، H_0 ويمكننا القول إن نوع الدفع له تأثير كبير على إكمال الحدث (الحصول على تقييم سيء أو جيد).

- وبالمثل، متغير النوع العمري: احتمال رفض H0 بينما يكون صحيحًا هو صفر (Sig = 0,000 < α = 0.005)، فنحن نرفض ، إذن ، H0 ويمكننا القول أن نوع الدفع له قيمة تأثير كبير على تحقيق حدثنا (الحصول على تقييم سيء أو جيد).

نجد أنه بالنسبة للمتغير الذي لديه بطاقة ائتمان AMEX فإن " Sig = 0,915 > α = 0.05". ومع ذلك، فإن الفرضية الصفرية التي تتوافق مع هذا الاختبار المذكورة أعلاه تؤدي إلى قبول H_0 ويمكننا، بعد ذلك، اعتبار أن متغيرنا ليس له تأثير كبير على تحقيق حدثنا "للحصول على تقييم جيد لا يساعد سوء الائتمان أو التصنيف الائتماني السيئ "في تقدير الفروق بين مجموعتنا.

8.1.1.5 جدول التصنيف

الترتيب: الخطوة الأخيرة في تفسير النتائج هي النسبة المئوية لترتيب نموذجنا.

الجدول: جدول التصنيف (أ، ب)

Table de classification[a]

			Prévisions		
		Observé	Cote de crédit		Pourcentage correct
			Mauvaise	Bonne	
Pas 1	Cote de crédit	Mauvaise	164	2	98,8
		Bonne	16	90	84,9
	Pourcentage global				93,4

a. La valeur de coupe est ,500

Classification Table[a,b]

			Predicted		
			cote de crédit		Percentage
	Observed		mauvaise	bonne	Correct
Step 0	cote de crédit	mauvaise	166	0	100,0
		bonne	106	0	,0
	Overall Percentage				61,0

a. Constant is included in the model.

b. The cut value is ,500

ذكر أن الحد الأقصى لفرصة المعيار النسبي = $(272/166)^2 + (272/106)^2 + 15\% = (0.61)^2 + (0.39)^2 + 15\% = 67\%$. يجب أن يكون نموذجنا قادرًا على تصنيف أكثر من 64٪ من الحالات.

Classification Table[a]

			Predicted		
			cote de crédit		Percentage
	Observed		mauvaise	bonne	Correct
Step 1	cote de crédit	mauvaise	165	1	99,4
		bonne	24	82	77,4
	Overall Percentage				90,8

a. The cut value is ,500

صنّف نموذجنا 165 عميلاً في مجموعة "التصنيف الائتماني السيئ"، أي 99.4٪ تصنيف جيد لهذه المجموعة (166/165) و82 عميلاً في مجموعة "التصنيف الائتماني الجيد". أي أن النموذج احتل المرتبة 77.4٪ من العملاء الذين ينتمون إلى هذه المجموعة جيدًا.

بشكل عام، صنف النموذج 90.8٪ من الشركات التي كانت موضوع هذه الدراسة. مع العلم أن عدد الحالات التي كانت موضوع تحليلنا يساوي 272 (وليس 323)، حصلنا على ترتيب جيد = 272/247 SBU = 91٪.

نجح نموذجنا في تصنيف 90.8٪ من العملاء (أكثر من التصنيف الساذج = 67٪)، وبالتالي يمكننا القول إن نموذجنا يتمتع بقوة تنبؤية عالية جدًا ويشرح جيدًا الفرق بين مجموعتنا (التقييم السيئ والجيد) التصنيف الائتماني).

2.1.5 السؤال 2

قارن بين التحليل التمييزي والانحدار؟

الإجابة

من أجل التحليل التمييزي والانحدار، حصلنا على تصنيف جيد جدًا للعملاء في مجموعتنا: التصنيف الجيد والتصنيف السيئ. معدل الترتيب مشابه تقريبًا.

- بالنسبة للتحليل المميز 88٪ = ((141 + 98) / 272)، 84.9٪ من العملاء ينتمون إلى مجموعة الائتمان السيئ و92.5٪ من العملاء ينتمون إلى مجموعة الائتمان الجيد.

- بالنسبة للانحدار اللوجستي 90.8٪ = ((165 + 82) / 272)، 99.4٪ من العملاء ينتمون إلى مجموعة الائتمان السيء و 77.4٪ من العملاء ينتمون إلى مجموعة الائتمان الجيد.

1.2.1.5إشارة

ومع ذلك، في سياق منح القرض، يود البنك تقليل التصنيف السيئ قدر الإمكان: أي أنه يجب أن يكون هناك أقل عدد ممكن من العملاء المصنفين في المجموعة. ذات تصنيف ائتماني جيد بينما يجب تصنيفها في مجموعة التصنيف الائتماني السيئ: هؤلاء هم العملاء الذين لن يتمكنوا من سداد قروضهم بعد حصولهم على تصنيف ائتماني جيد.

- من هذا المنظور: يفضل اختيار التحليل التمييزي "توقع النموذج 25 عميلاً مصنفين على أنهم ينتمون إلى مجموعة التصنيف الائتماني الجيد بينما كانوا في مجموعة التصنيف الائتماني السيئ. بالنسبة للانحدار اللوجستي، يتم تصنيف عميل واحد في هذه المجموعة. لذا فإن الهدف هو تقليل تكلفة الترتيب السيئ. في هذه الحالة، يكون التحليل المميز أكثر ثراءً، نظرًا لأنه تم إجراء تصنيف قوي.

- على نفس المنوال، إذا أخذنا مجموعة درجة الائتمان الجيدة، فإن الانحدار اللوجستي يصنف 24 عميلًا من الائتمان الجيد إلى سوء الائتمان. ومع ذلك، صنف التحليل التمييزي سوء الائتمان 8 في مجموعة الائتمان الجيدة. في هذه الحالة، يكون الانحدار اللوجستي أكثر تنبؤًا بترتيبها.

وجهة نظر أخرى، أهمية المتغيرات:

- في التحليل التمييزي، حددنا المتغيرات التي تميز أكثر المتغيرات بين مجموعتنا: نوع الدفع ونوع العمل وعمر العميل. جميع المتغيرات مهمة باستثناء الذي صاحبه لديه بطاقة AMEX.

- في الانحدار اللوجستي، لا يوجد سوى متغيرين مهمين: نوع الدفع وعمر العميل، وجميع المتغيرات الأخرى لا تساهم بشكل كبير في حدثنا. وعليه فإن أهمية المتغيرات التوضيحية تختلف باختلاف التحليلين وهما الانحدار اللوجستي والتحليل التمييزي.

- بالنسبة للتحليل التمييزي، تمكننا من تصنيف المتغيرات وفقًا لقوتها التمييزية، بينما بالنسبة للانحدار اللوجستي، حددنا المتغيرات التي تساهم في حدثنا. تسمح لنا هذه الطريقة بتقليل عدد المتغيرات في تحليلنا التمييزي.

- بالنسبة للانحدار اللوجستي، تمكننا من تحديد مساهمة المتغير في تحقيق الظاهرة. بالنسبة للانحدار، لا يهمنا الترتيب فقط، بل هناك أيضًا مساهمة واحتمال كل متغير مهم. في الواقع، في

التحليل التمييزي، تمكننا من تفسير الأوزان التمييزية من خلال مقارنتها بمتوسط درجات التمييز لكل مجموعة. ومع ذلك، لم نتمكن من تحديد احتمالية أو فرصة حدوث ذلك (الانتقال من سوء الائتمان إلى الائتمان الجيد).

2.5 الاستنتاج

- أخيرًا، تجدر الإشارة إلى أن الانحدار اللوجستي أقل تقييدًا من التحليل التمييزي "ليس من الضروري التحقق مما إذا كانت المتغيرات مستقلة أنها موزعة وفقًا لتوزيع عادي. هذه الطريقة أكثر مرونة، ومع ذلك، فإن الانحدار اللوجستي حساس للغاية لعلاقة خطية متداخلة متعددة.

- في الواقع وفي النهاية، نتائج التحليل التمييزي والانحدار اللوجستي قريبة.

- التحليل التمييزي، المتغيرات الأكثر تميزًا، لكن الانحدار اللوجستي المتغيرات التي تساهم بشكل كبير في تحقيق الظاهرة: الانتقال من التصنيف السيئ إلى التصنيف الائتماني الجيد.

- تستخدم الملاحظة الرئيسية، التحليل التمييزي، تحليل متوسطات الأشخاص المعنيين بظاهرة الدراسة، لكن الانحدار اللوجستي لا يستخدم التحليل عن طريق الموضوعات.

قائمة المراجع

مراجع رئيسية

- Amroune Boudjemaa (2008). Document sur des travaux pratiques sur les techniques statistiques d'analyse des données quantitatives. Université du Québec de Montréal, Ecole des Sciences de la Gestion, dans le cadre de la préparation de PhD en administration, le document n'est pas publié.

- Darren George et Mallery Paul. (2008). SPSS for Windows step by step. Pearson Education, Inc.

- Gavard Perret Marie-Laure, Gottelland Christophe, Haon Christophe et Jolibert Alain. (2008). Méthodologie de la recherche : Réussir son mémoire ou sa thèse en sciences de gestion. Pearson Education, France

- Zuccaro Cataldo. (2008). Séminaire sur les techniques statistiques sur les données multivariés pour les étudiants de doctorat. Ecole des sciences de la gestion, Université du Québec, document non publié.

مراجع ثانوية

- Baillargeon Gérald et Rainville Jacques. (1978). Statistique appliquée, Tome 2. Les éditions SMG, C. P., Trois-Rivières

- Gavard Perret Marie-Laure, Gottelland Christophe, Haon Christophe et Jolibert Alain. (2008). Méthodologie de la recherche : Réussir son mémoire ou sa thèse en sciences de

- Image 8 google. (2021). Google image. [En ligne] https://www.google.com/search?hl=fr&tbm=isch&source=hp&biw=1088&bih=474&ei=NG4IYOKdJuyflwS8kLuwAQ&q=logistic+regression&oq=logistic+regression&gs_lcp=CgNpbWcQDFAAWABgxxJoAHAAeACAAQCIAQCSAQCYAQCgAQKqAQtnd3Mtd2l6LWltZw&sclient=img&ved=0ahUKEwjih-qciqvuAhXsz4UKHTzIDhYQ4dUDCAc#imgrc=vR0j7p0nF5gVTM (page consultée le 20-01-2021).

- Plaisent Michel, Bernard Prosper, Zuccaro Cataldo et Daghfous Naoufel. (2004). SPSS 12.0 pour Windows: Guide d'auto formation. Presse de l'Université du Québec, Québec, Canada

- Spiegel Murray R. (1993). Statistique: cours et problème. McGraw-Hill Inc, Paris, France.