AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

تحسين جودة التوفيق لأنموذج الاستجابة الثنائي مع التطبيق العملي على المصابين بوباء Covid-19

أ.م.د. مشتاق كريم عبد الرحيم	علي محمد جواد
mushtag.k@uokerbala.edu.iq	ali.jawad@uokerbala.edu.iq

قسم الاحصاء - كلية الادارة والاقتصاد - جامعة كربلاء، كربلاء، العراق

المستخلص

تعد دراسة المتغيرات التابعة الثنائية من العمليات المهمة في وقتنا الحاضر لكثرة الظواهر التي توصف بهذه الطريقة ويعد انموذج الاستجابة الثنائي من اهم الوسائل لتمثيل هذا النوع من الظواهر، كما تعد عملية اختيار المتغيرات المستقلة التي تؤثر على المتغيرات التابعة الثنائية ضرورية جدا. تضمنت الدراسة استعمال اسلوبين، الاول هو الاسلوب التجريبي (محاكاة) والاسلوب الثاني هو التطبيقي على المصابين بوباء كورونا وقد تم استعمال ثلاث طرائق لاختيار افضل انموذج استجابة ثنائي وهي طريقة التقدم الامامي، و طريقة الحذف العكسي ،و طريقة مقترحة و هي طريقة التحليل العاملي مع اختبار جودة التوفيق للأنموذج النهائي الناتج عن كل طريقة باستعمال اختبارين هما اختبار الانحراف (D) واختبار هوزمر- ليمشو (H&L) و تمت المقارنة بين النماذج النهائية الناتجة عن الطرائق الثلاث باستعمال ثلاثة معايير و هي معيار نسبة الامكان الاعظم(MLR) و معيار المعلومات لـ اكاكي (AIC) و معيار المعلومات البيزي (BIC) و بينت النتائج ان للعوامل الناتجة عن التحليل العاملي قدرة على تقليل نسبة الامكان الاعظم افضل من اي تشكيلة مختارة من المتغيرات المستقلة باستعمال احدى طرائق الاختيار الاخرى، ومن ثم فان معايير اختيار افضل انموذج التي تعتمد على نسبة الامكان الاعظم وهما (AIC , BIC) تعطي افضلية للعوامل الناتجة عن طريقة التحليل العاملي على حساب المتغيرات المستقلة الناتجة عن طريقتي التقدم الامامي والحذف العكسي. كما بينت النتائج ايضا انه كلما كان عدد المتغيرات المستقلة التي لها تأثير معنوي على الانموذج كثيرة كلما اعطت العوامل نتائج افضل حسب المعايير المستخدمة. و أظهرت النتائج الاسلوب التطبيقي ان المتغيرين المستقلين عمر المريض والتدخين لهما اكثر تأثير على حياة المصابين بوباء كورونا.

معلومات البحث

تواريخ البحث

تاريخ تقديم البحث: 2021/8/16
تاريخ قبول البحث: 2021/10/29
تاريخ رفع البحث على الموقع: 2022/12/31

الكلمات المفتاحية

جودة التوفيق، انموذج الاستجابة الثنائي، وباء كورونا.

للمراسلة:
علي محمد جواد
ali.jawad@uokerbala.edu.iq

doi: https://doi.org/10.55562/jrucs.v52i1.539

1.1. المقدمة

يعد انموذج الاستجابة الثنائي من النماذج اللاخطية التي تصف العلاقة بين متغير تابع ثنائي القيمة اي يأخذ قيمتين هما الصفر لاحتمال عدم حدوث حدث معين و الواحد الصحيح لاحتمال حدوث ذلك الحدث والمتغيرات المستقلة التي تأخذ قيم وصفية او كمية، ويمكن تحويل انموذج الاستجابة الثنائي الى الصيغة الخطية وذلك باستعمال تحويل اللوجت، كما تعد طريقة الامكان الاعظم من اهم الطرائق التي تستعمل في تقدير معالمه. ولزيادة كفاءة الانموذج وخصوصا عندما يكون عدد المتغيرات المستقلة كبيراً يلجأ الباحثون الى استعمال طرائق اختيار المتغيرات المستقلة التقليدية مثل طريقة التقدم الامامي وطريقة الحذف العكسي ولكن هذه الطرائق مع انموذج الاستجابة الثنائي تكون قليلة الفائدة واحيانا غير دقيقة لأنها تعتمد على مقاييس قد تكون غير مهمة في هذا النوع من النماذج. لذلك تم اقتراح طريقة ثالثة لاختيار المتغيرات المستقلة وهي طريقة التحليل العاملي لأنها تعتمد على اساس بعيد عن اي مقياس في انموذج الاستجابة الثنائي.

1.2. أهمية البحث

تكمن اهمية هذا البحث في جانبيه النظري والتطبيقي في التعرف على العوامل او المتغيرات المستقلة التي لها تأثير كبير في المتغير التابع محل الدراسة فدراسة جميع العوامل او المتغيرات يحتاج وقت، جهد، مال.

1.3. هدف البحث

يهدف البحث الى تحسين جودة التوفيق لأنموذج الاستجابة الثنائي وذلك باستعمال اسلوب احصائي هو التحليل العاملي عن طريق اختزال عدد المتغيرات المستقلة الى عوامل يكون عددها اقل من عدد المتغيرات المستقلة اذ ان كل عامل يمثل مجموعة من المتغيرات المستقلة و يمثل متغير مستقل واحد في الانموذج الجديد. ومقارنة نتائجها مع طرائق اختيار افضل انموذج مثل طريقة التقدم الامامي والحذف العكسي عن طريق استعمال بعض معايير المفاضلة بين النماذج.

1.4. مشكلة البحث

تكمن مشكلة البحث في عدد المتغيرات المستقلة إذ ان كل ظاهرة تدرس هناك العديد من المتغيرات التي تؤثر بها وهذه المتغيرات بعضها له تأثير كبير وبعضها بدرجة اقل إذ يمكن ان يستبعد من الانموذج لأسباب عديدة منها ان تأثيره يمكن ان يكون مماثلاً لتأثير متغيرات اخرى في الانموذج.

2. الجانب النظري

2.1. مفهوم انموذج الاستجابة الثنائي

ان العديد من الظواهر الطبيعية عند دراسة سلوكها نجدها تسلك سلوكاً غير خطي ، ولكي نحلل تلك الظواهر فأننا نستعمل النماذج اللاخطية ، كما ان بعض الظواهر تأخذ قيمتين وهذا ما يسمى بثنائي الاستجابة وهذه القيم ليس لها معنى قائم بحد ذاته انما هي مجرد دلالة على وجود الحدث من عدمه . ان الانموذج الملائم لهذا النوع من البيانات هو انموذج الاستجابة الثنائي والذي يمكن تعريفه بأنه اسلوب احصائي لوصف العلاقة بين المتغير التابع الوصفي الذي يأخذ قيمتين هي الواحد الصحيح لحدوث الحدث والصفر لعدم حدوث الحدث ومتغيراً واحداً او اكثر من متغير مستقل وصفي او كمي ، ويفترض وجود علاقة خطية بين لوغاريتم نسبة الرجحان(Odd ratio) والمتغيرات المستقلة و عدم وجود علاقة تامة أو شبه تامة بين المتغيرات المستقلة مع بعضها البعض لتجنب الوقوع في مشكلة التعدد الخطي بين المتغيرات المستقلة وللحصول على تقديرات غير متحيزة للمعالم واخطاء معيارية صغيرة

يعد فرض المتغير التابع هو الفرض الاساس الذي يبنى عليه انموذج الاستجابة الثنائي إذ يرمز لوقوع الحدث بالرمز (1) وباحتمال (P) وعدم وقوعه بالرمز (0) وباحتمال (1-P) [8] و حدوثهما من عدمه يعد دالة للمتغير او المتغيرات المستقلة . و المتغير التابع يتبع توزيعاً احتماليا متقطعا وهو توزيع برنولي ويمكن كتابته بدالة وفق الانموذج اللوجستي (logistic model) بالشكل الآتي [11]:

$$\boldsymbol{P = \frac{e^{\beta_0+\sum \beta_j x_j}}{1 + e^{\beta_0+\sum \beta_j x_j}}} \qquad (1)$$

$$\boldsymbol{1 - P = 1 - \frac{e^{\beta_0+\sum \beta_j x_j}}{1 + e^{\beta_0+\sum \beta_j x_j}}}$$

$$\boldsymbol{q = \frac{1}{1 + e^{\beta_0+\sum \beta_j x_j}}} \qquad (2)$$

اذ ان:

$$\boldsymbol{q = 1 - p, \quad j=1,2,3,\ldots\ldots m}$$

(m) هي عدد المتغيرات المستقلة، β_0= الحد الثابت و β_j = معالم الانموذج وتكون مجهولة القيم يتم تقديرها من الانموذج

2.2. طرائق تقدير معالم انموذج الاستجابة الثنائي

تختلف طريقة تقدير معالم انموذج الاستجابة الثنائي عنه في الانحدار الخطي البسيط والمتعدد، ففي حالة انموذج الانحدار الخطي البسيط والمتعدد يتم التقدير باستعمال طريقة المربعات الصغرى الاعتيادية (Ordinary Least squares method) التي تهدف الى تقليل مجموع مربعات الخطأ الى اقل ما يمكن، اي يتم تقدير قيم المعالم التي ينتج عنها اقل مجموع مربعات خطأ بين قيم المتغير التابع الحقيقية والقيم التي تم تقديرها [8]. اما في حالة انموذج الاستجابة الثنائي (الذي يكون المتغير التابع ثنائي القيمة) فأنه من غير المفضل استعمال طريقة المربعات الصغرى الاعتيادية ، لأنه يقلل من مجموع مربعات الخطأ بين القيم المشاهدة والمتوقعة للمتغير التابع وهذا يعطي تقديرات لا تتمتع بالكفاءة في حالة كون المتغير التابع ثنائي القيمة [14]. كما انه في حالة بيانات المتغير التابع الثنائية والذي يتوزع توزيع برنولي وهو توزيع متقطع و الخطأ هو الاخر يمتلك توزيعاً متقطعاً هو توزيع ثنائي الحدين ومن ثم نواجه مشكلة ان الخطأ لا يتوزع توزيعاً طبيعياً واذا فرضنا انه يتوزع توزيعاً طبيعياً فان التباين يكون غير متجانس.

وتعد طريقة الامكان الاعظم (Maximum Likelihood method) واحدة من الطرائق التي تستعمل لتقدير المعالم في مثل هذه النماذج الرياضية ، ومن الملاحظ ان طريقتي الامكان الاعظم والمربعات الصغرى مع انهما طريقتان مختلفتان في

الاسلوب إلا انهما يعطيان النتائج نفسها في انموذج الانحدار الخطي البسيط والمتعدد بشرط ان بيانات المتغير التابع تتبع التوزيع الطبيعي ، ولهذا تعد طريقة الامكان الاعظم ملائمة للنماذج الخطية وغير الخطية وهذا ما يميزها عن طريقة المربعات الصغرى التي تعد ملائمة للنماذج الخطية فقط [12] .

2.3. اختبار جودة التوفيق للأنموذج الانحدار اللوجستي

2.3.1. اختبار كاي تربيع لنسبة الترجيح (الانحراف D)

يعتمد هذا الاختبار على مبدأ هو المقارنة بين القيم المشاهدة والقيم المتوقعة للأنموذج الذي يحتوي الحد الثابت فقط (constant) والأنموذج الذي يحتوي كافة المتغيرات إذ يعتمد في ترجيحه على الفرق بين نسبة الامكان الاعظم (MLR) للأنموذج الذي يحتوي الحد الثابت فقط و نسبة الامكان الاعظم للأنموذج الذي يحتوي على جميع المتغيرات [11] وحسب الفرضية الاحصائية الآتية:

H_0 : الانموذج المقدر غير معنوي
H_1: الانموذج المقدر معنوي (3)

ويمكن ان يحسب حسب الصيغة الآتية [11]:

$$D = MLR_{(\text{للانموذج مع المتغيرات})} - MLR_{(\text{للانموذج مع الحد الثابت فقط})} \quad (4)$$

إذ ان

D :الانحراف لنسبة الرجحان ويتبع توزيع كاي تربيع بدرجة حرية تساوي عدد المتغيرات المستقلة في الانموذج اذا كانت المتغيرات كمية وعدد اوجه المتغيرات المستقلة اذا كانت وصفية

$MLR_{(\text{للانموذج مع الحد الثابت فقط})}$: نسبة الامكان الاعظم للأنموذج الذي يحتوي الحد الثابت

$MLR_{(\text{للانموذج مع المتغيرات})}$: نسبة الامكان الاعظم للأنموذج الذي يحتوي على جميع المتغيرات

و إذا كانت قيمة (p-value) لـ (D)، اقل من مستوى المعنوية المحدد (α) نرفض فرضية العدم ويكون الانموذج المقدر معنوي.

2.3.2. اختبارهوزمر – ليمشو *Hossmer - Lemshow tast (H&L)*

هو احد اهم اختبارات جودة التوفيق لأنموذج الانحدار اللوجستي يعتمد على مدى قرب الاحتمالات المشاهدة والاحتمالات المتوقعة ويبنى على أساس تقسيم الحالات المدروسة الى عشر مجموعات على شكل أعمدة اما الصفوف فتقسم على اساس القيم المشاهدة للمتغير المعتمد وهما الصفر والواحد [9] وهو يختبر الفرضية الاحصائية الآتية:

H_0 : الانموذج المقدر يوافق البيانات بشكل جيد
H_1 : الانموذج المقدر لا يوافق البيانات بشكل جيد (5)

ويمكن ان يحسب من الصيغة الآتية [11]:

$$\mathbf{x^2} = \sum_{\mathbf{k=1}}^{\mathbf{10}} \frac{(\mathbf{O_{1k}} - \mathbf{n_k}\bar{\mathbf{p}}_\mathbf{k})^2}{\mathbf{n_k}\bar{\mathbf{p}}_\mathbf{k}(\mathbf{1} - \bar{\mathbf{p}}_\mathbf{k})} \quad (6)$$

O_{1k}: التكرار المشاهد للمجموعة (k) عند الاستجابة y=1

$\bar{p}_k$: متوسط الاحتمالات المتوقعه للمجموعة (k)

n_k : مجموع التكرار المشاهد في المجموعة (k) و يساوي $(O_{0k} + O_{1k})$ k=1,2,…..10

وتتم مقارنة قيمة (p-value) لـ (x^2) بدرجة حرية (k-2) مع مستوى المعنوية المحدد(α) فأذا كانت قيمة (p-value) اقل من مستوى المعنوية فيتم رفض فرضية العدم اما اذا كانت قيمة (p-value) اكبر من مستوى المعنوية فلا نرفض فرضية العدم وهذا يعني ان الانموذج يوافق البيانات بشكل جيد .

2.4. معايير اختيار افضل انموذج استجابة ثنائي

2.4.1. معيار نسبة الامكان الاعظم *Maximum Likelihood Ratio (MLR)*

يعتمد هذا المعيار في تقييمه لأداء الانموذج على امكانية حدوث نتائج البيانات التي اعتمدناها اذا ما عرفنا تقديراً لمعلمات الانموذج مع اختيار تقديراً للمعلمات التي تعظم امكانية الحصول على النتائج المشاهدة (المعالم المقدرة بطريقة الامكان الاعظم) وبما ان قيمة الامكان الاعظم دائما تكون صغيرة جداً وهي اقل من الواحد، لذلك نأخذ لها اللوغاريتم الطبيعي ونضربها (-2)

للتخلص من القيمة السالبة الناتجة من لوغاريتم دالة الامكان الاعظم ، وبهذا يكون معيار المفاضلة هو (2-) مضروب في لوغاريتم الامكان الاعظم والمسمى بنسبة الامكان الاعظم ويمكن حسابه كما في المعادلة الآتية:

$$\mathrm{MLR} = -2\ln\prod_{i=1}^{n}\hat{p}_i^{yi}\,(1-\hat{p}_i)^{1-yi} \qquad (7)$$

إذ ان $\hat{p}_i$ هي نسبة الاستجابة المتوقعة لكل مشاهدة من مشاهدات المتغير المعتمد [11] .
وكلما كانت قيمة دالة الامكان كبيرة تكون قيمة معيار نسبة الامكان الاعظم صغيرة يدل هذا على جودة الأنموذج المقدر اي ان كلما كانت قيمة معيار نسبة الامكان الاعظم اقل كان الانموذج افضل [9] .

2.4.2. معيار المعلومات البيزي (*BIC*) *Bayesian Information Criterion*

اقترح هذا المعيار من لدن (Schwarz) عام 1978 إذ تناول كيفية اختيار انموذج واحد من عدة نماذج غير متساوية بعدد المتغيرات المستقلة وذلك عن طريق ايجاد الحل البيزي لها (Bayes Solution) وتم توسعة الحل البيزي باستعمال نظرية بيز وحسب الصيغة الآتية [17] .

$$\boldsymbol{BIC = -2\,LL(\hat{\beta}) + k\,\log(n)} \qquad (8)$$

إذ ان:
$LL(\hat{\beta})$: لوغاريتم الامكان الاعظم .
k : عدد المعالم المقدرة في الانموذج مع الحد الثابت
إذ يتم اختيار الانموذج الذي تكون فيه قيمة هذا المعيار اصغر [5] .

2.4.3. معيار المعلومات لـ أكاكي (*AIC*) *Akaike's Information Criterion*

اقترح هذا المعيار من لدن (Akaike) وهو يعتمد بشكل أساس على مقياس K-L (kulback - Leibler) الذي يقيس المسافة أو المعلومات المفقودة بين التوزيع الحقيقي غير المعروف للمتغيرات العشوائية والتوزيع المفترض ويسمى معيار المعلومات (IC) والذي يمكن ان يحسب من الصيغة الآتية:

$$\boldsymbol{AIC = -2\,LL(\hat{\beta}) + 2K} \qquad (9)$$

إذ ان:
AIC : معيار المعلومات لأكاكي
$LL(\hat{\beta})$: لوغاريتم الامكان الاعظم K : عدد المعالم المقدرة مع الحد الثابت
ويتم اختيار الانموذج الذي يقلل من قيمة هذا المعيار **[5]**.

2.5. طرائق اختيار انموذج استجابة ثنائي

2.5.1. طريقة التقدم الامامي (*Forward Selection Method*)

تفترض هذه الطريقة في البداية ان الانموذج خال من المتغيرات المستقلة اي يحتوي على الحد الثابت (Constant) فقط ثم نبدأ بإدخال المتغيرات بشكل تدريجي، اذ يتم اختيار المتغير المستقل الذي له اقل (p-value) للفرق بين (MLR) للحد الثابت و(MLR) للحد الثابت وهذا المتغير بدرجة حرية تساوي واحد اذا كان المتغير كمي وعدد اوجه ذلك المتغير اذا كان وصفي بشرط ان تكون قيمة (p-value) اقل من مستوى المعنوية المحدد (α)، ويثبت المتغير في الانموذج اذا كان الانموذج معنوي حسب اختبار الانحراف (D) . ثم ننتقل الى الخطوة الثانية وهي اختيار المتغير الذي له اقل (p-value) للفرق بين (MLR) للأنموذج الناتج من الخطوة السابقة و (MLR) للحد الثابت مع المتغير المضاف من الخطوة السابقة واحد المتغيرات البقية، ويتم تثبيت المتغير اذا كان الانموذج معنوي حسب اختبار الانحراف (D)، ونستمر بإضافة المتغيرات التي لها اقل (p-value) ويتم التوقف عن الاضافة اذا كانت قيمة اصغر (p-value) للمتغيرات غير الداخلة في الانموذج اكبر من مستوى المعنوية المحدد (α) او كان الانموذج غير معنوي حسب اختبار الانحراف (D) [11].

2.5.2. طريقة الحذف العكسي (*Backward Elimination Method*)

على عكس الطريقة السابقة فان هذه الطريقة تبدأ بافتراض الانموذج يحتوي على جميع المتغيرات المستقلة ويتم حذف المتغير الذي له اكبر (p-value) للفرق بين (MLR) للأنموذج الذي يحتوي على جميع المتغيرات المستقل عدا احد المتغيرات و (MLR) للأنموذج الذي يحتوي على جميع المتغيرات ، بشرط ان تكون اكبر (p-value) اكبر من مستوى المعنوية المحدد (α)، يحذف المتغير اذا كان الانموذج للمتغيرات البقية معنوي حسب اختبار الانحراف (D)، وهكذا نستمر بحذف المتغيرات التي لها

اكبر (p-value) ويتم التوقف عن الحذف اذا كانت اكبر (p-value) للمتغيرات الباقية في الانموذج اصغر من مستوى المعنوية المحدد (α) او اذا كان الأنموذج غير معنوي حسب اختبار الانحراف (D) [11] .

2.5.3. طريقة التحليل العاملي *Factor Analysis Method*

هو اسلوب احصائي يهدف الى اختزال عدد كبير من المتغيرات الى عدد اقل من العوامل إذ يكون لكل عامل من هذه العوامل دالة تربطه مع بعض او كل المتغيرات ، وتتلخص فكرة التحليل العاملي عن طريق استخلاص مجموعة من العوامل لها ارتباط بالمتغيرات الاصلية إذ تفسر هذه العوامل اكبر قدر ممكن من التباين في المتغيرات الاصلية [10] وقد وردت العديد من التعاريف للتحليل العاملي منها. فقد عرفه (باهي وآخرون)[3] بأنه " اسلوب احصائي يساعد الباحث على دراسة المتغيرات المختلفة بقصد ارجاعها الى اهم العوامل التي اثرت فيها " وقد عرفاه (الأسدي و فارس) [1] بأنه " العملية التي يتم فيها تلخيص مجموعة كبيرة نسبياً من المتغيرات المترابطة بأقل عدد ممكن من العوامل غير المترابطة " كما عرفه (صفوت فرج) [6] بأنه " اسلوب احصائي يستعمل بيانات متعددة ارتبطت فيما بينها بدرجات مختلفة من الارتباط لتلخص في صورة تصنيفات مستقلة قائمة على اسس نوعية للتصنيف " ومهما تعددت التعريفات فأن التحليل العاملي هو اسلوب احصائي يهدف الى التعرف على عدد صغير من العوامل الكامنة (Latent Factors) التي تعبر بصورة جيدة عن علاقات متشابكة بين عدد كبير من المتغيرات . هذه المتغيرات يمكن مشاهدتها وقياسها ويرمز لها $(X_1, X_2, X_3, \dots \dots . X_m)$ ومن ثم يمكن تحويلها الى متغيرات قياسية (معيارية) $(Z_1, Z_2, Z_3, \dots \dots . Z_m)$ حسب الصيغة الآتية:

$$\boldsymbol{Z_{ij}} = \frac{\mathbf{X_{ij}} - \overline{\boldsymbol{X}}_{\boldsymbol{j}}}{\boldsymbol{S_j}} \qquad (10)$$

إذ ان s_j هي الانحراف المعياري للمتغير (j) و $\overline{X}_j$ هو المتوسط الحسابي للمتغير (j)

والصيغة العامة لمعادلة التحليل العاملي تشبه الى حد كبير الصيغة العامة لأنموذج الانحدار المتعدد، إذ ان كل متغير يمكن التعبير عنه بعلاقة خطية لجميع العوامل وكما في الانموذج الآتية [9] :

$$\left.\begin{aligned} \boldsymbol{z_1} &= \boldsymbol{a_{11}F_1 + a_{12}F_2 + \cdots \cdots + a_{1r}F_r + u_1} \\ \boldsymbol{z_2} &= \boldsymbol{a_{21}F_1 + a_{22}F + \cdots \cdots + a_{2r}F_r + u_2} \\ &\vdots \\ &\vdots \\ &\vdots \\ \boldsymbol{z_m} &= \boldsymbol{a_{m1}F + a_{m2}F_2 + \cdots \cdots + a_{mr}F_r + u_m} \end{aligned}\right\} \qquad (11)$$

إذ ان :

m هو عدد المتغيرات $\quad j = 1,2,3, \dots . m$ $\quad r$ هي عدد العوامل $\quad k = 1,2,3, \dots . r$

a_{kj} : تشبع الدرجة المعيارية للعامل (k) على الدرجة المعيارية للمتغير (j)

F_k : الدرجة المعيارية للعامل (k)

u_j : العامل الوحيد او النوعي وهي عوامل تختص بنوع واحد من المتغيرات اي يوجد فيها متغير واحد فقط وتكون غير مرتبطة فيما بينها من جهة وغير مرتبطة مع العوامل العامة من جهة اخرى.

ان العوامل الناتجة عن عملية التحليل العاملي ما هي ألا متغيرات مشابهة للمتغيرات الاخرى مع فرق بسيط وهو ان معظم المتغيرات يمكن ان تقاس بصورة مباشرة اما العوامل فهي متغيرات افتراضية تم اشتقاقها من تحليل بيانات مجموعة متغيرات تقاس قياساً مباشراً، إذ ان كل عامل من العوامل يمكن التعبير عنه بعلاقة خطية بدلالة بعض المتغيرات او جميعها وكما يلي [4]:

$$\left.\begin{aligned} \boldsymbol{F_1} &= \boldsymbol{b_{11}X_1 + b_{12}X_2 + \cdots + b_{1m}X_m} \\ \boldsymbol{F_2} &= \boldsymbol{b_{21}X_1 + b_{22}X_2 + \cdots + b_{2m}X_m} \\ &\vdots \\ &\vdots \\ &\vdots \\ \boldsymbol{F_r} &= \boldsymbol{b_{r1}X_1 + b_{r2}X_2 + \cdots + b_{rm}X_m} \end{aligned}\right\} \qquad (12)$$

إذ ان: F_k : العامل k $\quad b_{rj}$: وزن المتغير j على العامل r $\quad x_j$: المتغير j

$k = 1,2,3, \ldots . r \quad j = 1,2,3, \ldots . m$

ولتطبيق التحليل العاملي هناك مجموعة من الفروض الواجب توفرها.

1. يفترض ان تكون المتغيرات متوزعة توزيعاً طبيعياً وان لا يكون توزيعها ملتوياً التواء شديداً او متعدد المنوال ولاختبار التوزيع الطبيعي هناك العديد من الاختبارات و اشهرها اختبار كولمجروف - سميرنوف (kolmogorov- smirnov) وفي حالة الالتواء الشديد او التوزيع ليس طبيعياً نلجأ الى زيادة حجم العينة [4] .
2. العينة المسحوبة يجب ان تكون ممثلةً تمثيلاً صحيحاً للمجتمع الذي سحبت منه وغير متحيزة و على درجة كافية من التجانس والتي نحصل عليها عن طريق اختبار كايز- ماير – اولكن Kaiser-mayar- olkin (KMO) [9]
3. وجود علاقة بين المتغيرات حتى يمكن استنتاج مكونات مشتركة بين المتغيرات و العوامل وتجنب استعمال المتغيرات غير المترابطة ويمكن التأكد من وجود علاقة من معاملات الارتباط بين المتغيرات [4].
4. يقوم التحليل العاملي على امكانية تجميع المتغيرات وفقاً لمصفوفة الارتباطات بين المتغيرات ويجب ان لا تكون هذه المصفوفة من النوع المفرد (singular matrix) اي ان محددها لا يساوي صفراً ، ومختلفة عن مصفوفة الوحدة ولمعرفة انها تختلف عن مصفوفة الوحدة او لا نستعمل اختبار بارتليت (Bartletts test) [9].

ولغرض استعمال التحليل العاملي كطريقة لاختيار المتغيرات نتبع الخطوات الاتية :

1. اجراء التحليل العاملي لجميع المتغيرات في الانموذج .
2. استعمال العوامل التي تكون فيها قيمة الجذر المميز اكبر من او يساوي الواحد الصحيح بدل المتغيرات المستقلة في الانموذج وتطبيق انموذج الاستجابة الثنائي عليها .
3. ان كل عامل من العوامل الداخلة في انموذج الاستجابة الثنائي ممثل بمجموعة من المتغيرات وهذه المتغيرات يمكن ان تحدد عن طريق تشبعها بالعامل إذ يتم قبول المتغير الذي تشبعه اكبر من او يساوي مطلق (0.3) .

3. الجانب العملي

تم توليد بيانات عشوائية باستعمال البرنامج الاحصائي الجاهز (Minitab 18) على شكل متغيرات عشوائية بما يتناسب مع طبيعة الدراسة ويمكن توضيح طريقة التوليد من البرنامج كما يلي:

(Minitab 18) ⟶ Calc ⟶ Random data ⟶ ثم نختار التوزيع المطلوب للبينات

اذ تم توليد واحد وعشرين متغيراً منها واحد تابع يتبع توزيعاً احتمالياً هو توزيع برنولي (Bernolli) والبقية تم تعريفها على انها متغيرات مستقلة تتبع توزيعات احتمالية معينة وبحجم عينة (200) مشاهدة ، والمتغيرات التي تم توليدها كلاً حسب معلمته والتوزيع الاحتمالي الخاص به كما يأتي:

$y \sim Bernoulli\ (0.25) \quad x_1 \sim Binomial\ (5\ ,\ 0.3) \quad x_2 \sim Binomial\ (6\ , 0.37) \quad x_3 \sim Binomial\ (7\ , 0.5)$
$x_4 \sim Geomrtric\ (\ 0.4) \quad x_5 \sim Geomrtric\ (\ 0.8) \quad x_6 \sim Geomrtric\ (0.23)$

$x_7 \sim Negativ\ Binomial\ (\ 0.35\ , 6) \quad x_8 \sim Negativ\ Binomial\ (\ 0.55\ , 4) \quad x_9 \sim Negativ\ Binomial\ (\ 0.1, 3)$

$x_{10} \sim Poissan\ (5) \quad x_{11} \sim Poissan\ (10) \quad x_{12} \sim Poissan\ (13) \quad x_{13} \sim chi - square\ (9)$

$x_{14} \sim chi - square\ (11) \quad x_{15} \sim chi - square\ (7) \quad x_{16} \sim Normal\ (3.5\ , 1.2)$

$x_{17} \sim Normal\ (7\ , 2) \quad x_{18} \sim Normal\ (11\ , 3.6) \quad x_{19} \sim Uniform\ (1\ , 5) \quad x_{20} \sim Uniform\ (1.5\ , 8)$

تم استعمال المتغيرات المذكورة انفا لتطبيق انموذج الاستجابة الثنائي التالي:

$$\ln\left(\frac{p}{q}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots \ldots \ldots + \beta_m x_m \qquad (13)$$

وبثلاث حالات الاولى عندما يتضمن الانموذج (10) متغيرات والثانية (15) متغيراً والثالثة (20) متغيراً وباستخدام طرائق الاختيار الثلاث وهن طريقة التقدم الامامي والحذف العكسي والطريقة المقترحة وهي طريقة التحليل العاملي ،وللحصول على النتائج تم استخدام البرنامج الاحصائي الجاهز (SPSS 23) لإيجاد المؤشرات الاحصائية التي تستعمل لتقويم جودة الانموذج ، وقد تم استعمال ثلاثة معايير لتحديد الانموذج الافضل وهي معيار نسبة الامكان الاعظم (MLR) ،ومعيار المعلومات لـ اكاكي (AIC)،و معيار المعلومات البيزي (BIC). وتم اختبار جودة التوفيق للأنموذج بشكل عام باستعمال اختبارين هما اختبار هوزمر-ليمشو (H&L)، واختبار الانحراف (D) . وبمستوى ثقة مقداره (95%) وكانت النتائج كما في الجدول الاتي :

جدول (1): نتائج طرائق اختيار الانموذج المناسب

الطريقة	عدد المتغيرات الكلي	عدد المتغيرات في الانموذج النهائي	عدد الدورات للوصول الى النموذج النهائي	معايير اختيار افضل انموذج			اختبار جودة التوفيق للأنموذج					
				MRL	AIC	BIC	H&L	df	p-value	x^2 model	df	p-value
الاختيار الامامي	10	2	3	207.59	213.59	214.49	11.09	6	0.086	8.12	2	0.017
	15	2	3	207.59	213.59	214.49	11.09	6	0.086	8.12	2	0.017
	20	2	3	207.59	213.59	214.49	11.09	6	0.086	8.12	2	0.017
الحذف العكسي	10	2	9	207.59	213.59	214.49	7.30	8	0.504	8.12	2	0.017
	15	3	13	204.17	212.17	213.37	10.78	8	0.214	11.54	3	0.009
	20	3	18	204.17	212.17	213.37	10.78	8	0.214	11.54	3	0.009
التحليل العاملي	10	3	2	62.29	70.29	71.49	3.08	8	0.93	153.43	3	0.000
	15	5	2	75.23	87.23	89.04	6.34	8	0.61	140.48	5	0.000
	20	5	2	65.60	77.60	79.41	16.50	8	0.36	149.49	5	0.000

نلاحظ من الجدول (1) في حالة استعمال طريقة التقدم الامامي لاختيار المتغيرات التي تدخل في انموذج الاستجابة الثنائي وعندما كان عدد المتغيرات (m=10) فأن الانموذج الناتج يحتوي على متغيرين فقط وكما في المعادلة الآتية:

$$\ln\left(\frac{p}{q}\right) = -2.73 + 0.08X_6 - 0.15\,X_8 \qquad (14)$$

كانت قيمة نسبة الامكان (MLR) للأنموذج مع الحد الثابت فقط في الدورة الحسابية الرابعة تساوي (215.71) اما في حالة استعمال طريقة الحذف العكسي لاختيار المتغيرات على انموذج الانحدار اللوجستي وبعدد متغيرات (m=10) فان عدد المتغيرات في الانموذج النهائي لم يختلف عن طريقة التقدم الامامي وهو متغيرين وكما في المعادلة (14). اما وفق طرقة التحليل العاملي فكان ثلاثة عوامل حلت محل المتغيرات المستقلة وكما في المعادلة الآتية:

$$\mathbf{\ln\left(\frac{p}{q}\right) = -3.4 + 2.62F_1 + 2.86F_3 + 2.02F_5} \qquad (15)$$

نلاحظ من الجدول (1) ان نسبة الامكان الاعظم (MLR) انخفضت من (207.59) في طريقتي التقدم الامامي والحذف العكسي الى (62.29) في طريقة التحليل العاملي، اما معيار (AIC) فكانت قيمته (213.59) في طريقتي التقدم الامامي والحذف العكسي اما في طريقة التحليل العاملي فنجد ان قيمته انخفضت لتصل الى (70.29) ، وكانت النتائج وفق معيار (BIC) مشابه تقريبا لنتائج معيار (AIC). اما عند استعمال عدد متغيرات (m=15 اوm=20) وفق طريقة التقدم الامامي فان عدد المتغيرات النهائية في الانموذج لم يتغير عما هو عليه في حالة عدد المتغيرات (10) وبقي بمتغيرين فقط ، اما وفق طريقة الحذف العكسي فقد زاد الانموذج النهائي متغيراً واحداً وهو (X_{11}) في حالة عدد المتغيرات (15) او حالة عدد المتغيرات (20) ، اما وفق طريقة التحليل العاملي فان عدد العوامل في الانموذج النهائي هو (5) عوامل في كلا الحالتين . وكانت قيم معايير اختيار افضل انموذج كما مبينة في الاعمدة (5 – 7) من الجدول (1) . ودلت نتائج اختبار جودة التوفيق للنماذج الموضحة في الاعمدة (8 ــ 13) من الجدول (1) على معنوية جميع النماذج المقدرة بمستوى ثقة (95%) .

4. الاسلوب التطبيقي

في الاسلوب التجريبي تم توليد البيانات عشوائياً باستعمال اسلوب المحاكاة وطُبقت طرائق اختيار افضل انموذج استجابة ثنائي عليها، وبعد التأكد من افضلية طريقة التحليل العاملي في تحسين جودة التوفيق لأنموذج الاستجابة الثنائي سيتم تطبيقها بشكل عملي على ظواهر ذات المتغير التابع الثنائي، ونظراً لأهمية الجانب الصحي وتأثيره في حياة الانسان لذا تم اختيار التطبيق في هذا الجانب، إذ سيتم التطبيق على جائحة، وسيتم تطبيق انموذج الاستجابة الثنائي على المصابين بوباء كورونا الراقدين في مستشفى سوق الشيوخ العام لسنة 2020 الذين بلغ عددهم (439) لمعرفة العوامل التي لها اكثر تأثير على وفيات هؤلاء المصابين.

4.1. وصف المتغيرات

تم استعمال تسعة متغيرات منها متغير واحد تابع (y) والذي يأخذ قيمتين هما الصفر لاحتمال عدم حدوث الوفاة والواحد الصحيح لاحتمال حدوث الوفاة وثمانية متغيرات مستقلة سيتم تعريفها كما في الجدول ادناه.

جدول (2): تعريف المتغيرات المستقلة

رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة	رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة
X_1	متغير وصفي يمثل جنس المريض الراقد	1	ذكر	X_3	متغير وصفي يمثل مهنة المريض الراقد	1	طفل
		2	انثى			2	طالب
X_2	متغير وصفي يمثل التحصيل الدراسي للمريض الراقد	1	امي			3	موظف
		2	ابتدائية			4	اعمال حرة
		3	متوسطة			5	ربة بيت
		4	اعدادية			6	متقاعد
		5	دبلوم وبكالوريوس			7	عاجز
		6	ماجستير ودكتوراه	X_6	متغير وصفي يمثل وجود مرض مزمن او لا	1	لا يوجد مرض مزمن
X_4	متغير وصفي يمثل الحالة الاجتماعية	1	أعزب			2	يوجد مرض مزمن
		2	متزوج	X_7	متغير كمي يمثل فترة رقود المريض بالأيام		
		3	**أرمل او ارملة او مطلق او مطلقة**	X_8	متغير وصفي يمثل التدخين	1	مدخن
X_5	متغير كمي يمثل عمر المريض بالسنوات					2	غير مدخن

4.2. النتائج

تم استعمال البرنامج الاحصائي (SPSS 23) لإيجاد انموذج الاستجابة الثنائي والنتائج التي تم الحصول عليها والتي سيتم مناقشتها مع تبويبها في جداول ليسهل تحليلها وفهمها . وأظهرت النتائج ان قيمة نسبة الامكان الاعظم مع الحد الثابت فقط عند الخطوة التكرارية الخامسة هي (281.447). وسنعرض نتائج طرائق الاختيار الثلاث مع المقارنة بينها عن طريق معايير اختيار افضل انموذج وسنستعمل مستوى ثقة مقداره (95%) لترشيح المتغير للدخول الى الانموذج ومستوى ثقة مقداره (99%) لتثبيت المتغير في الانموذج.

- **اولا : طريقة التقدم الامامي**

وفق هذه الطريقة اصبح الانموذج النهائي الذي يحتوي على الحد الثابت مع المتغيرات المستقلة الآتية (x_8, x_5) وهما عمر المريض و التدخين اذ تم التوقف عن اضافة المتغيرات عند الخطوة الثانية وكانت نتائج المعلمات المقدرة كما مبينة في المعادلة ادناه.

$$\ln\left(\frac{p}{1-p}\right) = -14.64 + 0.109X_5 + 4.38X_{8(1)} \qquad (16)$$

ولاختبار معنوية الانموذج المقدر في المعادلة (16) بشكل عام تم استعمال اختبار هوزمر – ليمشو (H&L) والذي كانت قيمته التي تم حسابها على وفق الصيغة (6) وكما مبين بالجدول التالي:

جدول (3): اختبار هوزمر- ليمشو لطريقة التقدم الامامي عند الخطوة الثانية

Step	Chi-square	df	p.value
2	.929	7	.996

وهذا يعني عدم رفض فرضية العدم (7)، وان الانموذج المقدر يمثل البيانات بشكل جيد .

- **ثانيا: طريقة الحذف العكسي**

وفق هذه الطريقة فان الانموذج النهائي وهو ما تحقق عنده الخطوة السابعة إذ أن اكبر قيم (p-value) هي (0.039) التابعة للمتغير (X_5) عمر المريض هي اقل من (0.05) لذلك لم يتم حذفه من الانموذج وبهذا كان الانموذج المقدر هو الانموذج الناتج عن الخطوة السادسة و يمكن كتابة الانموذج الناتج كما في المعادلة الآتية:

$$\ln\left(\frac{p}{1-p}\right) = \quad -14.64 + 0.109X_5 + 4.38X_{8(1)} \qquad (17)$$

وهو الانموذج نفسه الذي نتج من طريقة التقدم الامامي وكما تبين النتائج في الجدول ادناه:

جدول (4): اختبار هوزمر- ليمشو لطريقة الحذف العكسي عند الخطوة السابعة

Step	Chi-square	Df	p.value
7	.929	7	.996

- **ثالثا : طريقة التحليل العاملي**

قبل البدء بطريقة التحليل العاملي يجب التأكد من تحقق فروض التحليل العاملي على البيانات ، و أول فرض هو التوزيع الطبيعي وبما ان عدد المشاهدات كبير فأن مشكلة التوزيع الطبيعي للبيانات قد حلت اعتماداً على نظرية النهاية المركزية، اما الفرض الثاني المتعلق بكفاية عدد المشاهدات فتم استعمال اختبار كايز- ماير – اولكن Kaiser-mayar- olkin (KMO) وكانت قيمته (0.606) وهي اكبر من (0.5) وهذا يدل على كفاية العينة المستعملة، علماً ان محدد مصفوفة الارتباطات يساوي (0.2) وهذا يعني ان المصفوفة ليست مفردة. ولمعرفة المتغيرات التي يمثلها كل عامل يجب حساب التشبع العاملي للمتغيرات بالعوامل الاربعة التي تم استخلاصها ، وباستعمال طريقة المركبات الاساسية و اهمال المتغيرات ذات التشبع الاقل من مطلق (0.3)، فقد تبين ان العامل الاول (F_1) له ثلاثة متغيرات تشبعاتها اكبر من مطلق (0.3) وهي (x_3, x_4, x_5) اما العامل الثاني (F_2) فضم متغيرين وهما (x_1, x_6,) و العامل الثالث (F_3) له متغيران هما (x_2, x_8) والعامل الرابع (F_4) له متغير واحد فقط وهو (x_7) وباستعمال العوامل الاربعة التي تم استخلاصها كمتغيرات مستقلة بدلا عن المتغيرات المستقلة الاصلية في انموذج الاستجابة الثنائي تم الحصول على:

$$\ln\left(\frac{p}{1-p}\right) = \quad -216.75 - 2.1F_1 - 291.7F_2 + 167.5F_3 - 72.9F_4 \qquad (18)$$

نلاحظ من المعادلة تم تخفيض المتغيرات المستقلة في انموذج الاستجابة الثنائي من (8) متغيرات الى اربعة عوامل. ولاختبار جودة التوفيق للأنموذج بشكل عام سنستعمل اختبارين الاول هو اختبار الانحراف (D) الذي يمثل الفرق بين نسبة الامكان الاعظم للأنموذج الحد الثابت فقط التي تساوي (MLR=281.44712) ونسبة الامكان الاعظم للأنموذج مع العوامل في الانموذج النهائي التي تساوي (MLR=0.00012) بهذا تكون قيمة (D) تساوي (281.447) وتكون قيمة (p-value) بدرجة حرية (4) هي (0.000) وهي اقل من مستوى المعنوية (α=0.05) لذلك يعتبر الانموذج المقدر معنوي حسب هذا الاختبار ، اما الاختبار الثاني فهو اختبار هوزمر – ليمشو (H&L) فكانت نتائجه كما مبينة في الجدول ادناه :

جدول (5): اختبار هوزمر- ليمشو

Step	Chi-square	df	P.Value.
1	.000	8	1.000

نلاحظ ان قيمة (p.value) تساوي (1) وهي اكبر من مستوى المعنوية المحدد (α=0.05) وهذا يعني عدم رفض فرضية العدم (7) و ان الانموذج المقدر يمثل البيانات بشكل جيد.

4.3. المقارنة بين الطرائق

كل طريقة من الطرائق الثلاث افرزت لنا انموذج استجابة ثنائي بعدد من المتغيرات التي تعدها الطريقة مهمة ويجب وجودها في الانموذج . ولغرض المقارنة بين الطرائق الثلاث يتم استعمال معايير المقارنة بين النماذج ، ونبدأ من المعيار الاول وهو نسبة الامكان الاعظم (MLR) إذ نلاحظ ان قيمتها في طريقة التقدم الامامي والحذف العكسي هي (132.475) اما وفق بطريقة التحليل العاملي كانت تساوي (0.00012) وإذ ان كلما قلت القيمة يعتبر الانموذج هو الافضل وفق هذا المعيار لذلك يعتبر الانموذج الناتج من طريقة التحليل العاملي هو الافضل، اما المعيار الثاني فهو معيار المعلومات لـ اكاكي (BIC) فكانت قيمته لطريقة التقدم الامامي والحذف العكسي هي (144.64) اما وفق طريقة التحليل العاملي فكانت (24.34)وهو يدل على ان طريقة التحليل العاملي هي الافضل ايضا، اما معيار (AIC) فكانت نتائجه هي الاخرى تدل على افضلية طريقة التحليل العاملي الذي كانت قيمته اقل من الطريقتين الاخريتين وقيمته لطريقة التقدم الامامي و الحذف العكسي تساوي (136.475) ولطريقة التحليل العاملي تساوي (8.00012).

5. الاستنتاجات والتوصيات

من تطبيق الطرائق الثلاث لاختيار افضل انموذج الاستجابة ثنائي، يمكن الإشارة الى بعض الاستنتاجات التي توصل اليها الباحث والتوصيات التي يعتقد انها ضرورية.

5.1. الاستنتاجات

أ. ان جميع نماذج الاستجابة الثنائية النهائية والناتجة من الطرائق الثلاث لاختيار افضل انموذج انحدار لوجستي سواءً في الجانب التجريبي أم الجانب التطبيقي قد اجتازت اختباري جودة التوفيق للأنموذج بشكل عام وهما اختبار الانحراف (D) واختبار هوزمر – ليمشو (H&L).

ب. من نتائج الجانبين التطبيقي والتجريبي تبين ان للعوامل الناتجة من التحليل العاملي قدرة على تقليل نسبة الامكان الاعظم افضل من اي تشكيلة مختارة من المتغيرات المستقلة باستعمال اي من طرائق الاختيار الاخرى، وهذا ما اوضح ان

معياري الاختيار اللذان يعتمدان على نسبة الامكان الاعظم وهما (AIC, BIC) يعطيان افضلية للعوامل على حساب المتغيرات المستقلة .

ت. فضلاً عن افضلية النتائج التي تميزت بها طريقة التحليل العاملي عن طريقتي التقدم الامامي والحذف العكسي فهناك ميزة اخرى وهي ان عدد الدورات للوصل الى الانموذج النهائي اقل من الطريقتين الأخريتين.

ث. من نتائج الجانب التطبيقي وعند استعمال طريقتي التقدم الامامي و الحذف العكسي تبين ان اكثر المتغيرات تأثير هي (عمر المريضX_5 والتدخين X_8) ، وهما صاحبي التأثير الاكبر في وفيات هؤلاء المصابين.

5.2. التوصيات

أ. ضرورة اختيار المتغيرات المستقلة بعناية تامة من لدن الباحث حتى يتم الحصول على النتائج بصورة دقيقة.

ب. اعتماد طريقة التحليل العاملي في تحسين جودة التوفيق لأنموذج ثنائي الاستجابة.

ت. اعتماداً على ما توصلت له الدراسة فان من اكثر العوامل المؤثرة في الانموذج هو التدخين لذا يوصي الباحث بتوعية المجتمع بعدم التدخين لما له من تأثير قوي على حياتهم.

ث. ضرورة تشجيع و توسيع البحوث التي تخص الجانب الصحي ولاسيما البحوث التي تتعلق بالأمراض التي لها تأثير في حياة الناس بصورة مباشرة .

المصادر

- **اولاً: المصادر العربية**

[1] الأسدي، سعد جاسم و فارس، سندس عزيز (2015)، الاساليب الاحصائية في البحوث، دار صفاء للنشر والتوزيع-عمان، الطبعة الاولى.

[2] بابطين ، عادل احمد حسن (2009)، "الانحدار اللوجستي وكيفية استعماله في بناء نماذج التنبؤ للبيانات ذات المتغيرات التابعة ثنائي القيمة"، اطروحة دكتوراه، جامعة ام القرى، كلية التربية، قسم علم النفس، الرياض، المملكة العربية السعودية.

[3] باهي، مصطفى حسين و عبد الفتاح، محمود و عز الدين، حسني محمد (2002)، التحليل العاملي: النظرية - التطبيق، مركز الكتاب للنشر، القاهرة، مصر.

[4] سامي، بلبخاري (2009)، "استعمال التحليل العاملي للمتغيرات في تحليل استبيانات التسويق- دراسة تطبيقية على بعض البحوث"، رسالة ماجستير، جامعة العقيد الحاج لخضر - باتنة ، كلية العلوم الاقتصادية والتسيير، الجزائر.

[5] السباح، شروق عبد الرضا سعيد (2009)، "بناء انموذج انحدار لوجستي معدل لحياة الاطفال الخدج في محافظة كربلاء"، اطروحة دكتوراه، جامعة بغداد، العراق .

[6] صفوت فرج، محمد حافظ (1991)، التحليل العاملي في العلوم السلوكية، القاهرة، دار الفكر العربي، الطبعة الثانية.

[7] عباس، علي خضر (2012)، "استعمال نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات التابعة النوعية"، مجلة جامعة كركوك للعلوم الادارية والاقتصادية، المجلد 2، العدد 2.

[8] الفرهود، سهيلة حمود عبد الله (2014)، "استعمال الانحدار اللوجستي لدراسة العوامل المؤثرة على اداء الاسهم"، مجلة جامعة الأزهر، سلسلة العلوم الطبيعية، المجلد 16، العدد 1.

[9] فهمي، محمد شامل بهاء الدين، (2005)، الاحصاء بلا معاناة - المفاهيم مع التطبيقات على برنامج SPSS، الادارة العامة للطباعة والنشر، معهد الادارة العامة، الرياض، المملكة العربية السعودية.

[10] محمد، حيدر فتحي (2013)، "دراسة العوامل المرتبطة بمرض السكري عن طريق مواءمة اسلوبي التحليل العاملي وتحليل المسار"، رسالة ماجستير، الجامعة المستنصرية - بغداد.

- **ثانياً : المصادر الاجنبية**

[11] Hosmer , D .W , and Lemeshow, S. (2000), Applied Logistic Regression, 2nd edition, John Wiley and Son, New York .

[12] Husssain, J. N. & Low, H. C. (2009), "An alternative method to construct goodness-of-fit test for multinomial Logistic Regression model", Proceedings of the 5th Asian Mathematical Conference, Malaysia

[13] Kleinbaum, D. & Klein , M. (2010), Logistic Regression: A Self-Learning Text (Statistics for Biology and Health), 3rd edition, Springer, USA.

[14] Liu , Y. (2007), "On Goodness Of Fit Of Logistic Regression", Doctoral thesis in Statistics, University of Kansas.

[15] Pample, F. C. (2000), Logistic Regression: A Primer (Quantitative Applications in the Social Sciences) 2nd edition, Beverly Hills, CA, USA.

[16] Rat Kowsky, D. A. (1983), Nonlinear Regression Modeling: A Unified Practical Approach, 1st edition, Marcel Dekker, New York.

[17] Schwarz , G. (1978), “Estimating the Dimension of a model”, The Annals of statistics, Vol. 6, No. 2, P461-464 .

[18] Walker, J. (1996), "Methodology Application: Logistic Regression Using the codes Data", Department of Transportation, National Highway Traffic Safety Administration (NHTS), Publication/ Report Number: HS-808 460; Washington DC.

Journal of AL-Rafidain University College for Sciences (2022); Issue 52; 20-31

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Improve the Fitment Quality of the Binary Response Model With Practical Application to those Infected With the Covid-19 Epidemic

Mushtak K. Abdulraheem	Ali M. Jawad
mushtag.k@uokerbala.edu.iq	ali.jawad@uokerbala.edu.iq
Dep. of statistics, College of Administration and Economics, University of Karbala, Karbala, Iraq.	

Article Information

Article History:
Received: August, 16, 2021
Accepted: October, 29, 2021
Available Online: December, 31, 2022

Keywords:
Matchmaking quality, binary response model, Corona epidemic.

Correspondence:
Ali M. Jawad
ali.jawad@uokerbala.edu.iq

Abstract

The study of binary dependent variables is considered one of the important processes because of increasing phenomena that are described by this process. Therefore, the Binary Response model is considered one of the important methods that represent this type of phenomena. Also, the process of choosing independent variables, which affect the Binary Independent variables is considered very necessary. The study included the use of two approaches, the first is the experimental approach (simulation), and the second is the practical application to those infected with the Corona epidemic. Three methods were used to choose the best binary Response model, which are the forward method, backward method and the proposed method (factor analysis method) with testing the quality of the model fit after applying each method, by using the Deviation (D) test and Hossmer-Lemshow (H&L) test. The comparison of the final results based on three criteria; Maximum Likelihood Ratio (MLR), Akiake information criteria (AIC) and Bayesian information criteria (BIC). The final results show that the factors resulted from the proposed method (Factor Analysis) have an ability to decrease the MLR better than any group of variables chosen by other methods. Consequently, AIC and BIC criteria, which are based on MLR gave preferences for the factors that resulted from the proposed method (Factor Analysis) better than the other two methods (forward and backward). The results also showed that the greater the number of independent variables that have a significant effect on the model, the more the factors gave better results according to the criteria used. The results showed that the two independent variables, age of the patient and smoking, have the most impact on the lives of those infected with the Corona epidemic.

doi: https://doi.org/10.55562/jrucs.v52i1.539