المجلة العراقية للعلوم الإحصائية

www.stats.mosuljournals.com

التحليل الاحصائي لمتغير الاستجابة الرتبي: دراسة مقارنة

لقاء ياسين أحمد و زيد طارق صالح الخالدي

قسم الاحصاء والمعلوماتية ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، الموصل ، العراق

معلومات النشر

تاريخ المقالة:

تم استلامه في 27 ايار 2022

تم القبول في 28 حزيران 2022

متاح على الإنترنت في 1 كانون الاول 2022

الكلمات الدالة:

متغير الاستجابة الترتيبية ، الانحدار المتعدد ، الانحدار اللوجيستي الترتيبي ، المتغيرات الفئوية

المراسلة:

لقاء ياسين أحمد

Leqaa.20csp116@student.uomosul.edu.iq

الخلاصة

تتنوع متغيرات الاستجابة في الظواهر الحيوية بين ثلاثة أنواع: متغيرات الاستجابة العددية، متغيرات الاستجابة الفئوية الرتبية، ومتغيرات الاستجابة الفئوية الاسمية. في الدراسات الاحصائية، يختلف التعامل مع المتغيرات الرتبية وفقا لمنظور الطريقة الاحصائية لمتغير الاستجابة. حيث يمكن اعتماده كمتغير فئوي اسمي وذلك باهمال الخاصية الرتبية لفئات المتغير. ويمكن أيضا معاملته كمتغير فئوي رتبي (متغير متقطع)، وفي هذه الحالة يمكن الاستفادة من معلومات الترتيب في بناء الانموذج التنبؤي. تم في هذه الدراسة الوقوف على اهم الطرائق الاحصائية التي من الممكن توظيفها لتحليل البيانات ذات متغير استجابة رتبي. ومن هذه الطرائق، طريقة الانحدار المتعدد، وطريقة الانحدار اللوجستي الرتبي. حيث تم توضيح آلية بناء النماذج وتقدير المعالم نظرياً، فضلا عن كيفية قراءة المدلول الإحصائي لمعاملات الانحدار في كل النماذج المدروسة. وتم التطبيق على عينة حقيقية خاصة بمرضى هشاشة العظام. حيث تم بناء نماذج متعددة للوقوف على اهم العوامل المؤثرة في احتمالية الاصابة بالمرض. وتمّ تشخيص الانموذج الأفضل وفقاً لمعيار معلومات أكياكي ومعيار المعلومات البيزي. أثبتت نتائج التحليل الإحصائي للبيانات تفوق انموذج الانحدار اللوجستي الرتبي على انموذج الانحدار الخطي المتعدد في تفسيره للعلاقة بين بين متغير الاستجابة والمتغيرات التوضيحية.

DOI: https://doi.org/10.33899/iqjoss.2022.176204 , ©Authors, 2022, College of Computer and Mathematical Science, University of Mosul.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0).

1. المقدمة Introduction

يعد تحليل الانحدار ركناً أساسياً من أركان علم الإحصاء، وأسلوب مهم من أساليب الإحصاء التطبيقي عند دراسة بعض الظواهر الاقتصادية أو الاجتماعية ويمكن استخدام الطرائق الإحصائية أينما وجد البحث العلمي سواء كان في مجال الاقتصاد أم في الطب أم في غيرها من المجالات (كاظم، 1988). وأن الانحدار المتعدد يحقق اغلب اهداف البحث العلمي لتحليل البيانات ووصف العلاقة الدالية بين المتغيرات الا أنه يكون غير دقيق في تحديد وتفسير العلاقة بين متغير الاستجابة مما ادت الحاجة الى تطوير اساليب احصائية لها قوة الانحدار المتعدد في التوصل الى افضل المعادلات توفيقا وكذلك تعالج تعذر مشكلة تطبيق الانحدار المتعدد الخطي وغير الخطي في حالة التي يكون متغير الاستجابة من النوع الفئوي (اسمي)، وذلك لعدم توفر فروض تحليل الانحدار المتعدد. أما في الحالات التي يكون فيها المتغير المعتمد من النوع الرتبي (فئوي كمي)، فبالامكان استخدام اسلوب الانحدار المتعدد الخطي أو اللاخطي، لكن مع ضرورة توفر شروط محددة في هيكلية والتوزيع التكراري للمتغير المعتمد (Kutner, et. al.; 2004). ولقد ظهرت العديد من نماذج الانحدار غير الخطي التي تتعامل مع البيانات رتبية الاستجابة، ومنها أنموذج الانحدار اللوجستي الرتبي (Ordinal Logistic Regression)، والذي سيكون محور الرئيسي في بحثنا هذا.

وتكمن مشكلة البحث في إيجاد افضل أنموذج رياضي يلائم البيانات اذا كان لدينا مجموعة من البيانات يكون فيها متغير الاستجابة (Y) متغير رتبي. ومع أن تحليل الانحدار يحقق أغلب اهداف البحث العلمي فهناك أسباب عديدة لعدم ملاءمة أنموذج الانحدار الخطي وأنموذج الانحدار اللوجستي الثنائي عندما يكون متغير الاستجابة من النوع الرتبي.

1. هدف البحث

لذلك يهدف هذا البحث الى تحديد الاسلوب الاحصائي الأمثل لتحليل ونمذجة البيانات التي يكون فيها متغير الاستجابة من النوع الرتبي وذلك بتحليلها ومقارنة النتائج. ومن الأساليب الاحصائية التي من الممكن توظيفها لتحليل هكذا نوع من البيانات هي اسلوب الانحدار الخطي المتعدد، واسلوب الانحدار اللوجستي الرتبي. حيث يهدف البحث لمقارنة النمذجة الإحصائية بالاعتماد على الطريقتين المذكورتين آنفاً.

2. انواع المتغيرات

في علم الإحصاء، هنالك نوعين من المتغيرات: النوع الاول هو المتغيرات الكمية وهي المتغيرات التي يمكن عدها، وتكون إما متقطعة أو مستمرة. والنوع الثاني هو المتغيرات الفئوية. وعندما يكون هنالك فئتين فقط في المتغير الفئوي فعندئذٍ يعرف بأنه متغير ثنائي. ويمكن التعبير عن المتغيرات الفئوية الثنائية رياضيا باستخدام متغير دليلي indicator variable، حيث يتم ترميز قيم المتغير الثنائي بالعددين 0 و 1. ويمكن أن يحتوي المتغير الفئوي على ثلاثة فئات مختلفة فأكثر، وعندها يرمز باستخدام تقنية المتغيرات الوهمية. وتكون المتغيرات الفئوية على نوعين: النوع الاول المتغيرات الاسمية، وفيه تكون المشاهدات فئوية غير رتبية، مثلا دراسة تأثير التداخل الجيني البشري على لون الشعر (اسود، اشقر، بني، احمر، أبيض). والنوع الثاني المتغيرات الرتبية، وفيه تكون فئات المتغير الفئوي قابلة للترتيب، مثلا دراسة تأثير بعض العوامل البيئية والحيوية على مستوى ضغط الدم لدى الانسان (منخفض، طبيعي، مرتفع) (Agresti, 2003).

3. الأساليب الإحصائية لتحليل متغير الاستجابة الرتبي:

بالامكان توظيف الطرائق الاحصائية الاخرى لتحليل المتغيرات الرتبية، وذلك بالرغم من وجود بعض المحددات التي تتعلق بفروض التحليل. وعادة ما يلجأ الباحثون الى مقارنة الطرق المتاحة فيما بينها للوصول الى الطريقة المثلى لتحليل العينة الخاصة بهم. وتختلف الطريقة الاحصائية باختلاف منظورنا الى متغير الاستجابة، بغض النظر عن حقيقة كونه مستمراً أم فئوياً. سنوضح في الفقرات الآتية الخيارات المتاحة للباحثين في تعاملهم مع المتغيرات الرتبية، فضلاً عن تحديد الايجابيات والسلبيات الممكنة لكل حالة (Lall, 2005).

3.1.المتغير الرتبي كمتغير فئوي اسمي متعدد الفئات:

أحد خيارات التعامل مع المتغير الرتبي هي اعتباره متغيراً اسمياً متعدداً للفئات. وينطوي على ذلك إهمال الطبيعة الرتبية لفئات المتغير قيد الاهتمام. ويكون هذا الفرض منطقيا في بعض الاحيان، فعلى سبيل المثال عندما يكون هناك عدد قليل من الفئات ولا يكون ترتيب الفئات ذات أهمية محورية في مشكلة البحث. أهم ميزة لهذا النهج هو عدم إنتهاك فرضيات التحليل، خصوصا في حال كون حجم العينة مناسب لتطبيق الإجراءات الاحصائية. هناك العديد من الخيارات لتحليل المتغيرات الاسمية، حيث يمكن استخدام اختبار مربع كاي للاستقلالية بين متغير الاستجابة والمتغيرات التوضيحية بعد تبويبها لتشخيص العلاقة بين الاستجابة والتنبوء. ويمكن ايضا توظيف اختبار Cochran–Mantel–Haenszel test (CMH)، وهو اسلوب لاختبار العلاقة بين متغير الاستجابة الاسمي والمتغيرات التوضيحية الفئوية. كلا الاختبارين المذكورين يعتمدان على تكوين جدولة تقاطعية cross tabulation لبيانات متغيري الاستجابة والتوضيحي. كما يمكن استخدام الانحدار اللوجستي متعدد الفئات Polytomous Logistic Regression، والذي لا يتطلب وجود أي نوع من القيود على متغير الاستجابة. ويتميز الاجراء الأخير عن سابقيه في كون أن التحليل ينمذج العلاقة بين متغير الاستجابة وجميع المتغيرات التوضيحية في آن واحد. ويترتب على معاملة متغير الاستجابة الرتبي كمتغير اسمي فقدان المعلومات الخاصة بالترتيب، مما قد يؤدي الى بناء استدلالات غير دقيقة (Ananth & Kleinbaum, 1997; Scott, et. al., 1997).

3.2.المتغير الرتبي كمتغير ثنائي:

يمكن معاملة المتغير الرتبي متعدد الفئات كمتغير ثنائي وذلك بعد دمج عدد من الفئات وخفض عددها الى فئتين فقط. وفي هذه الحالة يمكن استخدام الانحدار اللوجستي الثنائي لنمذجة العلاقة بين متغير الاستجابة والمتغيرات التوضيحية. ويكون هذا الاجراء مقبولا غالباً عندما يكون ترتيب الفئة التي تصنف على أنها الحالة الطبيعية (أو السيطرة) في وسط الفئات الأخرى، وكان اهتمام الدراسة مركزاً في أحد الاتجاهين غير الطبيعيين. ورغم مقبولية هذا الاسلوب في بعض الأحيان، إلا أنه يتسبب في ضياع معلومات الترتيب الفئوي،

وكذلك صعوبة تحديد نقطة القطع cut–point لتحديد الفئات الجديدة. وهذا بالتأكيد سيؤدي إلى عدم دقة تمثيل العلاقة بين المتغيرات قيد الدراسة (Strömberg 1996).

3.3.المتغير الرتبي كمتغير كمي:

الخيار الآخر المتاح في التعامل مع المتغير الرتبي هو اعتباره متغيراً كمياً. وفي هذه الحالة يتم التعامل مع متغير الاستجابة الفئوي إما كمتغير كمي مستمر، أو كمتغير كمي متقطع. وفي كلتا الحالتين يتم الاحتفاظ بالخاصية الرتبية لبيانات المتغير قيد الاهتمام. ففي الحالة الأولى، يصار إلى اعتبار أن المتغير قادم من توزيع مستمر ومتناظر، وذلك لضمان توفر فرضية التوزيع الطبيعي لبيانات متغير الاستجابة. ويتطلب هذا الاجراء توفر عينة بحجم كبير نسبياً، وكذلك الحال بالنسبة لعدد الفئات. فكلما زادت الفئات، كلما اقترب سلوك المتغير الرتبي من المتغير المستمر. ويفضل أيضا أن تكون المسافات بين الفئات متساوية، وذلك لتلافي حدوث مشكلة الالتواء في توزيع متغير الاستجابة (Armstrong et al., 1989). وفي هذه الحالة يمكن توظيف اسلوب الانحدار المتعدد (الخطي وغير الخطي) لنمذجة العلاقة بين المتغيرات. وأيضاً يمكن استخدام تحليل الاتجاه Trend Analysis لاستقصاء العلاقة بين متغير الاستجابة والمتغيرات التوضيحية (Siegel, 1988). أما في حال اعتبار المتغير الكمي متقطعاً، فعندئذٍ يمكن استخدام الانحدار اللوجستي الرتبي Ordinal Logistic Regression لتوصيف العلاقة بين المتغيرات. ويعد أنموذج الانحدار اللوجستي الرتبي أحد نماذج الانحدار اللوجستي متعددة الفئات، مع الأخذ بنظر الاعتبار الطبيعية الرتبية لفئات متغير الاستجابة، مما يعطي أفضلية نظرية لهذا الاسلوب مقارنة بالاساليب الأخرى. (Ananth & Kleinbaum, 1997; Agresti, 2003).

4. الانحدار الخطي المتعدد:

يعرف أنموذج الانحدار الخطي المتعدد بأنه عملية تقدير العلاقة الخطية بين متغير استجابة واحد مع عدد من المتغيرات التوضيحية (الراوي، 1987). وهدف الانحدار الخطي المتعدد هو استخدام العلاقة بين متغير الاستجابة والمتغيرات التوضيحية للتنبؤ أو توضيح سلوك متغير الاستجابة.

ومعادلة الانحدار الخطي المتعدد التي تحوي على $p-1$ من المتغيرات التوضيحية تكون بالشكل الآتي:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i \quad (1)$$

حيث أن y_i تمثل المشاهدة i للمتغير المعتمد، x_{ij} تمثل المشاهدة i للمتغير التوضيحي j، β_0 الحد الثابت، β_j معامل الانحدار الخاص بالمتغير التوضيحي X_j، وأن ε_i تمثل القيمة i للخطأ العشوائي، ويفترض أنه يتوزع توزيعا طبيعيا مستقلاً ومتماثلاً بمتوسط صفر وتباين ثابت. ويمكن كتابة المعادلة (1) بدلالة المصفوفات:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (2)$$

حيث أن $\boldsymbol{Y}$ يمثل متجه الاستجابة، $\boldsymbol{X}$ تمثل مصفوفة البيانات، $\boldsymbol{\beta}$ يمثل متجه المعالم، و$\boldsymbol{\varepsilon}$ يمثل متجه الاخطاء العشوائية [الراوي، 1987].

4.1. تقدير متجه معالم الانحدار الخطي المتعدد

يمكن إيجاد مقدر متجه معالم الانحدار الخطي المتعدد وفقاً المربعات الصغرى الاعتيادية $\widehat{\boldsymbol{\beta}}$ من خلال اشتقاق مجموع مربعات الاخطاء الموضح في المعادلة (3) بالنسبة لكل معلمة من معالم أنموذج الانحدار الخطي المتعدد ومن ثم الحل بالنسبة لـ $\widehat{\boldsymbol{\beta}}$:

$$\sum_{i=1}^{n} \varepsilon_i = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta) \quad (3)$$

$$\frac{\partial \varepsilon'\varepsilon}{\partial \beta} = X'Y - X'X\hat{\beta} := 0 \quad (4)$$

ومنها نحصل على مقدر متجه معالم الانحدار الخطي المتعدد، وكالآتي:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (5)$$

علما أن طريقة الامكان الأعظم في التقدير تعطي الصيغة الموضحة في المعادلة (5) نفسها، وذلك بالاعتماد على فرضية التوزيع الطبيعي المستقل والمتطابق للأخطاء العشوائية.

5. الانحدار اللوجستي:

يستخدم الانحدار اللوجستي لنمذجة العلاقة بين متغير معتمد واحد كمي او نوعي وعدد من المتغيرات التوضيحية. مع ذلك، فإن الانحدار اللوجستي غالبا ما يستخدم لنمذجة البيانات التي تحوي على متغير استجابة نوعي، أو ما يسمى بمتغير الاستجابة الفئوي،

وتحديدا متغير الاستجابة الثنائي، وحينها يسمى بالانحدار اللوجستي الثنائي. علما أن الانحدار اللوجستي يمكن أن يستخدم لنمذجة البيانات التي تحوي على متغير استجابة فئوي ذات فئات متعددة، وعندها يسمى بالانحدار اللوجستي المتعدد. ومن نماذج بالانحدار اللوجستي المتعدد هو أنموذج الانحدار اللوجستي الرتبي، والذي يفترض أن متغير الاستجابة عبارة عن متغير فئوي ذات فئات قابلة للترتيب.

5.1. الانحدار اللوجستي الثنائي:

لنفرض أنه لدينا متغير الاستجابة Y_i من النوع الثنائي، أي أن قيم متغير الاستجابة تكون $(1,0)$. عندئذٍ تكون صيغة أنموذج الانحدار اللوجستي الثنائي كالآتي:

$$Y_i = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} + \varepsilon_i \qquad (6)$$

حيث أن Y_i يمثل متغير الاستجابة، X_i يمثل الصف i من مصفوفة البيانات X، β يمثل متجه معالم الانحدار، وأن ε_i يمثل الخطأ العشوائي الخاص بالمشاهدة i، ويتوزع بمتوسط صفري وتباين ثابت.

ويعتبر Y_i وفقا للانحدار اللوجستي الثنائي بأنه متغير عشوائي يتوزع وفقا لتوزيع برنولي بمعلمة π_i، والتي تمثل توقع Y_i. وبذلك تكون دالة الاستجابة اللوجستية الثنائية كالآتي:

$$E(Y_i) = \pi_i = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \qquad (7)$$

إذن نحن نقدر احتمالية أن $Y_i = 1$، وليس قيمة Y_i نفسها (Kutner, et. al., 2004).

5.2. الانحدار اللوجستي الرتبي:

لنفرض أنه لدينا متغير الاستجابة Y_i من النوع الفئوي يحتوي على K من الفئات القابلة للترتيب. يتم إعطاء رتب لفئات المتغير Y_i وفقا لطبيعة المتغير والفئات الخاصة به. وبذلك تكون القيم الممكنة لمتغير الاستجابة مرتبة بشكل تصاعدي ابتداءً من العدد 1 ولغاية العدد K. وبما أن فئات متغير الاستجابة قابلة للترتيب فعليه سيتم حساب دالة الاستجابة اللوجستية الرتبية وفقاً للاحتمالية التراكمية كالآتي:

$$p(Y_i \le k) = \frac{\exp(\alpha_k + X_i\beta)}{1+\exp(\alpha_k + X_i\beta)} \qquad (8)$$

وذلك بفرض أن المتغير Y_i يعتمد على $p-1$ من المتغيرات التوضيحية. حيث أن $k = 1,2,\dots,K-1$. لاحظ أن $p(Y_i \le K) = 1$، وذلك لكون أن K تمثل العدد الأكبر في الترميز لفئات متغير الاستجابة الرتبي. وتمثل α_k معلمة المقطع لكل فئة من فئات متغير الاستجابة. وأن X_i يمثل متجها صفيا لقيم المتغيرات المستقلة عند المشاهدة i، وأن β يمثل متجه معاملات المتغيرات التوضيحية.

وبذلك يكون لدينا $K-1$ من دوال الاستجابة اللوجستية الرتبية، وتختلف هذه الدوال فيما بينها باختلاف معلمة المقطع فقط، فيما تكون باقي المعالم الخاصة بالمتغيرات التوضيحية ثابتة في جميع دوال الاستجابة. وبتطبيق تحويل اللوجيت على المعادلة (8) نحصل على (Agresti, 2003):

$$\log_e \frac{p(Y_i \le k)}{1-p(Y_i \le k)} = \alpha_k + X_i\beta \qquad (9)$$

حيث يطلق على المعادلة (9) أعلاه بلوغارتيم أرجحية odds كون أن $Y_i \le k$.

5.2.1. تقدير معالم أنموذج الانحدار اللوجستي الرتبي:

لاحظ أن كل فئة من فئات متغير الاستجابة تعامل كمعاملة متغير ثنائي عند احتساب الارجحيات. ولتوضيح ذلك نفرض أن:

$$Y_{i(k)} = \begin{cases} 1 & \text{if } Y_i = k \\ 0 & \text{if } Y_i \ne k \end{cases}$$

عندئذ يمكن كتابة احتمالية حدوث $Y_{i(k)} = 1$ من خلال الاحتمالات التراكمية وكالآتي:

$$p\left(Y_{i(k)} = 1\right) = p(Y_i = k) = p(Y_i \le k) - p(Y_i \le k-1) = \pi_{i(k)}$$

حيث أن $k = 1,2, \dots, K$. لاحظ أن $p(Y_i \leq 0) = 0$ و $p(Y_i \leq K) = 1$.

ويمكن التعبير عن الاحتمالية أعلاه بدلالة الاحتملات الخاصة بالقيم الأخرى للمتغير Y_i:

$$p(Y_{i(k)} = 1) = \prod_{k=1}^{K} [\pi_{i(k)}]^{Y_{i(k)}} \tag{10}$$

وذلك بعد إعادة صياغتها كالآتي (8)وبما أن الاحتمالات المذكورة أعلاه يمكن كتابتها بدلالة الانموذج اللوجستي الموضح في المعادلة

$$\pi_{i(k)} = [\exp(-(\alpha_k + x_i\beta)) + 1]^{-1} - [\exp(-(\alpha_{k-1} + x_i\beta)) + 1]^{-1} \tag{11}$$

عليه، يكون لوغارتيم دالة الامكان كالآتي:

$$\log_e L(\alpha, \beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} y_{i(k)} \log_e [\pi_{i(k)}] \tag{12}$$

، وكما موضح في المعادلتين 10 و 11 $p(Y_{i(k)} = 1)$، والتعابير المشتقة لـ $p(Y_i \leq K) = 1$ و $p(Y_i \leq 0) = 0$وبالتعويض بـ نحصل على الصيغة النهائية للوغارتيم دالة الامكان:

$$\begin{aligned} \log_e L(\alpha, \beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} y_{i(k)} \log_e \Big[& [\exp(-(\alpha_k + x_i\beta)) + 1]^{-1} \\ & - [\exp(-(\alpha_{k-1} + x_i\beta)) + 1]^{-1} \Big] \end{aligned} \tag{13}$$

بحل منظومة المعادلات اللاخطية $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{K-1}, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{p-1}$وبذلك يمكن الحصول على مقدرات الامكان الاعظم الآتية عددياً، علما أن عدد المعدلات اللاخطية هو $K + p - 2$:

$$\begin{aligned} \frac{\partial \log_e L}{\partial \alpha_j} = \sum_{i=1}^{n} y_{i(j)} & \frac{\left[\exp\left(-(\alpha_j + x_i\beta)\right) + 1\right]^{-2} \exp\left(-(\alpha_j + x_i\beta)\right)}{\left[\exp\left(-(\alpha_j + x_i\beta)\right) + 1\right]^{-1} - \left[\exp\left(-(\alpha_{j-1} + x_i\beta)\right) + 1\right]^{-1}} \\ & - y_{i(j+1)} \frac{\left[\exp\left(-(\alpha_j + x_i\beta)\right) + 1\right]^{-2} \exp\left(-(\alpha_j + x_i\beta)\right)}{\left[\exp\left(-(\alpha_{j+1} + x_i\beta)\right) + 1\right]^{-1} - \left[\exp\left(-(\alpha_j + x_i\beta)\right) + 1\right]^{-1}} := 0 \end{aligned}$$

، مع ملاحظة أن:$j = 1,2, \dots, K-1$حيث أن

$$[\exp(-(\alpha_K + x_i\beta)) + 1]^{-1} = 1$$

$$[\exp(-(\alpha_0 + x_i\beta)) + 1]^{-1} = 0$$

$$\begin{aligned} \frac{\partial \log_e L}{\partial \beta_j} = \sum_{i=1}^{n} \sum_{k=1}^{K} y_{i(k)} \Bigg[& \frac{[\exp(-(\alpha_k + x_i\beta)) + 1]^{-2} \exp(-(\alpha_k + x_i\beta)) x_{ij}}{[\exp(-(\alpha_k + x_i\beta)) + 1]^{-1} - [\exp(-(\alpha_{k-1} + x_i\beta)) + 1]^{-1}} \\ & - \frac{[\exp(-(\alpha_{k-1} + x_i\beta)) + 1]^{-2} \exp(-(\alpha_k + x_i\beta)) x_{ij}}{[\exp(-(\alpha_k + x_i\beta)) + 1]^{-1} - [\exp(-(\alpha_{k-1} + x_i\beta)) + 1]^{-1}} \Bigg] := 0 \end{aligned}$$

حيث أن $j = 1,2, \dots, p-1$.

5.2.2. المدلول الإحصائي لمعاملات الانحدار اللوجستي الرتبي:

يكون تفسير المدلول الاحصائي لمعاملات الانحدار اللوجستي الرتبي مختلفا عن مفهومه في الانحدار الخطي والذي يتمثل بالميل أو التغيير الحاصل في الاستجابة عند زيادة المتغير التوضيحي بوحدة واحدة. وهنا يكون التفسير معقدا نسبيا ويتطلب فهما لموضوع الأرجحيات odds الخاصة بكل متغير توضيحي. حيث أن الأرجحيات تحسب كالآتي:

$$\text{odds} = \frac{p(Y_i \leq k)}{1 - p(Y_i \leq k)} \tag{14}$$

وبالاعتماد على المعادلة (9)، وبفرض أن جميع المتغيرات التوضيحية ثابتة باستثناء المتغير X_j، حيث أنه يزداد بوحدة واحدة فقط، عندئذٍ يمكن القول بأن:

$$\hat{\beta}_j = \log_e\left(\frac{\text{odds}(x_j+1)}{\text{odds}(x_j)}\right) \tag{15}$$

أي أن معامل الانحدار اللوجستي الرتبي للمتغير المعتمد Y على المتغير التوضيحي X_j يمثل لوغارتيم نسبة الأرجحيات log of odds ratio (OR). وبذلك تكون نسبة الأرجحيات المقدرة كالآتي:

$$\widehat{OR} = \frac{\text{odds}(x_j+1)}{\text{odds}(x_j)} = \exp(\hat{\beta}_j) \tag{16}$$

فإذا كانت نسبة الأرجحيات المقدرة أكبر من 1، فذلك يعني أن أرجحية أن تكون $Y_i \leq k$ تزداد كلما زادت قيمة المتغير التوضيحي، أي يكون التأثير طرديا. من ناحية أخرى، إذا كانت نسبة الأرجحيات المقدرة أقل من 1، فذلك يعني أن أرجحية أن تكون $Y_i \leq k$ تنخفض كلما زادت قيمة المتغير التوضيحي، أي يكون التأثير عكسيا. وأخيرا، فإن نسبة الارجحيات تساوي 1 فذلك يعني عدم وجود تأثير للمتغير المستقل على الاستجابة.

6. الجانب التطبيقي:

تم إجراء تطبيق عملي على بيانات تخص مرض هشاشة العظام Osteoporosis لدى عينة من النساء في مدينة الموصل. حيث أن بيانات العينة، المكونة من 344 مشاهدة، مأخوذة من وحدة اختبار دكسا في مستشفى ابن سينا التعليمي، وكذلك من وحدة هشاشة العظام في مستشفى السلام العام في مدينة الموصل في الفترة الواقعة ما بين 15 تشرين الأول 2009 ولغاية 30 أيلول 2010 (الجميلي، 2010). تم إجراء جميع التحليلات الإحصائية باستخدام برنامج IBM SPSS Statistics v26 (IBM Corp. Released 2019)، وبرنامج R v4.1.3 (R Core Team, 2022).

6.1. وصف واستكشاف البيانات احصائياً:

ويمثل المتغير المعتمد القياسات الخاصة بالكثافة المعدنية للعظام Bone Mineral Density (BMD) للأشخاص الخاضعين لاختبار دكسا في المستشفيين المذكورين. علما أن اختبار دكسا يعني اختبار امتصاص الأشعة السينية ثنائية الطاقة Dual Energy X-ray Absorptiometry (DEXA). ويقيس هذا الاختبار مستوى المعادن المهمة في تركيب العظام باستخدام الاشعة السينية، ومنه يمكن الاستدلال على مستوى تطور المرض لدى الاشخاص الخاضعين للاختبار. ويوضح الجدول (1) المتغيرات الداخلة في الدراسة.

جدول رقم (1): المتغيرات الداخلة في الدراسة

ت	المتغيرات	الرمز	نوع المتغير	الفئات
1.	**الكثافة المعدنية للعظام BMD**	Y	مستمر، أو فئوي (رتبي أو أسمي)	1 = طبيعي 2 = ترقق العظام 3 = هشاشة العظام
2.	**العمر Age**	X_1	فئوي	0 = أقل من 60 سنة 1 = 60 سنة فأكثر
3.	**عنوان السكن Address**	X_2	فئوي	0 = الريف 1 = الحضر
4.	**المهنة Job**	X_3	فئوي	0 = لا تعمل 1 = تعمل
5.	**مقياس كتلة الجسم BMI**	X_4	مستمر	---
6.	**التدخين Smoking**	X_5	فئوي	0 = كلا 1 = نعم
7.	**التعرض لأشعة الشمس Sun Exposure**	X_6	فئوي	0 = كلا ،لا تعمل 1 =نعم ،تعمل
8.	**عدد مرات الحمل Pregnancy**	X_7	فئوي	0 = أقل من 5 مرات 1 = 5 مرات فأكثر
9.	**سن اليأس Menopause**	X_8	فئوي	1 = لم تبلغ سن اليأس

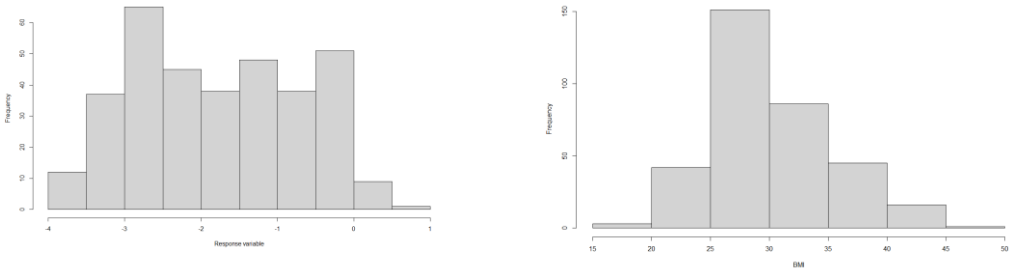
				2 = بلغت قبل سن الخمسين 3 = بلغت بعد سن الخمسين
.10	**موانع الحمل Contraceptive**	X_9	فئوي	0 = كلا 1 = نعم
.11	**الكسور Fractures**	X_{10}	فئوي	0 = لا يوجد 1 = يوجد كسور
.12	**تاريخ المرض Family History**	X_{11}	فئوي	0 = كلا 1 = نعم
.13	**الأدوية المنشطة Steroids**	X_{12}	فئوي	0 = كلا 1 = نعم
.14	**شرب الحليب Milk**	X_{13}	فئوي	1 = كلا 2 = قدح واحد في اليوم 3 = أكثر من قدح واحد في اليوم
.15	**المشروبات الغازية Beverages**	X_{14}	فئوي	1 = كلا 2 = قدح واحد في اليوم 3 = أكثر من قدح واحد في اليوم

وتم إجراء احصاءات وصفية بسيطة لمتغير الاستجابة المستمر ومتغير كتلة الجسم وكما موضح في الجدول (2). وأيضا تم رسم التوزيع التكراري (histogram) لبيانات المتغيرين المذكورين.

جدول (2): الوصف الاحصائي للمتغيرات المستمرة في بيانات الدراسة

Variables	Minimum	Maximum	Mean	SD	Variance
BMD (continuous)	-3.99	.90	-1.7684	1.10164	1.214
BMI	19	48	30.65	5.054	25.547

من الجدول (2) يبين الحد الادنى والحد الاعلى وقيمة الوسط والانحراف المعياري والتباين للمتغير BMD والمتغير BMI، وكانت القيمة التباين لمتغير BMD منخفضة هذا يدل على وجود تجانس أو تقارب بين قيم المتغير BMD. واضح أنه من المدرجين في الشكل (1) أن توزيع كلا المتغيرين متناظر نوعا ما. مما يعني توفر خاصية التوزيع الطبيعي لبيانات المتغير المعتمد، خصوصا مع توفر عدد كبير من المشاهدات.



الشكل (1): المدرجين التكراريين لمتغير كتلة الجسم ومتغير الاستجابة

ولغرض التأكد من عدم وجود مشكلة ترابط خطي بين أعمدة مصفوفة البيانات $\boldsymbol{X}$، تمّ حساب رتبة المصفوفة. وكانت رتبة المصفوفة تساوي 14، لذلك يمكن القول بأن أعمدة المصفوفة $\boldsymbol{X}$ مستقلة خطيا. حيث أن رتبة المصفوفة تمثل عدد الأعمدة المستقلة خطيا فيها (Strang, 1993). تم أيضا حساب عامل تضخم التباين للمتغيرات التوضيحية الداخلة في الدراسة. ويحسب من خلال المعادلة $(VIF)_j = \left(1 - R_j^2\right)^{-1}$، حيث أن R_j^2 تمثل معامل التحديد المتعدد للأنموذج الذي يعتبر أن المتغير X_j معتمداً على باقي المتغيرات التوضيحية (Kutner, et. al., 2004). وكانت قيم معاملات تضخم التباين كالآتي:

جدول (3): عوامل تضخم التباين للمتغيرات التوضيحية الداخلة في دراسة

Variable	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{X_4}$	$\boldsymbol{X_5}$	$\boldsymbol{X_6}$	$\boldsymbol{X_7}$
VIF	1.394	1.073	1.196	1.112	1.079	1.061	1.119
Variable	$\boldsymbol{X_8}$	$\boldsymbol{X_9}$	$\boldsymbol{X_{10}}$	$\boldsymbol{X_{11}}$	$\boldsymbol{X_{12}}$	$\boldsymbol{X_{13}}$	$\boldsymbol{X_{14}}$
VIF	1.300	1.055	1.156	1.030	1.044	1.063	1.111

وبما أن جميع قيم عوامل تضخم التباين أقل من 5، فيمكن القول بأنه لا توجد مشكلة تعدد العلاقة الخطية Multicollinearity problem بين المتغيرات التوضيحية (Forthofer, et. al; 2006).

6.2. بناء ومقارنة النماذج:

تمّ بناء انموذج الانحدار اللوجستي الرتبي على اعتبار أن متغير الاستجابة متغيرا فئويا قابلا للترتيب. حيث تم تقدير معالم الانموذج واختبارالفرضية الآتية للوقوف على معنوية كل متغير توضيحي:

$$H_0: \beta_j = 0 \quad vs. \quad H_A: \beta_j \neq 0$$

جدول (4): انموذج الانحدار اللوجستي الرتبي

Variables	Coefficients	Estimates	Std. Error	t value	p value
1|2	α_1	-2.500	0.942	-2.654	0.008
2|3	α_2	-0.121	0.931	-0.130	0.896
Age (>60)	β_1	0.812	0.275	2.946	0.003
Address	β_2	0.057	0.417	0.136	0.892
Job	β_3	-0.293	0.404	-0.725	0.468
BMI	β_4	-0.083	0.024	-3.515	0.000
Smoking	β_5	1.821	0.708	2.572	0.010
Sun Exposure	β_6	-0.863	0.301	-2.864	0.004
Pregnancy (>5)	β_7	0.653	0.245	2.662	0.008
Menopause 2	$\beta_{8.2}$	1.891	0.333	5.683	0.000
Menopause 3	$\beta_{8.3}$	1.356	0.326	4.163	0.000
Contraceptive	β_9	-0.318	0.262	-1.215	0.225
Fracture	β_{10}	1.493	0.381	3.919	0.000
Family History	β_{11}	0.572	0.587	0.973	0.330
Steroids	β_{12}	0.685	0.530	1.292	0.196
Milk 2	$\beta_{13.2}$	-0.399	0.379	-1.054	0.292
Milk 3	$\beta_{13.3}$	-2.134	1.037	-2.058	0.040
Beverages 2	$\beta_{14.2}$	0.945	0.335	2.820	0.005
Beverages 3	$\beta_{14.3}$	1.295	0.660	1.962	0.050

يوضح الجدول (4) المعالم المقدرة للانموذج اللوجستي الرتبي مع اختبار t للوقوف على معنوية كل معلمة من معالم الانموذج. حيث أن المتغيرات (العمر، BMI، التدخين، التعرض للأشعة الشمس، الحمل، سن اليأس، الكسور، شرب اكثر من كوب واحد من الحليب، المشروبات الغازية) معنوية ولها تأثير على الاصابة بمرض هشاشة العظام، وذلك لكون أن الـ p-value أقل من 0.05. من ناحية أخرى، فإن المتغيرات (السكن، العمل، موانع الحمل، تاريخ الكسور، الادوية المنشطة، شرب كوب واحد من الحليب) غير معنوية اي ليس لها تأثير على هشاشة العظام. علما أن المتغيرات التوضيحية ذات الفئات أكثر من اثنتين تم معاملتها كمتغيرات وهمية وحسب عدد الفئات في المتغير لتوضيحي.

جدول (5): نسبة الارجحيات (Odds Ratio) وحدود الثقة الخاصة بها للنموذج اللوجستي

Coefficients	Odds Ratios	Lower Limits	Upper Limits
Age (>60)	2.252	1.317	3.883
Address	1.059	0.466	2.404
Job	1.34	0.607	2.97
BMI	0.92	0.878	0.963
Smoking	6.177	1.697	29.679
Sun Exposure	0.422	0.231	0.757
Pregnancy (>5)	1.921	1.191	3.118
Menopause 2	6.626	3.484	12.859
Menopause 3	3.881	2.063	7.414
Contraceptive	0.728	0.435	1.215
Fracture	4.45	2.156	9.678
Family History	1.776	0.569	5.796
Steroids	1.983	0.716	5.797
Milk 2	0.671	0.317	1.406
Milk 3	0.118	0.012	0.807

Beverages 2	2.573	1.345	5.022
Beverages 3	3.644	1.081	15.047

من الجدول (5) ان احتمالية الاصابة بمرض هشاشة العظام مرتبطة طرديا مع كل من المتغيرات (العمر، السكن، العمل، التدخين، الحمل، سن اليأس، الكسور، تاريخ المرض، والأدوية المنشطة، المشروبات الغازية). بينما تكون احتمالية الاصابة بمرض هشاشة العظام مرتبطة عكسيا مع كل من المتغيرات (BMI، التعرض للأشعة الشمس، موانع الحمل، شرب الحليب). مع ذلك، فإن المتغيرات (السكن، العمل، موانع الحمل، تاريخ المرض، الأدوية المنشطة) تكون ذات تأثير غير مهم وذلك لكون أن حدود الثقة الخاصة بنسبة الارجحيات تضمنت العدد واحد ضمن مداها.

وتمّ في هذا الجانب أيضا تقدير انموذج الانحدار الخطي المتعدد وذلك من خلال تقدير دالة انحدار المتغيرات التوضيحية على المتغير المعتمد قبل تحويله إلى فئات، أي ان متغير الاستجابة يعتبر متغيرا مستمرا في هذه الحالة. ثم بعد ذلك، تم بناء نموذج انحدار خطي متعدد لانحدار المتغيرات التوضيحية على متغير الاستجابة وذلك بعد تحويله إلى فئات واعتباره متغيراً كمياً متقطعاً.

جدول (6): انموذج الانحدار الخطي المتعدد باعتبار أن متغير الاستجابة هو متغير مستمر

Variables	Coefficients	Estimates	Std. Error	t value	p value
Intercept	β_0	-1.782	0.420	-4.245	0.000
Age (>60)	β_1	-0.273	0.122	-2.242	0.026
Address	β_2	-0.180	0.182	-0.992	0.322
Job	β_3	0.047	0.179	0.263	0.793
BMI	β_4	0.028	0.010	2.744	0.006
Smoking	β_5	-0.617	0.238	-2.588	0.010
Sun Exposure	β_6	0.271	0.126	2.141	0.033
Pregnancy (>5)	β_7	-0.209	0.109	-1.917	0.056
Menopause 2	$\beta_{8.2}$	-0.779	0.142	-5.466	0.000
Menopause 3	$\beta_{8.3}$	-0.545	0.144	-3.784	0.000
Contraceptive	β_9	0.108	0.116	0.925	0.355
Fracture	β_{10}	-0.542	0.145	-3.737	0.000
Family History	β_{11}	-0.245	0.254	-0.965	0.335
Steroids	β_{12}	-0.211	0.211	-0.999	0.318
Milk 2	$\beta_{13.2}$	0.085	0.167	0.510	0.611
Milk 3	$\beta_{13.3}$	0.594	0.421	1.411	0.159
Beverages 2	$\beta_{14.2}$	-0.391	0.142	-2.75	0.006
Beverages 3	$\beta_{14.3}$	-0.608	0.234	-2.598	0.010

الجدول (6) أعلاه يوضح مقدرات معالم انموذج الانحدار الخطي المتعدد عند اعتبار أن Y متغيرا مستمراً. وأيضا تم إدراج نتائج اختبار t للوقوف على معنوية معاملات الانحدار. حيث كانت المتغيرات (BMI، التدخين، اشعة الشمس، سن اليأس، الكسور، المشروبات الغازية) معنوية اي لها تأثير على هشاشة العظام. بينما كانت المتغيرات (عنوان السكن، المهنة، الحمل، موانع الحمل، تاريخ المرض، الادوية المنشطة، شرب الحليب) غير معنوية وليس لها تأثير على هشاشة العظام، وذلك وفقا للانموذج الخطي المتعدد.

جدول (7): انموذج الانحدار الخطي المتعدد باعتبار أن متغير الاستجابة هو متغير كمي متقطع

Variables	Coefficients	Estimates	Std. Error	t value	p value
Intercept	β_0	2.301	0.282	8.145	0.000
Age (>60)	β_1	0.247	0.083	2.966	0.003
Address	β_2	0.013	0.125	0.101	0.920
Job	β_3	-0.094	0.123	-0.763	0.446
BMI	β_4	-0.023	0.007	-3.234	0.001
Smoking	β_5	0.422	0.163	2.586	0.010
Sun Exposure	β_6	-0.262	0.087	-3.034	0.003
Pregnancy (>5)	β_7	0.19	0.075	2.55	0.011
Menopause 2	$\beta_{8.2}$	0.592	0.098	6.068	0.000
Menopause 3	$\beta_{8.3}$	0.441	0.099	4.474	0.000
Contraceptive	β_9	-0.076	0.08	-0.95	0.343
Fracture	β_{10}	0.419	0.099	4.223	0.000

Family History	β_{11}	0.207	0.174	1.191	0.235
Steroids	β_{12}	0.146	0.144	1.014	0.311
Milk 2	$\beta_{13.2}$	-0.113	0.114	-0.991	0.322
Milk 3	$\beta_{13.3}$	-0.526	0.288	-1.824	0.069
Beverages 2	$\beta_{14.2}$	0.258	0.097	2.653	0.008
Beverages 3	$\beta_{14.3}$	0.317	0.16	1.982	0.048

الجدول (7) أعلاه يوضح مقدرات معالم انموذج الانحدار الخطي المتعدد عند اعتبار أن النسخة الفئوية من متغير الاستجابة Y متغيرا كميا متقطعاً. وأيضا تم إدراج نتائج اختبار t للوقوف على معنوية معاملات الانحدار. حيث كانت المتغيرات (العمر، BMI، التدخين، التعرض لأشعة الشمس، الحمل، سن اليأس، الكسور، شرب اكثر من كوب واحد من الحليب، المشروبات الغازية) معنوية ولها تأثير على هشاشة العظام، في حين المتغيرات (عنوان السكن، المهنة، موانع الحمل، تاريخ المرض، الأدوية المنشطة، شرب كوب واحد من الحليب) غير معنوية وليس لها تأثير على هشاشة العظام.

ولغرض المقارنة بين النماذج الثلاثة، تمّ اعتماد معياري أكياكي AIC والمعلومات البيزي BIC، حيث يمكن ملاحظة النتائج في الجدول (8). علماً أن الانموذج المفضل وفقا لهذين المعيارين هو الانموذج الذي يحقق أقل قيمة في المعيارين. ونلاحظ من الجدول (8) أعلاه ان أنموذج الانحدار اللوجستي الرتبي يعطي أقل قيمة AIC و BIC، وهذا يدل على انه أفضل أنموذج لتمثيل هذه البيانات، يليه أنموذج الانحدار الخطي المتعدد (المتقطع)، وأخيراً أنموذج الانحدار الخطي المتعدد (المستمر).

جدول (8): مقارنة النماذج أعلاه باستخدام معيار معلومات أكياكي ومعيار المعلومات البيزي

النماذج	عدد المعالم المقدرة	AIC	BIC
الانحدار اللوجستي الرتبي	19	611.414	684.387
الانحدار الخطي المتعدد (مستمر)	18	935.524	1008.497
الانحدار الخطي المتعدد (فئوي)	18	674.605	747.577

من الناحية الطبية، فإن الدراسات الحيوية تشير إلى أن المتغيرات الداخلة في الدراسة تكون ذات تأثير متفاوت على ارتفاع نسب الاصابة بمرض هشاشة العظام. فبعض المتغيرات المعتمدة في الدراسات تصنف على أنها عوامل طردية، أي أن زيادتها تزيد من احتمالية الاصابة بمرض هشاشة العظام، أو على الأقل تؤدي إلى ترقق العظام. والمتغيرات المصنفة عالمياً على أنها ذات تأثير طردي على فرض ظهور المرض هي (العمر خصوصا بعد الستين عاما، التدخين، الحمل لأكثر من خمسة أطفال، بلوغ سن اليأس، الكسور، المشروبات الغازية، تاريخ المرض، الأدوية المنشطة). فيما يصنف عدد من المتغيرات التوضيحية على أنها متغيرات ذات تأثير سلبي على المرض، أي تنخفض فرص الاصابة بالمرض كلما ارتفعت قيمة أو فئة المتغير التوضيحي المقصود. (مؤشر كتلة الجسم BMI، التعرض لأشعة الشمس). وأخيرا، فإن الدراسات السريرية تضاربت بشأن تأثير بقية المتغيرات التوضيحية (العنوان، المهنة، موانع الحمل، شرب الحليب) فيما إذا كان سلبيا أم ايجابيا على تطور المرض. وتحديدا فيما يخص موانع الحمل وشرب الحليب، حيث أشارت الدراسات أن استخدام أنواع محددة من موانع الحمل مرتبط بانخفاض كثافة العظم المعدنية (وهو المتغير المعتمد في هذه الدراسة). بينما أدى استخدام أنواع أخرى من أدوية منع الحمل إلى تحسين كثافة العظام لدى النساء. وكذا الحال بالنسبة للحليب، حيث تشير الدراسات أن الحليب الغني بالبروتينات مرتبط بتدهور حالة المرض، فيما أشارت دراسات أخرى أن الحليب الخالي من اللاكتوز مفيد لصحة العظام (Kröger, et al; 1994).

7. الاستنتاجات:

وبناءاً على ما تقدم، نلاحظ توافق النماذج الثلاثة في تشخيص اتجاه العلاقة (إيجابي سلبي) بين المتغيرات التوضيحية والمتغير المعتمد. كما ويلاحظ أيضا تطابقها مع ما هو متعارف عليه دولياً في مجال الدراسات الطبية الخاصة بمرض هشاشة العظام. وبالرغم من توافق النماذج الثلاثة في تحديد اتجاه العلاقة، إلا أن الأفضلية تميل لصالح الانموذج اللوجستي الرتبي، وذلك لكونه لا يتطلب فروضا مسبقة بخصوص توزيع الاخطاء، وأيضا لتفوقه في معايير المقارنة المذكورة آنفاً. وهذا يدل على أفضلية تحويل المتغيرات المستمرة الى متغيرات فئوية رتبية قبل تحليلها. فعندما يكون الاهتمام في مدى محدد من البيانات الخاصة بمتغير الاستجابة، فإن تحويل المتغير المستمر إلى متغير فئوي (رتبي أو اسمي) يعطي أفضلية ودقة لتشخيص وتفسير العلاقة بين المتغيرات التوضيحية والمتغير المعتمد مقارنة في حالة اعتماده كمتغير مستمر. وذلك واضح من خلال مقارنة نتائج هذه الدراسة مع الدراسات العالمية المنجزة بخصوص مرض هشاشة العظام. من ناحية أخرى، تتفوق طريقة الانحدار اللوجستي الرتبي على الطرق الأخرى المتعارف

عليها في نمذجة البيانات المتضمنة على متغير استجابة رتبي. ولوحظ أيضا أن دقة تشخيص العلاقة بين المتغيرات التوضيحية الفئوية والمتغير المعتمد الرتبي تتأثر بعدد التكرارات في الجداول التقاطعية بين كل متغير توضيحي مع المتغير المعتمد، وهذا يفسر سبب عدم معنوية (او تخبط) علاقة بعض المتغيرات التوضيحية والتي أثبتت الدراسات السابقة تأثيرها فعليا على مرض هشاشة العظام.

8. المصادر:

1. AL-Jumayli ,Hana Husian,(2010)," Evaluation of osteoporosis among women in the city of Mosul",MSc. Thesis, College of Nursing ,University of Mosul, Mosul ,Iraq.
2. AL-Rawi, Kashe Mahmoud,(1987),"Introduction to Regression Analysis", Directorate of Book House for Printing and Publishing, University of Mosul .,Mosul, Iraq.
3. Kadem, Amori Hade, (1988) " Introduction to Linear Regression Analysis ",Baghdad University Iraq.
4. Agresti, A. (2003). Categorical data analysis. John Wiley & Sons.
5. Ananth, C. V., & Kleinbaum, D. G. (1997). Regression models for ordinal responses: a review of methods and applications. International journal of epidemiology, 26(6), 1323-1333.
6. Armstrong, B. G., & Sloan, M. (1989). Ordinal regression models for epidemiologic data. American Journal of Epidemiology, 129(1), 191-204.
7. Forthofer, R. N., Lee, E. S., & Hernandez, M. (2006). Biostatistics: a guide to design, analysis and discovery. Elsevier.
8. IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.
9. Kröger, H., Tuppurainen, M., Honkanen, R., Alhava, E., & Saarikoski, S. (1994). Bone mineral density and risk factors for osteoporosis—a population-based study of 1600 perimenopausal women. Calcified Tissue International, 55(1), 1-7.
10. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Wasserman, W. (2004). Applied linear regression models. New York: McGraw-Hill/Irwin.
11. Lall, R. (2004). The application of ordinal regression models in quality of life scales used in gerontology (Doctoral dissertation, University of Sheffield).
12. Siegel S, Castellan N. Non-parametric Statistics for the Behavioral Sciences. Second Edition. McGraw-Hill International Editions, 1988.
13. Scott, S. C., Goldberg, M. S., & Mayo, N. E. (1997). Statistical assessment of ordinal outcomes in comparative studies. Journal of clinical epidemiology, 50(1), 45-55.
14. Strang, G., (1993). Introduction to linear algebra (Vol. 3). Wellesley, MA: Wellesley-Cambridge Press.
15. Strömberg, U. (1996). Collapsing ordered outcome categories: a note of concern. American journal of epidemiology, 144(4), 421-424.
16. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Statistical Analysis of Ordinal Response Variable: A Comparative Study

Leqaa Yaseen Ahmad Zaid Thariq Saleh Alkhaledi
leqaa.20csp116@student.uomosul.edu.iq zaid.**alkhaledi@uomosul.edu.iq**
Department of Informatics & Statistic, College of Computer & Mathematical Science, University of Mosul, Mosul, Iraq

Abstract
Response variables in biological phenomena vary between three types: numerical response variables, ordinal categorical response variables, and nominal categorical response variables. In statistical studies, handling ordinal variables varies in accordance with the perspective of the statistical approach to the response variable. Ordinal variables can be adopted as nominal categorical variables, which neglect the ordinal property of the categories. Ordinal variables can also be treated.

as an ordinal categorical variable (discrete variable), in which case the ranking information can be utilized in establishing the predicted models. In this study, the most important statistical methods that can be used to analyze data with an ordinal response variable have been investigated. Among these methods are the Multiple Regression Method, and The Ordinal Logistic Regression Method. The mechanism of building models and parameter estimations were theoretically exhibited, as well as reading the statistical significance of the regression coefficients in all the models in the study. The application was carried out on a real sample of patients with osteoporosis. Where multiple models were built to determine the most important factors affecting the likelihood of developing the disease. The best model was diagnosed according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The results of the statistical analysis demonstrated the superiority of the ordinal logistic regression model over the multiple linear regression model in its explanation of the relationship between the response variable and the covariates.

Keywords: Ordinal Response Variable, Multiple Regression, Ordinal Logistic Regression, Categorical Variables.