AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

توظيف طريقة الانحدار المعكوس المُجزّأ مع طريقة معكوس لاسو لتحديد أهمّ العوامل المؤثّرة على مرضى Covid-19

https://doi.org/10.29124/kjeas.1547.14

أ.م. سيف حسام رحيم [1]
saif.hosam@qu.edu.iq
جامعة القادسية / كلية الإدارة والاقتصاد

أ. د. جاسم ناصر حسين[2]
Jasim.nasir@uokerbala.edu.iq
جامعة كربلاء / كلية الإدارة والاقتصاد

أ.م. د. ايناس عبد الحافظ محمد[3]
enas.albasri@uokerbala.edu.iq
جامعة كربلاء / كلية الإدارة والاقتصاد

المستخلص

يعدّ فايروس كورونا Covid-19 من أخطر الأمراض التي واجهها العالم ، إذْ واجه العالم العديد من المخاطر و الخسائر بالأرواح والخسائر اقتصادية. ولمعرفة أهمّ العوامل المؤثّرة في المرضى المصابين تَمّتْ دراسة مجموعة من العوامل التي تؤثّر في مرضى فيروس كورونا وتحليلها باستعمال إحدى طرائق اختزال الأبعاد، وهي طريقة الانحدار المعكوس المُجَزّأ (sliced inverse regression)، و توظيفها مع إحدى طرائق اختيار المتغيّرات، وهي طريقة معكوس لاسو (Reciprocal Lasso) . تَمَّ جمع البيانات الحقيقية للمرضى الراقدين في مستشفى كربلاء في محافظة كربلاء للمدّة (2021-2022). وتَمَّ تصميم استمارة بهذا الخصوص، تَمَّ جدولة بيانات الاستمارة التي احتوت عيّنة حجمها 130 مريضًا. وتَمَّ اختيارهم بشكل عشوائي، وباستعمال برنامج لغة R تَمَّ التوصّل إلى أنّ أهمّ العوامل المؤثّرة في مدّة البقاء في المستشفى للمرضى الراقدين.

الكلمات المفتاحية : اختيار المتغيّرات ، Covid-19، لاسو ، الانحدار المعكوس المجزئ ، معكوس لاسو .

المقدّمة

تؤثّر العديد من العوامل على المصابين بفيروس كورونا، بما في ذلك: العمر إذْ يبدو أنّ الأشخاص الذين يتقدّمون في العمر هم الأكثر عرضة للإصابة بأعراض خطيرة لفيروس كورونا؛ وذلك لأنّ نظامهم المناعي يكون ضعيفًا بشكل عام. وكذلك

الحالات الصحّية الأساس، مثل الأشخاص الذين يعانون من حالات صحّية مزمنة مثل: أمراض القلب، والرئة، والسكري، فهم الأكثر عرضة لتطوّر أعراض خطيرة بعد الإصابة بفيروس كورونا. و من الممكن المستوى كذلك الاجتماعي والاقتصادي: فيمكن للفقر وعدم الاستقرار الاجتماعي والاقتصادي أنْ يوثّر على فرص الأشخاص في الحصول على الرعاية الصحّية اللازمة والوصول إلى الموارد الضرورية للتعافي بسرعة. وكذلك التطعيم، ويُعدّ التطعيم ضدّ فيروس كورونا من أهمّ العوامل التي تؤثّر على المصابين، فالأشخاص الذين يتلقّون التطعيم يصابون بأعراض أخفّ، ويخفّفون من خطر الإصابة بالأعراض الخطيرة. وهنالك مجموعة أخرى من العوامل المؤثّرة على المصابين بمرض كورونا.

مشكلة البحث:

تعدّ الإصابة بفايروس كرونا من أبرز الأمراض التي يعاني منها العالم وأخطرها، و بسبب نتائج هذا المرض على صحّة الإنسان، وكذلك الاقتصاد، جاءت مشكلة البحث.

هدف البحث:

يهدف البحث لتحديد أهمّ العوامل التي تؤثّر على المصابين بمرض كورونا في محافظة كربلاء من خلال جمع البيانات من مستشفى كربلاء خلال العام (2021-2022)، وتحليلها باستعمال إحدى طرائق اختزال الأبعاد (SIR) وتوظيفها مع إحدى طرائق اختيار المتغيّرات، و هي طريقة معكوس لاسو (RLasso).

فايروس كورونا [1,2,3] (COVID-19):

فيروس كورونا هو فيروس يسبّب مجموعة متنوّعة من الأمراض التنفسية، وينتمي إلى عائلة الفيروسات الكورونية التي تشمل عددًا من الفيروسات التي تصيب الإنسان والحيوانات. ظهر فيروس كورونا الجديد (COVID-19) لأول مرّة في مدينة (ووهان) في الصين في ديسمبر 2019، وانتشر بسرعة في أنحاء العالم جميعًا. وقد أعلنت منظّمة الصحّة العالمية أنّ كورونا فيروسٌ جديدٌ لم يكن معروفًا سابقًا للإنسان. تنتقل العدوى بالفيروس عبر قطرات اللعاب المنتشرة عند السعال والعطس، أو عند لمس سطح ملوّث ثُمّ لمس العين أو الفم أو الأنف. وتظهر الأعراض عادة بعد مدّة حضانة قد تصل إلى 14 يومًا، وتشمل الأعراض الشائعة: الحمّى، والسعال، وضيق التنفس. وفي بعض الحالات يمكن أن يؤدّي الفيروس إلى أمراض خطيرة مثل التهاب الرئة والفشل الكلوي والوفاة. يمكن تشخيص الإصابة بفيروس كورونا عن طريق فحص العيّنات الحيوية، مثل عيّنات البلغم أو الدم. ويعالج الفيروس حالياً بواسطة العناية الداعمة والعلاج الدوائي. وتتضمّن الإجراءات الوقائية من الإصابة بفيروس كورونا: الالتزام بارتداء الكمّامات، والتباعد الاجتماعي، وتجنّب التجمّعات الكبيرة، وغسل اليدين بانتظام، وتنظيف الأسطح بشكل منتظم. وتشجّع المنظّمات الصحّية العالمية على تطعيم الأفراد للوقاية من الإصابة بفيروس كورونا.

اختيار المتغيّرات [4,5] **Variables selection**

يَتمحور اهتمامٌ كبيرٌ حاليًا حول اختيار المتغيّرات (Variables selection) ، إذْ يَتُمُّ استعماله لتقليل عدد المتغيّرات التي يَتُمُّ استعمالها في التحليل الإحصائي وإزالة المتغيّرات التي لا تؤثّر بشكل كبير على النتائج. ويُسهمُ ذلك في توفير الجهد والتكلفة

وتحسين جودة التحليل. يمكن استعمال اختيار المتغيّرات في العديد من المجالات، مثل: المجال الطبي، والهندسة الوراثية، والزراعة، والبيئة؛ وذلك للحصول على نتائج دقيقة وشاملة وتقليل التكاليف والجهد. ويمكن تطبيق هذه الطريقة في العديد من الأبحاث والدراسات والتحليلات الإحصائية للوصول إلى نتائج دقيقة وموثوقة. يوجد نوعان رئيسان من طرائق اختيار المتغيّرات، منها ما يسمّى بالطرائق التقليدية، وتتضمّن طريقة الحذف العكسي، وهي طريقة تعتمد على اعتماد المتغيّرات جميعها في معادلة الانحدار، ثم يَتُمُّ حذف تلك المتغيّرات تباعًا باستعمال قيمة F الجدولية. وطريقة الاختيار الأمامي المباشر، والتي يَتُمُّ من خلالها إدخال المتغيّرات واحدة تلو الأخرى إلى معادلة الانحدار. كما يتضمّن النوع الثاني طريقة الاختيار التدريجي، فضلًا عن معيارين مهمّين، وهما: معيار اكاكي للمعلومات، ومعيار بيز للمعلومات. أمّا النوع الثاني من الطرائق فيسمّى بطرائق التنظيم، إذْ تفرض هذه الطرائق حدّاً للجزاء على المَعْلَمات؛ لتسهيل عملية تقدير المَعْلَمات و تقليص المتغيّرات. وبعض تلك الطرائق هي طريقة لاسو التي اقترحها العالم (Tibshirani, 1996)، و طريقة مجموعة لاسو (Yuan and Lin, 2006)، وكذلك طريقة لاسو التكيفي ,(Zou, 2006)، و طريقة معكوس لاسو (Song and ,2015 Liang).

طريقة لاسو Lasso Method [6,7]

تُعدّ طريقة لاسو (Lasso) إحدى طرائق التنظيم، والتي تعني (إضافة دالّة جزاء لغرض التحكّم بتعقيد النموذج و الذي يمثّل له بـ(Penalty) بحيث يكون عمل هذا الجزء تقليل تباين النموذج مع الزيادة الكبيرة في التحيّز). وكلمة Lasso هي اختصار لـ (Least Absolute Shrinkage and Selection Operator) الذي تَمَّ اقتراحه من قبل العالم Tibshirani عام (1996). وهي من الطرائق الفعّالة إذْ أنّها تعمل على اختيار المتغيّرات، وكذلك تقدير المعالم. وتُعدُّ طريقة لاسو من عائلة المربعات الصغرى الجزائية . لكن في العام 2001 وَضّحَ الباحثان Fan & Li أنّ طريقة لاسو لا تتمتّع بخصائص اوراكل.

وخصائص اوراكل مجموعة من الخصائص التي يمتلكها المقدّر منها: (الاتّساق، التناثر، تقارب الأمثلية) إذْ يمكن جمع تلك الخصائص بالقول بأنّ الطريقة تمتلك تلك الخصائص إذا كان لديها القدرة على اختيار النموذج الحقيقي باحتمال يساوي (1).

من العيوب الأخرى لطريقة لاسو عندما يكون عدد المتغيّرات المستقلّة P أكبر من حجم العيّنة n، وهذا يعني $P > n$ إذْ سيَتُمُّ اختيار n من المتغيّرات المستقلّة وإهمال الباقي منها. ويَتُمُّ الحصول على مقدّر لاسو من خلال إضافة دالّة جزاء لاسو إلى دالّة الخسارة لطريقة المربعات الصغرى وكما في المعادلة الآتية:

$$\hat{\beta}_{lasso} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} |\beta_k|$$

إذْ إنّ λ تمثّل مَعْلَمَة التنظيم، والتي تكون قيمتها أكبر من الصفر أو مساوية له $\lambda \geq 0$، إذْ أنّها تتحكّم أو تنظم كمّية الجزاء أو العقوبة، فكلّما زادت قيمة λ أدّى هذا إلى أعلى مستوى للانكماش. وفي حالة كون قيمة الجزاء صغيرة فإنّ هذا يؤدّي إلى أنْ يكون للنماذج المختارة تباينٌ كبيرٌ و بالمقابل تحيّز صغيرٌ. أمّا في حالة كبر مقدار قيمة الجزاء فسيَتُمُّ اختيار نماذج فيها متغيّرات أقلّ و بتباين أقلّ. أكّد الباحثون على ضرورة الاختيار الدقيق لقيمة λ ؛ وذلك لغرض الحصول على نتائج دقيقة

يعوّل عليها ، إذْ يمكن تحديد قيمتها من خلال GCV (Generalized Cross Validation) والموضّحة في المعادلة الآتية:

$$GCV=\frac{RSS}{n\{1-p(\lambda)/nx\}^2}$$

إذْ إنّ:

$$RSS=\sum_{i=1}^{n}(y_{i-\theta^T x_i})^2$$

$P(\lambda)$ تمثّل عدد المَعْلَمات، إذْ إنّ كبر قيمتها يؤدّي إلى المزيد من الجزاء.

P : تمثّل عدد المتغيّرات .

n يمثّل حجم العيّنة.

طريقة معكوس لاسو [9,10] **Reciprocal lasso Method** تَمَّ اقتراح طريقة معكوس (rLasso) لاسو في العام 2015 من قبل Song & Ling؛ وذلك لتجنّب النماذج التي تعاني من overfitted ، إذْ تُعدّ هذه الطريقة من طرائق اختيار المتغيّرات وتقدير المعالم في الوقت نفسه. ومِنْ مزايا طريقة معكوس لاسو أنّها مستَمرّة عند الصفر، وتتناقص في المدّة $(0,\infty)$. وبتلك المزايا للطريقة أصبحت مرغوبة لاختيار النماذج، ويمكن تَمَثيل مقدر الطريقة من خلال الصيغة الآتية:

$$\hat{\beta}_{r\text{lasso}} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda\sum_{k=1}^{p}\frac{1}{|\beta_k|} I\{\beta_k \neq 0\} \qquad (2)$$

إذْ إنّ:

λ : تمثّل مَعْلَمَة الضبط و التي تتحكّم بقدر درجة العقوبة أو الجزاء.

$I(.)$: تمثّل indicator function.

قيمة λ إذا كانت في قيمتها الدنيا سوف تعطي أعلى مستوى من الانكماش، وبذلك تعطي ما يقارب صفر من المعاملات. وبالمقارنة مع دالّة جزاء لاسو فإنّها لا تتناقص في المدّة $(0,\infty)$. وعندما تكون λ مساوية للصفر في فستعطي ما يقارب صفر من المعاملات، بعكس طريقة المعكوس التي تعطي صفر من المعاملات عند العقوبات التي تقترب من اللانهاية. فمن الناحية النظرية تَمَثلك طريقة معكوس لاسو خاصية اوراكل، وعلية فهي أفضل من حيث الأداء من طريقة لاسو.

الانحدار الشرائحي المجزئ [11,12] **Sliced Inverse Regression**

اقترحت طريقة الانحدار الشرائحي المُجَزّأ في عام 1991 من قبل العالم (Li)، وهي واحدة من طرائق اختزال الأبعاد للبيانات التي تحوي مشكلة البعدية (CD)، إذْ تقدّر هذه الطريقة الفضاء المركزي (central subspace). ومصطلح الانحدار المعكوس (Inverse Regression) يشير إلى التوقّع الشرطي ($E(X/Y)$)، واستعملت كلمة معكوس (Inverse) لعكس التوقّع الشرطي، ففي الحالة الاعتيادية يكون التوقّع الشرطي بالشكل الآتي ($E(Y/X)$). أمّا كلمة مُجَزّأ (Sliced) فتشير إلى تقدير التوقّع الشرطي ($E(X/Y)$) بأخذ المدّة ل (Y). وعلي فإنّ طريقة الانحدار الشرائحي المُجَزّأ تعمل على عكس العلاقة السببية بين متغيّر الاستجابة (Y) والمتغيّرات التوضيحية (X'S). أيْ سَيتُمّ جعل المتغيّر (Y) يمثّل المتغيّر المستقلّ، والمتغيّر (X) سيمثّل المتغيّر المعتمد. وقد لاقتْ هذه الطريقة انتشارا واسعا من حيث التطبيق إذْ يَتُمُّ تطبيقها في مجالات مختلفة منها: التمويل، والاقتصاد، و المعلومات الحيوية، والمجالات الطبية. يتلخّص عمل هذه الطريقة على تقسيم النموذج إلى مجموعة من الشرائح وحسب قيم المتغيّر المعتمد (Y)، ويَتُمَّ التعامل مع كُلّ شريحة بشكل منفصل من خلال عمليات إحصائية مختلفة، ثُمَّ يَتُمُّ دمج معلومات الشرائح جميعًا للحصول على الجذور الكامنة، واختيار أكبرها لغرض تمثيل متّجه تقليل البعد المؤثّر (EDR) لطريقة الانحدار الشرائحي المُجَزّأ.

في حالة البيانات عالية الأبعاد وفي ظلّ وجود مشكلة لعنة الأبعاد سيكون من الصعب تقدير معالم الانحدار الاعتيادي. ويمكن للانحدار المعكوس أنْ يتعامل مع تلك المشكلات من خلال تقسيم الانحدار إلى P من الشرائح ($E(X_i/Y)$) إذْ إنّ (i=1,2,…,P)، ففي هذا الحالة سيهمل مشكلة لعنة الأبعاد. أيْ أنّنا لو فرضنا أنّ ($\overline{X}$) يمثّل الوسط الحسابي للمتغيّر (X)، وفرضنا ($\hat{Z} = \hat{\Sigma}^{-\frac{1}{2}}(X - \overline{X})$) نسخة مبّسطة من ($Z$)، إذْ إنّ ($\hat{\Sigma}$) تمثّل مصفوفة التغاير للمتغيّر (X). وعلى فرض أنّ h تمثّل عدد الشرائح، و n_y تمثّل عدد المشاهدات من y^{th} شريحة. وعليه فإنّ $\widehat{M} = \sum_{y=1}^{h} \hat{f}_y \hat{Z}_y \hat{Z}_y^T$ تمثّل نسخة مبسّطة من (M) ، إذْ إنّ $\hat{f}_y = \frac{n_y}{n}$ ، و ($\hat{Z}_y$) تمثّل معدّل (Z) من الشريحة (y). ولو فرضنا ($\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_p \geq 0$) الذي يمثّل القيم الذاتية المقابلة للمتّجهات الذاتية ($\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_p$) من ($\widehat{M}$). فإذا كان الفضاء d من $S_{Y|Z}$ معروف، فالنموذج الذي تعتمد علية $span(\hat{\beta}) = span(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$ هو مقدّر متّسق لـ $S_{Y|X}$ ، عندما $\hat{\beta}_i = \hat{\Sigma}^{-\frac{1}{2}}\hat{v}_i$. طريقة الانحدار الشرائحي المُجَزّأ مشابه لنموذج الانحدار شبة المَعْلَمي:

$$y = f\ (\beta_1^T X, \beta_2^T X, \ldots, \beta_k^T X, \varepsilon)$$

إذْ إنّ ($\beta's$)متّجه المَعْلَمات، ويكون غير معروف، و(ε) تمثّل حدّ الخطأ العشوائي المستقلّ عن (X)، و(f) تمثّل دالّة غير معلومة. فعندها سَيتُمُّ جمع معلومات الشرائح كَافّة للحصول على الجذور الكامنة، ثُمَّ بعد ذلك سَيتُمُّ اختيار أكبرها لتمثيل متّجهات تقليل البعد المؤثّر (EDR) لـ(SIR)، والذي يمثّل الشكل الجديد للبيانات التي تمثّل المَعْلَمات ($\beta_k's$). ويَتُمُّ تحويل البيانات إلى نموذج مختزل، واستبدالها بالبيانات الأصلية لسهولة التعامل معها. وفي هذه الحالة تَمّتْ معالجة مشكلة البعدية بطريقة الانحدار المعكوس المُجَزّأ. ويمكن تمثيل نموذج الانحدار المعكوس المُجُزّأ بالشكل الآتي:

$$F(A,C) = \sum_{y=1}^{h} \left\| \hat{f}_y^{1/2} \hat{Z}_y - AC_y \right\|^2$$

طريقة Sparse Sliced Inverse Regression Reciprocal Lasso

يمكن الحصول على نموذج طريقة SSIR-Rlasso من خلال تصغير

$$\sum_{y=1}^{h} \left\| \hat{f}_y^{1/2} \hat{Z}_y - \hat{\Sigma}^{\frac{1}{2}} diag(\hat{B}\hat{C}_y)\alpha \right\|^2 + \lambda \sum_{k=1}^{p} \frac{1}{|\beta_k|}$$

إذْ إنّ:

p تمثّل عدد المتغيّرات (j =1,2,…,p).

λ تمثّل مَعْلَمَة الضبط.

h تمثّل عدد الشرائح في النموذج (y = 1,2,…, h).

البيانات الحقيقية:

تَمَّ جمع البيانات الحقيقية للمرضى الراقدين في مستشفى كربلاء في محافظة كربلاء للمدّة من (2021-2022)، وقَدْ تَمَّ تصميم استمارة بهذا الخصوص، ثُمَّ تمّتْ جدولة بيانات الاستمارة التي احتوت عيّنة حجمها 130 مريضًا، وقَدْ تَمَّ اختيارهم بشكل عشوائي. إذْ تَمَّ اعتماد المتغيّر المعتمد (Y)، والذي يمثّل مدّة الرقود في المستشفى، و تَمَّ اعتماد المتغيّرات المستقلّة، والتي عددها (47) متغيّرًا والمبيّنة تفاصيلها في الجدول رقم (1). إذْ تَمّ اعتماد برنامج لغة R في التحليل.

الجدول رقم (1)

يبيّن المتغيّر المعتمد و المتغيّرات التوضيحية

رمز المتغيّر	اسم المتغيّر	رمز المتغيّر	اسم المتغيّر	رمز المتغيّر	اسم المتغيّر
Y	مدّة الرقود في المستشفى	X16	الأمراض المزمنة	X32	Zocin
X1	العمر	X17	نسبة الإصابة	X33	Loratadine
X2	جنس المريض	X18	Favipiravir	X34	Salbutamol

Budesonide	X35	Clexane	X19	السمنة	**X3**
Albuterol	X36	Decadron	X20	المهنة	**X4**
Tiotropium	X37	Paracetamal	X21	الإصابة بأمراض التنفس مثل الربو والتهاب الشعب الهوائية المزمن	**X5**
Fluticasone	X38	Vit. D3	X22	التدخين	**X6**
Amoxicillin	X39	Vit. Zinc	X23	ألم الراس	**X7**
Azithromycin	X40	Vit. C	X24	درجة الحرارة	**X8**
N-Acetylcysteine	X41	Cetirizine	X25	سعال	**X9**
Penicillin	X42	Ceftriaxone	X26	نسبة الاوكسجين	**X10**
Cefuroxime	X43	Solvodin	X27	ارتفاع معدّل الكولسترول في الدم	**X11**
Clarithromycin	X44	Remdesivir	X28	معدّل النبض	**X12**
Doxycycline	X45	Meropenium	X29	ضغط الدم	**X13**
Levofloxacin	X46	Azithromycin	X30	الإصابة بأمراض الدم مثل الثلاسيميا	**X14**
Moxifloxacin	X47	Becozym	X31	التقيّؤ	**X15**

الكشف عن مشكلة التعدّد الخطّي:

للكشف عن مشكلة التعدّد الخطّي سَيتُمُّ حساب معامل التضخّم لغرض تقويم درجة تأثير التعدّد الخطّي بين المتغيّرات التوضيحية في نموذج الانحدار. إذْ سَيتُمُّ حساب معامل التضخّم لكلّ متغيّر معتمد في النموذج. فإذا كانت قيمة معامل التضخّم

تساوي (1) فإن ذلك يَدُلّ على عدم وجود تضخّم للتباين عند هذا المتغيّر، أمّا إذا كان معامل التضخّم بين (1 – 5) فهنالك ارتباط معتدّل. أمّا إذا كانت قيمة معامل التضخّم أكبر من (5)، فهذا يدلّ على وجود ارتباطٍ قويّ بين المتغيّرات. والجدول رقم (2) يوضّح قيمة معامل التضخّم للمتغيّرات التوضيحية.

الجدول رقم (2)

يبيّن قيم معامل التضخّم VIF للمتغيّرات التوضيحية

VIF	المتغيّر	VIF	المتغيّر	VIF	المتغيّر	VIF	المتغيّر
7.99	X37	6.26	X25	3.12	X13	8.36	X1
5.56	X38	5.32	X26	7.65	X14	2.58	X2
11.37	X39	5.33	X27	1.88	X15	7.95	X3
5.98	X40	4.36	X28	8.54	X16	4.33	X4
1.36	X41	2.36	X29	9.25	X17	8.52	X5
4.02	X42	5.36	X30	6.25	X18	2.56	X6
5.85	X43	7.31	X31	7.98	X19	1.89	X7
4.69	X44	7.22	X32	2.36	X20	5	X8
8.98	X45	9.45	X33	4.36	X21	9.85	X9
7.69	X46	5.36	X34	1.32	X22	9.32	X10
8.88	X47	8.96	X35	2.22	X23	4.78	X11
		4.58	X36	7.36	X24	1.89	X12

نلاحظ من الجدول رقم (2) بأنّ أغلب المتغيّرات التوضيحية مرتبطة ارتباطا قويًّا فيما بينها، أيْ إنّ المتغيّرات تعاني من مشكلة التعدّد الخطّي، وهذا يَدُلّ على وجود مشكلة لعنة الأبعاد (CD).

تَمَّ استعمال الطريقة المقترحة SSIR-Rlasso مع البيانات الحقيقة، وتَمَّ الحصول على المعالم المقدّرة للمتغيّرات التوضيحية، إذْ تَمَّ تصفير بعض المتغيّرات غير المهمّة وكما في الجدول رقم (3) الآتي:

الجدول رقم (3)

يبيّن قيم المعاملات المقدّرة بطريقة SSIR-Rlasso

المتغيّر	قيمة المَعْلَمَة المقدّرة	المتغيّر	قيمة المَعْلَمَة المقدّرة	المتغيّر	قيمة المَعْلَمَة المقدّرة	المتغيّر	قيمة المَعْلَمَة المقدّرة
X1	0.5422	X13	0.0003	X25	0.4399	X37	0.6286
X2	0.8485	X14	0.5524	X26	0.7970	X38	0.4470
X3	-0.3170	X15	-0.0001	X27	-0.0003	X39	0.0006
X4	0.4867	X16	0.7006	X28	0.0001	X40	0.1128
X5	0.4290	X17	0.8225	X29	0.6868	X41	0.3086
X6	0.2227	X18	-0.7401	X30	0.0006	X42	0.3600
X7	0.0004	X19	0.0111	X31	-0.0012	X43	-0.7330
X8	0.2258	X20	-0.0008	X32	0.0222	X44	0.5764
X9	0.0002	X21	0.6950	X33	0.0014	X45	0.9217
X10	0.3964	X22	0.7737	X34	0.9016	X46	-0.0811
X11	0.1811	X23	-0.1160	X35	0.4296	X47	0.0139
X12	0.0001	X24	0.1339	X36	0.7754		

الجدول من إعداد الباحث: بالاعتماد على نتائج برنامج لغة R.

نلاحظ من الجدول رقم (3) بأنّ طريقة SSIR-Rlasso قامت باستبعاد 12 متغيّرًا وهي: (ألم الراس، سعال، معدّل النبض ، ضغط الدم ، التقيّؤ، Decadron، Solvodin، Remdesivir، Azithromycin، Becozym، Loratadine، Amoxicillin).

الاستنتاجات

1- بالإمكان تحليل البيانات التي تعاني من مشكلة لعنة الأبعاد من خلال توظيف إحدى طرائق اختزال الأبعاد مع إحدى طرائق اختيار المتغيّرات.
2- من خلال الطريقة المقترحة تَمَّ التوصّل إلى مجموعة من العوامل المؤثّرة في مدّة رقود المريض في المستشفى.
3- من العوامل المؤثّرة التي تَمَّ التوصّل إليها من خلال البحث هي: (العمر، وجنس المريض، والسمنة، والمهنة) والإصابة بأمراض التنفس مثل: (الربو، والتهاب الشعب الهوائية المزمن، والتدخين، ودرجة الحرارة، ونسبة الاوكسجين، وارتفاع معدّل الكولسترول في الدم)، والإصابة بأمراض الدم مثل: (الثلاسيميا، والأمراض المزمنة، ونسبة الإصابة).
4- ومِنْ العوامل الأخرى التي لها تأثير في مدّة الرقود في المستشفى هي: (Favipiravirو Clexane و Paracetamal و Vit. D3 و Vit. Zinc و Vit. C و Cetirizine و Ceftriaxone و Meropenium و zocin و Salbutamolو Budesonide و Albuterol و Tiotropium و Fluticasone و Azithromycin و N-Acetylcysteineو Penicillin و Cefuroxime و Doxycycline و Clarithromycin و Levofloxacinو Moxifloxacin).

التوصيات:

1- اعتماد النموذج الرياضي الذي تَمَّ التوصّل إليه في المستشفيات لغرض تحديد أهمّ العوامل التي تؤثّر على المرضى.
2- تعميم النموذج على باقي المستشفيات لغرض الاستفادة من دقّته في تحديد العوامل المؤثّرة.

المصادر:

1- Yang, L., Liu, S., Liu, J., Zhang, Z., Wan, X., Huang, B., ... & Zhang, Y. (2020). COVID-19: immunopathogenesis and Immunotherapeutics. Signal transduction and targeted therapy, 5(1), 128.
2- Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W. C., Wang, C. B., & Bernardini, S. (2020). The COVID-19 pandemic. Critical reviews in clinical laboratory sciences, 57(6), 365-388.
3- Ndwandwe, D., & Wiysonge, C. S. (2021). COVID-19 vaccines. Current opinion in immunology, 71, 111-116.
4- Andersen, Charlotte Møller, and Rasmus Bro. "Variable selection in regression—a tutorial." Journal of chemometrics 24.11-12 (2010): 728-737.

5- Allen, David M. "The relationship between variable selection and data agumentation and a method for prediction." technometrics 16.1 (1974): 125-127.

6- Zou, Hui. "The adaptive lasso and its oracle properties." Journal of the American statistical association 101.476 (2006): 1418-1429.

7- Roth, Volker. "The generalized LASSO." IEEE transactions on neural networks 15.1 (2004): 16-28.

8- Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.

9- Song, Qifan, and Faming Liang. "High-dimensional variable selection with reciprocal l 1-regularization." Journal of the American Statistical Association 110.512 (2015): 1607-1620.

10- Mallick, Himel, et al. "The reciprocal Bayesian lasso." Statistics in Medicine 40.22 (2021): 4830-4849.

11- Li, K. C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association, 86(414), 316-327.

12- Li, L., & Yin, X. (2008). Sliced inverse regression with regularizations. Biometrics, 64(1), 124-131.