جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل/كلية العلوم الصرفة
قسم الرياضيات

الانحدار اللوجستي

Logistic Regression

بحث تخرج تقدمت به الطالبة

هدى مجيد مهدي

إلى كلية العلوم الصرفة – قسم الرياضيات جامعة بابل وهو جزء من متطلبات درجة بكالوريوس الرياضيات

اشراف

كريمه عبد الكاظم مخرب الخفاجي

بِسْمِ اللَّهِ الرَّحْمنِ الرَّحِيمِ

وَقُل رَّبِّ زِدْنِي عِلْمًا

[سورة طه /آية 114]

الاهداء

إلى سيّدتي ومولاتي السيّدة فاطمة الزهراء عليها السلام وأهل بيت العصمة والطهارة ، أُهدي هذا العمل القليل المتواضع ، راجياً بذلك التقرُّب إلى الله والى عائلتي وأساتذتي الذين تفضلوا بالإشراف والمساهمة في إنجازه.

الشكر والتقدير

الحمد لله الذي أنار لنا درب العلم والمعرفة وأعاننا على أداء هذا الواجب ووفقنا الى انجاز هذا العمل لكني سأحاول رد جزء من جميلكم ..

أتقدم أولا بالشكر والامتنان الى

كريمه عبد الكاظم مخرب الخفاجي

التي تفضلت بالا شراف على بحثي هذا فجزاها الله خير الجزاء على جهودها والتي وقفت بجانبي وزودتني بالمعلومات اللازمة وتقديم كل ما هو مفيد لإتمام هذا البحث فلها مني كل الشكر والتقدير والعرفان ..

فلكم مني كل الثناء والتقدير ..

المحتويات

المحتويات	الصفحة
الآية القرآنية	أ
الاهداء	**ب**
الشكر والعرفان	**ج**
الخلاصة	
التمهيد	**1**
الفصل الاول : التعريف بمنهج البحث	
1-1- اهمية الانحدار اللوجستي	**2**
1-2- اسباب اختيار فكرة الموضوع	**2**
الفصل الثاني : خصائص الانحدار اللوجستي	**3**
2-1- التمهيد	**3**
2-2- مفهوم الانحدار اللوجستي	**5-3**
2-2-1- انموذج الانحدار اللوجستي	**6**
2-2-2- شروط تطبيق الانحدار اللوجستي	**8-7**
2-2-3- انموذج الانحدار اللوجستي المتعدد	**12-9**
الفصل الثالث : تحليل الانحدار اللوجستي	**13**
3-1-تحليل الانحدار اللوجستي	**15-13**
3-2- الرياضيات وراء الانحدار اللوجستي	**19-16**
اولا : الاستنتاجات	**20**
ثانيا : التوصيات	**21**
المصادر	**22**

الخلاصة

شكل علم الاحصاء يداً ممتدة تقدم خدماتها عبر الزمن الى العلوم الأخرى ورافداً يساهم في تحليل نتائج ما تتوصل اليه الدراسات في شتى المجالات ، تعتبر التجارب المختبرية عنصراً رئيسياً يستند اليه الباحثون في الجانب الحياتي من حيث التحليل والاستنتاج من أجل الوصول الى أبرز العوامل المؤثرة في تحسين كفاءة التجربة وتحديد العوامل المؤثرة على مجريات التجربة وتطويرها والوصول الى النتائج المطلوبة ،.

1- تمهيد

شهدت السنوات الأخيرة كثرة استعمال النماذج الإحصائية الذكية في تحليل البيانات الوصفية لاسيما في المجالات الاجتماعية والزراعية والطبية وغيرها.

أن تحليل الانحدار هو أحد تلك النماذج التي تستخدم في وصف العلاقة بين متغيرين أو أكثر حيث تقسم هذه النماذج الى قسمين الأول النماذج الخطية و الثاني النماذج اللاخطية ، وسوف يتم التركيز في دراستنا هذه على النماذج اللاخطية انموذج الانحدار اللوجستي المتعدد)، حيث بعد أنموذج الانحدار اللوجستي المتعدد من النماذج الإحصائية التي شهدت اهتماما كبيراً وملحوظا في العديد من الدراسات، وتم استخدامه بشكل واسع في التجارب الحياتية اذ يعد من أهم الأساليب الإحصائية التي تدرس تحليل البيانات المصنفة ولاسيما في حالة متغير الاستجابة الذي يرتبط مع المتغيرات المستقلة من النوع الأسمي أو الرتبي ويتألف من مستويين فأكثر. والهدف الأساس من تحليل الانحدار اللوجستي هو إيجاد أفضل توفيق للنموذج، حيث يوضح العلاقة بين المتغير التابع والمتغير المستقل واحد أو أكثروينقسم الموذج الانحدار اللوجستي إلى قسمين Binary Logistic Regression Modeالأول هو أنموذج الانحدار اللوجستي الثنائي و يستخدم في حالة كون متغير الاستجابة مؤلف من مستوين فقط والثاني هو أنموذج الانحدار اللوجستي المتعدد (Multiple Logistic Regression Mode)يستخدم في حالة كون متغير الاستجابة مؤلف من ثلاثة مستويات فأكثر، وسوف يتم التركيز على النوع الثاني من أنموذج الانحدار اللوجستي المتعدد.

لذا سلطنا الضوء في بحثنا المتواضع هذا على نموذج الانحدار اللوجستي لأهميته... بالتعرف عليه وأهميته وأسباب استخدامه وشروط استخدامه وبعض أنواعه وذكر بعض الأمثلة الواقعية

الفصل الأول
التعريف بمنهج البحث

1-2 – اهمية الانحدار اللوجستي

1. تحليل العلاقة بين المتغيرات: يتيح الانحدار اللوجستي الرياضي فهم العلاقة المحتملة بين المتغير التابع الثنائي والمتغيرات التفسيرية. يمكن استخدامه لتحديد مدى تأثير المتغيرات المستقلة على احتمالية حدوث الحدث أو النتيجة المرتبطة بالمتغير التابع.
2. تنبؤ النتائج: يمكن استخدام الانحدار اللوجستي الرياضي للتنبؤ بالنتائج المحتملة بناءً على القيم المتغيرة للمتغيرات التفسيرية. قد يساعد هذا في اتخاذ القرارات المستندة إلى التوقعات وفهم العوامل المؤثرة في النتائج المرتقبة.
3. تقييم الأثر النسبي: يسمح الانحدار اللوجستي الرياضي بتقييم الأثر النسبي لكل متغير تفسيري على المتغير التابع. يمكن تحديد أي المتغيرات المستقلة لها تأثيرات قوية ومعنوية على النتائج والتركيز عليها في الدراسات المستقبلية.
4. فهم العلاقات غير الخطية: يمكن للانحدار اللوجستي الرياضي التعامل مع العلاقات غير الخطية بين المتغيرات. قد يتم تحقيق ذلك من خلال استخدام تحويلات البيانات أو إضافة متغيرات تفسيرية تعكس العلاقات غير الخطية.

بشكل عام، يعتبر الانحدار اللوجستي الرياضي أداة مهمة في تحليل البيانات وفهم العلاقات بين المتغيرات ثنائية القيمة. يوفر نموذجًا إحصائيًا قويًا للتنبؤ والتحليل واتخاذ القرارات في مجالات متنوعة.

1-3- أسباب اختيار فكرة الموضوع

هناك عدة أسباب لاستخدام الانحدار اللوجستي الرياضي في التحليل الإحصائي، وتشمل:

1. التعامل مع المتغيرات الثنائية: الانحدار اللوجستي الرياضي يعتبر أداة فعالة لتحليل المتغيرات الثنائية (مثل نجاح/فشل، نعم/لا)، حيث يمكن استخدامه لفهم العلاقة بين هذه المتغيرات والمتغيرات التفسيرية الأخرى.
2. التعامل مع الاحتماليات والنسب: الانحدار اللوجستي الرياضي يعتمد على توزيع لوجستي، وهو توزيع احتمالي يمكن استخدامه لتقدير الاحتماليات والنسب فيما يتعلق بالمتغيرات التابعة والمتغيرات التفسيرية.

3. تـوفير تحليـل تـوقعي: يمكـن اسـتخدام الانحـدار اللوجسـتي الرياضـي للتوقـع والتنبـؤ بالنتـائج المحتملـة بنـاءً علـى القـيم المتغيـرة للمتغيـرات التفسـيرية. يسـاعد هـذا فـي اتخاذ القرارات وتقدير المخاطر والتخطيط للأحداث المستقبلية.

الفصل الثاني
خصائص الانحدار اللوجستي

2-1-التمهيد

مع التطور السريع للعلوم والتكنولوجيا، هناك المزيد والمزيد من الحالات التي تحتاج إلى التعامل مع البيانات عالية الأبعاد في العمل البشري والإنتاج. في هذه الحالات يمكننا استخدام طريقة اختيار المتغير لحلها. سواء كانت مشكلة الانحدار أو التصنيف، يتم تحقيق طريقة التحليل التقليدية باستخدام الانحدار التدريجي مع معيار المعلومات. تثبت هذه الممارسة أن طريقة التحليل التقليدية عملية، ولكن هناك العديد من أوجه القصور في نفس الوقت. ليس من المستقر استخدام طريقة الانحدار التدريجي لتحديد المتغيرات، وعندما يكون هناك العديد من المتغيرات البديلة للمعلمات، يكون عبء عمل الحساب ضخمًا، مما يؤدي إلى خطأ حسابي كبير، لا يمكنه تلبية متطلبات الحل.

2-2- مفهوم الانحدار اللوجستي

تحليل الانحدار اللوجستي هو أسلوب إحصائي متعدد المتغيرات للتصنيفات. كتحليل تمييزي تمثيلي طريقة لتصنيف مجموعة من الملاحظات إلى فئات محددة مسبقًا فيما يتعلق بعدة متغيرات.

يعتمد على واحد أو أكثر من المتغيرات التنبؤية (المستقلة أو التوضيحية). الاحتمالات التي تصف النتيجة المحتملة تم تصميم تجربة واحدة، كدالة للمتغيرات التوضيحية، باستخدام وظيفة لوجستية بسبب عدم الافتراض فيما يتعلق بتوزيع المتغيرات المتوقعة، يعتبر الانحدار اللوجستي قويًا نسبيًا ومرنًا وسهل الاستخدام إنه يفسح المجال لتفسير ذي معنى. في الممارسة العملية، عادة ما يستخدم الانحدار اللوجستي كمصنف، أي اللوجستي ، مصنف الانحدار، للتصنيف الثنائي أو متعدد المتغيرات الاحتمالي. معيار اختيار الفئة في الخدمات اللوجستية يقع الانحدار على أعلى احتمال يتم إنشاؤه من دالة لوجستية [1]

الوظيفة اللوجستية (السيجمويدية): الوظيفة اللوجستية هي وظيفة غير خطية تُستخدم في الانحدار اللوجستي الرياضي لتحويل القيم المستمرة إلى احتماليات. صياغتها الرياضية هي:

$$P(y=1|x) = 1 / (1 + e^{(-z)})$$

حيث:

P(y=1|x)هو الاحتمالية المشروطة لحدوث الحدث (النجاح) مع القيمة المعطاة للمتغيرات التفسيرية e .(x) هو العدد الأساسي الطبيعي z.هو المعادلة الخطية التي تتضمن المتغيرات التفسيرية ومعاملات النموذج.

الوظيفة اللوجستية :تقوم بتحويل القيم المستمرة إلى نطاق بين 0 و1، حيث يمثل 0 الاحتمالية الدنيا و1 الاحتمالية العالية.

هذه بعض القواعد الأساسية حول الانحدار اللوجستي الرياضي:.

1. قواعد التصنيف: في الانحدار اللوجستي الرياضي، يتم تحديد قواعد التصنيف بناءً على قيمة احتمالية الحدث المحسوبة من الوظيفة اللوجستية. عادةً، يتم تعيين قيمة قطع الاحتمالية(threshold)، مثل 0.5، للتمييز بين الفئتين الممكنتين للمتغير التابع. إذا كانت قيمة الاحتمالية أعلى من قطع الاحتمالية، يتم تصنيف الحدث كـ 1، وإلا يتم تصنيفه كـ 0.
2. تعامل مع المتغيرات التفسيرية: يمكن استخدام المتغيرات التفسيرية في الانحدار اللوجستي الرياضي بصورة متعددة، مثل المتغيرات الرقمية والمتغيرات الثنائية والمتغيرات الفئوية. يتم تحويل المتغيرات الفئوية إلى متغيرات وهمية (dummy variables)لاستخدامها في النموذج. يتم تضمين المتغيرات التفسيرية في المعادلة اللوجستية لتحديد تأثيرها على الاحتمالية
3. العينات وحجم العينة: يجب أن يكون لديك عينة كافية من البيانات لتدريب وتقييم النموذج اللوجستي. حجم العينة الكافي يعتمد على عدة عوامل، مثل عدد المتغيرات التفسيرية وعدد الحالات الناجحة والفاشلة في النموذج. عادةً ما يكون لزيادة حجم العينة تأثير إيجابي على أداء النموذج وقدرته التنبؤية.
4. التعامل مع الانحدار اللوجستي غير الخطي: في بعض الحالات، قد يحدث الانحدار اللوجستي غير الخطي، حيث لا يمكن تمثيل العلاقة بين المتغيرات التفسيرية والاحتمالية بشكل خطي. في مثل هذه الحالات، يمكن استخدام تقنيات متقدمة مثل تحويلات البولينوميال أو الانحدار اللوجستي العائلي (Generalized Additive Logistic Regression)للتعامل مع العلاقات غير الخطية.
5. التعويضات ومقاومة الضغط: قد يواجه الانحدار اللوجستي التحديات مثل التعويضات(collinearity) بين المتغيرات التفسيرية ومشكلة الضغط (overfitting)عندما يكون لديك عدد كبير من المتغيرات وعدد قليل من

الحـالات. يمكـن اسـتخدام تقنيـات مثـل اختيـار المتغيـرات وتحسـين النمـوذج للتغلـب على هذه المشكلات وتحسين أداء النموذج

تحديد احتمالية الإصابة بنوبة قلبية – يستخدم الباحثون الطبيون البيانات لفهم العلاقة بين المتغيرات المتوقعة لتقدير ما إذا كان الفرد سيصاب بنوبة قلبية أم لا .تخبر نتائج النموذج الباحثين بالضبط كيف تؤثر التغيرات في التمارين والوزن (المتغيرات المتوقعة) على احتمال إصابة فرد معين بنوبة قلبية .يتم استخدام نموذج الانحدار اللوجستي المجهز هنا .وهذا يقدم لنا صورة واضحة عن أهمية علم البيانات في مهنة الطب وتطبيق المعادلة

فهم إمكانية الحصول على القبول في إحدى الجامعات – يرغب مجمعو طلبات التقديم في الكليات في تحديد كيف تساعد المتغيرات مثل درجات CGPA و GMATو TOEFLفي تحديد احتمالية القبول في جامعة معينة .ولهذا السبب، يقوم المجمعون بإجراء انحدار لوجستي لفهم العلاقة بين متغيرات التوقع واحتمال الحصول على القبول.

Gmailوصناديق البريد الوارد الأخرى التي تحدد "رسائل البريد الإلكتروني العشوائية – "أحد الأمثلة الأكثر وضوحًا لكيفية عمل ذلك، هو التصفية التي تقوم بها صناديق البريد الإلكتروني .يتم تحديد ما إذا كانت اتصالات البريد الإلكتروني ترويجية/رسائل غير مرغوب فيها من خلال فهم متغيرات التوقع وتطبيق خوارزمية الانحدار اللوجستي للتحقق من صحتها ، الآن بعد أن فهمنا الرياضيات الأساسية وراء الانحدار اللوجستي وكيف تتصرف الدالة اللوغاريتمية، إلى جانب الخطوات التي يجب أن نأخذها في الاعتبار أثناء التعامل مع مجموعة البيانات مع الانحدار اللوجستي، كخطوة تالية، سوف نتعلم كيف يمكننا تنفيذ هذه الخوارزمية في بايثون، وكيف يمكن أن تولد نتائج إيجابية.

ويعتبر الانحدار اللوجستي نموذج إحصائي يستخدم عادةً لنمذجة متغير تابع ثنائي بمساعدة الوظيفة اللوجستية .اسم آخر للوظيفة اللوجستية هو وظيفة السيني ويعطى بواسطة:

$$F(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

تساعد هذه الوظيفة نموذج الانحدار اللوجستي في ضغط القيم من (−k,k) إلى (0,1)يستخدم الانحدار اللوجستي بشكل رئيسي لمهام التصنيف الثنائي؛ ومع ذلك، يمكن استخدامه لتصنيف متعدد الفئات السبب وراء الانحدار" اللوجستي هو أنه مثل الانحدار الخطي، يبدأ الانحدار اللوجستي من معادلة خطية .ومع ذلك، تتكون هذه المعادلة من احتمالات السجل التي يتم

تمريرها أيضًا من خلال دالة سينية تضغط على ناتج المعادلة الخطية إلى احتمال يتراوح بين 0 و 1. ويمكننا تحديد حدود القرار واستخدام هذا الاحتمال لإجراء مهمة التصنيف.

2-2-1- أنموذج الانحدار اللوجستي

Logistic Regression Model

أنموذج الانحدار اللوجستي هو أنموذج احصائي شائع ينتمي إلى نماذج الانحدار اللاخطية الذي يمتاز بالمرونة العالية، ويعرف بانه أحد الأساليب الإحصائية المهمة التي تستخدم للتنبؤ باحتمالية وقوع حدث ما وذلك من خلال ملاءمة البيانات والمعلومات على المنحى اللوجستي. حيث يعمل الانحدار اللوجستي عادة لوصف طبيعة العلاقة بين الاستجابة وبين متغير أو مجموعة من المتغيرات المستقلة. لذلك يعتبر أنموذج الانحدار اللوجستي حالة خاصة من نماذج الانحدار الاعتيادي بسبب طبيعته الاسمية التي يحملها متغير الاستجابة مثل الحياة الموت). ويستخدم أنموذج الانحدار اللوجستي بشكل واسع في تحليل البيانات والتطبيقات الإحصائية. دم هذا النوع عندما يأخذ متغير الاستجابة قيمتين وهما (0.1) وكذلك الانحدار اللوجستي و هو امتداد النموذج الانحدار (Multinomial Logistic Regression) متعدد الاستجابة اللوجستي الثنائي عندما يقع المتغير التابع في أكثر من فئتين، أما الانحدار اللوجستي الرتبي أو يدعى الرتبوي (Ordinal Logistic Regression) ويستخدم هذا النوع من الانحدار في حالة متغير الاستجابة فيكون متغيراً رتبياً سوف نقتصر في العرض النظري على الانحدار اللوجستي متعدد الاستجابة.

2-2-2- مجالات استعمال الانحدار اللوجستي :

1- يهدف الانحدار اللوجستي الى التنبؤ وشرح قيم متغيرات كيفية، حيث يعتبر أفضل أسلوب احصائي في الدراسات التي تهدف لبناء نماذج تنبؤية.

2- يستخدم الانحدار اللوجستي في المجال الصحي والطبي لذلك فهو يستعمل أيضا في مجالات حياتية كثيرة ويشاع توظيفه بشكل واسع في الطب والعلوم الاجتماعية.

2-2-3- شروط تطبيق الانحدار اللوجستي.

يرتكز أنموذج الانحدار اللوجستي على مجموعة من الافتراضات التي تعتبر شروطاً أساسية ويجب توفيرها قبل تطبيق وهي:

1- يجب أن يكون المتغير التابع متغير اسميا ثنائيا بالنسبة لصنف الانحدار اللوجستي ثنائي الاستجابة، أو يكون تصنيفي بالنسبة للانحدار اللوجستي المتعدد، أو قد يكون رتبياً بالنسبة للانحدار اللوجستي الرتبي

2- يجب أن يكون أنموذج الانحدار اللوجستي في المستوى المئوي أو النسبي أو التصنيفي أو الترتيبي، محتويا على أكثر من متغير مستقل.

3- يجب أن يسلك متغير الاستجابة سلوكا احتمالياً واحداً لكل مشاهدة بمعنى لا يمكن أن تكون المشاهدة في فئتين في الوقت نفسه.

4- يجب إن تكون هنالك علاقة بين المتغيرات المستقلة النسبية والتحويل اللوغاريتمي للمتغير التابع ويظهر هذا الشرط عند تطبيق أنموذج الانحدار اللوجستي.

5- يجب ألا توجد هنالك قيم شاذة في المتغيرات المستقلة ويمكن أن تحدد القيم الشاذة من خلال spssالمتوفر على برنامج test mahala Nobis اختبار

2-2-4-أنواع الانحدار اللوجستي:

ينقسم الانحدار اللوجستي إلى ثالثة نوعين من النماذج وهي: أنموذج الانحدار اللوجستي ثنائي الاستجابة.

Binary Logistic Regression

يعرف أنموذج الانحدار اللوجستي الثنائي على أنه أحد نماذج الانحدار اللاخطية حيث ببنى على فروض أساسية لذلك فإن متغير الاستجابة الذي نهتم بدراسته يتبع توزيع برنولي (Bermoulli)باحتمال نجاح عندما (1) و احتمال فشل 1- عندما (0) بذلك تكون صيغة دالة الكثافة الاحتمالية تكتب كالاتي

$$P(y_i \backslash X_i) = [\pi (X_i)]^{y_i} [1-\pi(X)]\ \pi(Xi)]. ^{1- y} \ldots..$$

سيكون (1) فان y = وعند تعويض قيم 0.1

$$P(Y)=\begin{cases} \pi_i & \text{عند حدوث الاستجابة } Y=1 \text{ عندما} \\ 1-\pi_i & \text{عند عدم حدوث الاستجابة } Y=0 \text{ عندما} \end{cases}$$

ويمثل احتمال الفشل 1- πi

وعندما yi=1

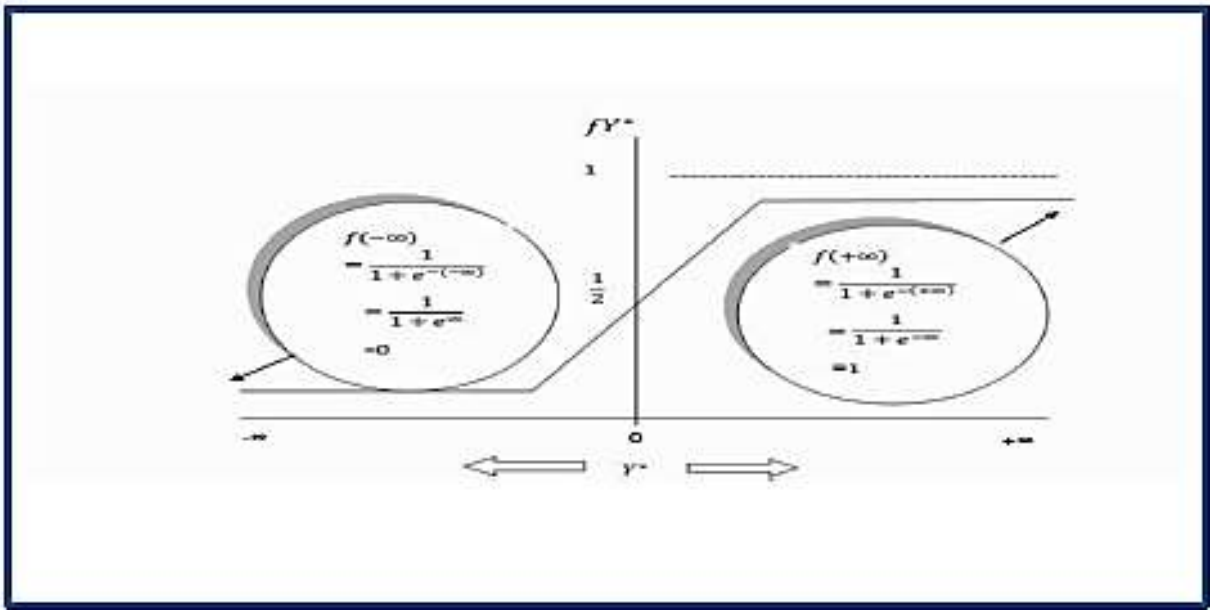
$$\pi i = \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}}$$

وعندما yi=0 نحصل على

$$1 - \pi i = 1 - \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}}$$

$$1 - \pi i = \frac{1}{1 + e^{B_0 + B_1 X_1}}$$

ويمكن عرض الدالة اللوجستية بالشكل الاتي :



الشكل (1) شكل الدالة اللوجستية

من الشكل أعلاه ويتضح بان الدالة اللوجستية (*y) هي دالة احتمالية تكون محصورة ضمن القيم (0.1)

2-2-5 أنموذج الانحدار اللوجستي المتعدد

Multinomial Logistic Regression

يعتبر أنموذج الانحدار اللوجستي متعدد الاستجابة من النماذج الاحصائية ذات الأهمية الكبيرة في تحليل البيانات المصنفة، حيث يستخدم بصورة عامة في حالة إذا كان متغير الاستجابة يعود المتغيرات التي تكون من النوع الاسمي (Nominal) أو الرتبي (Ordinal)والمكونة من تصنيفين أو مستويين في أكثر .ويعد أنموذج الانحدار اللوجستي المتعدد الاستجابة امتداداً بسيطا لأنموذج الانحدار اللوجستي الثنائي الاستجابة. ويعتمد أنموذج الانحدار اللوجستي متعدد الاستجابة بصورة اساسية على التوزيع المتعدد المحدد.

$$P_r(r_{i1}, r_{i2}, \ldots\ldots, r_{ij}) = \binom{Ni}{r_{ij}, \ldots\ldots, r_{ij}} \pi_{i1}^{r_{i1}} \ldots \pi_{ij}^{r_{ij}}$$

نماذج دوال التحويل الخطى :

Linear Transformation Function

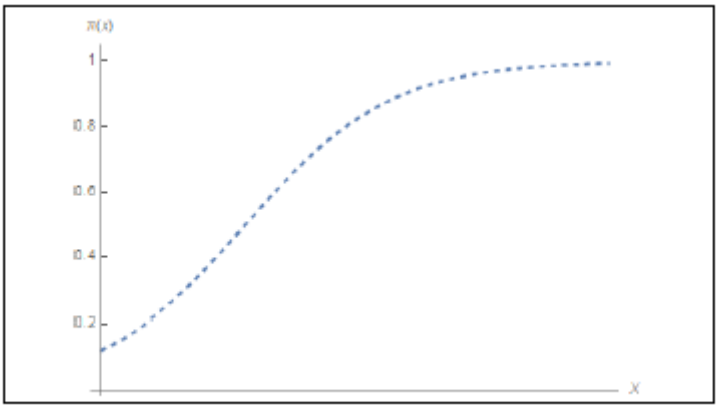
هنالك عدة تحويلات يمكن استخدامها لتحويل أنموذج من الانحدار اللاخطي الى الانحدار الخطي، إن الغاية الأساسية من اجراء عملية التحويل هو لكي يتغير شكل التوزيع التكراري ليلائم الافتراضات النظرية في أنموذج، وهنالك عدة تحويلات يمكن استخدامها لتحويل

أنموذج من الانحدار اللاخطي الى الانحدار الخطي، فقد تم استخدام أنموذج (Logit) لتحويل أنموذج اللوجستي الى معادلة خطية يمكن استعمالها والحصول على قيم المعلمات.

2-2-6-أنموذج لوغاريتم النسبة المضافة

Models of odds Ratio(logit)

من المعلوم إن طبيعية العلاقة التي تربط المتغير متعدد الاستجابة مع مجموعة من المتغيرات التوضيحية تكون علاقة غير خطية في الشكل(2)



شكل (2) العلاقة غير الخطية (2)

وتمثل المعادلة التالية :

$$\text{Logit}(\pi_i) = \text{Ln} \frac{\frac{e^{B_O + \sum_i^n B_i X_i}}{1 + e^{B_O + \sum_i^n B_i X_i}}}{\frac{1}{1 + e^{B_O + \sum_i^n B_i X_i}}}$$

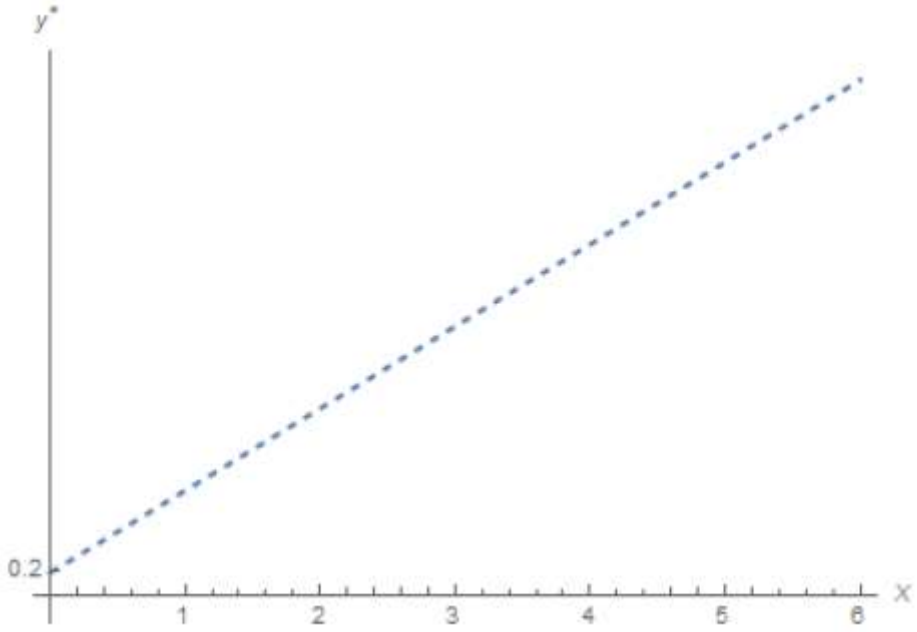
$$\text{Logit}(\pi_i) = B_0 + \sum_i^n B_i X_i$$

$$Y^* = B_0 + \sum_i^n B_i X_i$$

حيث ان: i=1.2..... n

لإحيث تمثل متغير اللوجستيك الجديد والتي قيمته تساوي (1) عند حدوث الاستجابة وتساوي صفر عند عدم حدوث الاستجابة وبذلك سوف تكون قيم الانحدار اللوجستي محصورة $0< Y < 1$

ستكون طبيعة العلاقة في معادلة (3-8) ما بين متغير الاستجابة ومجموعة المتغيرات التوضيحية علاقة خطية عند تطبيق تحويل(logit) ، ويوضح ذلك لنفترض إنَّ $Ln = y \pi 1-\pi$* فإنَّه يمكن اعادة كتابة y* بالشكل الآتي [2]



شكل (3) العلاقة الخطية

2-2-7- طرائق تقدير أنموذج الانحدار اللوجستي:

بعد أن تم وصف أنموذج وتحويله من اللاخطي الى الخطي، بعد ذلك تأتي مرحلة تقدير أنموذج حيث يتم تقدير أنموذج ببعض الطرائق التي تكون خاصة بالمتغيرات متعدد المستويات، والتي تعد أحد أهم الجوانب في الاستدلال الإحصائي لأغلب النماذج، ومن طرائق التقدير هذه طريقة تقدير الإمكان الأعظم (Maximum Likelihood Method)من خلال تطبيق خوارزمية نيوتن رافسون (Newton Raphson .. حيث يتم تعظيم أنموذج الانحدار اللوجستي من خلال دالة الإمكان الأعظم (MLE) بدلالة معلمات L (8) يتم الحصول على معادلة:

$$L(\beta \cdot \pi_{i1} \ldots \pi_{ij}) = \prod_{i=1}^{n} \prod_{j=1}^{J} [\pi_{ij}(\underline{x})]^{r_{ij}}$$

وبأخذ اللوغاريتي لطرفي المعادلة أعلاه سوف يتم الحصول على: (تحصل على x) Tij : وبالتعويض عن وباشتقاق دالة الإمكان بالنسبة للمعلمات:(B)

$$\mathrm{Log}\, L(\beta, \pi_{i1} \ldots \pi_{ij}) = \sum_i \sum_j r_{ij} Log\, \pi_{ij}(\underline{x})$$

$$\mathrm{Log}(\beta, \pi_{i1} \ldots \pi_{ij}) = \sum_i \sum_j r_{ij}\, \mathrm{Log} \frac{e^{B_0 + \sum_i^n B_i X_i}}{1 + e^{B_0 + \sum_i^n B_i X_i}}$$

$$= \sum_i \sum_j r_{ij}(B_0 + B_i X_i) - \sum_i \sum_j r_{ij}\, Log(1 + e^{-B_0} + \sum_i^n B_i X_i)$$

ويمكن اشتقاق دالة الامكان بالنسبة المعلمات (β)

$$\frac{\partial \mathrm{LogL}(\beta)}{\partial(\beta)} = \frac{\partial \sum_i \sum_j r_{ij}(B_0 + B_i X_i)}{\partial(\beta)} - \frac{\partial \sum_i \sum_j r_{ij}\, \mathrm{Log}\left(1 + e^{-B_0} + \sum_i^n B_i X_i\right)}{\partial(\beta)} \ldots .$$

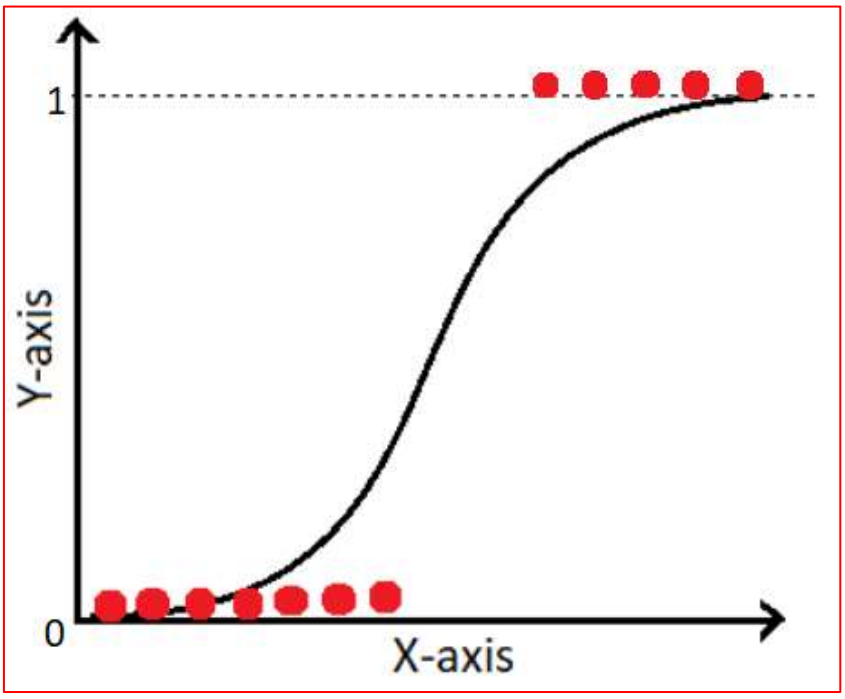
تحصل على معادلات طبيعية، وإن حلول وعند مساواه المشتقة بالصفر ($\frac{\partial \mathrm{LogL}(\beta)}{\partial(\beta)} = 0$) . هذه المعادلات تكون عبارة عن قيم المعلمات التقديرية التي تنتج من خلال عملية التعظيم إلى استخدام أحد الأساليب التكرارية ولكون المعادلات التي تم الحصول عليها غير خطية نلجأ a(B) والتي تنتج تقديرات مناسبة ومن ضمن هذه الأساليب هي نيوتن رافسون [3]

الفصل الثالث
تحليل الانحدار اللوجستي

3-1- تحليل الانحدار اللوجستي

1-تحليل الانحدار اللوجستي لحساب الاحتمال الشرطي للكائنات في جدول المعلومات.

يمكن أن يكون الانحدار اللوجستي ثنائيًا أو متعدد الحدود. يشير الانحدار اللوجستي الثنائي أو ذو الحدين إلى مثال يمكن أن تحتوي فيه النتيجة المرصودة على نوعين محتملين فقط (على سبيل المثال، "ميت" مقابل "حي" أو "النجاح" مقابل "الفشل"، أو "نعم" مقابل "لا"). يشير الانحدار اللوجستي متعدد المتغيرات أو متعدد الحدود إلى الحالات التي يمكن أن تحتوي فيها النتيجة على ثلاثة أو أكثر الأنواع المحتملة (على سبيل المثال، "أفضل" مقابل "لا تغيير" مقابل "أسوأ") باعتباره نهج تحليل الانحدار اللوجستي المشترك، ثنائي يتم استخدام أساليب الانحدار اللوجستي والانحدار اللوجستي متعدد الحدود للتعامل مع مشكلة التصنيف لأن يمكنهم حساب احتمالية وقوع حدث ما بشكل مباشر.[1]



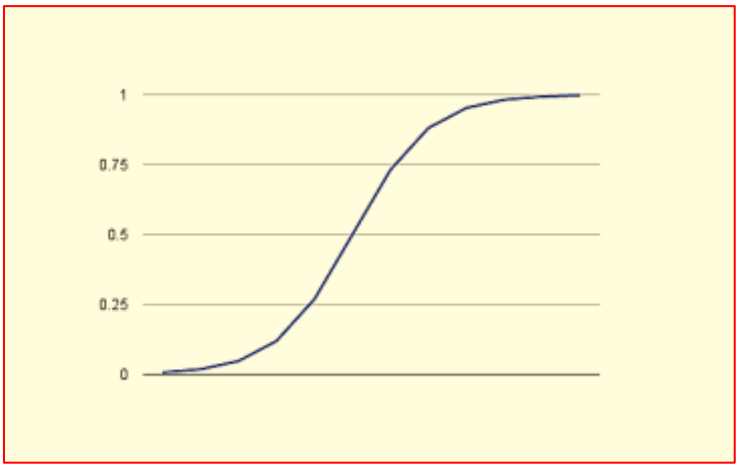
وبناءً على مجموعة معينة من المتغيرات المستقلة .بكلمات بسيطة، يتنبأ الانحدار اللوجستي باحتمالية وقوع حدث ما عن طريق ملاءمة البيانات لوظيفة السجل (ومن هنا جاء اسم الانحدار LOGIsTic يتنبأ الانحدار اللوجستي بالاحتمالية، وبالتالي تقع قيم مخرجاته بين 0 و1. أن الأساس وراء صيغة الانحدار اللوجستي تم تسمية الانحدار اللوجستي على اسم الوظيفة المستخدمة في جوهر الطريقة، وهي الوظيفة اللوجستية .الدالة اللوجستية أو الدالة السيني هي منحنى على شكل حرف S يمكن أن يأخذ أي رقم ذو قيمة حقيقية ويعينه في قيمة بين 0 و 1، ولكن ليس بالضبط عند تلك الحدود.

1 / (1 + قيمة e^)

أين:

1. "e"هو أساس اللوغاريتمات الطبيعية
2. "القيمة" هي القيمة العددية الفعلية التي تريد تحويلها

ويمكن تحويل الأرقام الموجودة ضمن نطاق معين إلى نطاق من 0 إلى 1 باستخدام دالة لوجستية/سينية .على سبيل المثال، قمنا بتطبيق الدالة اللوجستية بين نطاق من −10 إلى +10، وهذا ما يبدو عليه الرسم البياني لدينا:



يستخدم الانحدار اللوجستي معادلة كتمثيل يشبه إلى حد كبير معادلة الانحدار الخطي .في المعادلة، يتم دمج قيم المدخلات خطيًا باستخدام الأوزان أو قيم المعاملات للتنبؤ بقيمة المخرجات .يتمثل الاختلاف الرئيسي عن الانحدار الخطي في أن قيمة الإخراج التي يتم تصميمها هي قيمة ثنائية (0 أو 1) وليست قيمة رقمية .فيما يلي مثال على معادلة الانحدار اللوجستي: [22]

1. xهي قيمة الإدخال
2. yهو الناتج المتوقع

3. b0هو مصطلح التحيز أو الاعتراض
4. b1هو معامل قيمة الإدخال الفردية(x)

في المعادلة، يحتوي كل عمود في بيانات الإدخال الخاصة بك على معامل b مرتبط (قيمة حقيقية ثابتة) يجب تعلمه من بيانات التدريب الخاصة بك.

ويمكننا أيضًا تحويل هذه المعادلة إلى: [23]

$$\ln(y / 1 - y) = b0 + b1 * X$$

تحليل معادلة الانحدار اللوجستي:

1. المتغير التابع في الانحدار اللوجستي يتبع توزيع برنولي
2. ويتم التقدير من خلال أقصى احتمال
3. لا يوجد مربعR ، يتم حساب لياقة النموذج من خلال توافقKS-Statistics

عند تنفيذ نموذج الانحدار اللوجستي

يجب تقدير المعاملات (قيم بيتا ب) لخوارزمية الانحدار اللوجستي من بيانات التدريب الخاصة بك باستخدام تقدير الاحتمال الأقصى .ستؤدي أفضل قيم بيتا إلى نموذج يتنبأ بقيمة قريبة جدًا من 1 للفئة الافتراضية وقيمة قريبة جدًا من 0.

افتراضات الانحدار اللوجستي

1. في حين أن الانحدار اللوجستي يبدو وكأنه خوارزمية بسيطة إلى حد ما لاعتمادها وتنفيذها، إلا أن هناك الكثير من القيود حول استخدامها .على سبيل المثال، لا يمكن تطبيقه إلا على مجموعات البيانات الكبيرة .وبالمثل، يجب وضع افتراضات متعددة في مجموعة البيانات لتتمكن من تطبيق خوارزمية التعلم الآلي.
2. يجب أن يكون المتغير التابع ثنائيًا في معادلة لوجستية ثنائية
3. يجب أن يمثل مستوى العامل 1 للمتغير التابع النتيجة المرجوة
4. قد يؤدي تضمين متغيرات غير ذات معنى إلى حدوث أخطاء .قم بتضمين المتغيرات الضرورية فقط والتي قد تظهر ارتباطًا
5. يجب أن يحتوي النموذج على علاقات خطية قليلة أو معدومة – يجب أن تكون المتغيرات المستقلة مستقلة تمامًا عن بعضها البعض
6. ترتبط المتغيرات المستقلة خطيًا باحتمالات السجل

2-الرياضيات وراء الانحدار اللوجستي

يمكننا أن نبدأ بافتراض أن p(x) هي الدالة الخطية .ومع ذلك، المشكلة هي أن p هو الاحتمال الذي يجب أن يختلف من 0 إلى 1 بينما p(x) هي معادلة خطية غير محدودة .لمعالجة هذه المشكلة، لنفترض أن السجل p(x) هو دالة خطية لـx ، ولربطها بين نطاق (0,1)، سنستخدم التحويل اللوغاريتمي .لذلك، سننظر في السجل .((p(x)/(1-p(x بعد ذلك، سنجعل هذه الدالة خطية: [30-31]

$$\log \frac{p(x)}{1-p(x)} = \alpha_0 + \alpha . x$$

بعد حل:p(x)

$$p(x) = \frac{e^{\alpha_0+\alpha}}{e^{\alpha_0+\alpha}+1}$$

$$L(\alpha_0, \alpha) = \prod_{I=1}^{n} p(x_i)^{y_i}(1-p(x_i)^{1-y_i}$$

يمكن تحويل الضرب إلى مجموع عن طريق أخذ السجل:

$$l(\alpha_0, \alpha) = \sum_{i=0}^{n} \mathrm{y_i} \log p(x_i) + (1 - y_i) log 1 - p(x_i)$$

$$= \sum_{i=0}^{n} \log 1 - p(x_i) + \sum_{i=0}^{n} y_i log \frac{p(x_i)}{1 - p(x_i)}$$

علاوة على ذلك، بعد وضع قيمة:p(x)

$$l(\alpha_0, \alpha) = \sum_{i=0}^{n} -\log 1 + e^{\alpha_0 + \alpha} + \sum_{i=0}^{n} y_i(\alpha_0 + \alpha . x_i)$$

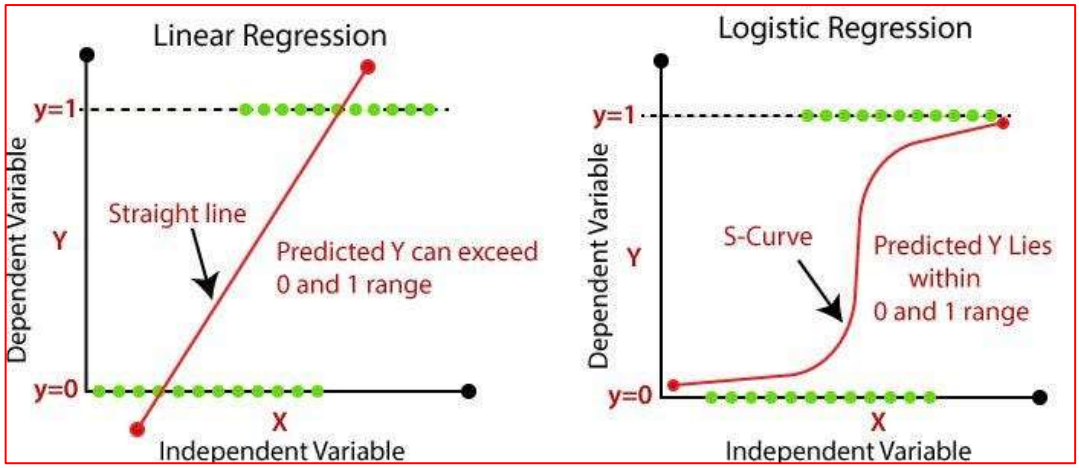
الخطوة التالية هي اتخاذ الحد الأقصى لوظيفة الاحتمالية المذكورة أعلاه لأنه في حالة الانحدار اللوجستي يتم تنفيذ صعود التدرج (عكس نزول التدرج.(

3-تقدير الاحتمالية القصوى(MLE)

طريقة لتقدير معلمات التوزيع الاحتمالي عن طريق تعظيم دالة الاحتمالية، من أجل زيادة احتمال حدوث البيانات المرصودة .يمكننا العثور على MLE من خلال التمييز بين المعادلة أعلاه فيما يتعلق بالمعلمات المختلفة وتعيينها على صفر .على سبيل المثال، المشتق بالنسبة لأحد مكونات المعلمةalpha ، أيa_j ، يُعطى بواسطة: [36]

$$\frac{\partial l}{\partial \alpha_j} = \sum_{i=0}^{n} \big(y_i - p(x_i; \alpha_0, \alpha)\big) x_{ij}$$

لماذا يسمى "الانحدار اللوجستي" وليس "التصنيف اللوجستي"؟



الانحدار اللوجستي، على الرغم من اسمه، هو خوارزمية تعلم آلي مستخدمة على نطاق واسع لمهام التصنيف الثنائية .يمكن أن يكون الاسم مضللاً إلى حد ما، نظرًا لأنه يستخدم في المقام الأول للتصنيف بدلاً من الانحدار .

وفي تحليل الانحدار اللوجستي، يتم استخدام مجموعة من المتغيرات المستقلة (المفسرة)، سواء كانت رقمية أو نوعية، لتوقع قيمة المتغير النوعي المراد تحليله. يتم تقدير الاحتماليات الشرطية لحدوث القيمة النوعية بناءً على القيم المعروفة للمتغيرات المستقلة على سبيل مثال:

لنفترض ان لدينا متغير تابع (مستجيب) Y ذو مشاهده ثنائيه وليكن المرض (مصاب بكورونا_غير مصاب بكورونا) ولدينا متغير اخر مستقل (مؤثر) X ذو مشاهده ثنائيه تمثل (التعرض لخطر الاصابه بالفيروس_عدم التعرض لخطر الاصابه بالفيروس) وبالتالي فان عامل Y للاستجابة للمرض يتاثر بعامل X للتعرض لخطر الاصابه.ومن اجل ايجاد الخطورة النسبيه نقوم بإنشاء جدول(2×2) يمثل تكرارات التقاطعات قد لا يكون المتغير التابع Y (متغير الاستجابه) ثنائي القيمه الاحتمالية كما في مدة بقاء المريض في المستشفى فمثلا مدة بقاء المراءه في المستشفى لاجل الولاده في هذه الحاله يتم تقسيم الفتره بناء على المده الزمنيه المتوقعه لبقائها في الحالات الاعتياديه فنقول (اصغر او يساوي 48 ساعه _اكبر من 48 ساعه بالتحليل اللوجستي، قيمة الـ p تعبر عن الاحتمالية أو الاحتمال النسبي لحدوث حدث معين. في تحليل الانحدار اللوجستي، يتم استخدام قيمة الـ pلتقدير العلاقة بين المتغيرات المستقلة والمتغير التابع.لنفترض أن لدينا دراسة

تهدف إلى فهم عوامل الخطر المحتملة للإصابة بمرض معين، مثل السكري. يمكننا استخدام تحليل الانحدار اللوجستي لتحديد العوامل المستقلة التي تؤثر في احتمالية الإصابة بالسكري. في هذا المثال، قد تكون لدينا المتغيرات المستقلة التالية:

- ✓ العمر
- ✓ الجنس
- ✓ مؤشر كتلة الجسم(BMI)
- ✓ مستوى النشاط البدني

ويكون لدينا المتغير التابع هو احتمالية الإصابة بالسكري.بعد تنفيذ تحليل الانحدار اللوجستي، يتم حساب قيمة الـ p لكل متغير مستقل لتقدير تأثيره على احتمالية الإصابة بالسكري. قيمة الـ p تشير إلى مدى تأثير المتغير المستقل على المتغير التابع، حيث تكون القيمة أقل من 0.05 تعني وجود تأثير إحصائي معنوي.على سبيل المثال، إذا كانت قيمة الـ p لمتغير العمر تساوي 0.02، فإن ذلك يعني أن هناك تأثير إحصائي معنوي للعمر على احتمالية الإصابة بالسكري.بالإضافة إلى ذلك، يمكن استخدام قيمة الـ p لمقارنة تأثير المتغيرات المستقلة المختلفة. على سبيل المثال، إذا كانت قيمة الـ p لمتغير الجنس تساوي 0.10، في حين أن قيمة الـ p لمتغير BMI تساوي 0.01، فإن ذلك يعني أن لدينا دليل إحصائي أقوى على تأثير BMI على احتمالية الإصابة بالسكري مقارنة بالجنس.

اولا : الاستنتاجات

بصياغة النموذج: في الانحدار اللوجستي، تتم نمذجة العلاقة بين متغير الاستجابة (الثنائي) والمتغيرات التوضيحية المتعددة باستخدام الدالة اللوجستية. تقوم الوظيفة اللوجستية بتعيين المجموعة الخطية للمتغيرات التوضيحية إلى قيمة احتمالية تتراوح بين 0 و1. ويمكن التعبير عن معادلة النموذج على النحو التالي:

1- $P(Y=1|X) = 1 / (1 + \exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)))$

أين:يمثل $P(Y=1|X)$ احتمال أن يكون متغير الاستجابة 1 بالنظر إلى قيم المتغيرات التوضيحيةX_1.، X_2، ...، X_pهي المتغيرات التوضيحيةβ_0.، β_1، β_2، ...، β_pهي معاملات الانحدار التي سيتم تقديرها.

2- تقدير: يقوم نموذج الانحدار اللوجستي بتقدير معاملات الانحدار باستخدام تقنيات تقدير الاحتمالية القصوى .(MLE) الهدف هو العثور على قيمβ_0 ، β_1، β_2، ...، β_pالتي تزيد من احتمالية مراقبة مجموعة البيانات المحددة.

3- تفسير نسبة الأرجحية: في الانحدار اللوجستي، تمثل معاملات المتغيرات التوضيحية التغير في احتمالات متغير الاستجابة لتغيير وحدة واحدة في المتغير التوضيحي المقابل، مع ثبات المتغيرات الأخرى. يمكن حساب نسبة الأرجحية عن طريق مضاعفة المعامل. على سبيل المثال، إذا كانت نسبة الأرجحية للمتغير X_1 هي 1.5، فهذا يعني أن زيادة وحدة واحدة في X_1 تؤدي إلى احتمالات أعلى بمقدار 1.5 مرة لأن يكون متغير الاستجابة 1.

4- اختيار المتغير: عند التعامل مع متغيرات توضيحية متعددة، يصبح اختيار المتغير حاسما لتحديد أهم المتنبئين. يمكن استخدام تقنيات مثل الانحدار التدريجي، أو انحدار لاسو، أو معايير المعلومات) على سبيل المثال، AIC،

(BICلتحديد المجموعة الفرعية المثالية من المتغيرات التي تشرح متغير الاستجابة بشكل أفضل.

ثانيا : التوصيات

1- التحليل التفاضلي: قبل استخدام الانحدار اللوجستي، يجب إجراء تحليل تفاضلي للبيانات لفهم العلاقات والارتباطات بين المتغيرات المستقلة والمتغير المعتمد. يمكن استخدام أدوات مثل الرسوم البيانية والتحليل التفاضلي لاستكشاف العلاقات المحتملة وتحديد المتغيرات التي تستحق دراسة الانحدار اللوجستي.

2- التعامل مع التعددية والتداخل: عند استخدام متغيرات متعددة في الانحدار اللوجستي، يجب مراعاة التعددية والتداخل بين المتغيرات المستقلة. يمكن استخدام تحليل الانحدار المتعدد لتحديد العلاقات المستقلة الفعلية بعد تصحيح التأثيرات المتداخلة.

3- التفسير السليم: يجب التفسير السليم للمعاملات المقدرة في نموذج الانحدار اللوجستي. ينبغي فهم أن المعاملات تمثل التغير في الصعوبة المنطقية (oddsللمتغير المعتمد لكل وحدة تغير في المتغير المستقل، مع الأخذ في الاعتبار المتغيرات الأخرى.

المصادر

[1] ادريس، حسن نوار ، (2009) ، " استخدم النموذج اللوجستي المتعدد لتحديد العواملالمؤثرة في مرض الجلكوم " كلية العلوم / قسم الإحصاء التطبيقي / جامعة السودان للعلوم والتكنولوجيا.

[2] الحيالي، ليث علي 2017، "اختيار أفضل طرائق تقدير المتغير المعتمد متعدد الاستجابة" رسالة ماجستير كلية الإدارة والاقتصاد، جامعة كربلاء.

[3] الرديني ساره عادل 2019" استعمال الخوارزمية الجينية في تقدير معلمات انموذج الانحدار اللوجستي الثنائي مع تطبيق عملي رسالة ماجستير كلية الإدارة والاقتصاد جامعة بغداد.