استخدام نموذج الانحدار اللوجستي

لتحديد أعراض الإصابة بفيروس كورونا

الدكتور

عبدالرحيم عوض عبدالخالق

دكتوراه في الاحصاء

المستخلص:

يهدف البحث إلى تحديد أهم الأعراض المصاحبة للإصابة بفيروس كورونا، وتكوين نموذج انحدار لوجستي ثنائي للتنبؤ باحتمالية إصابة أو عدم إصابة الشخص بالفيروس، وتم ذلك من خلال نموذج الانحدار اللوجستي الثنائي ذو متغير تابع نوعي (مصاب أو غير مصاب)، ومجموعة من المتغيرات المستقلة وهي السن ودرجة الحرارة والاسهال والصداع وضيق التنفس والتهاب الحلق والأمراض المزمنة ومخالط أو غير مخالط وحاسة الشم والتذوق وتوصل الباحث إلى معنوية تأثير درجة الحرارة والسن والاسهال والصداع وحاسة الشم والتذوق ومخالط أو غير مخالط وعدم معنوية كلاً من التهاب الحلق وضيق التنفس والأمراض المزمنة، والسعال وكان ترتيب المتغيرات حسب الأهمية النسبية وتأثيرها في الإصابة بالفيروس، هي حاسة الشم، والتذوق، بنسبة ٤٧,٦٪ يليها الاسهال ١٤,٦٢٪ يليها السن ٩,٢٨٪ يليها درجة الحرارة ٩,١٢٤٪ يليها الصداع ٤,٩٧٧٪ وأخيرًا مخالط أو غير مخالط بنسبة ٣,٩٢٪، ويتبين إن المتغيرات المستقلة تساهم بنسبة ٧٩,٤٪ في شرح التغيرات التي تحدث في المتغير التابع كما تساهم بنسبة ٥٩٪ في تحسين جودة التنبؤ للنموذج وتساهم بنسبة ٨٩,٦٪ في التصنيف الصحيح للأشخاص.

Abstract:

The main goal of the research is to identify the factors affecting infection with the covid-19, and to form a binary logistic regression model to predict the probability of the person being infected or not with the virus. This was done through the binary logistic regression model through a qualitative dependent variable (infected or not infected), and a set of independent variables such as age, temperature, diarrhea, headache, shortness of breath, sore throat, chronic diseases, mixtures or non-mixtures, sense of smell, taste and cough. The researcher found the significance of the effect of temperature, age, diarrhea, headache, smell and taste, Contact with a sick or non-contact person, and nonsigificance of both sore throats,

shortness of breath, chronic diseases, and cough are shown. The order of the variables according to the relative importance and their impact on infection with the virus was the first place, the sense of smell and taste, at ٪٤٧,٦, followed by diarrhea ٪١٤,٦٢, followed by age ٪٩,٢٨, followed by temperature ٪٩,١٢٤, followed by headache ٪٤,٩٧٧and finally contact or non-contact with a rate of ٪٣,٩٢. The independent variables contribute by ٪٧٩,٤in explaining the changes that occur in the dependent variable and contribute by ٪٥٩to improving the quality of the prediction of the model and contribute ٪٨٩,٦to the correct Classification of people.

استخدام نموذج الانحدار اللوجستي لتحديد أعراض الإصابة بفيروس كورونا

١ – مُقدمة:

لا أحد يستطيع أن يغفل عن حقيقة أن الوباء الأسرع انتشارًا في العالم والذي أطلق عليه كوفيد ١٩ أو كورونا أضر البلدان الغربية المتقدمة في العالم مثل الولايات المتحدة وإيطاليا واسبانيا والمانيا وغيرها أكثر بكثير مما فعل بالبلدان التي تعد متراجعة من الناجية التنموية. جاءت كورونا لتغير مفاهيم وأفكار العالم فقد غيرت الخريطة السياسية والاقتصادية في العالم.

وتعد الإصابة بفيروس أو وباء كورونا أحد مسببات الوفاة الأوسع انتشارًا حول العالم وتعد بيانات الإصابة بالوباء هي عبارة عن مجموعة من اعرض الإصابة بالفيروس، درجة الحرارة، الصداع، ضيق التنفس، الاسهال، التهاب الحلق، مخالطة حالة كورونا، الامراض المزمنة، السن، النوع، إذا كان يعمل بالقطاع الصحي.

أما المتغير التابع، متغير الاستجابة مصنف إلى مصاب أو غير مصاب، وبنمذجة هذه البيانات باستخدام الأساليب الإحصائية المناسبة يمكن معرفة اهم اعراض الإصابة أو عدم الإصابة بفيروس كورونا. وذلك من خلال بناء نموذج إحصائي يوصف البيانات وتحت افتراضات معينة للتعرف على أكثر الأعراض التي تساعد الطبيب على سرعة تشخيص الإصابة بالفيروس. ويعد التشخيص بالإصابة بالفيروس باستخدام نماذج إحصائية متقدمة مهم لكونه وباء من الأوبئة الخطيرة التي تفتك بحياة المصاب، كما أن التشخيص المبكر بالإصابة بالفيروس يساعد على انقاذ حياة المصاب ويمنع من تدهور حالته.

ومن النماذج الإحصائية المستخدمة والتي اثبتت نجاح كبير في المجال الطبي هو نموذج الانحدار اللوجستي ثنائي الاستجابة (Binary Logistic Regression)، وهو نموذج انحدار غير خطي (Non-Linear Regression)، ويستخدم في حالة كون المتغير التابع (متغير الاستجابة) متغير وصفي يتبع توزيع ذو حدين الاحتمالي.

٢- مُشكلة البحث:

يُعد القطاع الصحي أحد القطاعات الخدمية الذي يتولى مهمة توفير الخدمات الصحية للمجتمع من خلال المؤسسات الصحية من أجل حماية المجتمع وتحقيق رفاهية الأفراد. ومما لا شك فيه أن الصحة العامة من المؤشرات المهمة على تقدم البلدان وتطورها، وللجانب الصحي أهمية مميزة على صعيد التنمية لكي يتمكن العنصر البشري من مواكبة التقدم والتطور الحضاري ومن المعروف أن ظهور وباء مثل وباء كورونا أربك العالم بأجمعه وأهدر أموالاً طائلة دون تخطيط وهنا ظهرت الحاجة الماسة إلى الاستعانة ببعض الأساليب الإحصائية التي تساعد على دراسة مثل هذه المشاكل.

وتُعد نماذج الانحدار أحد أهم الأساليب الإحصائية التي تستخدم لمثل هذا النوع من الدراسات. لكن ليس بشكل تقليدي إنها بشكل مطور كي تلائم حالة متغير الاستجابة (المتغير التابع)، عندما يكون ثنائي الحدين (مصاب أو غير مصاب) كما تظهر الحاجة إلى وجود نموذج ذو قدرة تنبؤية عالية باحتمال الإصابة أو عدم الإصابة بفيروس كورونا ومن هنا بدأت مشكلة البحث في الظهور.

٣- أهمية البحث:

يستمد البحث أهميته من خطورة وباء كورونا في مثل هذه الآونة، حيث أدى إلى زيادة عدد الوفيات حول العالم وأدخل حالة من الفزع والرعب في الشعوب اجمعها ولم يفرق بين غني وفقير وعالم وجاهل.

كما أن معظم الدراسات كانت تستخدم تحليل البيانات باستخدام نماذج الانحدار الخطي مع وجود مشكلة واحدة من المشاكل التي تخص النماذج الخطية.

ويُعد الانحدار اللوجيستي من الأساليب الإحصائية المهمة التي تهتم بتحليل البيانات المصنفة ولاسيما في حالة أن يكون متغير التابع (الاستجابة)، والذي يعود لمتغيرات من النوع الأسمي أو الرقمي والمؤلف من تصنيفين أو أكثر وأن الهدف من الانحدار اللوجستي هو التنبؤ بوجود أو عدم وجود صفة معينة بالاعتماد على قيمة متغير أو مجموعة من المتغيرات المستقلة الأخرى التي لها علاقة بالمتغير التابع.

كما تأتي أهمية هذا البحث في بناء نموذج إحصائي فعال ملاءم للبيانات وذو كفاءة عالية في التنبؤ باحتمال إصابة أو عدم الإصابة بوباء كورونا معتمدًا على مجموعة من الاعراض الدالة للإصابة بوباء كورونا.

ومن خلال هذا النموذج يُمكن أدخال أو قياس هذه العوامل لشخص جديد في النموذج بحيث يمكن تقدير احتمالية الإصابة بالوباء من عدمه ومن ثم تقليل عدد الوفيات الناتجة عن هذا الوباء من خلال علاج المرضى في المراحل الأولى من الإصابة لسرعة التعامل مع الحالة في مراحلها الأولى.

وتتركز أهمية البحث في توفير أداة فاعلة في يد المسؤولين لاتخاذ الإجراءات اللازمة ومن خلال هذه الأداة الوصول لاحتمال الإصابة أو عدم الإصابة بالوباء وبالتالي سرعة التعامل مع الحالة.

٤- أهداف البحث:

يُعد الهدف الأساسي لنموذج الانحدار اللوجستي هو إيجاد أفضل توفيق لنموذج يوصف العلاقة بين متغير الاستجابة (التابع) والمتغير المستقل الواحد أو متغيرات مستقلة وطريقة الانحدار اللوجستي لا تفترض شروطًا تخص توزيع المتغيرات المستقلة وهي طريقة أكثر كفاءة.

كما أن هناك أهداف أخرى من استخدام النموذج الانحدار اللوجستي وهي:

١) تحديد أهم أعراض الإصابة بالوباء كورونا.

٢) إيجاد تأثير كل عرض من هذه الاعراض من خلال نموذج الانحدار اللوجستي الثنائي.

٣) ترتيب هذه الأعراض حسب الأهمية النسبية ودرجة تأثيرها في الإصابة بوباء كورونا.

٤) تحديد أوزان لأعراض الإصابة بالفيروس والتعرف على كيفية تغيرها في نموذج الانحدار اللوجستي الثنائي.

٥) التنبؤ باستخدام نموذج الانحدار اللوجستي الثنائي لمعرفة مدى احتمال الإصابة أو عدم الإصابة بالوباء كورونا.

٥- متغيرات البحث:

يتكون البحث من مجموعة من المتغيرات وهي:

١- مُتَغَير تابع (متغير الاستجابة)، وهو نوعي ثنائي (مصاب أو غير مصاب).

٢- مجموعة من المتغيرات المستقلة وهي:

- التهاب الحلق.
- حاسة الشم والتذوق.
- مخالط أو غير مخالط.
- الامراض المزمنة.
- السعال.
- السن.
- درجة الحرارة.
- ضيق التنفس.
- الاسهال.
- الصداع.

مصادر البيانات:

أخذت البيانات من سجلات مستشفى الصدر بجامعة إسكندرية ومستشفى الصدر بكفر الشيخ على عينة من (٢٦٠) شخص وتم سؤالهم عن متغيرات الدراسة وهي السن ودرجة الحرارة وضيق التنفس، والاسهال، والصداع، والتهاب الحلق، وحاسة الشم والتذوق، ومخالط والأمراض المزمنة والسعال.

٦- الدراسات السابقة:

نستعرض الدراسات ذات الصلة بالنموذج الإحصائي وهو تحليل الانحدار اللوجستي الثنائي وهي على النحو التالي:

١- دراسة غانم ،عدنان ، الجاعونى ،فريد خليل "استخدام تقنية الانحدار اللوجستي ثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الأسرة" (٢٠١١).

تضمنت هذه الدراسة أهم محددات كفاية دخل الأسرة وتحليلها من وجهة نظر إحصائية والتعرف على أهم المحددات المعنوية المؤثرة باستخدام تحليل الانحدار اللوجستي ثنائي الاستجابة وقد شملت عينة الدراسة ١٦١ أسرة سحبت بشكل عشوائي من الأسواق الشعبية في مدينة دمشق وريفها، وتناول البحث سبع متغيرات مستقلة من النوع الثنائي (Binary) وقام باختيار الفرض العدمي.

$$H_0 : b_i = 0 \qquad (i = 1,2, \ldots\ldots\ldots 7)$$

وتشير bi إلى معاملات الانحدار للمتغيرات المستقلة وتم رفض الفرض العدمي عند $\alpha = 0.05$

وتم التوصل للنتائج الآتية:

١- معنوية كلاً من المتغيرات التالية وترتيبها حسب معنوية تأثيرها في المتغير التابع وهو كفاية دخل الأسرة والمتغيرات هي وجود طلبة جامعيين في الأسرة يليه الدخل الشهري للأسرة يليه ملكية السكن ويليه حجم الأسرة.

٢- عدم معنوية كلاً من المتغيرات المتمثلة فى المستوى التعليمي ومكان الاقامة والحالة الوظيفية عند مستوى معنوية $\alpha = 0.05$.

٢- دراسة احمد ،شمس الدين أحمد علي "تقدير وتحليل العوامل المؤثرة في الرضا الوظيفي باستخدام الانحدار اللوجستي" (٢٠١٣).

هدفت الدراسة إلى استخدام نموذج الانحدار اللوجستي لتقدير وتحليل العوامل التي تؤثر في الرضا الوظيفي وكانت مشكلة الدراسة عدم وجود نموذج إحصائي قادر على التنبؤ بمستوى الرضا الوظيفي للعاملين بجامعة شيندي بالسودان ومعرفة العوامل المؤثرة على مستوى الرضا ونسبة تأثير كُل عامل وفحص النموذج واستخدامه في التقدير والتنبؤ بمستوى الرضا واستخدامه في اتخاذ القرارات.

وبعد تجميع البيانات من مفردات عينة الدراسة والتي قوامها ٢٠١ شخص تم استخدام مربع كاي x^2 للاستقلالية لاختبار معنوية العلاقة بين الرضا الوظيفي والمتغيرات المستقلة (التعامل الإداري، الأجر والحوافز، بيئة العمل والتعامل مع الزملاء).

وتوصلت الدراسة إلى أن نموذج الانحدار اللوجستي الذي تم الحصول عليه هو نموذج معنوي وقادر على التنبؤ بمستوى الرضا الوظيفي للعاملين بجامعة شندي ووجد ان التعامل الإداري والأجر والحوافز وبيئة العمل والتعامل مع الافراد، هي عوامل تؤثر في مستوى الرضا الوظيفي.

ووجد أن التعامل مع الزملاء داخل وخارج إطار العمل تساهم بنسبة كبيرة جدًا على مستوى الرضا الوظيفي سلبية التعامل الإداري ثم الأجر والحوافز وأخيرًا بيئة العمل.

أما بالنسبة للجانب التطبيقي على وباء كورونا نلاحظ أن كل الدراسات السابقة في هذا المجال ركزت على العوامل المؤثرة على الإصابة ولم تركز على التنبؤ وأن أغلبها دراسات نظرية ولا توجد دراسة اهتمت ببناء نموذج إحصائي يعمل على التنبؤ بتلك الإصابة.

وتأتي هذه الدراسة لتناول جوانب أخرى تتعلق بتقدير وبناء نموذج إحصائي فعال يهدف إلى التنبؤ باحتمال الإصابة أو عدمها التي تحدث للشخص وعن طريقه يُمكن معرفة عوامل الخطورة المهمة بحيث يعمل كأداة للاختبار المبكر حول تلك العوامل الخطرة، مما يتيح المجال للشخص من اتخاذ ما يلزم من إرشادات وتوصيات من الطبيب لمواجهة هذه العوامل. وقد استخدمت الدراسة اسلوبًا احصائيًا وهو تحليل الانحدار اللوجستي للتنبؤ باحتمال الإصابة أو عدم الإصابة لوباء كورونا وهذا يمثل إضافة جديدة خاصة بعد أن يتم اختبار متغيرات بيانات شخص جديد.

٣- دراسة Keener (٢٠١٣):

وتهدف الدراسة إلى تحديد نموذج للتنبؤ بالتعثر المالي في قطاع تجارة التجزئة وقد تم استخدام أسلوب تحليل الانحدار اللوجيستي على عينة الدراسة المكونة من ٣٥٩٥ شركة وذلك من (٢٠٠٥-٢٠١٢) واهم النتائج التي تم التوصل إليها نموذج مقترح للتنبؤ بالفشل مكون من خمس نسب مالية وغير مالية وهي: عدد الموظفين، العائد على الأصول، هامش التدفق النقدي، نسبة الديون إلى حقوق المساهمين، نسبة النقدية إلى الخصوم المتداولة.

٤- دراسة الفرهود، سهيلة حمود عبد الله "استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على أداء الأسهم" (٢٠١٤).

تناولت هذه الدراسة، أسلوب الانحدار اللوجستي الثنائي لدراسة أهم المؤثرات المالية المؤثرة على أداء الأسهم في سوق الكويت للأوراق المالية، وقد توصلت الدراسة إلى مجموعة من النتائج أهمها كفاءة النموذج بشكل كلي للتنبؤ بأداء الأسهم بالإضافة إلى ان النسب التي تمثل أفضل متنبئات بالأداء كانت ستة وقد تمكنت من تصنيف أداء الأسهم للشركات قيد البحث إلى نوعين (جيد ورديء) وذلك بدقة عالية بلغت ٧٤,٢٪ وكان ترتيب النسب المالية المؤثرة معنويًا على أداء الأسهم بالسوق الكويتي وفقًا لأهميتها النسبية كالتالي، ربحية الأسهم، نسبة الدين إلى حقوق الملكية، نسبة السعر السوقي إلى القيم الدفترية، نسبة التداول، نسبة صافي الإيرادات البنكية إلى الأصول الكلية، وأخيرًا العائد على حقوق الملكية.

٥- دراسة صالح ،عائدة هادي "تحليل الانحدار اللوجستي لدراسة زمن البقاء لمرضى سرطان الدم (٢٠١٤)".

تكونت الدراسة من متغير تابع ثنائي الحدين وهي مصاب (١) وغير مصاب (٠) أما العوامل المستقلة فهي العمر والجنس ونوع المرض ونوع العلاج ومنطقة السكن وحالة فقر الدم وهدفت الدراسة إلى إمكانية استخدام نموذج الانحدار اللوجستي في دراسة زمن البقاء للمرضى المصابين بمرض سرطان الدم (اللوكيميا) وتحديد المتغيرات المؤثرة في حالة المصاب والمرشحة للبقاء في النموذج ومن نتائج هذه الدراسة أن أكثر المصابين بمرض اللوكيميا هم من عمر أقل من ٢٠ سنة إذ تبلغ نسبتهم ٤٢,٤٧٪ وأن أكثر المصابين بمرض اللوكيميا هم من الاناث إذا تبلغ نسبتهم ٤٣,٨٪ وإن أكثر نسبة من الوفيات هو من نوع سرطان الدم النخاعي الحاد إذا تبلغ نسبتهم ٦٣,٦٣٪ وإن نوع العلاج الذي يعطي

للمريض يختلف باختلاف نوع المرض ومرحلة المرض وعند وصول المريض لحالة متقدمة يستخدم العلاج الكيميائي، أكثر حالات الوفيات هم الذين يعالجون بالعلاج الكيميائي وتبلغ نسبتهم ٧٢,٧٢٪ من الوفيات.

٦- دراسة خليفة، صهيب الشيخ عبد المحيد "تطبيق الانحدار اللوجستي لمعرفة أهم العوامل الاجتماعية والاقتصادية للتدخين" (٢٠١٤).

أهتمت هذه الدراسة إلى تحديد أهم المتغيرات الاجتماعية والاقتصادية المؤثرة على التدخين وفحص مدى ملاءمة النموذج اللوجستي مع بيانات التدخين في السودان واتضح من نتائج هذه الدراسة ان النموذج اللوجستي الثنائي الذي يحتوي على المتغيرات المستقلة يصنف الطلاب إلى مدخنين وغير مدخنين ووفق بيانات التدخين بدرجة جيدة وأهم العوامل التي تدفع الطلاب للتدخين هي العوامل الاجتماعية والاقتصادية والعلاقات العاطفية واتضح أيضًا ان الدوافع الاجتماعية تأتي في مقدمة الترتيب بنسبة أفضلية ٨٤,٢٥٪ وأن العلاقة العاطفية في المرتبة الثانية بنسبة أفضلية ٢,٤٨٪ وظهرت في مؤخرة الترتيب من حيث الأهمية المتغيرات الديموغرافية وهي مكان السكن والحالة الزواجية والدخل الشهري بنسبة أفضلية ٠,٣٨، ١٩٩، ٢٩٨، على التوالي.

٧- دراسة Sinamnuayphol and chanchardat (٢٠١٥):

تهدف هذه الدراسة إلى التعرف على العوامل المؤثرة في فشل الشركات المالية خلال الفترة (٢٠٠٤-٢٠١٥) اعتمدت الدراسة على ١٧ نسبة مالية لعينة من ٢٠ شركة متعثرة و٣٩ شركة غير متعثرة وكل هذه الشركات هي مؤسسات مالية مسجلة في بورصة تايلاند تم استخدام الانحدار اللوجستي والتوصل إلى أربع نسب مالية تكون هي النموذج المقترح هي: مخصص خسائر القروض إلى إجمالي القروض، صافي الدخل إلى إجمالي حقوق الملكية، الفائدة على القروض إلى إجمالي القروض، فوائد الدخل إلى الدخل بدون فوائد. تم اختبار هذا النمو ج وتبين أنه قادر على التنبؤ بالتعثر في المؤسسات المالية المدرجة في بورصة تايلاند بدقة تعادل ٩٤,٩٠٪ في العينة ولذلك توصلت الدراسة بأهمية النموذج لحماية الشركات من الافلاس مستقبلاً.

٨- دراسة ادريس، عبد الله جمعة فضل المولى "استخدام الانحدار اللوجستي الثنائي لدراسة أهم العوامل الاقتصادية والاجتماعية للآباء وأثرها على تحصيل التلاميذ" (٢٠١٥).

عرفت هذه الدراسة نموذج الانحدار اللوجستي على إنه أحد نماذج الانحدار التي تكون فيها العلاقة بين المتغير التابع والمتغيرات المستقلة لا خطية وهدفت الدراسة إلى بناء نموذج رياضي لقياس أثر الخصائص الاقتصادية والاجتماعية للآباء على التحصيل الأكاديمي للأبناء واستخدمت الدراسة النموذج اللوجستي ثنائي الاستجابة لبيان تأثير المتغيرات المستقلة على المتغير التابع وتوصلت الدراسة إلى عدد من النتائج أهمها وجود مكتبة في المنزل وعمل الأم وتعليم الأب وعدم هجرة الوالد تؤثر بصورة فعالة على مستوى تحصيل التلاميذ وأوصت الدراسة على الاهتمام بعمل الأم وتعليم الآباء وتوفير المكتبات.

٩- دراسة مهران ،سحر"التنبؤ بمراحل تأثير الأزمة المالية على أداء المصارف باستخدام الانحدار اللوجستي" (٢٠١٧).

توصلت هذه الدراسة إلى ثلاث نماذج تحتوي على مؤشرات مالية تتسم بالقدرة التنبؤية بالأزمات المالية قبل وقوعها في المصارف العاملة في دول مجلس التعاون الخليجي من أجل التعرف على وضعية تلك المصارف وإعطاء مؤشرًا مبكرًا للجهات المعينة للتدخل واتخاذ الإجراءات التصحيحية المناسبة ولتحقيق ذلك تم حساب إحدى عشر نسبة مالية لعينة من ٤١ مصرف من خلال المعطيات الربع سنوية التي تمت في المدة من سنة ٢٠٠٦ إلى ٢٠١٤، تم تحليل هذه النسب باستخدام الانحدار اللوجستي كأسلوب إحصائي لبناء ثلاث نماذج بحيث كل نموذج يمثل مرحلة من مراحل الازمة المالية العالمية وتم التواصل إلى نموذج لوجستي للمرحلة الأولى قادر على التنبؤ بدخول المصارف في تلك الأزمة المالية بجودة تصنيف النموذج ٨٠,٦٪ يتكون من ثلاث نسب مالية مرتبة حسب أهميتها كالتالي نسبة "إجمالي حقوق الملكية إلى إجمالي الأصول"، ونسبة دخل التشغيل إلى المصروفات التشغيل ونسبة إجمالي القروض إلى إجمالي الودائع.

كما تم التوصل إلى نموذج لوجستي للمرحلة الثانية قادر على التنبؤ تحسين الأوضاع المالية وخروج المصارف من تأثير الأزمة العالمية وقد بلغت جودة التصنيف ٧٩,٣٪ ويتكون من ثلاث نسب مالية مرتبة حسب أهميتها كالتالي نسبة إجمالي القروض إلى إجمالي الودائع "نسبة" دخل التشغيل إلى مصروفات التشغيل ونسبة "صافي الربح إلى إجمالي حقوق الملكية".

وتم التوصل إلى نموذج لوجستي للمرحلة الثالثة قادر على التنبؤ بنمو ربحية المصارف بعد خروجها من تأثير الأزمة العالمية وقد بلغت جودة التصنيف للنموذج ٨٢٪ ومكون من خمس نسب مالية مرتب حسب أهميتها كالتالي:

نسبة دخل التشغيل إلى إجمالي الأصول، نسبة إجمالي حقوق الملكية إلى إجمالي الأصول، نسبة إجمالي القروض إلى إجمالي الودائع، ونسبة "النقدية + الأوراق المالية" إلى إجمالي الأصول ونسبة مصروفات التشغيل إلى إجمالي الأصول".

١٠- دراسة ابودومة ،حيدر جميل الله محمد "استخدام تحليل الانحدار اللوجستي والتحليل التمييزي للعوامل المؤثرة على الإصابة بأمراض القلب" (٢٠١٩).

تهدف الدراسة إلى تحديد أهم العوامل المؤثرة على الإصابة بأمراض القلب ومعرفة تأثير كل عامل وترتيب حسب الأفضلية في النموذج والقدرة على التصنيف والتمييز باستخدام كلاً من نموذج الانحدار اللوجستي الثنائي والتحليل التميزي ومن أهم ما توصلت إليه الدراسة من خلال الانحدار اللوجستي الثنائي الذي يضم العوامل المؤثرة مثل، ضغط الدم، الجنس، السكري، الكوليسترول، والوراثة، والوزن، وتبين أن نموذج الانحدار اللوجستي الثنائي ذو كفاءة تصنيفيه عالية على التنبؤ ٩١,٨٪ ونسبة مساهمة العوامل على الإصابة بأمراض القلب ٨٢,٣٪.

حيث ساهم ضغط الدم بنسبة أكبر وهي ٥٤,٥٪ سلبية الجنس بنسبة ١٥,٧٪ ثم السكري بنسبة ١٤,٥٪ ثم الكوليسترول بنسبة ٨,٣٪ ثم الوراثة بنسبة ٤,٤٪ وأخيرًا الوزن بنسبة ٤,١٪.

المبحث الثاني

الانحدار اللوجستي Logistic Regression

(٢/١) تمهيد:

كثير ما يواجه الباحثين عند دراستهم للعديد من الظواهر وجود عدد كبير من المتغيرات المتعلقة بالدراسة، وهذه المتغيرات قد تكون متداخلة أو مختلفة أحيانًا في طبيعتها (كمية أو نوعية)، مما يصعب عملية الوصول إلى فهم وتغير هذه الظواهر، لذلك فإننا نلجأ إلى البحوث العلمية لفهم وتفسير هذه الظواهر باستخدام التحليل الإحصائي، إن إحدى طرق التحليل الإحصائي في مجال دراسة الظواهر المختلفة، الانحدار اللوجستي وهو احد أبرز مواضيع تحليل الإنحدار ولعل أهميته ترجع إلى إمكانية استخدامه من خلال متغيرات يمكن الحصول عليها بسهولة لغرض تصنيف المتغير التابع إلى فئات غير متداخلة.

مفهوم الانحدار اللوجستي:

يستخدم نموذج الانحدار اللوجستي في تحليل طبيعة شكل العلاقة بين المتغيرات عندما يكون المتغير التابع يأخذ الطابع النوعي ويأخذ قيمتان أو أكثر بينما المتغيرات المستقلة يمكن أن تكون وصفية أو كمية.

فروض الانحدار اللوجستي:

- أن يكون المتغير التابع اسميًا ثنائيًا للانحدار اللوجستي الثنائي، أو تصنيفيًا بالنسبة للانحدار اللوجستي المنفرد أو رئيسيًا بالنسبة للانحدار اللوجستي الترتيبي.
- أن يحتوي النموذج على أكثر من متغير مستقل في المستوى التصنيفي أو الترتيبي أو الفئوي أو النسبي.
- أن يحتمل المتغير التابع احتمالاً واحدًا لكل فرد، بمعنى لا يُمكن أن يكون الفرد في فئتين في نفس الوقت، كان ينتمي الفرد إلى مرضين في نفس الوقت.
- عدم وجود مشكلة الازدواج الخطي Multi collineanty.
- وجود علاقة خطية بين المتغيرات المستقلة ولوغاريتم المتغير التابع.

(٢/٢) – أنواع الانحدار اللوجستي regression:

١- **الانحدار اللوجستي الثنائي Binary Logistic**: مُتغير تابع نوعي مقسم إلى حدين ومتغير أو أكثر من متغير مستقل.

٢- **الانحدار اللوجستي المُتعدد Multiple Logistic regression**: يُعد امتداد للانحدار الثنائي إلا أن المتغير التابع تم تقسيمه إلى أكثر من حدين.

ويُعد التصنيف من أهم مزايا الانحدار اللوجستي، كذلك نستعرض بعض النقاط الخاصة بالتصنيف وهي:

١) جدول التصنيف Classification table:

وهو جدول مُشترك يوضح عدد الحالات المشاهدة التي تمتلك صفة ما وعدد الحالات المشاهدة التي لا تمتلك تلك الصفة، وكذلك عدد الحالات المتوقعة التي تمتلك الصفة وعدد الحالات المتوقعة التي لا تمتلك تلك الصفة، بحيث يوضح الجدول عدد الحالات التي تم تصنيفها بطريقة صحيحة، وعدد الحالات التي تم تصنيفها بطريقة خاطئة.

(**Cox**, **D**.**R**. (**1996**))

أما الشكل العام لجدول التصنيف هو:

شكل (٢/١)

التصنيف	المتوقع		المجموع
	الموجب P	السالب N	
الموجب P	الموجب صحيح TP	السالب الخاطئ FN	P
السالب N	الموجب خطأ FP	السالب الصحيح TN	p̀
المجموع	Q	Q̀	1

- **الحساسية (S) Sensitivity**:

تحسب من جدول التصنيف وهي قيمة الاحتمال بأن يكون التصنيف المتوقع موجبًا للحالة التي تكون موجبة وتحسب كالتالي:

$$S = \frac{TP}{TP+FN} = \frac{TP}{P} \qquad (2\text{-}2\text{-}1)$$

- **النوعية (S_P) Specificity:**

تحسب من جدول التصنيف وهي قيمة احتمال أن يكون التصنيف المتوقع سالبًا للحالة التي تكون فعلاً سالبة كما يلي:

$$S_p = \frac{TN}{FP+TN} = \frac{TN}{\grave{p}} \qquad (2\text{-}2\text{-}2)$$

- **نسبة التصنيف الصحيح Hit Ratio:**

تحسب من جدول التصنيف وتعرف بأنها قيمة لاحتمال التصنيف الصحيح كما تعرف أيضًا بنسبة الكفاءة Efficiency.

$$(\text{hit Ratio})\ E_t = \frac{TP+TN}{P+\grave{P}} = \frac{Tp+TN}{Q+\grave{Q}} \qquad (2\text{-}2\text{-}3)$$

- **معامل الترجيح Odds:**

احتمال حدوث شيء مقارنة باحتمال عدم حدوثه

$$Odds = \frac{P}{\grave{P}} \qquad (2\text{-}2\text{-}4)$$

- **الأرجحية: OR (نسبة الترجيح) Odds Ratio**

تعني احتمال حدوث حدث معين مقسومًا على احتمال عدم حدوثه ونسبة الترجيح هي عبارة عن النسبة بين معامل الترجيح $Odds_1$ لمتغير ما (Q_1) ومعامل الترجيح $Odds_2$ لمتغير أخر (Q_2)

$$OR = \frac{Q1}{Q2} \qquad (2\text{-}2\text{-}5)$$

(٢/٣) مبادئ الانحدار اللوجستي Logistic regression Principles:

عندما يكون المتغير التابع (Y) هو متغير اسمي ذو حدين ولدينا متغير مستقل واحد (X) فأننا يمكن أن نطلق على أحد القيمتين اللتين يأخذهما المتغير التابع نجاح (Success) وتأخذ القيمة

(1) والقيمة الأخرى فشل (Failure) وتأخذ القيمة (o) ويرمز لاحتمال النجاح Pعندما تأخذ (Y) القيمة (1).

$$P = \frac{e^{Bo+B1\times}}{1+e^{Bo+B1\times}} \qquad (2\text{-}3\text{-}1)$$

واحتمال الفشل 1-P وعندما تأخذ y القيمة (0):

$$1\text{-}P = 1 - \frac{e^{Bo+B1\times}}{1+e^{Bo+B1\times}}$$

$$\therefore 1 - P = \frac{1}{1+e^{Bo+B1\times}} \qquad (2\text{-}3\text{-}2)$$

$$Odds = \frac{P}{1-P} = \frac{\frac{e^{Bo+B1\times}}{1+e^{Bo+B1\times}}}{\frac{1}{1+e^{Bo+B1\times}}} = e^{Bo+B1\times} \qquad (2\text{-}3\text{-}3)$$

$$Log\ (Odds) = B_o + B_1 \times$$

$$Logit\ (Y) = B_o + B_1 \times \qquad (2\text{-}3\text{-}4)$$

وفي حالة وجود أكثر من متغير مستقل:

$$Logit\ (Y) = Log\ (Odds) = Log\ (\frac{P}{1-P}) = B_o + B_1\ x_1 + B_2\ X_2 + \ldots\ldots + B_k\ x_k \quad \textbf{(2-3-5)}$$

(2/4) اختبارات تحليل الانحدار اللوجستي:

أولاً: الحكم على الانحدار:

بعد تقدير النموذج الانحدار اللوجستي لابد من اختبار معنوية المتغيرات الداخلة في النموذج والتي يجب أن تبقى في النموذج وأي منها غير معنوي ويجب ان يحذف وهذه الاختبارات هي:

١- اختبار نسبة الامكان الأعظم Likelihood–ratio test لاختبار معنوية معلمات الانحدار اللوجستي:

$$H_0 \qquad B_i = 0$$

$$H_1 \qquad B_i \neq 0$$

إحصائية الاختبار:

$$G = -2\ \text{Ln}\left[\frac{Like\ lihood\ without\ the\ variable}{Like\ lihood\ with\ the\ vanable}\right]$$

= - 2 ln Likelihood without the variable + 2 Ln Likelihood with The variable (2/4/1)

ومن خلال SPSS تعطى قيم لوغاريتم الإمكان مضروبًا في 2.

١- القيمة الجدولية:

$$\chi^2_{(P)}$$

يوجد حد ثابت حيث P هي عدد المتغيرات وتقارن بين G والقيمة الجدولية χ^2 بدرجات حرية (P) لو G أكبر من القيمة الجدولية χ^2 رفض الفرض العدمي.

(Mitchell, 1992)

٢-اختبار Wald:

يُعرف هذا الاختبار بإحصائية والد (Wald Statistic) وهي إحصائية تتبع توزيع χ^2 وتستخدم لمعرفة المساهمة المنفصلة لكل من المتغيرات المستقلة في نموذج الانحدار اللوجستي وكلما كانت قيمة Wald كبيرة تعني نسبية مساهمة المتغير الذي يقابلها أكبر.

$$H_0 \qquad B_1 = o$$

$$H_1 \qquad B_1 \neq O$$

إحصائية Wald:

$$W = \frac{B}{Se_B} \qquad (2\text{-}4\text{-}2)$$

وتقارن بالقيمة الحرجة من جدول التوزيع الطبيعي.

(2/5): اختبارات ملاءمة جودة النموذج اللوجستي Goodness Of Fit:

مدى جودة التوفيق للنموذج تعني مدى اقتراب القيم المشاهدة من خط التقدير والملائمة تعني هل النموذج الإحصائي ملائم لبيانات العينة الدراسية وجودة الملاءمة تقيس الاقتراب بين القيم المشاهدة والقيم المتوقعة للنموذج ومن اختبارات جودة المطابقة.

١-اختبار مربع كاي x^2 Chi – Square:

وهو مقترح من قبل Pearson ويعرف بالصيغة الآتية:

$$x^2 = \sum_{i}^{n} \frac{(O_i - E_i)}{E_i} \qquad (2\text{-}5\text{-}1)$$

O_i : القيم المشاهدة.

E_i: القيم المتوقفة.

n:عدد المشاهدات

القيمة الحرجة $x_{\propto}^2$ (n-p)

P هي عدد معلمات النموذج، أما إذا كانت قيمة إحصاء الاختبار أكبر من مستوى المعنوية المحدد أي ليست ذات دلالة معنوية مما يؤدي إلى أن القيم المشاهدة تتساوى مع القيم المتوقعة وهذا يدل على جودة التوفيق للنموذج أي أن النموذج مناسب للبيانات.

٢-اختبار معامل هزمر وليمشو R_L^2 Hosmer and Lemeshow Test:

نستخدم إحصائية (Hosmer and Lemeshaw) لجودة التوثيق لفرضية أن البيانات المشاهدة مختلفة بشكل كبير عن القيم المتوقعة من قبل النموذج.

$$R_L^2 = \frac{x^2}{-2 \ln (\text{orignal})} \qquad (2\text{-}5\text{-}2)$$

حيث:

x^2: تمثل الفرق بين لوغاريتم الاحتمال في نموذج الحد الثابت فقط ولوغاريتم الاحتمال عندما يتضمن النموذج بعض المتغيرات التي تمت إضافتها.

2ln (Original) -: لوغاريتم الاحتمال (دالة الترجيح) عندما يتضمن النموذج الحد الثابت.

R_L^2: تمثل الانخفاض النسبي في القيمة المطلقة اللوغاريتم الاحتمال وهي مقياس لمدى تحسن سوء التوافق نتيجة تضمن المتغيرات التنبؤية وتنحصر القيمة بين الصفر والواحد.

$R_{L=o}^2$ تعني المتغيرات التنبؤية عديمة الفائدة في متغير الاستجابة.

$R_{L=1}^2$ تعني أن المتغيرات المئوية ذات فائدة عالية.

(Hosmer D. W Lemshow, S. and Klor, J (1988))

٣- معامل التحديد للنموذج اللوجستي (R^2) (Coefficient of determination):

تستخدم لاختبار قوة النموذج اللوجستي، أي نسبة مساهمة العوامل المؤثرة التي يتضمنها النموذج المقدر على متغير النتيجة التابع (الإصابة) من خلال المعاملات الآتية:

١. معامل كوكس وسنيل Cox and Snell (R^2_{cs})

ويحسب كالتالي:

$$R^2_{cs} = 1 - e^{\left[\frac{2}{n}[Ln\ (new) - Ln\ (baseline)]\right]}$$

$$R^2_{cs} = 1 - \left[e^{\frac{1}{2}(new - basline)}\right]^{\frac{2}{n}} \qquad (2\text{-}5\text{-}4)$$

Ln (new): لوغاريتم الاحتمال للنموذج الجديد بعد إدخال المتغيرات.

Ln (baseline): لوغاريتم الاحتمال النموذج الاصلي بدون متغيرات بالحد الثابت.

n: حجم العينة.

ويشتهر هذا المعامل بعدم وصوله إلى القيمة (١) فهناك النسخة المعدلة منه وهي:

٢. مُعامل Negelkerke (R^2_N):

$$R^2_N = \frac{R^2_{Cs}}{1 - e^{\frac{[2ln(baseline)]}{n}}}$$

$$R^2_N = \frac{R^2_{Cs}}{1 - \left[e^{-\frac{baneline}{2}}\right]}\ \frac{2}{n} \qquad (2\text{-}5\text{-}5)$$

تعتبر قيمة Exp (B) من القيم الحرجة في تفسير النموذج اللوجستي وهي مؤشر على تغير الارجحية ناتج عن تغير المتغير التنبؤي بمقدار الوحدة وتفسر كما في الجدول.

جدول (٢/٢)

قيد اللوجت B	نسبة الترجيح Exp (B)	التفسير
$B > o$	Exp (B) > 1	زيادة درجة واحدة في المتغير التنبؤي (التفسيري) يرفع من احتمالية ان يكون الفرد مصاب على ان يكون غير مصاب
$B = o$	Exp (B)= 1	تتساوى احتمالية أن يكون الفرد مصاب مع غير المصاب
$B < o$	Exp (B) < 1	نقصان درجة واحدة للمتغير التنبؤي (التفسيري) تزيد احتمال عدم الإصابة من الإصابة.

سيتم استخدام حزمة SPSS، لتحليل البيانات

المبحث الثالث

الجانب التطبيقي

(3/1) تمهيد:

تستهدف الدراسة إلى استخدام الانحدار اللوجستي في التنبؤ باحتمال الإصابة وعدم الإصابة بفيروس كورونا وتحديد العوامل المؤثرة في الإصابة بالفيروس والأهمية النسبية لكل عامل وتم ذلك من خلال متغيرات البحث التالية:

المتغير التابع (Y) (مصاب = ١، غير مصاب = ٠)

المتغيرات المستقلة وهي:

١- السن (X_1)

٢- درجة الحرارة (X_2) (عالية =١، طبيعي = ٠)

٣- ضيق التنفس (X_3) (يوجد = ١، لا يوجد = ٠)

٤- الاسهال (X_4) (يوجد = ١، لا يوجد = ٠)

٥- الصداع (X_5) (يوجد = ١، لا يوجد = ٠)

٦- التهاب الحلق (X_6) (يوجد = ١، لا يوجد ٠)

٧- حاسة الشم والتذوق (X_7) (فقدان الحاسة = ١، طبيعي = ٠)

٨- مخالطة أو عدم مخالطة حالة مصابة (X_8) (مخالط = ١، غير مخالط = ٠)

٩- الأمراض المزمنة (X_9) (يوجد = ١، لا يوجد = ٠)

١٠- السعال (الكحة)(X_{10}) (يوجد = ١، لا يوجد = ٠)

وقبل استخدام نموذج الانحدار اللوجستي الثنائي لابد من التحقق من شرط مهم وهو عدم وجود ازدواج خطي بين المتغيرات المستقلة Mullicollineanty وتعني عدم وجود ارتباط فتوى بين المتغيرات المستقلة وللتحقق من هذا الشرط وباستخدام برنامج SPSS.

جدول (١-٣)

Variables	Collinearity Statistics	
	Tolerance	VIF
x_1	٠,٧٥٧	1.321
X_2	٠,٨٦٥	1.156
X_3	٠,٩٤١	1.063
X_4	٠,٨٩٠	1.123
X_5	٠,٨٧٢	1.147
X_6	٠,٩٦٣	1.038
X_7	٠,٧٦٤	1.309
X_8	٠,٨٧١	1.148
X_9	٠,٨٠٧	1.239
X_{10}	٠,٩٢٧	1.078

من الملاحظ أن جميع قيم Variance inflation Factor (VIF) وهي قيمة عامل التضخم أصغر من ١٠ وجميع إحصائيات التسامح Tolerance أعلى من ٠,٢ ولذلك يُمكن الاستنتاج عدم وجود مشكلة ازدواج خطي بين المتغيرات المستقلة وبحساب متوسط (VIF) بجمعها وقيمتها على عدد المتغيرات يساوي 1.16 وهي قريبة جدًا من الواحد وهذا ما يؤكد خلو النموذج من مشكلة الازدواج الخطي بين المتغيرات المستقلة وبالتالي إمكانية استخدام نموذج الاستخدام اللوجستي.

حيث أن الهدف من البحث استخدام الانحدار اللوجستي في دراسة وتصنيف وتشخيص الأفراد إلى مصابين أو غير مصابين بفيروس كورونا وتحديد العوامل المؤثرة في الإصابة والمرشحة للبقاء في النموذج باستخدام لوغاريتم دالة الإمكان الأعظم من خلال الخطوات الآتية:

الخطوة الأولى: النموذج يحتوي على الحد الثابت فقط (Blocko):

جدول (٢-٣)

Iteration History

Iteration		-2 Log Likelihood	Coefficient
			Constant
step (o)	1	354.259	0.308
	2	354.258	0.310
	3	354.258	0.310

في النموذج الذي يحتوي على الحد الثابت تم الوصول لأقل قيمة لضعف دالة الامكان الأعظم عند التكرار الثالث حيث قيمة التغير أقل من 001. وبلغت القيمة -2 Log Likelihood = 354.258.

أما بالنسبة لجدول التصنيف في النموذج الذي يحتوي على الحد الثابت فقط.

Observed		Predicted		Percentage Correct
		Y		
		No	Yes	
Step (o)	No غير مصاب ٢	0	110	0
	Yesمصاب	0	150	100
				57.7

نجد أن هناك ١٥٠ شخص قد أصيبوا بفيروس كورونا و١١٠ غير مصاب بالفيروس ولذلك إذا تنبأ النموذج أن جميع الأفراد مصابين بالفيروس فأن هذا التنبؤ سليم وصحيح بنسبة ٥٧,٧٪ أي ١٥٠ من أصل ٢٦٠ وأما إذا كان توقع أن لا يصابوا بفيروس كورونا يكون صحيح بنسبة ٤٢,٣٪ أي أن نسبة التصنيف الصحيح ٥٧,٧٪.

ومن جدول المتغيرات للنموذج المبْدَئيّ بالحد الثابت فقط.

جدول (٤-٣)

	B	S.E	Wald	dF	Sig	EXP (B)
Step (o) Constant	0.310	0.126	6.150	1	0.13	1.364

من الملاحظ أن نموذج الحل المبْدَئيّ بدون متغيرات غير معنوي، حيث Sig (0.13) أكبر من 0.05 وبالتالي النموذج غير معنوي أو ليس له دلالة إحصائية، وأن المتغيرات الموجودة في المعادلة هي:

جدول (٥-٣)

		Score	dF	Sig
Step (٠)	x_1	43.442	1	00
	X_2	43.871	1	00
	X_3	9.455	1	002
	X_4	42.166	1	000
	X_5	28.755	1	000
	X_6	0.978	1	0.٠٠٠
	X_7	131.650	1	000
	X_8	33.728	1	000
	X_9	14.784	1	0.000
	X_{10}	10.905	1	0.001
Over all statistics		171.019	10	000

ومن خلال إحصائية Overall Statistics وقيمة $x^2 = 171.019$ وهي ذات دلالة إحصائية وذلك لأن (٠٠٠) sig أقل من ٠,٠٥ وذلك يعني معنوية إضافة هذه المتغيرات إلى النموذج وأن الإضافة سوف تحسن قدرة النموذج على التنبؤ.

الخطوة الأولى: Block (1)

إضافة المتغيرات المستقلة إلى النموذج المَبْدَئيّ بالحد الثابت فقط وتم التوصل إلى النتائج التالية:

جدول (٦-٣)

Step	-2 Log Likelihood	Cox& Snell R Square	Nagelkerke R Square
1	122.190	0.590	0.794

نُلاحظ أن إضافة المتغيرات التفسيرية لنموذج المبدئي بالحد الثابت ($X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}$) أدت إلى تحسن النموذج وانخفاض قيمة سالب مضاعف دالة الامكان من ٣٥٤,٢٨٨ في Block (0) إلى ١٢٢,١٩٠ في (١) Block والفرق يساوي ٢٣٢,٠٦٦ وهي قيمة x^2 في جدول (٧-٣) وهذا تحسن معنوي وذو دلالة إحصائية عند مستوى معنوية ٠,٠٥ كما هو موضح بالجدول (٧-٣) التالي:

	Chi - Square	Df	Sig
Step (1) Step	232.069	10	000
Block	232.069	10	000
Model	232.069	10	000

مما يدل على أن النموذج الذي يحتوي المتغيرات التفسيرية أفضل بكثير من النموذج المبدئي بالحد الثابت فقط، كما أن إضافة المتغيرات التفسيرية تساهم بنسبة ٥٩٪ في تحسن قدرة النموذج على التنبؤ وتساهم في الاختلافات التي تسبب الإصابة بنسبة ٧٩,٤٪ ومن جدول (٧-٣) نلاحظ فيه $x^2 = 232.069$ هي مقدار التحسن في النموذج وهو ناتج

$$x^2 = \text{- Loglik hood (Block1)}$$
$$\text{- (-2 Loglik hood (Blocko)}$$

حيثُ أنها معنوية وذات دلالة إحصائية حيث Sig (000) أقل من 0.05 وذلك يعني أفضلية النموذج في حالة وجود المتغيرات المستقلة وقد حسن النموذج التصنيف ليصل لنسبة التصنيف الصحيح إلى ٨٩,٦٪ يمكن تصنيفها بشكل صحيح في الحالات كما في جدول (٨-٣)

Observed		Predicted		Percentage Correct
		Y		
		No	Yes	
Step (١)$_Y$	No غير مصاب	٩٧	13	88.2
	Yesمصاب	١٤	150	90.7
Over all percentage				89.6%

من خلال جدول التصنيف (٨-٣) تقوم المتغيرات المستقلة بالتصنيف الصحيح لعدد ٩٧ فرد غير مصابين وتصنيف غير صحيح لغير مصابين وعددهم ١٣ أي بنسبة ٨٨,٢٪ من الحالات بشكل صحيح.

أما بالنسبة للذين أصيبوا بالفيروس كورونا تم التصنيف الصحيح ١٣٦ و١٤ بشكل غير صحيح والنسبة التصنيف الصحيحة لهم ٩٠,٧٪ أي نسبة التصنيف الإجمالية الصحيحة ٨٩,٦٪ أي أن ادخال المتغيرات أدى إلى تحسن في التصنيف من ٥٧,٧٪ إلى ٨٩,٦٪ أي إن المتغيرات ساهمت بنسبة ٦٥,٥٪ في تخفيض قيمة لوغاريتم دالة الإمكان طبقًا لإحصائية ليمشو.

$$R_l^2 = \frac{354.258-122-190}{354.258} = 65.5\%$$

وفي جدول التالي (٩-٣):

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	x1	.063	.021	9.280	1	.002	1.065	1.023	1.109
	x2(1)	-2.225	.736	9.145	1	.002	.108	.026	.457
	x3(1)	-.405	.494	.671	1	.413	.667	.253	1.757
	x4(1)	-1.922	.503	14.626	1	.000	.146	.055	.392
	x5(1)	-1.241	.556	4.977	1	.026	.289	.097	.860
	x6(1)	-.278	.483	.330	1	.565	.758	.294	1.953
	x7(1)	-4.088	.592	47.648	1	.000	.017	.005	.054
	x8(1)	.940	.473	3.951	1	.047	2.559	1.013	6.463
	x9(1)	-.321	.521	.380	1	.538	.725	.261	2.015
	x10(1)	-.041	.492	.007	1	.933	.960	.366	2.516
	Constant	1.372	1.372	1.001	1	.317	3.945		

بالنظر إلى قيمة (B) لكل من المتغيراتx_{10} ,x_9, x_8, x_7, x_6, x_5, x_4, x_3, x_2 فمن الواضح وجود علاقة عكسية محتملة بالإصابة بالفيروس وهذا يعني عند التغيير بمقدار الوحدة تسوء الارجحية للإصابة كما نلاحظ معنوية كلاً من المتغيرات x_8, x_7, x_5, x_4, x_3, x_2, x_1 حيثُ Sig أقل من ٠,٠٥ وعدم معنوية كلاً من x_3, x_6, x_9, x_{10} ومن خلال إحصائية Wald تتوصل إلى أن أكثر المتغيرات تأثيرًا في احتمالية الإصابة بفيروس كورونا

هو حاسة الشم والتذوق (x_8) حيث تأتي في المرتبة الأولى بنسبة ٤٧,٦٥٪ يليها في المرتبة الثانية الاسهال (x_4) بنسبة ١٤,٦٢٦٪ وفي المرتبة الثالثة السن (x_1) بنسبة ٩,٢٨٪ وفي المرتبة الرابعة درجة الحرارة (x_2) بنسبة ٩,١٤٥٪ وفي المرتبة الخامسة الصداع (x_5) بنسبة ٤,٩٧٧٪ أما المرتبة السادسة فهي المخالطة وعدم المخالطة (x_8) في ٣,٩٥١٪ أما باقي المتغيرات وهي ضيق التنفس، التهاب الحلق، الأمراض المزمنة والسعال ليس لها لترمعنوي من احتمالية الإصابة بفيروس كورونا.

أما بالنسبة للعمود EXP (B) تمثل نسبة الارجحية حيث أن قيمة EXP (B) بالنسبة للمتغير x_1 (السن) عن تغير x_1 بوحدة فأن أرجحية الإصابة بالفيروس تزيد لأن EXP (B) أكبر من الواحد وهي ١,٠٢٣ ويتراوح مجال النسبة لهذه القيمة ١,٠٢٣، ١,١٠٩ ولذلك يمكن أن تكون واقوت بنسبة ٩٥٪ بأن قيمة EXP (B) تقع بين هاتين القيمتين وبما أن القيمة أكبر من الواحد فأن العلاقة بين السن والإصابة صحيحة.

أما بالنسبة لقيمة EXP (B) للتغير x_2 (درجة الحرارة) تساوي ٠,١٠٨ تُشير إلى أن إذا انخفضت درجة الحرارة x_2 من (١) درجة حرارة عالية إلى (0) درجة حرارة طبيعية فإن ارجحية الإصابة تنخفض لأن EXP (B) أقل من (١) وبما أن مجال الثقة لا يتقاطع مع الواحد فأن العلاقة بين درجة الحرارة والإصابة صحيحة بنسبة ٩٥٪ في القيمة المأخوذة من المجتمع.

وذلك يعني انخفاض درجة الحرارة بوحدة فأن الإصابة بالفيروس تنخفض بمقدار ١,١٠٨ مرة. أما بالنسبة لقيمة EXP (B) للمتغير x_4 (الاسهال) والتي تساوي ٠,١٤٦ تُشير إلى إذا انخفضت قيمة x_4 من (١) يوجد اسهال إلى (٠) لا يوجد إسهال (فإن ارجحية الإصابة تنخفض لأن EXP (B) أقل من الواحد وبما أن مجال الثقة لا يتقاطع مع الواحد فالعلاقة بين الاسهال والإصابة صحيحة بنسبة ٩٥٪ أي كل انخفض الصداع وحدة انخفضت الإصابة ١,٢٨٩ مرة.

أما بالنسبة لقيمة EXP (B) للمتغير x_5 (الصداع)، ويساوي ٠,٢٨٩ تُشير إلى انه لو انخفض الصداع من (١) لا يوجد إلى (٠) يوجد انخفضت ارجحية الإصابة، حيث إن قيمة EXP (B) أقل من (١) والعلاقة بين الصداع والإصابة صحيحة بنسبة ٩٥٪ أي كُلما أنخفض الصداع وحدة انخفضت الإصابة ١,٢٨٩ مرة.

أما بالنسبة لقيمة EXP (B)للمتغير x_7 (حاسة الشم والتذوق) وقيمتها ٠,٠١٧ إذا انخفضت حاسة الشم من (١) فاقد الحاسة إلى (٠) طبيعي انخفضت ارجحية الإصابة لأن القيمة أقل من الواحد وبما أن فترة الثقة لا تحتوي

على الواحد فالعلاقة بين حاسة الشم والتذوق والإصابة علاقة صحيحة بنسبة ٩٥٪ أي كلما انخفضت الحاسة الشم والتذوق من (١) إلى (٠) انخفضت الإصابة ١,٠١٧ مرة.

أما بالنسبة لقيمة EXP (B) للمتغير x_8 (مخالط أو غير مخالط) وهي ٢,٢٢٩ تعني عن زيادة x_8 من (٠) غير مخالط إلى (١) مخالط تؤدي إلى زيادة الأرجحية الإصابة لأن EXP (B) أكبر من الواحد. ومجال الثقة ١,٠١٣، ٦,٤٦٣ العلاقة صحيحة بين حاسة الشم والتذوق والإصابة كلما زادت x_8 بوحدة زادت ارجحية الإصابة ٢,٥٥٩ مرة.

كما يُمكن صياغة نموذج الانحدار اللوجستي بوحدات اللوجت

$$\text{Log}\left(\frac{P}{1-P}\right) = 1.372+0.063x_1 - 2.225x_2 - 0.405x_3 -1.922x_4 - 1.241x_5 -0.278x_6 -4.088x_7 +0.940x_8 - 0.321x_9 -0.041x_{10}$$

ولاختبار صورة ملاءمة النموذج اللوجستي الثنائي، Hosmer& Lemshow.

جدول (١٠-٣)

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	6.611	8	.579

نُلاحظ أن Sig (0.579) أكبر من ٠,٠٥ وبالتالي قبول الفرض العدمي بملاءمة النموذج للبيانات.

أما في حالة التنبؤ باحتمالية الإصابة فالنموذج كالتالي:

$$\text{P (Y)} = \frac{1}{1+e^{-(Bo+B1X1+B2X2+B4X4+BsXs+B7X7+B8X8)}}$$

$$\text{P (Y)} = \frac{1}{1+e^{-(1.372+0.63x1-2.225x2-1.922x4-1.241x5-4.088x7+0.940x8)}}$$

وهذه المتغيرات الإحصائية ذات علاقة بالإصابة بالفيروس كورونا وتحسن التنبؤ بها

$$R^2_{Cs} = 1 - (e^{\frac{1}{2}(122.190-354.258)\frac{2}{260}} = 59\%$$

وإن هذه المتغيرات فاقت بتفسير التغيرات في المتغير التابع بنسبة ٧٩,٤٪

$$R^2_N = \frac{0.59}{1-(e^{354.258})^{\frac{2}{260}}} = 79.4\%$$

وإن المتغيرات قامت بالتصنيف الصحيح بنسبة ٨٩,٦٪

من خلال النموذج السابق لو هناك شخص عمره (٥٥) درجة حرارته عالية (١) ولا يوجد إسهال (٠) ولا يوجد صداع (٠) ولدينه حاسة الشم والتذوق (٠) وغير مخالط (٠) فأن احتمال إصابته:

$$P(Y) = \frac{1}{1+e^{(1.372+0.063(55)-2.225(1)-1.922(0)-1.241(0)-4.088(0)+0.940(0))}} = 93.16\%$$

هُناك احتمال إصابته بالفيروس ٩٣,١٦٪ أن تكون شخص مصاب أما لو كانت درجة حرارته عادية (٠) مع ثبات باقي العوامل.

$$P(Y) = \frac{1}{1+e^{-(1.372+0.063(55)-2.225(0)-1.922(0)-1.241(0)-4.088(0)+0.940(0))}} = 99.2\%$$

فأن هناك احتمال عدم إصابته بفيروس كورونا بنسبة ٩٩,٢٪ وذلك يوصل أهمية درجة الحرارة وتأثيرها على الإصابة بفيروس كورونا.

إذا كان لدينا شخص عُمره (٧٠) سنة ودرجة حرارة عادية (٠) ويعاني من الاسهال (٠) ولا يعاني من الصداع (٠) وغير مخالط (٠) وفاقد حاسة الشم والتذوق (١) فاحتمال الإصابة.

$$P(Y) = \frac{1}{1+e^{-(1.372+0.063(70)-2.225(0)-1.922(0)-1.241(0)-4.088(1)+0.940(0)}} = 97.2\%$$

فهناك احتمال إصابته بنسبة ٩٧,٢٪ لو نفس الشخص وغير فاقد حاسة الشم والتذوق (٠).

$$P(Y) = \frac{1}{1+e^{-(1.372+0.063(70)-2.225(0)-01.922(0)-1.241(0)-4.088(1)+0.940(0)}} = 9٩.٧\%$$

احتمال عدم إصابته بالفيروس ٩٩,٧٪ وذلك يعكس أهمية حاسة الشم والتذوق في تأثيرها على الإصابة بالفيروس كورونا.

النتائج والتوصيات:

يهدف البحث إلى استخدام نموذج الانحدار اللوجستي الثنائي في أعراض فيروس كورونا Covid-19 وتكوين نموذج إحصائي للتنبؤ باحتمالية الإصابة أو عدم الإصابة وتحديد الأهمية النسبية للأعراض المؤثرة في الإصابة بالفيروس وتم التوصل إلى النتائج الآتية:

١. تأكد خلو البيانات من مشكلة الازدواج الخطي وبالتالي إمكانية استخدام نموذج الانحدار اللوجستي الثنائي.

٢. تم التأكد من معنوية إضافة متغيرات المستقلة إلى معادلة الانحدار وذلك باستخدام x^2 Chi- Square.

٣. تأكد من معنوية كلاً من العوامل الآتية حاسة الشم والتذوق ودرجة الحرارة والاسهال والصداع ومخالط او غير مخالط والسن وكان ترتيب المتغيرات حسب الأهمية النسبية وتأثيرها في احتمالية الإصابة كالتالي:

المتغير	نسبة الأهمية
١- الشم والتذوق	٪٤٧,٦
٢- الاسهال	٪١٤,٦٢
٣- السن	٪٩,٢٨
٤- درجة الحرارة	٪٩,١٢٤
٥- الصداع	٪٤,٩٧٧
٦- المخالطة وغير المخالطة	٪٣,٩٥

٤. تم التأكد من عدم معنوية كلاً من المتغيرات الآتية، ضيق التنفس والتهاب الحلق والأمراض المزمنة والسعال وتم استبعادهم من النموذج.

٥. كما تم التوصل إلى ان المتغيرات تساهم بنسبة ٥٩٪ في جودة تنبؤ النموذج كما تساهم بنسبة ٧٩,٤٪ في التغيرات التي تحدث في الإصابة بالفيروس كورونا.

٦. ووفقًا لمعيار Hosmer& Lemshow تبين مدى جودة مطابقة النموذج للبيانات حيث Sig أقل من ٠,٠٥ .

٧. المتغيرات التي في النموذج تساهم تصنيف صحيح بنسبة ٨٩,٦٪ للحالات.

التوصيات:

يوصي الباحث بالآتي:

١- التوسع في استخدام الانحدار اللوجستي في المجالات الاقتصادية وعدم اقتصاره على المجالات الطبية والاجتماعية لما اظهره من جوده في التنبؤ والتصنيف.

٢- إدخال اعراض أخرى تؤثر في احتمالية الإصابة مثل النوع، والوقت، والتدخين، أمراض الرئة، وهكذا.

٣- يوصي باستخدام النموذج اللوجستي المقدر في الدراسة لتحديد احتمالية إصابة أو عدم إصابة الأشخاص بالفيروس كورونا واستخدامه في المستشفيات .

٤- الاهتمام بتسجيل البيانات للمرضى لما لهُ من أثر هام في الدراسات والبحوث الإحصائية.

٥- إمكانية استخدام طريقة تحليل أخرى في تحديد الأثار والأعراض المهمة للإصابة بالفيروس كورونا ومقارنتها بالنتائج التي تم التوصل إليها للوصول إلى حزمة من العوامل المؤثرة في احتمالية الإصابة بالفيروس ومن خلال يُمكن السيطرة على هذا الوباء.

المراجع الأجنبية:

1- Agresti, A., (2007) "**Categorical Data Analysis**" New York. Johnson wiley& sons. Inc.

2- Cizek, G., J& Fit, Z. S. M., (1999). "Methods, Plain speaking: An introdution to logistic regression "**Measurement& Evaluation in counseling and Development**" Vol. 31. January. 1999.

3- Cox, D.R., (1996) "**Some Procedures associated with the logistic qualitative response curves**" in research paper in statistics. Birkbeck College, University of London P.P 55-71.

4- Dallal,G.,(2001)"Logistic regression" Available at: **www.tufts.edu/galallal/logistic.htm**.

5- Hilstan, Keener Mary. (2013) "**Predicting the Financial Failure of retail companies in the United states**" Journal of Business& Economics Research, Vol-11, No 8, PP.370-373.

6- Hosmer D.W., Lemshow, S. and Klor, J (1988) "**Goodness of fit testing for the logistic model when the estimated Probability are small**" 2nd edistion. New York, Johnson Wiley& Sans. Inc.

7- Hosmer, D., &lemeshow, S., (2000). "**Applied Logistic regression**" 2nd edistion. New York, Johnson Wiley& Sans. Inc.

8- Nakarin, Sinamnuayphol and Chancharat Nangnit (2015)"**The Financial Failure Predictions Model of Financial Insinuations in the stock Exchange of Thailand**" Australian Journal of Basic and Applied Sciences, 9 (24), Special, PP.1-7.

9- Rencher, A. C., (2002) "**Methods of Multivariate analysis**" second edition, wiley, interscience published simullaneously in Canada.

المراجع العربية:

١- أبو دومة، حيدر جميل الله محمد (٢٠١٩) "استخدام أسلوب تحليل الانحدار اللوجستي والتحليل التميزي للعوامل المؤثرة على الإصابة بأمراض القلب. كلية الدراسات العليا – جامعة السودان للعلوم والتكنولوجيا.

٢- أحمد، شمس الدين أحمد علي (٢٠١٣) "تقدير وتحليل العوامل المؤثرة في الرضا الوظيفي باستخدام الانحدار اللوجستي" كلية الدراسات العليا والبحث العلمي – جامعة شندي – السودان.

٣- إدريس، عبد الله جمعة (٢٠١٥) استخدام تحليل الانحدار اللوجستي الثنائي لدراسة أهم العوامل الاقتصادية والاجتماعية للآباء وأثرها على تحصيل تلاميذ الصف الثامن "دراسة تطبيقية، الدراسات العليا والبحوث، جامعة الجزيرة، السودان.

٤- خليفة، صهيب الشيخ عبد المجيد (٢٠١٤) تطبيق الانحدار اللوجستي لمعرفة أهم العوامل الاجتماعية والاقتصادية للتدخين" الدراسات العليا، والبحث العلمي – جامعة الجزيرة، السودان.

٥- صالح، عائدة هادي (٢٠١٤) "تحليل الانحدار اللوجستي لدراسة زمن البقاء لمرضى سرطان الدم" جامعة كربلاء، العراق، مجلس الإدارة والاقتصاد، المجلد الثالث – العدد التاسع.

٦- غانم، عدنان والجاعوني، فريد خليل (٢٠١١) "استخدام تقنية الانحدار اللوجستي الثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الأسرة" مجلس دمشق للعلوم الاقتصادية والقانونية – المجلد السابع والعشرين العدد الأول – جامعة دمشق.

٧- الفرهود، سهيلة حمود عبد الله (٢٠١٤) "استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على أداء الأسهم"، دراسة تطبيق على سوق الكويت للأوراق المالية، كلية الدراسات التجارية – مجلس الأزهر.

٨- مهران، سحر (٢٠١٧) "التنبؤ بمراحل تأثير الأزمة المالية على أداء المصارف باستخدام الانحدار اللوجستي" دراسة تحليلية بالتطبيق على قطاع المصارف في دول مجلس التعاون الخليجي –كلية التجارة – جامعة عين شمس – المجلة العربية للإدارة.