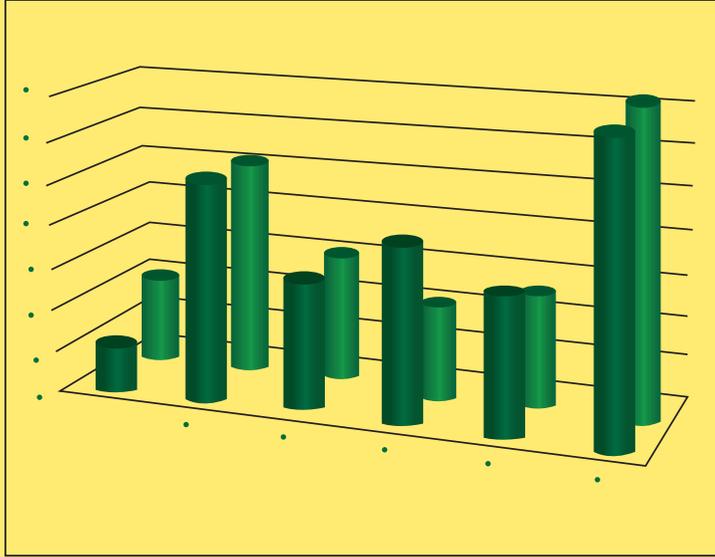


المعهد العربي للتدريب والبحوث الإحصائية



# مجلة العلوم الإحصائية



العدد رقم 21

مجلة علمية محكمة  
يصدرها المعهد العربي للتدريب والبحوث الإحصائية

معتمدة في قائمة المجلات العلمية Ulrich's  
[www.ulrichsweb.com](http://www.ulrichsweb.com)

مصنفة في معامl التأثير والاستشهادات المرجعية العربي (أرسيف)  
[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)

# مجلة العلوم الإحصائية

مجلة علمية محكمة

هيئة التحرير

رئيس هيئة التحرير

الأستاذ الهادي السعيد

أمين التحرير

الدكتور لحسن عبد الله باشيوه

أعضاء هيئة التحرير

أ. د. فيصل الشرعي  
أ. د. عيسى مصاروه  
أ. م. د. حسان أبو حسان  
أ. د. عبد الخالق التهامي  
أ. م. د. سلوى محمود عسار  
أ. م. د. حميد بوزيدة

أعضاء الهيئة الاستشارية

د. قاسم الزعبي  
د. ضياء عواد  
د. لؤي شبانه  
د. نبيل شمس  
د. خليفة البرواني  
د. غازي رحو  
د. علا عوض  
أ. د. عوض حاج علي  
أ. د. ميثم العيبي اسماعيل  
أ. م. د. محمد حسين علي الجنابي

معتمدة في قائمة المجلات العلمية Ulrich's

[www.ulrichsweb.com](http://www.ulrichsweb.com)

مصنفة في معامل التأثير والاستشهادات المرجعية العربي (أرسيف)

[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)

## شروط النشر في مجلة العلوم الإحصائية

- 1 - تنشر-المجلة البحوث والدراسات العلمية في المجالات الإحصائية والمعلوماتية المكتوبة باللغة العربية والانكليزية والفرنسية على أن لا يكون البحث المقدم للنشر. قد نشر. او قدم للنشر في مجلات او دوريات أخرى او قدم ونشر في دوريات لمؤتمرات أو ندوات.
- 2 - ترسل البحوث والدراسات الى أمين التحرير على أن تتضمن اسم الباحث او الباحثين وألقابهم العلمية وأماكن عملهم مع ذكر عنوان المراسلة وأرقام الهواتف والبريد الالكتروني. وان يرسل البحث المراد نشره الكترونياً (على قرص او بالبريد الالكتروني) وفق المواصفات أدناه:
  - أ - أن يكون مطبوعاً على ورق حجم A4 وان يكون على شكل عمود واحد ويستخدم للغة العربية نوع حرف (Simplified Arabic) و(Times New Roman) للإنجليزية والفرنسية وبحجم خط (12). وباستخدام Microsoft Word وعلى وجه واحد للورقة.
  - ب - الهامش مسافة 2.5 سم لجميع جوانب الورقة.
  - ج - يرفق الباحث ملخصاً عن بحثه باللغتين العربية والانجليزية والفرنسية بما لا يزيد عن صفحة واحدة.
  - د - يتم الإشارة الى المصادر العلمية في متن البحث وفي نهايته، مع مراعاة أن لا يتضمن البحث سوى المصادر التي تم الإشارة إليها في المتن ووفق الأصول المعتمدة في ذلك (اسم المؤلف، سنة النشر، عنوان المصدر، دار النشر، البلد).
  - هـ - ترقم الجداول والرسوم التوضيحية وغيرها حسب ورودها في البحث، كما توثق المستعارة منها بالمصادر الأصلية.
  - و- أن لا يزيد عدد صفحات البحث او الدراسة عن (25) صفحة.
- 3 - يتم إشعار الباحث باستلام بحثه خلال مدة لا تتجاوز يومين عمل من تاريخ استلام البحث.
- 4 - تخضع كافة البحوث المرسله الى المجلة للتقييم العلمي الموضوعي ويبلغ الباحث بالتقييم والتعديلات المقترحة إن وجدت خلال مدة لا تتجاوز اسبوعان من تاريخ استلام البحث.
- 5 - لهيئة تحرير المجلة الحق في قبول او رفض البحث ولها الحق في إجراء أي تعديل او إعادة صياغة جزئية للمواد المقدمة للنشر- بما يتماشى والنسق المعتمد في النشر- لديها بعد موافقة الباحث.
- 6 - يصبح البحث المنشور ملكاً للمجلة ولا يجوز إعادة نشره في أماكن أخرى.
- 7 - تعبر المواد المنشورة بالمجلة عن آراء أصحابها، ولا تعكس وجهة نظر المجلة او المعهد العربي للتدريب والبحوث الإحصائية.
- 8 - ترسل البحوث على العنوان الالكتروني للمجلة:

[journal@aitrs.org](mailto:journal@aitrs.org) / [Info@aitrs.org](mailto:Info@aitrs.org)

## المحتويات

رقم الصفحة	اسم البحث والباحث	ت
1	مقارنة بعض الطرائق الحصينة لتقدير انحدار بواسون بأستخدام المحاكاة أم اسيل عبد الرزاق رشيد/ الجامعة المستنصرية / كلية الادارة والاقتصاد الباحث محمد غازي جواد / مديرية تربية محافظة بغداد / الكرخ الثانية	1
16	العوامل الاقتصادية والاجتماعية للطلب على القروض الشخصية في محافظة عمان خلال جائحة فيروس كورونا (2020-2022) الدكتورة رانية عدنان السطل / اختصاصي مساعد/ قسم مرصد المشاريع الاقتصادية الصغيرة والمتوسطة / المؤسسة الأردنية لتطوير المشاريع الاقتصادية	2
36	A statistical analysis of the Type II Tobit model maximum Likelihood in case of non-ignoring missing data Mona Emad El-Din Mohamed, Mervat El-Gohary Department of Statistics- Faculty of commerce-Al-Azhar University Ahmed Amin El-Sheikh\ Dept. of Applied Statistics and Econometrics Faculty Cairo University	3
51	A Five State Progression Model: Estimation and Application Ayat Ahmed, Nahed Helmy, Mervat Elgohary Al- Azhar University Faculty of Commerce – Girls' Branch Department of Statistics.	4
70	Effect of COVID 19 on women and child care in Iraq Dr. Arjwan Marawan Shaaban\Family physician, Director of department of specialized centers and hospitals at MOH. Dr.Eman A. Al-Tae\ Family physician at Nineveh health directorate Dr.Raouf Tariq Raouf\ Director of the health statistical division at MOH	5



## مقارنة بعض الطرائق الحصينة لتقدير انحدار بواسون بأستخدام المحاكاة

الباحث محمد غازي جواد  
مديرية تربية محافظة بغداد / الكرخ الثانية  
أ.م اسيل عبد الرزاق رشيد  
الجامعة المستنصرية / كلية الادارة والاقتصاد

تاريخ استلام البحث: 2023/03/01

تاريخ قبول البحوث: 2023/03/17

نشر البحث في العدد الواحد والعشرون: ديسمبر / كانون اول 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/519.5

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/519.5

## مقارنة بعض الطرائق الحصينة لتقدير نموذج انحدار بواسون بأستخدام المحاكاة

الباحث محمد غازي جواد  
مديرية تربية محافظة بغداد / الكرخ الثانية  
أ.م اسيل عبد الرزاق رشيد  
الجامعة المستنصرية / كلية الادارة والاقتصاد

### المستخلص

يعد انموذج انحدار بواسون احد نماذج العد المهمة حيث استخدم لنمذجة العديد من الظواهر الاقتصادية والصحية وحوادث المرور وغيرها. يتم في هذا البحث تقدير انموذج انحدار بواسون في حالة وجود القيم الشاذة, حيث تم اجراء دراسة محاكاة مونت كارلو, لمقارنة المقدرات الحصينة (M, S, MM) مع المقدر الاعتيادي, وهي طريقة الامكان الاعظم (MLE), وقد اظهرت نتائج المحاكاة اعتمادا على معايير المفاضلة (MAE, MSE) ان طريقة (M) الحصينة اكثر كفاءة من باقي الطرائق الحصينة, وكذلك من الطريقة الاعتيادية للتقدير (MLE).

**الكلمات المفتاحية:** انحدار بواسون, طريقة مقدرات M, طريقة مقدرات S, طريقة مقدرات

MM

### Comparison of some robust methods for estimating the Poisson regression model using simulation

#### Abstract

The Poisson regression model is one of the important counting models, as it was used to model many economic and health phenomena, traffic accidents, and others. In this research, the Poisson regression model is estimated in the case of the presence of outliers, where a Monte Carlo simulation study was conducted, to compare the robust estimators (M, S, MM) with the usual estimator, which is the method of Maximum Likelihood Estimator (MLE), and the results of the simulation were shown depending on comparison criteria (MSE, MAE) The robust (M) method is more efficient than the rest of the robust methods, as well as the conventional method of estimation (MLE).

**Keywords:** Poisson regression, M estimators method, S estimators method, MM estimators method.

## المقدمة

يعتبر نموذج انحدار بواسون احد نماذج الانحدار غير الخطية، وكذلك احد اهم نماذج الانحدار اللوغارتمية الخطية وهو الاداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد. اذ يفترض هذا النموذج ان متغير الاستجابة يتبع توزيع بواسون بمتوسط وتباين متساويان. وفي حالة وجود القيم الشاذة في البيانات يصبح من الصعوبة الحصول على نتائج دقيقة لتقدير معالم النموذج بالطرائق الاعتيادية لذلك سوف نلجأ الى الطرائق الحصينة للتقدير وهي ( طريقة مقدرات M ، طريقة مقدرات MM ، طريقة مقدرات S) ونجري مقارنة بينها اعتمادا على معايير المقارنة لنعرف ماهي الطريقة الافضل لتقدير معالم نموذج انحدار بواسون.

## انموذج انحدار بواسون [9][10]

يعرف انموذج انحدار بواسون بأنه احد انواع نماذج الانحدار اللخطية والمستخدم في نمذجة البيانات المعدودة ويمكن كتابة انموذج انحدار بواسون على النحو التالي:

$$Y_i = E(Y_i) + u_i \\ = \lambda_i + u_i \quad \dots\dots\dots (4)$$

حيث ان  $i = 1, 2, \dots, n$

وان متغير الاستجابة  $Y_i$  يتبع توزيع بواسون بمتوسط وتباين يساوي  $(\lambda_i)$

$$\lambda_i = E(Y_i | X_i) = e^{\underline{X}_i' B} \quad \dots\dots\dots (5)$$

وعلى وفق هذه الصيغة فان المتغيرات التوضيحية تقترح اي قيمة حقيقية لمتوسط متغير الاستجابة (بمعنى ان المتوسط اما يكون موجب او سالب)، وهذا يتعارض مع خاصية معلمة توزيع بواسون وهي ان قيمة  $\lambda$  موجبة فقط، لذلك تم اقتراح نمذجة علاقة خطية لوغارتم المتوسط  $(\log \lambda)$  والمتغيرات التوضيحية  $(X_i)$  اي اعتماد النموذج الخطي اللوغارتمى للتخلص من القيم السالبة لمتوسط متغير الاستجابة  $(Y)$

$$\log (E(Y_i)) = \log (\lambda_i) = \underline{X}_i' B \quad \dots\dots\dots (6)$$

حيث ان :

$$\underline{B} : \text{يمثل متجه المعالم ذو الدرجة } 1 * (K+1) \text{ المطلوب تقديرها}$$

$$\underline{X}_i' : \text{يمثل متجه المتغيرات التوضيحية ذو الدرجة } 1 * (k+1)$$

اذ تقوم الفكرة الاساسية على ربط الدالة الاحتمالية لمتغير الاستجابة  $(Y_i)$  حيث  $i=1, 2, \dots, n$  و  $n$  هي عدد مشاهدات العينة بمتجه  $\underline{X}$  الذي هو عبارة عن مجموعة من المتغيرات التوضيحية  $(X_j)$  حيث  $(j=0, 1, 2, \dots, K)$  وبالاعتماد على هذا النموذج فان معاملات الانحدار  $B_j$  تمثل التغير المتوقع في لوغارتم المتوسط لمتغير الاستجابة  $Y$  لكل وحدة تغير حدثت في المتغير التوضيحي  $X_j$ .

وبأخذ  $\exp$  لطرفي العلاقة رقم (6) نحصل

$$\lambda_i = \exp (\underline{X}_i' B) \quad \dots\dots\dots (7)$$

وبالتالي فإن معامل الانحدار الاسي exponential coefficient regression يمثل التأثير المضاعف للمتغير التوضيحي  $X_j$  على متوسط الاستجابة  $Y$  حيث ان انموذج انحدار بواسون يستند على ثلاث فرضيات :

**الافتراض الاول** : متغير الاستجابة  $Y$  يتوزع وفق توزيع بواسون بمعلمة  $\lambda$  وان ( $\lambda > 0$ ) وحسب العلاقة (1) .

**الافتراض الثاني** : المعلمة  $\lambda$  والتي تمثل متوسط قيم متغير الاستجابة  $Y$  وتمثل التباين ايضا يمكن ان يعبر عنها استنادا الى متجه المتغيرات التوضيحية وكما مبين بالصيغة (7) .

**الافتراض الثالث** : ان الأزواج المرتبة للملاحظات  $(X_i, Y_i)$  لها توزيع مستقل .  
اجمالا يمكن دمج الافتراضين الاول والثاني الحصول على الدالة الاحتمالية الشرطية التالية :

$$f(y_i, x_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} = \frac{e^{-e(x_i/B)} e^{y_i x_i/B}}{y_i!} \quad y=0,1,2,\dots \quad \dots(8)$$

ان انموذج انحدار بواسون والذي يمتلك الخصائص التي ذكرناها سابقا يحقق دالة الوسط الحسابي الشرطية الاسية (الخطية اللوغارتمية) لان الانموذج يتحول بأخذ اللوغارتم الطبيعي الى انموذج خطي .

$$E(Y/X) = \lambda = \exp(x_i/B) \quad \dots(9)$$

وايضا دالة التباين الشرطية الاسية (الخطية اللوغارتمية)

$$\text{Var}(Y/X) = \lambda = \exp(x_i/B) \quad \dots(10)$$

وذكرنا سابقا عند مناقشة توزيع بواسون خاصة تساوي الوسط الحسابي والتباين، وهنا ايضا تنطبق هذه الخاصية في انموذج انحدار بواسون (التوزيع المتساوي) .

وعليه فإن الصيغة الخطية ل  $\lambda_i$  تصبح  $\log \lambda_i = x_i/B$

### طرائق التقدير الاعتيادية

يوجد الكثير من الطرق الاعتيادية (الكلاسيكية) للتقدير وبرزها والاكثر استخداما طريقة المربعات الصغرى الاعتيادية (OLS) وطريقة الامكان الاعظم (MLE) .

وسوف يتم التطرق الى طريقة الامكان الاعظم (MLE) بالتفصيل لتقدير انموذج انحدار بواسون والتي تهدف الى تعظيم الدالة الاحتمالية وجعل دالة الامكان في نهايتها العظمى ، حيث تتصف مقدرات هذه الطريقة بكونها لاتمتلك خصائص مرغوبة في العينات الصغيرة ، بينما تمتلك هذه المقدرات في العينات الكبيرة خصائص مرغوبة اضافة الى الاتساق [1] .

### طريقة الامكان الاعظم [9]

نأخذ عينة بحجم  $n$  من ازواج المشاهدات  $(y_i, x_i)$  المستقلة فان دالة الامكان سوف تمثل حاصل ضرب دوال الكتلة الاحتمالية الشرطية لكل مشاهدة  $(y_i)$  فتكون :

$$L(B; y_1, y_2, \dots, y_n, x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(y_i/x_i; B) \quad \dots(11)$$

وان مقدر الامكان الاعظم هو المقدر الذي يجعل دالة الامكان اكبر مايمكن وهنا سوف نعوض باللوغارتم في دالة الامكان لتسهيل العمليات الحسابية

$$L(B; Y, X) = \text{Log} \prod_{i=1}^n f(y_i/x_i; B) \quad \dots(12)$$

$$L(B; Y, X) = \sum_{i=1}^n \text{Log} f(y_i/x_i; B) \quad \dots(13)$$

$$L(B; Y, X) = \sum_{i=1}^n [-\exp(x_i/B) + y_i x_i/B - \text{Log}(y_i!)] \quad \dots(14)$$

الآن نشتق لوغارتيم دالة الامكان من خلال اشتقاق متجه المعلمات  $\underline{B}$  وبالتالي سوف نحصل على  $(k+1)$  من المشتقات :

$$A_n = \frac{dL(\underline{B}; \underline{Y}, \underline{X})}{d\underline{B}} = \sum_{i=1}^n [-\exp(\underline{x}_i' \underline{B}) \underline{x}_i + y_i \underline{x}_i'] \quad \dots (15)$$

$$A_n = \frac{dL(\underline{B}; \underline{Y}, \underline{X})}{d\underline{B}} = \sum_{i=1}^n [y_i - \exp(\underline{x}_i' \underline{B})] \underline{x}_i \quad \dots (16)$$

ثم نساوي المشتقة الاولى بالصفر

$$A_n = \frac{dL(\underline{B}; \underline{Y}, \underline{X})}{d\underline{B}} = 0$$

$$\sum_{i=1}^n [y_i - \exp(\underline{x}_i' \underline{B})] \underline{x}_i = 0 \quad \dots (17)$$

وإذا اردنا التأكد من صحة الاشتقاق نشتق المعادلة (15) مرة ثانية فنحصل :

$$T_n = \frac{d^2 L(\underline{B}; \underline{Y}, \underline{X})}{d\underline{B} d\underline{B}'} = -\sum_{i=1}^n \exp(\underline{x}_i' \underline{B}) \underline{x}_i \underline{x}_i' \quad \dots (18)$$

نلاحظ ان المشتقة الثانية تكون سالبة معناها اشتقاقنا صحيح . ان معلمات المعادلة (17) غير خطية وتحل باستخدام احدى الطرق التكرارية ، وسوف نستخدم طريقة نيوتن رافسون حيث نقوم بأعطاء قيم تقديرية اولية لمعالم الانموذج  $\hat{B}^0$  بحيث نحصل على تقريبات من الدرجة الثانية في لوغارتيم دالة الامكان  $L(\underline{B})$  حول القيم الاولية  $\hat{B}^0$  وكالتالي :-

$$L^*(\underline{B}) = L(\hat{B}^0) + A_n(\hat{B}^0)'(\underline{B} - \hat{B}^0) + \frac{1}{2}(\underline{B} - \hat{B}^0)' T_n(\hat{B}^0)(\underline{B} - \hat{B}^0) \approx L(\underline{B}) \dots (19)$$

وبتعظيم  $L^*(\underline{B})$  بدلا من  $L(\underline{B})$  بالنسبة للمعالم فنحصل على تقدير جديد لها وليكن  $\hat{B}^1$  والقيود الضرورية لتعظيم المعادلة يكون كالتالي :-

$$A_n(\hat{B}^0) + T_n(\hat{B}^0)(\hat{B}^1 - \hat{B}^0) = 0 \quad \dots (20)$$

ثم نحصل على  $\hat{B}^1$

$$\hat{B}^1 = \hat{B}^0 - [T_n(\hat{B}^0)]^{-1} A_n(\hat{B}^0) \quad \dots (21)$$

اذن صيغة نيوتن رافسون بأعتبار  $\hat{B}_0$  قيمة تقديرية اولية يكون كالتالي :-

$$\hat{B}^{k+1} = \hat{B}^k - [T_n(\hat{B}^k)]^{-1} A_n(\hat{B}^k) \quad \dots (22)$$

حيث  $k=0,1,2,\dots$

وعندما يكون الفرق بين  $[\hat{B}^{k+1} - \hat{B}^k]$  قريبة جدا من الصفر (مثلا عند النقطة 0.00001) نتوقف عن تكرار التقدير .

### مشكلة وجود القيم الشاذة في البيانات [2][3][6]

ان مشكلة وجود القيم الشاذة في البيانات قد افقد طرائق التقدير الاعتيادية اهميتها وبالتالي اصبحت غير دقيقة في تقدير المعلمات اي اننا نحصل من خلالها على معلمات ذات خصائص غير جيدة بسبب عدم توفر اهم شرط لتطبيق هذه الطرائق وهو(عدم وجود القيم الشاذة في البيانات) .

ويمكننا ان نعرف القيم الشاذة outliers بأنها القيم التي تشذ عن بقية المشاهدات ولا تنسجم معها فعند ترتيب البيانات بشكل تصاعدي تكون موقع القيم الشاذة اسفل او اعلى الترتيب وبعيدة عنه ايضا , حيث تتبع القيم الشاذة توزيع غير التوزيع الاصلي للبيانات , وتنتج القيم الشاذة من اخطاء القياس او اخطاء التسجيل او اخطاء المعاينة او تكون نتيجة حدث غير عادي مثل الحروب او الكوارث الطبيعية وغيرها .

يكون وجود القيم الشاذة في البيانات بثلاثة انواع  
 النوع الأول : وجود القيم الشاذة في مشاهدات متغير الأستجابة  $y$   
 النوع الثاني : وجود القيم الشاذة في مشاهدات المتغيرات التوضيحية  $x_j$   
 النوع الثالث : وجود القيم الشاذة في مشاهدات متغير الأستجابة  $y$  والمتغيرات التوضيحية  $x_j$   
 هناك عدة طرائق للكشف عن وجود القيم الشاذة في البيانات مثل طريقة الرسم الصندوقي والقطع المخططة ، وطريقة مصفوفة H ، وطريقة مسافات مهلنوبس التربيعية العادية ، وطريقة مسافات مهلنوبس التربيعية الحصينة وغيرها .  
 وسوف نكتفي هنا بالتطرق الى الطريقة الأولى للكشف عن وجود القيم الشاذة في البيانات

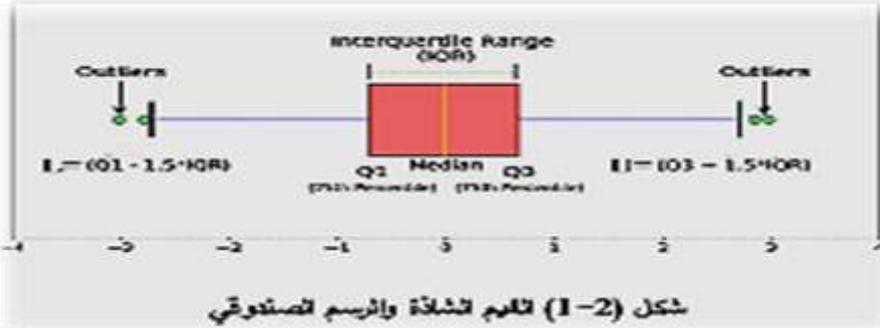
### طريقة الرسم الصندوقي والقطع المخططة [3]

هي أكثر الطرائق شيوعا وسهولة في الاستخدام وتعتمد على الربيعات حيث ان الربيع الثاني او نصف المدى الربيعي الذي يكشف 50% من انحرافات المشاهدات ونحدد الربيع الأدنى ( $Q_1$ ) ونحدد الربيع الأعلى ( $Q_3$ ) ونحدد اقل قيمة غير شاذة (L) ونحدد اكبر قيمة غير شاذة (U) وبالتالي اي قيمة تزيد عن U او تقل عن L نعتبرها قيمة شاذة .  
 نحدد قيمة L ، U من خلال المعادلات الآتية :

$$L = Q_1 - 1.5(Q_3 - Q_1) = Q_1 - 1.5 * IQR$$

$$U = Q_3 + 1.5(Q_3 - Q_1) = Q_3 + 1.5 * IQR$$

الخطوط العمودية الثلاثة للمربع تدل على الربيعات الثلاثة والخطوط المتباعدة تجاه اليمين واليسار تسمى بالشوارب معنى ذلك ان اي قيمة خارج هذه الشوارب تعتبر قيمة شاذة كما موضح بالشكل (1-2) .



بالتالي لتلافي مشكلة القيم الشاذة في البيانات وحتى لا تؤثر على عملية تقدير المعلمات سوف نستخدم التقديرات الحصينة للحصول على مقدرات ذات خصائص جيدة .

### طرائق التقدير الحصينة

اول من اطلق مصطلح المقدرات الحصينة هو الباحث Box في عام 1953 ثم توالى الدراسات بخصوصها من قبل عدة باحثين مثلا الباحث Johan Tukey في الفترة (1960-1962) تلاه الباحث Peter Huber في الفترة (1964-1967) ثم الباحث Frank Hampel في الفترة

(1971-1974) حيث اقترحوا عدة اساليب للأحصاءات الحصينة من خلال اصدار عدد من البحوث والدراسات في هذا المجال. [8]  
ان كلمة الحصانة تطلق على المقدرات التي لاتتأثر او تتحسن بسهولة لوجود مخالفة في احدى فرضيات التوزيع الطبيعي ، وان الهدف الاساسي من ايجاد طرائق التقدير الحصين هو لتقليل تأثير القيم الشاذة على المقدر. [2]

ان اتباع الطرق الكلاسيكية في تقدير المعالم سواء كان الانموذج خطيا او غير خطي ليس اكثر امانا من تطبيق الطرق الحصينة وذلك بسبب ان الفروض الواجب توفرها لتطبيق الطرق الكلاسيكية ليست بالسهلة مثل عدم وجود القيم الشاذة او اتباع الخطا العشوائي توزيعا غير التوزيع الذي يناسب الطريقة المعتمدة في التقدير. [5]

ان الهدف من الاحصاءات الحصينة هو تحسين الاجراءات التي يمكن الوثوق بها والتي تكون كفاءة في ظل وجود القيم الشاذة في البيانات، ان الاحصاءات الحصينة هي امتداد للاحصاءات المعلمية مع الاخذ بنظر الاعتبار ان النماذج المعلمية هي افضل وصف للواقع. [8]

يمكن استخدام نقطة الانهيار [6] كمقياس لحصانة المقدرات والتي تمثل الحد الذي يصف مقدار مقاومة المقدر للبيانات الشاذة (الملوثة) قبل ان يصبح بدون فائدة ، ويعد المقدر مقاوم اذا كانت نقطة انهياره اكبر من الصفر ، ويتصف المقدر بأنه الأكثر حصانة اذا كان يمتلك نقطة انهيار عالية، إذ ان أعلى نقطة انهيار لاتتجاوز 0.50 ، واذا تجاوزت يصبح من المستحيل التمييز بين الجزء الجيد وغير الجيد (الملوث) من العينة، وبالتالي من غيرالممكن التمييز بين التوزيع الاساس والتوزيع الملوث . ان ايسط مثال على المقدر غير الحصين (ذو نقطة الانهيار الواطئة) هو الوسط الحسابي حيث ان مشاهدة واحدة شاذة تجعله ينهار. بينما ايسط مثال على المقدر الحصين (ذو نقطة الانهيار العالية) هو الوسيط لانه لايعتمد على البيانات الشاذة. ويمكن ان نُعرّف المقدر الحصين هو المقدر الذي يحافظ على الخصائص المرغوبة في انموذج الانحدار عند انتهاك بعض فرضياته . وفي هذا البحث سوف نستخدم ثلاث طرق حصينة لتقدير انموذج انحدار بواسون وكالآتي :

### طريقة مقدرات M [3]

هي مقدر انحدار حصين امتداد لطريقة (MLE) وتستعمل على نطاق واسع وتسمى احيانا تقديرات (Huber) وتعتبر تقديرات قوية ضد القيم الشاذة في المتجه المعتمد (Y) ولكنها ليست قوية ضد القيم الشاذة في المتجه (X) وتكون مقدرات هذه الطريقة غير متحيزة وتمتلك اقل تباين ويكون هدف تقدير M هو تقليل البواقي للاخطاء .

$$\tilde{B}_M = \min p(y_i - \exp(\hat{X}B)) \quad \dots\dots\dots (23)$$

وبتقليل دالة البواقي الى اقل ما يمكن

$$\hat{B}_M = \min \sum_{i=1}^n P(u_i) = \min \sum_{i=1}^n P\left(\frac{e_i}{\hat{\sigma}}\right) = \min \sum_{i=1}^n P\left[\frac{y_i - \exp(\hat{X}B)}{\hat{\sigma}}\right] \quad \dots\dots\dots (24)$$

إذ ان  $e_i$  : تمثل البواقي

حيث يقدر التباين وفق الصيغة الاتية :

$$\hat{\sigma} = \frac{MAD}{0.6745} \quad \dots\dots\dots (25)$$

$$MAD = \text{median } |e_i - \text{median}(e_i)| \quad \dots\dots\dots (26)$$

إذ ان MAD : تمثل الوسيط لمطلق الفروقات بين البواقي ووسيطها

عند اشتقاق المعادلة (2-24) بالنسبة الى B نحصل :

$$\sum_{i=1}^n x_i \varphi \left[ \frac{y_i - \exp(XB)}{\hat{\sigma}_M} \right] = 0 \quad \dots\dots\dots (27)$$

إذ ان  $\varphi(\cdot)$  : تمثل مشتقة دالة P(.) وتسمى ايضا بدالة التأثير Influence function ويتم حساب الاوزان للقيم

$$W_i = \frac{\varphi \left[ \frac{y_i - \exp(XB)}{\hat{\sigma}_M} \right]}{\frac{y_i - \exp(XB)}{\hat{\sigma}_M}} \quad \dots\dots\dots (28)$$

وبالتالي يمكن صياغة المعادلة المشار إليها بالاعتماد على دالة توكي وكما يلي :

$$W_i = \begin{cases} \left[ 1 - \left( \frac{u_i}{c} \right)^2 \right]^2 & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases} \quad \dots\dots\dots (29)$$

حيث ان قيمة  $(c=4.685)$  حسب دالة Tukey

$$\text{Where } u_i = \frac{e_i}{\hat{\sigma}} \quad \sum_{i=1}^n x_i W_i (y_i - \exp(XB)) = 0 \quad \dots\dots\dots (30)$$

### 2-5 طريقة مقدرات S [3]

ان هذه الطريقة تعتمد على تقدير مقياس الاخطاء حيث نقوم بتقليل مجموع الاخطاء الى ادنى حد ممكن وهذه الطريقة قوية للغاية ضد القيم الشاذة الموجودة في بيانات المتجه X ، وضعيفة ضد القيم الشاذة في بيانات المتجه Y ، ويتم التقدير بالاعتماد على بواقي طريقة M وكالآتي :

$$\hat{B} = \min \hat{\sigma}_s (e_1(B), e_2(B), \dots\dots\dots , e_n(B)) \quad \dots\dots\dots (31)$$

وبأستخدام الحد الادنى من مقدار التباين الحصين  $\hat{\sigma}_s$

$$\min_{B_i} \sum_{i=1}^n p \left[ \frac{y_i - \exp(XB)}{\hat{\sigma}_s} \right] \quad \dots\dots\dots (32)$$

$$\hat{\sigma}_s = \sqrt{\frac{\sum_{i=1}^n W_i e_i^2}{nk}} \quad \dots\dots\dots (33) \quad \text{عندما}$$

$$K = 0.199 \quad , \quad W_i = \frac{p(u_i)}{u_i^2} \quad \text{إذ ان : } e_i \text{ تمثل البواقي}$$

ويكون التقدير الاول لمقدر التباين كالآتي :

$$\hat{\sigma}_s = \frac{\text{median } |e_i - \text{median}(e_i)|}{0.6745} = \frac{MAD}{0.6745} \quad \dots\dots\dots (34)$$

إذ ان MAD : تمثل الوسيط لمطلق الفروقات بين البواقي ووسيطها ثم نشتق المعادلة (32) بالنسبة الى B فيكون الناتج كالآتي :

$$\sum_{i=1}^n x_i \varphi \left( \frac{y_i - \exp(XB)}{\hat{\sigma}_s} \right) = 0 \quad \dots\dots\dots (35)$$

$$\varphi(u_i) = P(u_i) = \begin{cases} u_i \left[ 1 - \left( \frac{u_i}{c} \right)^2 \right]^2 & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases} \quad \dots\dots\dots (36)$$

وتكون دالة الاوزان حسب صيغة Tukey كالآتي : where  $c = 1.547$

$$W_i = \begin{cases} \left[ 1 - \left( \frac{u_i}{c} \right)^2 \right]^2 & , |u_i| \leq c \\ 0 & , |u_i| > c \end{cases} \quad \dots\dots\dots (37)$$

## طريقة مقدرات MM [4]

وهي إحدى الطرق الحصينة لتقدير معاملات نموذج انحدار بواسون والتي تعد تعديلا لطريقة M . والمقترحة من قبل الباحث (Yohai) عام 1987 .

$$\sum_{i=1}^n p(u_i)x_i = 0 \quad \dots\dots\dots (38)$$

$$\sum_{i=1}^n p\left(\frac{y_i - \exp(XB)}{S_{MM}}\right) x_i = 0 \quad \dots\dots\dots (39)$$

حيث  $S_{MM}$  هو الانحراف المعياري لبواقي طريقة (S) الذي يتم الحصول عليه من بقايا تقدير S و P هي دالة Tukey الموزونة .

$$\hat{\sigma}_{MM} = \begin{cases} MAD & iteration = 1 \\ \frac{\sum_{i=1}^n W_i e_i^2}{nk} & iteration > 1 \end{cases} \quad \dots\dots\dots (40)$$

إذ ان  $e_i$  تمثل البواقي

$$\text{Where } MAD = \text{median } |e_i - \text{median}(e_i)| \quad \dots\dots\dots (41)$$

إذ ان MAD : تمثل الوسيط لمطلق الفروقات بين البواقي ووسيطها

$$u_i = \frac{e_i}{\hat{\sigma}_{MM}} \quad \dots\dots\dots (42)$$

$$W_i = \begin{cases} (1 - \left(\frac{u_i}{c}\right)^2)^2 & |u_i| \leq c \\ 0 & |u_i| > c \end{cases} \quad \dots\dots\dots (43)$$

حيث  $c = 4.685$

$$p(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^2} & u_i \leq |c| \\ \frac{c^2}{6} & u_i > |c| \end{cases} \quad \dots\dots\dots (44)$$

## معايير المقارنة

حتى يتم اختيار افضل طريقة لتقدير معالم نموذج بواسون وبالتالي نحصل على نموذج ذو معنوية عالية ، سوف نعتمد على المعايير الآتية :

متوسط مربعات الخطأ (MSE) [11]

يعد هذا المعيار من معايير المفاضلة المهمة في نماذج الانحدار ويحسب كالآتي :

$$MSE = \frac{1}{p} \sum_{i=1}^p (\hat{B}_i - B)^2 \quad \dots\dots\dots (45)$$

حيث ان :

p : تمثل عدد التجارب

$\hat{B}_i$  : تمثل قيمة المعلمة i التقديرية

B : تمثل قيمة المعلمة الحقيقية

وهناك علاقة عكسية بين MSE ومعنوية الانموذج حيث كلما قل MSE زادت معنوية الانموذج والعكس بالعكس ايضا .

متوسط مطلق الخطأ (MAE) [11]

يشير هذا المعيار الى متوسط مدى قرب القيمة التقديرية الى القيمة الحقيقية ويستخدم للأشارة الى مدى دقة التقدير ايضا ويحسب كالآتي :

$$MAE = \frac{1}{p} \sum_{i=1}^p |\hat{B}_i - B| \quad \dots\dots\dots (46)$$

حيث p : تمثل عدد التجارب

$\hat{B}_i$  : تمثل قيمة المعلمة i التقديرية ، B : تمثل قيمة المعلمة الحقيقية

وهناك علاقة عكسية بين MAE ومعنوية الانموذج ، حيث كلما قل MAE زادت معنوية الانموذج والعكس بالعكس ايضا .

### مفهوم المحاكاة [6][7]

تعد المحاكاة من الاساليب المهمة التي من خلالها يمكن تقليد الواقع العملي ذلك عند تكوين مجتمع نظري يشابه المجتمع الحقيقي حيث نضع برنامج على الحاسبة الالكترونية يكون عدد من النماذج الاحصائية لتوليد البيانات ، ثم تكرر التجربة التي تسمح للباحث في كل مرة بتغيير المدخلات مثل حجم العينة وقيمة المعلمة المولدة ، وبالتالي يمكن استعمال اسلوب المحاكاة بديلا عن عدم توفر البيانات المطلوبة او صعوبة الحصول عليها او صعوبة تحليل بعض النظريات واشتقاقها .

هناك عدة طرائق للمحاكاة وهي الطريقة التناظرية والطريقة المختلطة وطريقة مونتي كارلو التي تعتبر من اهم واكثر الطرائق شيوعا كونها تمتاز بالمرونة من حيث تكرار العملية مرات عديدة للوصول الى النتائج المرجوة وتمتاز ايضا بالعشوائية حيث ان مجموعة البيانات العشوائية في التجربة الاولى مستقلة عن مجموعة البيانات العشوائية في التجربة الثانية وهكذا.

### مراحل تجربة المحاكاة

تم استعمال لغة البرمجة ماتلاب 2020 لكتابة برنامج المحاكاة والذي تم إرفاقه في الملحق (A)، ويتضمن البرنامج المكتوب أربعة مراحل أساسية لتقدير أنموذج الانحدار، وكما يأتي:

المرحلة الأولى: مرحلة تحديد القيم الافتراضية  
 إذ يتم في هذه المرحلة اختيار القيم الافتراضية للمعاملات باعتبار وجود متغيرين توضيحين، وقد تم اختيار القيم كما يأتي:  
 اختيرت القيم الافتراضية المختلفة للمعاملات وهذه القيم موضحة في الجدول (1).

جدول (1): القيم الافتراضية للمعاملات

Model	$\beta_0$	$\beta_1$	$\beta_2$
I	2	0.5	-1
II	0	1	1
III	0.5	-1	1.5
IV	2	0.9	0.8
V	3	2.9	2.5
VI	10	5	6

1. تم اختيار ثلاثة أحجام مختلفة للعينات (25، 50، 100).
2. اختيرت نسب مختلفة للتلوث (10%، 15%، 20%)
3. تم تكرار كل تجربة 1000 مرة.

### المرحلة الثانية: توليد البيانات

وهي مرحلة مهمة جداً لاعتماد الخطوات التي تليها عليها، إذ يتم فيها توليد المتغيرات التوضيحية باعتبارها مولدة من التوزيع المنتظم ومن ثم توليد المتغير التابع من توزيع بواسون.

### المرحلة الثالثة: التقدير

يتم في هذه المرحلة إجراء عملية التقدير لمعاملات الانحدار ومن ثم تقدير أنموذج الانحدار وذلك باستعمال طرائق التقدير الواردة في الفصل الثاني، وكما يأتي:

1. طريقة الإمكان الأعظم MLE.
2. طريقة مقدرات M.
3. طريقة مقدرات S.
4. طريقة مقدرات MM.

### المرحلة الرابعة: مرحلة المقارنة بين الطرائق

لغرض المقارنة بين طرائق التقدير المختلفة للمعاملات وإيجاد أفضل المقدرات تم استعمال معيار MSE و MAE حيث إن الطريقة التي تمتلك أقل قيمة MSE وأقل قيمة MAE تعتبر أفضل.

### نتائج عمليات المحاكاة

لغرض تطبيق طرائق التقدير لأنموذج الانحدار وتحديد الطريقة الأفضل، والتي ستستعمل في تقدير الأنموذج للبيانات الحقيقية في الجانب التطبيقي مستعملين لذلك برنامج تمت كتابته بلغة ماتلاب ، وسيتم عرض النتائج التي تمثل قيم MSE و MAE لكل طريقة وفقاً لأحجام العينات والقيم المختلفة للمعاملات ونسبة التلوث وكما يأتي:

جدول (2): قيم MSE و MAE للطرائق المختلفة وللحالة الأولى (I)

PR <sup>(1)</sup>	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	1.39608	0.93375	1.39256	1.14597
		MAE	1.68495	1.41146	1.6794	1.53438
	50	MSE	1.38106	0.64816	1.38217	1.03164
		MAE	1.66917	1.23404	1.67826	1.45804
	100	MSE	1.26681	0.49818	1.26530	0.98240
		MAE	1.52357	1.08589	1.52906	1.34659
15%	25	MSE	1.50028	1.04742	1.49599	1.38257
		MAE	1.74171	1.48303	1.74031	1.68178
	50	MSE	1.47195	0.85684	1.47007	1.33238
		MAE	1.7359	1.37523	1.73488	1.64577
	100	MSE	1.45851	0.65442	1.45963	1.29765
		MAE	1.71882	1.2758	1.7277	1.5233
20%	25	MSE	1.52413	1.15095	1.52627	1.46185
		MAE	1.74557	1.54687	1.7564	1.72727
	50	MSE	1.50953	1.0033	1.50619	1.36514
		MAE	1.75717	1.46147	1.74971	1.6609
	100	MSE	1.50169	0.81627	1.50134	1.28341
		MAE	1.74983	1.36588	1.73999	1.64722

<sup>1</sup> PR هو اختصار للكلمتين Pollution Rate والتي تعني نسبة التلوث والتي بدورها تعني نسبة وجود الشواذ في متغير الاستجابة وكذلك في المتغيرين التفسيريين.

جدول (3): قيم MSE و MAE للطرائق المختلفة وللحالة الثانية (II)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	2.3441	2.14388	2.2826	2.26486
		MAE	2.53997	2.39721	2.50445	2.48419
	50	MSE	2.27399	2.06369	2.20369	2.20181
		MAE	2.56184	2.40813	2.52054	2.50908
	100	MSE	2.23046	2.01891	2.15567	2.17831
		MAE	2.56769	2.40434	2.52533	2.53172
15%	25	MSE	2.40046	2.18482	2.32485	2.34095
		MAE	2.71563	2.56773	2.67315	2.67453
	50	MSE	2.28615	2.08078	2.21414	2.22948
		MAE	2.58631	2.44546	2.54586	2.54717
	100	MSE	2.27052	2.05638	2.20108	2.2299
		MAE	2.58274	2.42552	2.55423	2.5692
20%	25	MSE	2.43498	2.19929	2.36014	2.38575
		MAE	2.74042	2.58889	2.69938	2.7088
	50	MSE	2.31903	2.09456	2.24775	2.27214
		MAE	2.60992	2.46561	2.57084	2.57981
	100	MSE	2.30818	2.06587	2.23684	2.26741
		MAE	2.61503	2.46647	2.57619	2.59311

جدول (4): قيم MSE و MAE للطرائق المختلفة وللحالة الثالثة (III)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	2.93018	0.21094	2.86704	1.69096
		MAE	2.73804	0.61685	2.66965	1.90724
	50	MSE	2.79065	0.20090	2.73051	1.61044
		MAE	2.60765	0.58748	2.54252	1.81642
	100	MSE	2.65776	0.19133	2.60049	1.53375
		MAE	2.48348	0.55950	2.42145	1.72992
15%	25	MSE	3.18131	0.42035	3.11315	2.74179
		MAE	2.87539	0.80812	2.80877	2.61763
	50	MSE	3.02982	0.40033	2.96490	2.61123
		MAE	2.73847	0.76964	2.67502	2.49298
	100	MSE	2.88554	0.38127	2.82371	2.48689
		MAE	2.60807	0.73299	2.54764	2.37427
20%	25	MSE	3.34106	0.82904	3.27639	3.20643
		MAE	2.96280	1.17237	2.90029	2.88347
	50	MSE	3.18196	0.78956	3.12037	3.05374
		MAE	2.82171	1.11654	2.76218	2.74616
	100	MSE	3.03044	0.75196	2.97178	2.90832
		MAE	2.68734	1.06337	2.63065	2.61539

جدول (5): قيم MSE و MAE للطرائق المختلفة وللحالة الرابعة (IV)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	1.65152	1.34366	1.62324	1.54831
		MAE	2.31638	2.01915	2.29742	2.23542
	50	MSE	1.57288	1.27968	1.54594	1.47458
		MAE	2.20608	1.92300	2.18802	2.12897
	100	MSE	1.49798	1.21874	1.47232	1.40436
		MAE	2.10103	1.83143	2.08383	2.02759
15%	25	MSE	1.68235	1.48050	1.65199	1.60863
		MAE	2.34142	2.19273	2.32125	2.28776
	50	MSE	1.60224	1.41000	1.57332	1.53203
		MAE	2.22992	2.08831	2.21071	2.17882
	100	MSE	1.52594	1.34286	1.49840	1.45908
		MAE	2.12373	1.98887	2.10544	2.07507
20%	25	MSE	1.72722	1.51155	1.69010	1.67346
		MAE	2.37198	2.22116	2.34767	2.33575
	50	MSE	1.64497	1.43957	1.60962	1.59377
		MAE	2.25903	2.11539	2.23588	2.22452
	100	MSE	1.56664	1.37102	1.53297	1.51788
		MAE	2.15146	2.01466	2.12941	2.11859

جدول (6): قيم MSE و MAE للطرائق المختلفة وللحالة الخامسة (V)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	24.80634	12.83330	24.12417	21.56363
		MAE	8.97836	7.33459	8.86391	7.88489
	50	MSE	23.62509	12.22219	22.97540	20.53679
		MAE	8.55082	6.98532	8.44182	7.50942
	100	MSE	22.50009	11.64018	21.88133	19.55885
		MAE	8.14364	6.65269	8.03983	7.15183
15%	25	MSE	24.96664	17.33314	24.32649	23.90185
		MAE	9.00648	6.90739	8.89893	8.65880
	50	MSE	23.77775	16.50775	23.16809	22.76367
		MAE	8.57760	6.57847	8.47517	8.24648
	100	MSE	22.64548	15.72167	22.06485	21.67969
		MAE	8.16914	6.26521	8.07159	7.85379
20%	25	MSE	25.09571	19.30146	24.50446	24.50035
		MAE	9.03097	7.65601	8.93155	8.83676
	50	MSE	23.90068	18.38234	23.33758	23.33367
		MAE	8.60092	7.29144	8.50624	8.41596
	100	MSE	22.76255	17.50699	22.22627	22.22254
		MAE	8.19135	6.94423	8.10118	8.01520

جدول (7): قيم MSE وMAE للطرائق المختلفة وللحالة السادسة (VI)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	124.12461	42.31230	121.09400	71.23852
		MAE	19.92541	8.09803	19.70914	11.81062
	50	MSE	118.21391	40.29743	115.32762	67.84621
		MAE	18.97658	7.71241	18.77061	11.24821
	100	MSE	112.58468	38.37850	109.83583	64.61544
		MAE	18.07293	7.34515	17.87677	10.71258
15%	25	MSE	124.56729	49.03915	120.92441	77.63511
		MAE	19.93840	9.34597	19.70226	12.85716
	50	MSE	118.63551	46.70395	115.16610	73.93820
		MAE	18.98895	8.90092	18.76406	12.24491
	100	MSE	112.98620	44.47995	109.68200	70.41733
		MAE	18.08471	8.47707	17.87053	11.66182
20%	25	MSE	125.34374	63.73220	121.31534	91.12018
		MAE	19.94069	12.12219	19.84176	15.04016
	50	MSE	119.37499	60.69733	115.53842	86.78112
		MAE	18.99113	11.54494	18.89691	14.32396
	100	MSE	113.69047	57.80698	110.03659	82.64869
		MAE	18.08679	10.99518	17.99706	13.64187

### الاستنتاجات

من خلال تحليل نتائج المحاكاة وجدنا ان قيم متوسط مربعات الخطأ (MSE) وقيم متوسط مطلق الخطأ (MAE) لتقدير انحدار بواسون بالطريقة الأعتيادية (MLE) والطرائق الحصينة الموضحة في الجداول (2) الى (7) تشير الى :

- 1- افضلية طريقة M على الطرائق الحصينة الاخرى والتي بدورها تتفوق اغلب الاحيان على طريقة MLE لتقدير معلمات انحدار بواسون في حالة وجود القيم الشاذة .
- 2- ان قيم متوسط مربعات الخطأ MSE وقيم متوسط مطلق الخطأ MAE تقل كلما ازدادت حجوم العينات في كل الطرائق ولجميع النماذج المدروسة .
- 3- ان قيم متوسط مربعات الخطأ MSE وقيم متوسط مطلق الخطأ MAE تزداد كلما ازدادت نسبة التلوث في كل الطرائق ولجميع النماذج المدروسة .

### المصادر

- 1- سلطان، مها حسن، 2017 "طريقة الامكان الاعظم وبعض الطرائق اللامعلمية في تقدير انموذج انحدار بواسون" رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.
- 2- هادي، خديجة عبد الكريم، 2021 "مقارنة بعض الطرائق الحصينة لتقدير انموذج الانحدار للبيانات ذات التوزيع الطبيعي الملتوي" رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.
- 3- باهض، زهراء خالد، 2022 "التقدير الحصين لأنموذج الانحدار شبه المعلمي مع تطبيق عملي" رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.
- 4- محمد، رضا قاسم، 2022 "المقدرات الحصينة لمعلمات انموذج الانحدار بأخطاء عشوائية غير متجانسة التباين" رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.

- 5- الجشعبي، حسين علي، 2007 "مقارنة لبعض المقدرات الحصينة لمعالج النماذج اللاخطية"، اطروحه دكتوراه، كلية الادارة والاقتصاد، الجامعة المستنصرية.
- 6- صالح، طارق عزيز، 2009 "مقارنة بعض الطرائق الحصينة في تحليل الارتباط القويم الخطي بأستخدام المحاكاة مع تطبيق عملي" رسالة ماجستير، كلية الإدارة والاقتصاد، جامعة بغداد .
- 7- لعيبي، اماني عماد، 2018 "مقارنة بعض الطرائق الحصينة لتقدير معلمتي توزيع دالة القوى مع تطبيق عملي" رسالة ماجستير، كلية الأدارة والأقتصاد، الجامعة المستنصرية.
- 8- خليل، منال اسماعيل، 2011 "استخدام المقدرات الحصينة في التحليل العنقودي مع تطبيق عملي في مجال الفساد الاداري والمالي" رسالة ماجستير، كلية الأدارة والأقتصاد، جامعة بغداد.
- 9- Winklmann Rainer, 2008, " Econometric Analysis of Count Data" Fifth edition , springer-verlag Berlin Heidelberg .
- 10- Philip, N. and Sebastian, N., 2015, "Application of Poisson Regression on Traffic Safety"
- 11- Mohamed, R. A. and Omnia, M. S., 2020, "A comparative study of Robust Estimation for Poisson Regression Model with outliers" , Journal of statistics Application and probability, No. 2, PP. (279-286).

العوامل الاقتصادية والاجتماعية للطلب على القروض الشخصية في  
محافظة عمان خلال جائحة فيروس كورونا (2020-2022)

الدكتورة رانية عدنان السطل

اختصاصي مساعد

قسم مرصد المشاريع الاقتصادية الصغيرة والمتوسطة

المؤسسة الأردنية لتطوير المشاريع الاقتصادية

تاريخ استلام البحث: 2023/05/15

تاريخ قبول البحوث: 2023/06/22

نشر البحث في العدد الواحد والعشرون: ديسمبر / كانون اول 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/332.7

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/332.7

## العوامل الاقتصادية والاجتماعية للطلب على القروض الشخصية في محافظة عمان خلال جائحة فيروس كورونا (2020-2022)

د. رانية عدنان السطل

قسم الدراسات والبحوث / المؤسسة الأردنية لتطوير المشاريع الاقتصادية  
وزارة الصناعة والتجارة والتموين

### المستخلص

تبحث الدراسة عن العوامل الاقتصادية والاجتماعية على قرارات الحصول على قروض شخصية بمحافظة عمان. إذ لوحظ وجود طلب عليها خلال فترة جائحة كورونا (Covid-19 Crisis) والتي امتدت في الفترة (2020-2022)؛ مما جعلها قضية مثيرة للبحث عن ماهية العوامل التي أدت إلى ازدياد هذا الطلب. تهدف هذه الدراسة إلى التعرف على العوامل التي قد تؤثر في الطلب على القروض الشخصية في محافظة عمان، إذ اعتمدت في تحليلها على نموذج الانحدار اللوجستي ثنائي القيمة والذي يستخدم في تحليل البيانات الوصفية. تم دراسة العديد من المتغيرات والتي كان من المتوقع أن تؤثر في الطلب على القروض الشخصية كمتغيرات مستقلة؛ إلا أن نتائج التحليل الإحصائي أظهر أن حصول الفرد على تسهيلات ائتمانية ميسرة من الجهاز المصرفي له تأثير كبير وموجب على وجود القرض الشخصي وبنسبة ترجيح 8.59 مرة، إذ يأتي في المرتبة الأولى، وفي المرتبة الثانية يأتي تأثير جائحة فيروس كورونا على الفرد من حيث الأهمية في التأثير الموجب على الطلب على القروض الشخصية وبنسبة ترجيح 5.12 مرة، بينما يأتي المستوى التعليمي للفرد في المرتبة الثالثة في الأهمية وبتأثير موجب ونسبة ترجيح 9.09 مرة، وفي المرتبة الرابعة تأتي الحالة الاجتماعية للفرد وبتأثير موجب ونسبة ترجيح 28.03 مرة، وفي المرتبة الخامسة يأتي مكان إقامة الفرد في التأثير على الطلب على القروض الشخصية.

الكلمات المفتاحية: القروض، القروض الشخصية، الانحدار اللوجستي، نسبة التّرجيح (الفرصة)، عمان

### Socioeconomic Demand Factors for Personal loans in Amman Governorate through Covid-19 crisis ( 2020-2022)

Dr. Rania A. Al-satel

Studies and Researches Department /Jordan Enterprise Development Corporation  
(JEDCO) affiliated to Ministry of Industry and trade and supply

### ABSTRACT

This study examines the social and economic factors of decisions to submit for personal loans in Amman governorate. There was a demand for personal loans in Amman during Covid-19 crisis from (2020-2022); which made it an interesting issue to find out what factors led to this demand. This study aims to identify the factors that may affect the

demand for personal loans in Amman. In its analysis, it relied on the Binary logistic regression model that is used in the analysis of qualitative data. Several variables have been studied that were expected to affect the demand for personal loans as independent variables. However the results of econometrics analysis showed that the smoothly credit facilities from banking system has a significant and positive impact on the presence of personal loans with odds ratio 8.59 times, as it comes in the first place, then the second place comes to corona crisis in terms of importance in the positive impact on the on the presence of personal loans with odds ratio 5.12 times, while the Education level comes in the third place in importance and with a positive effect and odds ratio 9.09 times, and in the fourth place comes the Marital status and with a positive effect and odds ratio 28.03 times, in addition to the Residence place, which is ranked fifth in the positive impact on the demand for personal loans.

**Keywords:** Loans; Personal Loans; Logistic regression; Odds ratio; Amman

### المقدمة

تحتل القروض والتسهيلات الائتمانية المقام الأول بين توظيفات المصارف التجارية والمصارف غير التجارية، كما تُعد المصدر الأساسي لإيراداتها، ويتفاوت توزيع القروض لدى كل منها بحسب آجال الاستحقاق وطبيعة النشاط الاقتصادي، فالقروض قصيرة الاجل نجدها تتراجع من حيث الأهمية النسبية لدى البنوك التجارية التي تهتم أساسًا بالقروض متوسطة وطويلة الأجل استنادًا الى موارد مالية تتسم بالاستقرار النسبي من أجل ضمان حقوق المساهمين فيها. كذلك هنالك القروض المُوجهة نحو المشاريع الاستثمارية والقروض الاسكانية والقروض الشخصية...إلخ، إذ تختلف مُسميات القروض حسب النشاط الاقتصادي الذي يدعمه المصرف (مختار، 2005).

تُعد القروض الشخصية من أنواع القروض الأكثر شعبية وإقبالًا، إذ يمكن استخدامها في كثير من الأغراض، كإنشاء مشروع استثماري صغير أو متوسط الحجم، أو تغطية تكاليف الإجازات المتنوعة وحفلات الزفاف، أو تجديدات المنازل، أو تسديد الديون، أو تغطية مصاريف طارئة، وغيرها من الأغراض الاستهلاكية. وتختلف مدة سداد القروض الشخصية باختلاف جهات الاقراض، ولكنها تتراوح ما بين 24 وحتى 84 شهرًا في الغالب. ويتطلب الحصول على قرض شخصي أن تُقدم بعض الضمانات مثل الحصول على قرض بضمان المنزل أو الراتب الشهري حتى تؤمن الجهات المُقرضة أوضاعها وتضمن استرداد أموالها إذا فشل المُقرض في سداد القرض.

بلغ إجمالي حجم الديون (التسهيلات الممنوحة للقطاع الخاص) من قبل البنوك المرخصة الأردنية نحو 27559.0 مليون دينار أردني في عام 2021، وتشير هذه الأرقام إلى أن البنوك المختلفة في الأردن لا تزال تمنح قروضًا للعديد من المقترضين المختلفين في الأردن وبشكل

متزايد عن السنوات السابقة - بما في ذلك القروض الشخصية - وذلك رغم التحديات الاقتصادية والمالية التي تواجهها البلاد وخاصةً في ظل جائحة فيروس كورونا (البنك المركزي، 2022).

ومما لفت انتباه الباحثة وأثار مشكلة الدراسة أن الأهمية النسبية للتسهيلات الممنوحة من قبل الجهاز المصرفي الأردني للأفراد لغايات استهلاكية 2 (أي وفقًا للنشاط الاقتصادي) خلال أعوام جائحة كورونا 2020-2022 كانت نسب مرتفعة؛ إذ كانت خلال عام 2020 ما يقارب 23% للتسهيلات الممنوحة للأفراد لغايات استهلاكية وهي النسبة الأكبر بالنسبة لباقي النشاطات، ثم ارتفعت إلى 24% في سنة 2021 وأيضاً كانت نسبة مرتفعة بالنسبة لباقي النشاطات (وهي السنة التي كانت أزمة فيروس كورونا في ذروتها وكان الاقتصاد الأردني في وضع حرج 3)، وفي عام 2022 كان التغيير الأكبر لصالح التسهيلات الممنوحة للأفراد لغايات استهلاكية بالنسبة لباقي النشاطات الاقتصادية والذي بلغ 819.7 مليون دينار أردني (البنك المركزي، 2023).

إنّ القطاع المصرفي له دورٌ حيويٌّ في التنمية والتطور الاقتصادي في الدولة؛ وذلك لما يتميز به من أنشطة متنوعة خاصةً الأنشطة الخدمية المقدمة للأفراد، لذا تكمن أهمية الدراسة في كونها سلطت الضوء على العوامل الاقتصادية والاجتماعية المؤثرة على حجم الطلب على القروض الشخصية في محافظة عمّان خلال جائحة كورونا، - إذ لوحظ زيادة الطلب عليها خلال تلك الفترة - والتي بدورها قد تساعد البنك المركزي والجهات المعنية في الأردن عند رسم سياسات الاقتراض الخاصة بالقطاع المصرفي والمالي.

إن الهدف الذي يطمح إليه هذا البحث هو تحديد الأهمية النسبية للعوامل الاقتصادية والاجتماعية المؤثرة على حجم الطلب على القروض الشخصية في محافظة عمّان خلال جائحة كورونا عن طريق تطبيق نموذج إحصائي قادر على التنبؤ من خلال تحليل البيانات، وهو الانحدار اللوجستي الثنائي.

### الإطار النظري والدراسات السابقة

في عالم اليوم المتسارع يومًا بعد يوم يبقى السؤال والاستفسار حول القروض 4، وكيفية الحصول عليها حاضرًا بقوة بين الناس بمختلف فئاتهم، وتبقى القروض منجم من الذهب لكافة المصارف؛ بينما تتنافس فيما بينها محليًا في تقديم تسهيلات ائتمانية لمختلف العملاء ولمختلف الأغراض والتي قد تكون ميسرة أو مُشددة تبعًا لتعليمات وسياسات المصارف.

إن الوظيفة الرئيسية للمصارف التجارية تتركز في شقين رئيسيين: قبول الودائع المختلفة من المودعين، وتقديم القروض (التسهيلات الائتمانية) المتنوعة إلى المُقترضين، ويتفرع من هذين

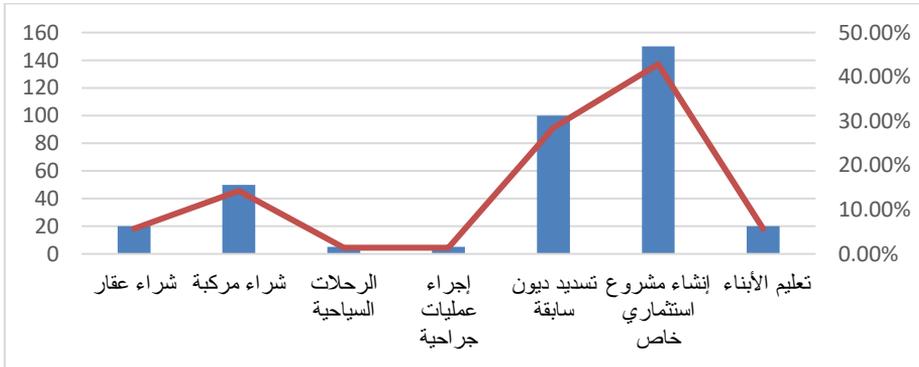
2 مُدرجة تحت بند أخرى في توزيع التسهيلات وفقًا للنشاط الاقتصادي في التقارير السنوية للبنك المركزي وأحيانًا تدرج تحت بند قروض شخصية.

3 مما لا شك فيه أن الإغلاقات والقيود التي فرضت على الأسواق خلال أزمة فيروس كورونا كان لها تأثير كبير على دخول الأفراد وللجوء إلى الاقتراض.

4 القروض أو (التسهيلات الائتمانية) عبارة عن واجبات يدين به أحد طرفي المعاملة (المدين) إلى الطرف الآخر والذي يسمى الدائن، وعادة ما يشير ذلك إلى الأصول التي كان الدائن قد أعطاها للمدين وفي نفس الوقت يمكن أن يستخدم هذا المصطلح على سبيل الاستعارة ليشمل الالتزامات الأخلاقية المترتبة على الطرفين وغيرها من التفاعلات غير القائمة على القيمة الاقتصادية.

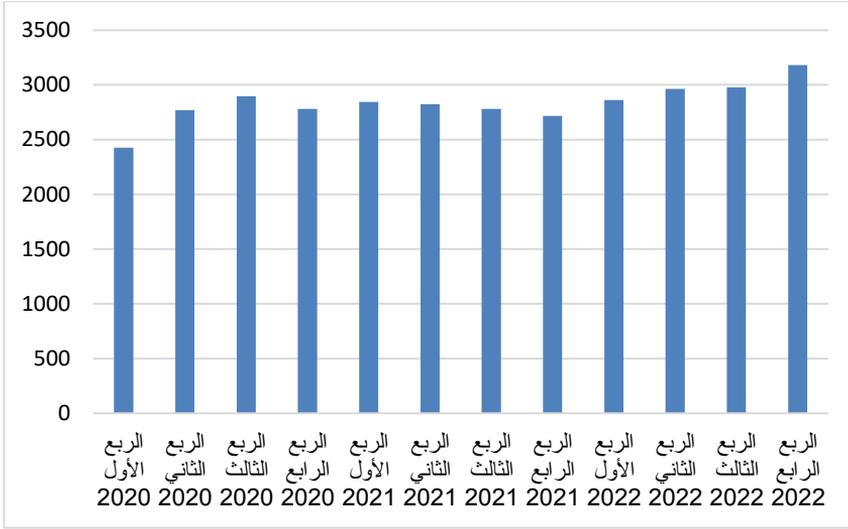
الشقين وبسببهما خدمات كثيرة يؤديها المصرف التجاري لكلا الفريقين والتي لها أهميتها الكبرى في النظام الاقتصادي بما يحقق للمصارف التجارية المساهمة في خلق الجو المناسب والتربة الصالحة للتنمية الاقتصادية، إذ تُعتبر المؤسسات التمويلية كافةً عضواً حيوياً في الحياة الاقتصادية للمجتمعات. إن منح الائتمان يجب أن يتفق مع توفير الأمان لأموال المودعين وبما يحقق الرفاهية للمجتمع ونموه وازدهاره. كما إن وظيفة منح الائتمان تُعتبر من أهم وأخطر وظائف المصارف التجارية؛ وذلك لأن الأموال التي تمنحها كتسهيلات ائتمانية ليست ملكاً لها بل هي أموال المودعين؛ لذلك تقوم إدارة المصرف برسم السياسة الائتمانية التي تحقق لها حسن وسلامة استخدام الأموال المتاحة لها مع تحقيق عائد مادي مناسب (السيسي، 1998).

إن قطاع التمويل الأصغر (والتي معظمها قروض شخصية تستخدم لتمويل مشاريع صغيرة) يُعد أداة فعالة للتنمية الاقتصادية تهدف إلى الحد من الفقر في المجتمع، وبما أن أصحاب الدخل المحدود يعانون من قلة الخدمات المالية بسبب الاستراتيجيات التقليدية للمصارف، فهذا القطاع يوفر خدمات مالية أساسية لشريحة واسعة من أصحاب المشاريع بالقليل من الضمانات وشروط الائتمان أو بدونها لتتوافق ومقدرتهم الاقتصادية البسيطة، وهذه المقدرّة تؤهلهم للحصول على خدمات مالية من المصارف، بحيث يتم تمويل مشاريع صغيرة توظف عشرة عمال فأقل بشروط وضمائم بسيطة سعياً لتنميتهم وتطويرهم (الغرباوي، 2019). وهذا ما أكدته التحليل الوصفي للعيينة، فقد بلغ عدد الأفراد الذين استخدموا القروض الشخصية لغرض إنشاء مشاريع استثمارية خاصة 150 فرداً أي ما نسبته 43% من إجمالي حجم العينة، في حين بلغ عدد الأفراد الحاصلون على القرض لغرض تسديد ديون سابقة 100 فرداً أي ما نسبته 28.5% من إجمالي حجم العينة المختارة. وهذا يعطى مؤشراً إلى التأثير السئ لجائحة كورونا على الأوضاع الاقتصادية للأردنيين (أنظر الشكل رقم 1).



الشكل (1): التوزيع النسبي لوحدة المعاينة والتحليل حسب أغراض القروض الشخصية

وبالرجوع إلى تقارير البنك المركزي تبين أن التسهيلات الائتمانية (القروض الشخصية الممنوحة لغرض إنشاء مشاريع صغيرة ومتوسطة) خلال الفترة 2020-2022 كانت في تزايد مستمر (أنظر الشكل 2).



الشكل (2): حجم التسهيلات الائتمانية الممنوحة من البنوك للشركات الصغيرة والمتوسطة خلال الفترة (2020-2022) (بالمليون دينار أردني)

وقد يتم استخدام القرض الشخصي لإنشاء مشروع صغير أو متناهي الصغر انتاجي، مثل الأعمال الزراعية والانتاج الحيواني والصناعات التحويلية، أو خدعي كخدمات الصحة والترفيه والتعليم وغيرها، وهناك المشاريع الانتاجية التي أساسها التحويل اي تحول المواد الخام الى منتج نهائي او وسيط، ومشاريع خدمات المواصلات والسياحة والاصلاح والتنظيف والأعمال المنزلية لربات المنازل وغير ذلك من الأنشطة التي يمكن أن تلبى هذا القروض الشخصية (سلمان والعبادي، 2015).

أما المصارف (البنوك) الأردنية المرخصة لمنح القروض والعامل في الأردن فهي:

أولاً: البنوك التجارية والتي تشمل: البنك العربي، بنك الأردن، البنك الأهلي الأردني، بنك القاهرة عمان، البنك الاردني الكويتي، البنك التجاري الأردني، بنك الاسكان للتجارة والتمويل، بنك المؤسسة العربية المصرفية (الأردن)، بنك المال الأردني، بنك الاستثمار العربي الأردني، البنك الاستثماري، بنك الاتحاد، بنك سوسيته جنرال الاردن.

ثانياً: البنوك الإسلامية والتي تشمل: البنك الإسلامي الأردني، البنك العربي الإسلامي الدولي، بنك صفوة الإسلامي (البنك المركزي، 2022). ومن الجدير بالذكر أن هناك بعض المصارف تُقدم تسهيلات مُشجعة للعميل حتى يقترض، وهناك من يضع المعايير أمامه للتشديد على موافقات منح القروض الشخصية.

من المعروف أن الائتمان - وتحديدًا الائتمان الشخصي - حتى ولو منح وفقاً لأسس مصرفية سليمة ينطوي على نوع من المخاطر لأن هناك احتمالاً أن الأموال التي يقرضها البنك للعملاء قد لا تسدد بسبب ظروف اقتصادية سيئة غير متوقعة تؤثر على المقترضين، ولأن هذه المخاطر لا يمكن للبنك تجنبها تمامًا إلا أنها مخاطر محسوبة ابتداءً عند منح القرض؛ ولذا يقال أن مخاطر الائتمان لا تنطوي على درجة كبيرة من الخطورة؛ فالخطورة فيها مألوفة في العالم

المصرفي. ومن المتعارف عليه أن ديون المصارف تنقسم إلى: ديون عادية لا يواجه البنك أية مشاكل في طرق استردادها وديون معدومة وهي التي استنفذ البنك بشأنها كافة وسائل المطالبة الودية والقانونية وتعذر عليه استردادها؛ إما لسبب فعلي أو لسبب قانوني، وديون مشكوك في تحصيلها أو شبه معدومة - وهي الديون المتوقعة جراء منح القروض الشخصية لأفراد ليسوا أصحاب ملاءة مالية واضحة عادةً- وهي تقف موقفاً وسطاً بين ديون المصرف العادية والديون المعدومة (النجار، 2000).

وهناك رأي آخر يعتبر الاقتراض آفة أو عرض من أعراض الاضطراب المالي، إذ يقولون أصحاب هذا الفكر الاقتصادي أن الاقتراض عامةً يجب أن لا يتم إلا لتمويل مشاريع إنتاجية كبيرة ولأفراد أو شركات تتميز بالملاءة المالية، لكن هذه النظرية أثبتت أنها ليست دقيقة عملياً؛ إذ إن التعثر المالي قد يحدث للجميع، وأن السبب الرئيسي في التعثر هو تقديم قروض سهلة من قبل بعض المصارف (بفرض فائدة أو عمولة قليلة نوعاً ما) وهذا في التقدير العام ليس ظاهرة ايجابية، إذ أن القروض السهلة قد تُعرض أحياناً على الإنفاق والاستهلاك، وعدم الالتزام والجديّة في السداد (المالكي، 2019).

يُعتبر جُموح الغلاء الحالي من الظواهر الاقتصادية العامة التي يشكو منها أغلبية المواطنين في الأردن، فبالرغم من الزيادات التي يحصلون عليها في دخولهم النقدية أصبحوا اليوم عاجزين عن الحصول على نفس الكميات من السلع التي كانوا يشترونها سابقاً بنقود قليلة، كما يعجبون كيف أن النقود تخسر من قوتها الشرائية بهذه السرعة وهو ما يعرف اقتصادياً بالتضخم المُتسارع؛ مما يدفع الأفراد إلى اللجوء إلى الاقتراض الشخصي لتلبية حاجاتهم المتزايدة (عجام، 2001).

ومن الجدير بذكره أنه عند منح القروض الشخصية تُعد شخصية العميل الركيزة الأساسية الأولى في القرار الائتماني، وهي الركيزة الأكثر تأثيراً في المخاطر التي قد تتعرض لها المصارف؛ وبالتالي فإن أهم مسعى عند إجراء التحليل الائتماني هو تحديد شخصية العميل بدقة؛ فكلما كان العميل يتمتع بشخصية أمينة ونزيهة وسمعة طيبة، وملتزمًا بكافة تعهداته، وحريصًا على الوفاء بالتزاماته؛ كان أقدر على إقناع البنك بمنحه الائتمان المطلوب والحصول على دعم المصرف له. إن قياس العوامل المعنوية كعامل الأمانة والنزاهة بدرجة دقيقة أمر تكتنفه بعض الصعوبات من الناحية العملية، إذ يتم التغلب على هذه الصعوبات من خلال الاستعلام الجيد وجمع البيانات والمعلومات عن العميل من محيطي العمل والعائلة؛ لمعرفة المستوى المعيشي للعميل وموارده المالية، والمشاكل المالية التي قد يعانيها، ومستواه الاجتماعي وسجل أعماله التي قام بها، وماضيه مع المصرف ومع الغير، وسابق تصرفاته مع المصارف الأخرى (عفانة، 2019).

إن الترويج المصرفي 5 من الأنشطة التسويقية الهامة والمعتمدة في المصارف والمؤسسات المالية بشكل عام؛ إذ إن إقبال الجمهور على القروض يكون نتيجة نجاح المصارف في تقديمها للائتمان بتكلفة مناسبة من خلال فروع المصرف المنتشرة جغرافياً في الدولة، وتواجد وسائل مختلفة للاتصال بالزبائن الحاليين والمرتبين وتعريفهم بطبيعة ومواصفات وتحديثات

4 الترويج المصرفي: هو كافة الأنشطة التسويقية التي ينطوي عليها عملية اتصال اقناعي يستهدف التأثير على جمهور المصرف بقصد استمالة استجابات سلوكية من أفراد الجمهور في الاتجاه المرغوب من قبل المصرف.

مختلف الخدمات المقدمة، وكيفية التقدم لها. إضافةً إلى دور الإعلانات التجارية في وسائل الاتصالات المختلفة والذي بات الأكثر تأثيرًا على الأفراد حاليًا (الصميدعي ويوسف، 2005).

تناولت العديد من الدراسات العربية موضوع القروض، لكنّ الباحثة قامت باستعراض البعض منها، فقد قام رجب وآخرون (2022) بدراسة عن أثر القروض المصرفية المتعثرة على مؤشرات الربحية، إذ هدفت الدراسة إلى التعرف على أثر القروض المصرفية المتعثرة في البنوك التجارية المصرية والتونسية والمدرجة في البورصة على مؤشرات الربحية، وقد تمحورت إشكالية الدراسة في التعرف على تأثير نسبة ( القروض المتعثرة إلى إجمالي القروض NPL) على مؤشرات الربحية ممثلة في معدل العائد على الأصول، ومعدل العائد على حقوق الملكية ونسبة هامش الفائدة الصافي. اعتمدت الدراسة على أسلوب تحليل البيانات المالية واستخدمت نموذج الانحدار الخطي. خلّصت الدراسة إلى أن هناك أثر سلبي لمستوى التعثر في القروض على مؤشرات الربحية في كلّ من البنوك المُدرجة في البورصة المصرية والبورصة التونسية. أوصت الدراسة بأن تعمل البنوك على وضع آليات من شأنها خفض درجة المخاطر التي قد تواجه العملاء وتؤدي إلى زيادة نسبة التعثر في سداد القروض، بجانب أن يكون هناك دعم من البنوك المركزية؛ لاسيما أن الدراسة ناقشت الفترة من 2011 حتى 2020 (وهي فترة شهدت أزمات اقتصادية حادة سواءً بالنسبة لدولة مصر أو دولة تونس).

أما دراسة عوينات (2017) فقد تناولت أثر القروض المتعثرة على الأداء المالي للبنوك التجارية (دراسة حالة البنوك (Agb, Bea, Baraka))، إذ هدفت الدراسة إلى إبراز أثر القروض المتعثرة على الأداء المالي في البنوك التجارية في الجزائر، إذ أصبحت القروض المتعثرة محل اهتمام كل البنوك التجارية وذلك لما لها من تأثير على أدائها المالي ومختلف أنشطتها، كما أن تأثيرها يشمل مختلف عناصر الأداء المالي والمتمثلة في مؤشرات السيولة والربحية وكفاية رأس المال. وقد تم إجراء الدراسة على بنك البركة، وبنك الجزائر الخارجي، وبنك الخليج الجزائري خلال الفترة 2010-2015، إذ تم استخدام أسلوب تحليل الانحدار البسيط للكشف عن أثر القروض المتعثرة على الأداء المالي وذلك باستخدام البرنامج الإحصائي Eviews 9. وقد توصلت الدراسة إلى مجموعة من النتائج أهمها وجود تأثير ذو دلالة إحصائية للقروض المتعثرة على بعض مؤشرات الأداء المالي.

في حين أكدت دراسة سارة عبد الحلیم (2012) على دور القروض المصرفية في تحقيق الربحية للبنوك التجارية (دراسة حالة وكالة القرض الشعبي الجزائري بعين مليلة)، إذ أوضحت الدراسة أن القروض البنكية مصدر من مصادر التمويل التقليدية التي تُشجع وتوجه النشاط الاقتصادي، وتساهم في انتعاش الاقتصاد الوطني، وتعمل على القضاء على جُل المشاكل الاقتصادية والاجتماعية. وحسب ما يمتلك البنك من مدخرات وودائع يقوم على أساسها بمنح القروض. وأكدت الباحثة في الدراسة أن القروض حتى ولو منحت وفق أسس ومعايير علمية سليمة، إلا أنها تنطوي على عنصر المخاطرة والذي لا يمكن للبنوك تفاديه كليًا؛ لأن تجنب الخطر يعني عدم قيام البنوك بنشاطها الرئيسي أساسًا والمتمثل في منح القروض.

أما دراسة أبو عبید (2003) فقد تناولت القروض المتعثرة لدى البنوك التجارية الأردنية من حيث الأسباب والآثار والحلول الممكنة، إذ هدفت الدراسة إلى التعرف على مفهوم القروض المتعثرة والتي يطلق عليها عادةً القروض غير العاملة، ومن ثم تحليل تطور هذه المشكلة لدى البنوك التجارية الأردنية خلال الفترة (2001-1992) لمعرفة ما إذا كانت تمثل ظاهرة أم لا.

بالإضافة على تحليل العلاقة بين القروض المتعثرة متمثلة بنسبة القروض غير العاملة إلى إجمالي حجم القروض والتسهيلات الائتمانية، وبين بعض المؤشرات الاقتصادية الكلية وبعض المؤشرات المصرفية، وتحديد أهم العوامل التي ساعدت على تطور ونمو مشكلة القروض المتعثرة وذلك من وجهة نظر الأطراف ذات العلاقة والمتمثلة في البنوك المقرضة، والمقترضين، ومدققى حسابات البنوك، وأخيرًا البنك المركزي الأردني. قامت الدراسة بتصميم استبانة تحتوي على 32 سؤالاً حول أسباب تعثر القروض، وتم توزيع هذه الاستبانة على عينة من موظفي الائتمان لدى البنوك التجارية الأردنية، وعينة من المقترضين، وعينة من العاملين في مجال التدقيق على حسابات البنوك، وعينة من مفتشى البنك المركزي الأردني. وقد تم تقسيم أسباب التعثر في الدراسة إلى أربع مجموعات رئيسية هي: مجموعة الأسباب المتعلقة بالبنوك وجاءت بالمرتبة الأولى من حيث تأثيرها على مشكلة القروض المتعثرة، ومجموعة الأسباب المتعلقة بالمقترضين وجاءت في المرتبة الثانية، ومجموعة الأسباب المتعلقة بالظروف الاقتصادية العامة وجاءت في المرتبة الثالثة، وأخيرًا مجموعة الأسباب المتعلقة بالجهات الرقابية وجاءت بالمرتبة الرابعة. وقد تبين وجود فروق ذات دلالة إحصائية بين أسباب تعثر القروض من وجهة نظر البنوك من جهة ومن وجهة نظر المقترضين من جهة أخرى، في حين لم تكن هناك فروق ذات دلالة إحصائية بين أسباب تعثر القروض من وجهة نظر البنك المركزي من جهة ومن وجهة نظر مدققى الحسابات من جهة أخرى. وقد تم التوصل إلى بعض التوصيات التي قد تساعد في حال تطبيقها في التخفيف من حدة المشكلة.

أما الدراسات الأجنبية التي تناولت القروض الشخصية فإنها عديدة أيضًا، منها ما قاموا به Gams and Correia, etc. (2020) حول تمويل القروض الشخصية والتجارية في شرق أوروبا، ودراسة حجم سوق التمويل الجماعي عبر الاتحاد الأوروبي؛ إذ أوضحت الدراسة إلى أن الائتمان الأصغر المخصص للأعمال التجارية والقروض الشخصية صعب الحساب والتقدير، وأن المقترضين النشيطين تزايد نسبهم باستمرار؛ إذ وصلت مبالغ القروض إلى مليار يورو في عام 2017 استنادًا إلى بيانات حول شرق أوروبا من المفوضية التابعة للاتحاد الأوروبي. طبقت الدراسة نموذج انحدار قادر على التنبؤ، والبيانات تم جمعها بين عامي 2018 - 2011 من خمس دول أوروبية، دولتان من فئة الدخل المتوسط الأدنى وهما: جمهورية مولدوفا وأوكرانيا، وثلاثة من فئة الدخل المتوسط الأعلى وهم: اليابان وأرمينيا وجورجيا. وأوضحت الدراسة كيفية التقدم بطلب للحصول على قروض شخصية سريعة التمويل مقارنة بقروض الأعمال، بالإضافة إلى ذلك توصلت النتائج إلى وجود علاقة على شكل U بين المعلومات المتوفرة وسرعة التمويل؛ ولكنها تنطبق فقط على قروض الأعمال، وأن التمويل الجماعي يتطلب تركيزًا أكبر عند منحه ويشكل الحلقة الأضعف في المشاكل الائتمانية، وأن المقترضين الأفراد لديهم رؤى طويلة الأجل خصوصًا أصحاب المشاريع الصغيرة الذين يستثمرون في التعليم والمهارات والتدريب.

كما قاما Bravo and Santamaria (2017) بدراسة حول تقييم أداء القروض لدى المكاتب البنكية وكيفية اتخاذ قرار الائتمان بناءً على عدم اليقين، إذ كانت معايير تقييم أداء القروض هي معدلات الخصم المختلفة التي يمكن استخدامها لحساب صافي القيمة الحالية (NPV) لكل تدفق خارج للقرض وتدفع للسداد، إذ تم تصنيف المكاتب المصرفية إلى مكاتب بنكية خاضعة لمعايير التقييم المحددة وأخرى غير خاضعة لمعايير (NPV) المتعددة، ونتيجة لذلك تم الحصول على ترتيب كامل لمكاتب البنوك من مؤشرات أدائها. وعلى الرغم من أن الطريقة المقترحة قابلة للتطبيق على أنواع مختلفة من القروض إلا إنه تم التأكيد على الفئة الخاصة

بالقروض الشخصية في الدراسة كحالة فعلية، وتم تقييم قروض هذه المجموعة من المكاتب المصرفية، وكيفية السداد خلال الفترة 2010-2001، وتوضيح كيفية جدولة البيانات الرقمية الخاصة بها.

وقد أجرت Shafinar Ismail (2013) دراسة حول محدّدات منح القروض الشخصية في ماليزيا. إذ يتمتع الماليزيون بمزايا ومنافع متعددة تقدمها الحكومة وتسهلها لهم، مثل قروض الاسكان، والمزايا الطبية، والحوافز المختلفة. إذ أظهرت الإحصائيات أن ديون الأسرة الماليزية تستمر في الزيادة نتيجة لتلك التسهيلات، حيث ارتفعت نسبة ديون الأسرة الى الناتج المحلي الإجمالي لماليزيا من 62.7 في عام 2008 الى 76.6 في عام 2011، كم أوضحت الدراسة إلى أن نسبة خدمة الدين المقبولة هي 30 %، ومع ذلك كانت نسبة خدمة ديون الأسرة الماليزية 9% في عام 2006 وارتفعت إلى 49% في عام 2009 وانخفضت بشكل طفيف إلى 48% في عام 2010، وهذا يشير إلى أن ما يقرب من نصف دخل الأسرة كان يستخدم لسداد ديونها، ونتيجة لذلك يصبح رصيد الإنفاق على التعليم والنقل والغذاء وحالات الطوارئ أقل. كما هدفت الدراسة إلى التحقيق في المحددات التي تساهم في زيادة القروض الفردية؛ حيث انها تساهم في زيادة ديون الأسرة الماليزية. تم تحديد العديد من المحددات المؤثرة في الطلب على القروض الشخصية منها: المعرفة حول القروض الشخصية، والوعي الإعلامي، والتصورات تجاه القروض الشخصية، وتأثير المعتقد الديني والأخلاقي. إذ تم توزيع الاستبيانات على 100 مُستجيب، وكان اجراء أخذ العينات المتبع هو أخذ العينات العشوائية الطبقة، وتم تحليل البيانات التي تم الحصول عليها باستخدام برنامج SPSS، وكانت نتائج الدراسة أن الوعي الإعلامي والمعتقد الديني والأخلاقي من أهم المحددات التي تؤثر على الإقراض الشخصي في ماليزيا.

وعند التّمعّن في الدراسات السابقة نراها قد تناولت موضوع القروض المختلفة من جوانب عدة، لكن هذه الدراسة تناولت القروض الشخصية حصريًا والطلب عليها في العاصمة الأردنية عمان خلال فترة محددة، واستخدمت أسلوب إحصائي كمي مُتقدم وهو الانحدار اللوجستي الثنائي لمعرفة أكثر العوامل أهمية في التأثير على قرار الطلب عليها خلال تلك الفترة.

### مجتمع وعينة الدراسة

يتكوّن مجتمع الدراسة من جميع الأشخاص النشيطين اقتصاديًا (قوة العمل) 6 في محافظة عمان، والذين بلغ عددهم 184388 شخصًا، وذلك حسب سجلات وزارة العمل لعام 2021، ومن أجل تحديد العينة تم الرجوع إلى الجدول الذي قدّمه كل من جونسون وكريستينسن لتسهيل قرار اتخاذ حجم العينة المطلوبة، اعتمادًا على حجم المجتمع الكلي وهامش الخطأ المسموح به (5%)، وبناءً عليه كان حجم العينة 384 شخص (Johnson & Christensen, 2014).

تم استخدام مفهوم العينة القصدية (الغرضية) في سحب العينة، وهي عينة غير احتمالية يتم اختيارها بعناية وتكون مناسبة في حالة توفر صفة معينة لدى مجموعة معينة ولا تتوفر في غيرهم، وقد تم ارسال 400 استبانة على موظفي بعض مؤسسات القطاع العام والخاص

وأصحاب الأعمال الحرة وممن لا يعملون في محافظة العاصمة الأردنية عمّان، وكان عدد الاستبيانات الصالحة للتحليل 350 استبيانياً، وقد تم توزيع الاستبانة 7 إلكترونيًا.

### منهجية الدراسة والنموذج القياسي

استخدمت الدراسة نموذج الانحدار اللوجستي ثنائي الاستجابة (logistic regression) من خلال برنامج SPSS لتحديد أهم العوامل المؤثرة في الطلب على القروض الشخصية، والذي يعتبر حالة خاصة من النموذج الخطي العام (Generalized Linear Model). علمًا أن دوال النموذج اللوجستي تُستخدم عندما يكون المتغير التابع المراد التنبؤ به هو متغير اسمي فيه صفتين فقط (ثنائي القيمة) والذي يتبع توزيع بيرنولي (Bernoulli) للمتغيرات ذات الحدين، ولتحديد حجم واتجاه التأثير النسبي لكل متغير من المتغيرات المستقلة والمقترحة في تأثيرها على المتغير التابع (Pallant, 2013).

ومن هنا، فإن النموذج اللوجستي الثنائي (Binary Logistic Model) هو الأنسب في التحليل كون بيانات المتغير التابع الثنائي تتبع التوزيع غير الطبيعي، والأكثر قوة من الأساليب الإحصائية الأخرى كالانحدار الخطي (البسيط والمتعدد)، وبخاصة في تحليل العوامل الأقوى في تفسير الظاهرة، وبالتالي فإن هذا يعني أنّ استخدام هذه المنهجية، وبنائها على بيانات الأفراد الخاضعين للدراسة؛ من المفترض أن تعطي معلومات أكثر دقة في التركيز على المتغيرات المعنوية وترتيبها حسب الأهمية، ومن ثم يتم تقدير معاملات النموذج بعد إدخال العوامل المحددة لقرار الطلب على القروض الشخصية باستخدام تقدير الإمكان الأعظم (Maximum Likelihood Estimation)، الذي يختار المعلمات التي تجعل احتمال حدوث القيم المشاهدة أعظمًا، فالهدف في طريقة (ML) هو تعظيم لوغاريتم دالة الإمكان (LLF) للحصول على قيم المعلمات المجهولة بحيث يكون احتمال مشاهدة قيم ال Y's كبيراً معظم بقدر الإمكان (Gujarati & Porter, 2009).

تُعبّر المعادلة (1) عن النموذج القياسي العام المستخدم في هذه الدراسة، وكما يأتي:

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + u_i \quad \dots (1)$$

علمًا أن:

$Y_i$ : المتغير التابع والذي يُعبّر عن احتمالية الطلب على القروض الشخصية، وهو متغير اسمي ثنائي القيمة، يأخذ القيمة (1) إذا كان الفرد حاصل على قرض شخصي، والقيمة (0) إذا لم يكن حاصل على قرض شخصي.

$X_{ij}$ : مجموعة من المتغيرات المستقلة الكمية والأسمية، والتي يتوقع أن يكون لها تأثير على قرار الحصول على قرض شخصي. وتشمل عوامل اقتصادية واجتماعية وثقافية مختلفة، كالمستوى التعليمي، ومجال عمل الفرد، وعدد أفراد العاملين في الأسرة، ووجود تسهيلات مُيسرة من قبل الجهاز المصرفي، ووجود تحويلات نقدية مستلمة من الخارج، وتأثر دخل الفرد بأزمة كورونا، ودخل الفرد، إضافة إلى متغيرات مستقلة أخرى.

$u_i$ : حد الخطأ.

ومن هنا، وبما أنّ المتغير التابع في مثل هذه النماذج من المتغيرات غير الكمية، فإن استخدام نموذج الانحدار الخطي لن يكون ملائمًا إحصائيًا في دراسة تحليل العوامل المؤثرة في قرار الطلب على القروض الشخصية؛ وهذا يناقض الافتراضات الأساسية لتحليل الانحدار الخطي، وأهمها افتراضيين رئيسيين، الأول اعتدالية التوزيع (Normality) والثاني تجانس التباين (Homoscedasticity)، إذ تنشأ هاتان المشكلتان بسبب الطبيعة الثنائية للمتغير التابع (Greene, 1993).

وبالتالي، فإن المتغير التابع سيأخذ فقط قيمتين (1,0) تعبران عن وجود القرض الشخصي من عدمه، أي أنّ هذا المتغير يأخذ القيمة (1) باحتمال (P) إذا كان لدى الفرد قرضًا شخصيًا، والقيمة 0 باحتمال (1-P) إذا لم يكن لديه. وبما أن قيمة  $0 \leq P_i \leq 1$  فإن هذا يعني أنّ  $\frac{P_i}{1-P_i}$  ستأخذ قيمة موجبة ومتصلة بين  $0 \leq \frac{P_i}{1-P_i} \leq \infty$  وبأخذ اللوغاريتم الطبيعي لهذه القيمة تصبح القيم عندها محصورة بين  $-\infty \leq \ln \frac{P_i}{1-P_i} \leq \infty$  وبالتالي، فإنه يمكن كتابة نموذج الانحدار في المعادلة (1) السابقة بصيغة الانحدار اللوجستي، وبالاعتماد على الافتراضات السابقة كما يأتي:

$$\ln \frac{P_i}{1-P_i} = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + v_i \quad \dots \dots \dots (2)$$

علمًا أن  $\frac{P_i}{1-P_i}$  تسمى نسبة التّرجيح أو الفرصة (Odds ratio).

ولتحليل المتغير التابع ثنائي القيمة يتم عادة استخدام النموذج اللوجستي، والذي يعتمد على دالة التوزيع التراكمي الطبيعي (Normal cumulative distribution function) التي تكون على شكل حرف S، والذي يفترض دالة الكثافة الاحتمالية التراكمية (Cumulative probability density function) الآتية:

$$P(Y) = \frac{1}{1+e^{-(\beta_0 + \sum_{j=1}^k \beta_j X_{ij})}} \quad \dots \dots \dots (3)$$

وتعبر معادلة (3) عن احتمال حدوث Y (أي احتمال أن تنتمي حالة ما إلى فئة معينة). وبالتالي فإن القيمة الناتجة عن المعادلة هي قيمة احتمال تتراوح بين 0 و1 وتعني القيمة القريبة من 0 أنه من غير المحتمل أن تحدث Y، أما القيمة القريبة من 1 فتعني أنه من المرجح جدًا حدوث Y.

وبالتالي، فإن التحليل القياسي لهذه الدراسة سيعتمد على تقدير معاملات النموذج اللوجستي الثنائي لاحتمالية الطلب على القروض الشخصية من قبل المواطنين الأردنيين في المعادلة التالية:

$$\text{PDP} = \beta_0 + \sum_{i=1}^4 \beta_{1i} \text{Edu}_i + \sum_{i=1}^4 \beta_{2i} \text{S}_i + \sum_{i=1}^3 \beta_{3i} \text{Wm}_i + \sum_{i=1}^3 \beta_{4i} \text{Inc}_i + \sum_{i=1}^4 \beta_{5i} \text{Rp}_i + \sum_{i=1}^4 \beta_{6i} \text{MS}_i + \beta_7 \text{Rem} + \beta_8 \text{Scf} + \beta_9 \text{Cor} + \beta_{10} \text{Gen} + \varepsilon_i \quad \dots \dots \dots (4)$$

إذ تمثل  $\beta_0$  إلى  $\beta_{10}$  معاملات الدالة اللوجستية الثنائية، وتعرف المتغيرات 8 في المعادلة السابقة كالتالي:

PDP: احتمالية طلب قرض شخصي وتأخذ القيمة بين الصفر والواحد.

Edui: المستوى التعليمي، وقد تم تقسيم هذا المتغير إلى متغيرات (مجموعات) فرعية كما يأتي:

Edu1: أقل من ثانوية عامة؛ Edu2: ثانوية عامة أو دبلوم متوسط؛ Edu3: بكالوريوس؛ Edu4: دراسات عليا (ماجستير، أو دكتوراه)

Js: طبيعة عمل الفرد، وقد تم تقسيم هذا المتغير إلى متغيرات (مجموعات) فرعية كما يأتي:

Js1: قطاع عام؛ Js2: قطاع خاص؛ Js3: أعمال حرة؛ Js4: لا يعمل

Wmi: عدد الأفراد العاملين في الأسرة، وقد تم تقسيم هذا المتغير إلى متغيرات (مجموعات) فرعية كما يأتي:

Wm1: فرد واحد؛ Wm2: 2-4 أفراد؛ Wm3: أكثر من أربعة أفراد

Inci: الدخل الشهري للفرد بالدينار الأردني، وقد تم تقسيم هذا المتغير إلى متغيرات (مجموعات) فرعية كما يأتي:

Income1: أقل من 500؛ Income2: 500 - 1000؛ Income3: أكثر من 1000

Rpi: مكان إقامة الفرد، وقد تم تقسيم هذا المتغير إلى متغيرات (مجموعات) فرعية كما يأتي:

Rp1: شرق عمان، Rp2: غرب عمان، Rp3: شمال عمان، Rp4: جنوب عمان

Msi: الحالة الاجتماعية للفرد، وقد تم تقسيم هذا المتغير إلى متغيرات (مجموعات) فرعية كما يأتي:

Ms1: متزوج، Ms2: أعزب، Ms3: منفصل، Ms4: أرمل

Rem: استلام تحويلات نقدية، إذ يأخذ هذا المتغير قيمة واحد إذا كان الشخص يتلقى تحويلات من الخارج، والقيمة صفر إذا كان لا يتلقى.

Scf: وجود تسهيلات ائتمانية مُبسرة من قبل الجهاز المصرفي، إذ يأخذ هذا المتغير قيمة واحد إذا كانت متوفرة، والقيمة صفر إذا كانت الإجابة بلا.

COR: وجود تأثير اقتصادي لجائحة فيروس كورونا، إذ يأخذ هذا المتغير قيمة واحد إذا كان يوجد، والقيمة صفر إذا كانت الإجابة بلا.

Gen: جنس الفرد، إذ يأخذ هذا المتغير قيمة واحد إذا كان الشخص ذكراً، والقيمة صفر إذا كان أنثى.

حد الخطأ:  $\epsilon$

مع مراعاة أن تكون قيمة كل متغير تفصيلي من المتغيرات المُصنفة السابقة = 1 إذا كان الفرد من ضمن الفئة المحددة، وخلاف ذلك فإن قيمته = 0 أثناء التحليل. تعتمد هذه الدراسة على الفرضية التالية:

الفرضية الرئيسية: لا يوجد أثر لأبرز المتغيرات الاقتصادية والاجتماعية على الطلب على القروض الشخصية في محافظة عمان. وتتفرع عن هذه الفرضية الفرضيات الفرعية التالية:

الفرضية الفرعية الأولى: لا يوجد أثر معنوي للمؤهل العلمي للفرد على طلب القروض الشخصية.

الفرضية الفرعية الثانية: لا يوجد أثر معنوي لطبيعة عمل الفرد على طلب القروض الشخصية.

الفرضية الفرعية الثالثة: لا يوجد أثر معنوي لعدد الأفراد العاملين في الأسرة على طلب القروض الشخصية.

الفرضية الفرعية الرابعة: لا يوجد أثر معنوي لدخل الفرد الشهري على طلب القروض الشخصية.

الفرضية الفرعية الخامسة: لا يوجد أثر معنوي لمكان إقامة الفرد على طلب القروض الشخصية.

الفرضية الفرعية السادسة: لا يوجد أثر معنوي لحالة الفرد الاجتماعية على طلب القروض الشخصية.

الفرضية الفرعية السابعة: لا يوجد أثر معنوي للتحويلات النقدية التي يستلمها الفرد من الخارج على طلب القروض الشخصية.

الفرضية الفرعية الثامنة: لا يوجد أثر معنوي لوجود تسهيلات ائتمانية مُيسرة من قبل الجهاز المصرفي على طلب القروض الشخصية.  
 الفرضية الفرعية التاسعة: لا يوجد أثر معنوي لحدوث جائحة فيروس كورونا على طلب القروض الشخصية.  
 الفرضية الفرعية العاشرة: لا يوجد أثر معنوي لجنس الفرد على طلب القروض الشخصية.

### نتائج تقدير النموذج القياسي

تظهر نتائج التّقدير كما هو موضح في الجدول رقم (1) التأثير النسبي لكل متغير من المتغيرات المستقلة، والتي من الممكن أن تؤثر في قرار الطلب على القروض الشخصية. في البداية أوضحت إحصائية كاي تربيع (Chi-square statistic) والتي تستخدم لفحص مدى ملائمة معاملات النموذج ككل (والتي هي بديل عن استخدام قيم F و R2 في حالة الانحدار الخطي)، إلى أنّ النموذج المستخدم ملائم، إذ بلغت قيمة هذه الإحصائية 146.35 وهي قيمة معنوية عند مستوى معنوية 0.000، كذلك اختبار (Hosmer & Lemeshow) الذي بلغت قيمته 9.603 عند مستوى معنوية 0.294 وبالتالي لا نستطيع رفض الفرضية العدمية التي تنص أن النموذج ملائم، كما يظهر في أسفل الجدول (1). كما وتشير قيم كل من (Cox & Snell R2) و (Negelkerkel) والتي تظهر القوة التفسيرية للنموذج إلى أنّ النموذج المُستخدم استطاع أن يفسر ما نسبته 54% بحسب قيمة (Negelkerke R2) و 34% بحسب قيمة (Cox & R2) Snell من التغيرات في احتمالية الطلب على القروض الشخصية.  
 إنّ هذه القيم والتي تقيس مدى ملائمة وجودة النموذج في قياس العوامل التي يمكن أن تؤثر في الطلب على القروض الشخصية، تدلّ على أنّ النموذج المفترض مقبول إحصائياً، ويمكن الاعتماد عليه في تفسير اتجاه وقوة المتغيرات المستخدمة في التّقدير. إذ أظهرت هذه الإحصائيات أنّ الانحدار اللوجستي ثنائي القيمة والمستخدم في هذا الجزء من الدراسة يعطي نتائج موثوقة نسبياً؛ مما يزيد فعالية الاعتماد على نتائج هذه الدالة.  
 جدول رقم (1): نتائج تقدير الدالة اللوجستية للعوامل الاقتصادية والاجتماعية للطلب على القروض الشخصية

Dependent Var. Personal Loan Demand Probability (PDP)					
No. of Observations: 350					
Variable	$\beta$	S.E.	Wald	Sig.	Exp( $\beta$ )
Inc			5.840	.054	
Inc1	-20.469	9809.825	.000	.998	.000
Inc2	-19.165	9809.825	.000	.998	.000
Rem	.583	1.324	.194	.660	1.792
Edu			14.164	.003	
Edu1	1.659	1.825	.826	.363	5.254
Edu2	2.207	1.840	1.439	.230	9.093
Edu3	.109	1.806	.004	.952	1.115
Scf	2.151	.453	22.527	.000	8.597
Cor	1.633	.556	8.635	.003	5.121
Rp			8.609	.035	
Rp1	-1.266	.934	1.834	.176	.282

Rp2	-2.094	.998	4.402	.036	.123
Rp3	.633	1.191	.282	.595	1.884
Gen	.019	.440	.002	.965	1.020
Ms			12.815	.005	
Ms1	3.333	1.404	5.635	.018	28.031
Ms2	1.289	1.461	.779	.377	3.630
Ms3	24.246	12183.632	.000	.998	3.38688
Jo			.856	.836	
Jo1	-21.178	12559.404	.000	.999	.000
Jo2	-20.615	12559.404	0.000	.999	.000
Jo3	-20.424	12559.404	0.000	.999	.000
Wm			2.377	.305	
Wm1	.923	.900	1.051	.305	2.517
Wm2	.275	1.049	.069	.793	1.317
Constant	37.092	15936.497	0.000	.998	4496.000
Goodness of Fit		Chi-square 146.354		Df	Sig.
				20	.0000
		Hosmer & Lemeshow 9.603		8	.294
Cox & Snell R <sup>2</sup>	.34	Nagelkerke R <sup>2</sup>		.54	

المصدر: إعداد الباحثة

أظهرت نتائج تقدير دالة الانحدار اللوجستي الثنائي أن هناك متغيرات مستقلة كانت ذات دلالة إحصائية عالية جدًا، مقارنة بمتغيرات مستقلة أخرى كانت دلالتها الإحصائية منخفضة. إذ كانت التسهيلات الائتمانية الميسرة الممنوحة من قبل الجهاز المصرفي، وحدوث جائحة فيروس كورونا، والمستوى التعليمي للفرد، والحالة الاجتماعية للفرد، ومكان إقامة الفرد من أبرز العوامل ذات الدلالة الإحصائية والتأثير المعنوي المرتفع على طلب القروض الشخصية. وفي المقابل فقد أشارت النتائج بشكل عام إلى أن الدخل الشخصي للفرد، وجنس الفرد، وطبيعة عمل الفرد بالإضافة إلى عدد العاملين في الأسرة كانت غير معنوية، ولا تؤثر بشكل كبير على طلب القروض الشخصية.

كما أظهرت نتائج التقدير أن الدالة اللوجستية المستخدمة قد استطاعت التنبؤ بقرابة 260 حالة من أصل 350 فرد بشكل صحيح، وبنسبة تنبؤ 84% من إجمالي الحالات، وهي نسبة مرتفعة تؤكد معنوية النموذج بشكل عام، كما هو مبين في الجدول رقم (2).

جدول رقم (2): النسبة المئوية للتصنيف الصحيح للنموذج

Observed	Predicted			
	Var.PDP		Percentage Correct	
		0.00	1.00	
Step1 Var.PDP	0.00	34	36	48.6
	1.00	20	260	92.9
Overall Percentage				84.0

\* The cut value is 0.5

المصدر: إعداد الباحثة

تشير النتائج القياسية للدالة اللوجستية إلى أن من أكثر المتغيرات أهمية في التأثير على طلب القروض الشخصية هو التسهيلات الائتمانية المُيسرة الممنوحة من قبل الجهاز المصرفي والذي جاء في المرتبة الأولى، إذ أظهرت النتائج أن تأثير هذا المتغير كان معنوياً بدرجة كبيرة وبقوة تأثير موجبة، إذ إن نسبة ترجيح الطلب على القروض الشخصية تزداد بمقدار 8.59 مرة في حالة أن يكون الفرد قد تلقى تسهيلات ائتمانية مُيسرة من قبل المصرف الذي يتعامل معه عن الفرد الذي لم يتلقى تسهيلات مُيسرة.

كما أشارت النتائج أن التأثير بجائحة فيروس كورونا يأتي بالمرتبة الثانية من ناحية القوة والأهمية في طلب القروض الشخصية، إذ تشير النتائج إلى أن نسبة ترجيح الطلب على القروض الشخصية تزداد بمقدار 5.12 مرة في حالة أن يكون الفرد قد تأثر بالجائحة مادياً عن الأفراد الذين لم يتأثروا.

وفي المقابل أظهرت النتائج أن المستوى التعليمي للفرد جاء في المرتبة الثالثة من ناحية القوة والأهمية في الطلب على القروض الشخصية، إذ تشير النتائج إلى أن نسبة ترجيح الطلب على القروض الشخصية من قبل الأفراد ضمن فئة أقل من ثانوية عامة تزداد بمقدار 5.25 مرات في حالة أن يكون الفرد ضمن هذه الفئة عن الأفراد الذين ضمن فئة الدراسات العليا، وبمعنوية إحصائية عالية وقوة تأثير موجبة.

أظهر متغير الحالة الاجتماعية للفرد أهمية في قرار الطلب على القروض الشخصية والذي جاء في المرتبة الرابعة، إذ أشارت النتائج إلى أن نسبة ترجيح الحصول على قرض شخصي تزداد بمقدار 28 مرة في حالة أن يكون الفرد ضمن فئة الأعزب عن الأفراد الأرامل، وبدلالة إحصائية مرتفعة وقوة تأثير موجبة.

وأخيراً كان تأثير متغير مكان إقامة الفرد من بين مجموعة المتغيرات المعنوية المؤثرة في احتمالية الطلب على القروض الشخصية وجاء في المرتبة الخامسة. إذ أظهرت نتائج التقدير القياسي أن تأثير هذا المتغير كان معنوياً بدرجة مقبولة. وبقوة تأثير موجبة وضئيلة. إذ إن نسبة ترجيح الطلب على القروض الشخصية من قبل الأفراد الذين يقيمون في شرق محافظة عمان تنخفض بمقدار 0.28 مرة عن الأفراد الذين يقيمون في جنوب المحافظة.

وعلى الجانب الآخر فقد أظهرت نتائج الدالة اللوجستية باستخدام عينة الدراسة من المواطنين الأردنيين أن عددًا من المتغيرات التي كان يُعتقد أنها مؤثرة في الطلب على القروض الشخصية كانت غير معنوية، ولا تؤثر في احتمالية الطلب على القروض، ومن أبرزها متغير الدخل الشخصي للفرد، وطبيعة عمل الفرد، وكذلك عدد العاملين في الأسرة.

### الخلاصة

خلُصت الدراسة إلى أن التسهيلات الائتمانية المُيسرة الممنوحة من قبل الجهاز المصرفي، وحدوث جائحة فيروس كورونا، والمستوى التعليمي للفرد، والحالة الاجتماعية للفرد، ومكان إقامة الفرد من أبرز العوامل المؤثرة على الطلب على القروض الشخصية. وفي المقابل فقد توصلت الدراسة إلى أن الدخل الشخصي للفرد، وجنس الفرد، وطبيعة عمل الفرد بالإضافة إلى عدد العاملين في الأسرة كانت عوامل غير مؤثرة على هذا الطلب.

بناءً على ما تقدم من تحليل ونتائج تم التوصل إليها، يُمكن وضع التوصيات والمقترحات الآتية:

- أظهرت نتائج الدراسة بأن الدخل الشخصي للفرد متغير لم يكن معنوياً وذلك خلافاً لما هو متوقع، مما يدل على أن إقبال الأفراد على القروض لا يرتبط مباشرة بنقص أو كفاية الدخل

المادي الشهري حسب نتائج الدراسة؛ لذلك توصي الدراسة بإجراء دراسات جديدة حول الدوافع الحقيقية للطلب على القروض الشخصية.

- أظهرت نتائج الدراسة بأن طبيعة عمل الأفراد لم يكن أيضاً معنوياً وذلك خلافاً لما هو متوقع، مما يدل على ابتعاد الأفراد وخصوصاً ممن يعملون بالقطاع العام عن تحمل أعباء الديون المادية المترتبة على القروض. إذ كان من المتوقع إقبال العاملين في القطاع العام خاصةً على الطلب على القروض لسهولة حصولهم عليها من قبل المصارف بسبب دخولهم الثابتة.
- أظهرت نتائج الدراسة بأن التسهيلات الائتمانية المُيسرة الممنوحة من قبل الجهاز المصرفي لها تأثير معنوي قوي على الطلب على القروض الشخصية، مما شجع العديد من الأفراد على الطلب عليها، ومن هنا توصي الدراسة بتشديد بعض الشروط والتعليمات الخاصة بالقروض وعدم التساهل في منحها خصوصاً للأفراد الذين قد يتعثرون في السداد، الأمر الذي قد يؤدي إلى تراكم الديون المتعثرة لدى تلك المصارف.
- أظهرت نتائج الدراسة بأن جائحة فيروس كورونا لها تأثير معنوي على الطلب على القروض الشخصية، وهذا يدل على تبعات الجائحة على الأوضاع الاقتصادية للأفراد، وإقبالهم على القروض. الأمر الذي يتطلب من الجهات المعنية وضع خطط للطوارئ للأفراد والمؤسسات في حال حدوث أي مُستجدات أو كوارث طبيعية.
- إجراء دراسات مستفيضة على الجهاز المصرفي من قبل البنك المركزي، خاصةً مع ظهور تحديات اقتصادية واجتماعية عديدة مؤخراً تستدعي البحث والاهتمام، وكذلك دراسات تقييمية حول الحاجة الفعلية للتسهيلات الائتمانية وتحديد سقفوف دنيا وعليا للقروض وخصوصاً الشخصية منها.

### المراجع العربية

1. أبو عبيد، جمال (2003)، القروض المتعثرة لدى البنوك التجارية الأردنية: الأسباب والآثار والحلول الممكنة: دراسة تحليلية، أطروحة دكتوراة، جامعة عمان العربية، الأردن.
2. البنك المركزي (2022) و(2023)، تقارير سنوية منشورة، الموقع الإلكتروني الرسمي للبنك المركزي، عمان، الأردن.
3. رجب، أيمن، صقر، أحمد، وآخرون (2022)، أثر القروض المصرفية المتعثرة على مؤشرات الربحية ( دراسة مقارنة بين البنوك في مصر وتونس)، مجلة البحوث البيئية والطاقة، مجلد 11، العدد 19، جامعة المنوفية، الجمهورية العربية المصرية.
4. سلمان، ميساء، العبادي، سمير (2015)، المشروعات الصغيرة وأثرها التنموي، مركز الكتاب الأكاديمي، عمان، الأردن.
5. السيسي، صلاح الدين (1998)، التسهيلات المصرفية للمؤسسات والأفراد، دار وسام للطباعة والنشر، بيروت، لبنان.
6. الصميدعي، محمد، يوسف، ردينة (2005)، التسويق المصرفي، دار المناهج للنشر والتوزيع، عمان، الأردن.
7. عجام، ميثم (2001)، نظرية التمويل والتمويل الدولي، دار زهران للنشر والتوزيع، عمان، الأردن.

8. عوينات، محمد (2017)، أثر القروض المتعثرة على الأداء المالي للبنوك التجارية: دراسة حالة البنوك (Agb ,Bea ,Baraka) فترة: 2010-2015، رسالة ماجستير، جامعة قاصدي مرباح- ورقلة، الجزائر.
9. عفانة، محمد (2019)، إدارة الائتمان المصرفي، دار اليازوري العلمية للنشر والتوزيع، عمان، الأردن.
10. عبد الحليم، سارة (2012)، دور القروض المصرفية في تحقيق الربحية للبنوك التجارية (دراسة حالة وكالة القرض الشعبي الجزائري بعين مليلة)، رسالة ماجستير، جامعة العربي بن مهيدي أم البواقي، الجزائر.
11. الغرباوي، شهدان (2019)، تمويل المشروعات الصغيرة، دار الفكر الجامعي للنشر والتوزيع، الاسكندرية، مصر.
12. المالكي، عبد الله (2019)، مضائق الاقتصاد الأردني، دار الفارس للنشر والتوزيع، عمان، الأردن.
13. مختار، إبراهيم (2005)، التمويل المصرفي (منظومة للتعامل بين البنوك والمقترضين)، الأنجلو المصرية للتوزيع والنشر، القاهرة، جمهورية مصر العربية.
14. النجار، فريد (2000)، إدارة الائتمان والقروض المصرفية المتعثرة، مؤسسة شباب الجامعة، جامعة الزقازيق، الجمهورية العربية المصرية.

### المراجع الأجنبية

1. Bravo, M., Santamaria, D., (2017), Evaluating loan Performance for Bank Offices: A Multicriteria Decision-Making Approach, INFOR: Information System and Operational Research, volume 50, 2012- Issue3, Canada.
2. Gama, A., Correia, R., Augusto, M., Duarte, F., (2020), Online microfinance in Eastern Europe: Personal versus business loan funding, European Microfinance Network.
3. Greene, H., (1993). Econometrics Analysis, 2<sup>nd</sup> ed. New York University, Prentice Hall International, Inc., p.589.
4. Gujarati, N., and Porter, C. (2009). Basic Econometrics, 5<sup>nd</sup> ed. New York: McGraw-Hill companies, Inc., p.589.
5. Ismail, SH. (2013). Determinants of Personal Loans Borrowing: An Empirical Study. Faculty of Business Management, University of Teknologi MARA, Malaysia.
6. Johnson, R., and Christensen, L. (2014). Educational research: Quantitative, qualitative, and mixed approaches, 5<sup>nd</sup> ed. South Alabama University, USA. ISBN 978-1-4522-4440-2.
7. Pallant, J., (2013). SPSS Survival Manual: A Step by step Guide to Data Analysis Using IBM Spss, 5<sup>nd</sup> ed. University of Melbourne, Australia, ISBN-13: 978-0335262588, Inc., p.211.

## المُلحق (1)

## أداة الدراسة (الاستبانة)

العوامل الاقتصادية والاجتماعية للطلب على القروض الشخصية في محافظة عمان خلال جائحة كورونا (2020-2022)  
دراسة موجهة إلى قوة العمل الأردنية في محافظة عمان

السيدة/..... المحترم/ة

تحية طيبة وبعد،،،

في إطار الدراسة الحالية حول الطلب على القروض الشخصية تم تصميم هذه الاستبانة لمعرفة آراء واتجاهات الأفراد حول هذه الظاهرة، ويتوقع ألا يستغرق إكمال تعبئة هذه الاستبانة أكثر من خمس دقائق، ولا حاجة لكتابة اسمك أو عنوانك، كما أنه لا داعي للتفكير كثيراً بإجابة الأسئلة حيث إنه لا توجد إجابة خاطئة أو صحيحة، فقط أجب حسبما يتبادر إلى تفكيرك أولاً، وتؤكد الباحثة أن المعلومات التي ستزودونها بها ستعامل بسرية تامة ولأغراض البحث العلمي فقط؛ لذا يرجى عند إكمال تعبئة الاستبانة إعادة إرسالها إلى الباحثة. مع خالص الشكر والتقدير لتعاونكم في دعم البحث العلمي.

الباحثة

رانيا عدنان

الرجاء وضع دائرة حول رمز الإجابة المناسبة:

1- الجنس	ذكر	أنثى
----------	-----	------

2- الحالة الاجتماعية	متزوج	أعزب	منفصل	أرمل
----------------------	-------	------	-------	------

3- المؤهل العلمي	أقل من ثانوية عامة	ثانوية عامة أو دبلوم متوسط	بكالوريوس	دراسات عليا (ماجستير، أو دكتوراه)
------------------	--------------------	----------------------------	-----------	-----------------------------------

4- العمر	دون 30 سنة	30- أقل من 40 سنة	40 سنة فأكثر
----------	------------	-------------------	--------------

5- مجال العمل الحالي	قطاع حكومي	قطاع خاص	أعمال حرة	لا أعمل
----------------------	------------	----------	-----------	---------

6- عدد الأفراد العاملين في الأسرة؟	فرد واحد	2-4 أفراد	أكثر من أربعة أفراد
------------------------------------	----------	-----------	---------------------

7- دخل الفرد الشهري (بالدينار الأردني)	أقل من 500	500- أقل من 1000	أكثر من 1000
--	------------	------------------	--------------

8- هل تستلم تحويلات نقدية منتظمة من أفراد يعملون في الخارج؟	نعم	لا
---	-----	----

9- هل أنت حاصل على قرض شخصي؟	نعم	لا
حدد المصرف		

10- هل تعتقد بوجود تسهيلات ميسرة لمنح القروض من قبل الجهاز المصرفي؟	نعم	لا
---	-----	----

11- هل جائحة كورونا أثرت على دخلك الشهري؟	نعم	لا
---	-----	----

شراء عقار	12- أسباب الطلب على القرض الشخصي ( يمكن اختيار أكثر من خيار)
شراء مركبة	
الرحلات السياحية	
عمليات جراحية	
تسديد ديون سابقة	
إنشاء مشروع استثماري خاص	
تعليم الأبناء	
غير ذلك ) حدد).....	

غرب عمان	شرق عمان	13- مكان الإقامة
شمال عمان	جنوب عمان	

خالص الشكر والتقدير للوقت الثمين في تعبئة الاستبانة،،،،،

### المُلحق (2)

#### قائمة الاختصارات و الرموز

ABBREVIATIONS OR SYMBOLS	WORD
PDP	Personal Loan Demand Probability
Edu	Education level
Wm	Workers members
Inc	Income
Rp	Residence place
Ms	Marital status
Rem	Remittances
Cor	Corona crisis
Scf	Smoothly credit facilities
Gen	Gender
Js	Job status
$\varepsilon$	Error term

## **A statistical analysis of the Type II Tobit model maximum Likelihood in case of non-ignoring missing data**

Mona Emad El-Din Mohamed, Mervat El-Gohary

Department of Statistics- Faculty of commerce-Al-Azhar University (Girls Campus), Egypt.

Ahmed Amin El-Sheikh. Dept. of Applied Statistics and Econometrics Faculty of Graduate Studies for Statistical Research Cairo University. Egypt.

تاريخ استلام البحث: 2023/05/12

تاريخ قبول البحث: 2023/06/22

نشر البحث في العدد الواحد والعشرين: ديسمبر / كانون اول 2023

2522-64X/511.8

رمز التصنيف ديوي / النسخة الالكترونية (Online):

2519-948X/511.8

رمز التصنيف ديوي / النسخة الورقية (Print):

## A statistical analysis of the Type II Tobit model maximum Likelihood in case of non-ignoring missing data

Mona Emad El-Din Mohamed, Mervat El-Gohary

Department of Statistics- Faculty of commerce-Al-Azhar University (Girls Campus), Egypt.

Ahmed Amin El-Sheikh. Dept. of Applied Statistics and Econometrics Faculty of Graduate Studies for Statistical Research Cairo University. Egypt.

### Abstract

Credit scoring is regarded as a core competence of commercial banks during the last few decades. A number of credit scoring models have been developed to evaluate credit risk of new loan applicants and existing loan clients. The main purpose of the present paper is to evaluate credit risk in banks using credit scoring models. Statistical techniques are used: maximum likelihood for one can use linear models and for, one can use Type II Tobit model, a Monte Carlo simulation study is employed, under non-ignorable missing data. The credit scoring task is performed on one bank's personal loans data-set. The results show that Tobit type-II model is more fitted than linear models.

**Key words:** Credit scoring, Type II Tobit, loan prediction, missing data, linear models, credit risk, maximum likelihood function.

التحليل الإحصائي للبيانات المفقودة باستخدام دالة إمكان نموذج توبيت من النوع الثاني  
المستخلص:

يعتبر نظام التصنيف الائتماني من الكفاءات الأساسية للبنوك التجارية خلال العقود القليلة الماضية. تم تطوير عدد من نماذج التصنيف الائتماني لتقييم مخاطر الائتمان لمقدمي طلبات القروض الجدد وعملاء القروض الحاليين. الغرض الرئيسي من هذه الورقة هو تقييم مخاطر الائتمان في البنوك باستخدام نماذج التصنيف الائتماني. يتم استخدام التقنيات الإحصائية: أقصى احتمالية يمكن للمرء استخدام النماذج الخطية ومن أجل، يمكن استخدام نموذج توبيت من النوع الثاني، يتم استخدام دراسة محاكاة مونت كارلو، في ظل بيانات مفقودة غير قابلة للتجاهل. يتم تنفيذ مهمة تسجيل الائتمان على مجموعة بيانات القروض الشخصية لأحد البنوك. أظهرت النتائج أن نموذج Tobit type-II أكثر ملاءمة من النماذج الخطية.

**الكلمات المفتاحية:** التصنيف الائتماني، نموذج توبيت من النوع الثاني، التنبؤ بالقروض، البيانات المفقودة، النماذج الخطية، مخاطر الائتمان، دالة الإمكان الأعظم.

### Introduction

Credit scoring is regarded as a core competence of commercial banks during the last few decades. A number of credit scoring models have been developed to evaluate credit risk of new loan applicants and

existing loan clients. The main purpose of the present study is to evaluate credit risk in Egyptian banks using credit scoring models. Three statistical techniques are used: discriminant analysis, probit analysis and logistic regression. The credit scoring task is performed on one bank's personal loans data-set. The results so far revealed that all proposed models gave a better average correct classification rate than the one currently used. Also both type I and type II errors had been calculated in order to evaluate the misclassification costs.

Latterly, credit risks have become one of the most important financial topics of interest, especially in the banking sector. The role of credit risks has changed dramatically over the last ten decades, from passive automation to a strategic device. The process of credit risk evaluation has the interest of many researchers nowadays. Recently, bankers have come to realize that banking operations affect and affected by the natural environment and that consequently the banks might have an important role to play in helping to raise environmental standards. Although the environment presents significant risks to banks, in particular environmental credit risk, it also perhaps presents profitable opportunities (Thompson, 1998).

Credit scoring is the use of statistical models to determine the likelihood that a prospective borrower will default on a loan. Credit scoring models are widely used to evaluate business, real estate, and consumer loans (Gup & Kolari, 2005,). Credit scoring models (see, for example: Lewis, 1992; Bailey, 2001; Mays, 2001; Malhotra & Malhotra, 2003; Thomas et al., 2004; Sidique, 2006; Chuang & Lin, 2009; Sustersic et al, 2009) are some of the most successful applications of research modelling in finance and banking. Harris (2015) investigated the practice of credit scoring and introduced the use of the clustered support vector machine (CSVM) for credit scorecard development. Abbod, et al. (2016) during the last few years there has been marked attention towards hybrid and ensemble systems development, having proved their ability more accurate than single classifier models. Kozodo, et al. (2019) Credit scoring models support loan approval decisions in the financial services industry.

Judgemental techniques and/or credit scoring models can support making a decision about accepting or rejecting a client's credit. The judgemental techniques rely on the knowledge and both past and present experience of credit analysts, who evaluate the required requisites, such as the personal reputation of a client, the ability to repay credit, guarantees and client's character. Due to the rapid increase in

fund-size invested through credit granted by Egyptian banks, and the need for quantifying credit risk, financial institutions including banks have started to apply credit-scoring models. Abdou, Etal (2009).

The present study is concerned with evaluating credit risk in banks using credit-scoring models. Statistical techniques are used: maximum likelihood for one can use linear models and for, one can use Type II Tobit model, a Monte Carlo simulation study is employed, under non-ignorable missing data. The credit scoring task is performed on one bank's personal loans data-set. The results show that Tobit type-II model is more fitted than linear models.

In this paper, a simulation study to examine the behaviour of the suggested methods: using linear model in case of ignoring missing and Type II Tobit model in case of non-ignoring missing data. Results of the Monte Carlo experiments show strange behavior that has never been reported before for the Type II Tobit MLE. A real life data is also presented.

### Type II Tobit Model Estimation

The models considered in this paper, classified as Type 2 Tobit models by Amemiya (1984), have the following structure:

$$Y_{1i} = X_{1i}\beta + \sigma\varepsilon_{1i} \quad (1)$$

Where  $(\varepsilon_{1i}, \varepsilon_{2i})$  is bivariate standard normal with correlation  $\rho_\epsilon$ . The first equation is a regression equation and the second a selection equation. In a typical economic application, the regression equation is a pricing or expenditure function, and the selection equation is a decision function that governs the occurrence of the transaction. Only qualitative information is available for the dependent variable in the selection equation,  $Y_{2i}$ . This is recorded as a binary variable,  $J_i$ , that takes the value one when  $Y_{2i}$  is positive. In addition, the dependent variable in the regression equation,  $Y_{1i}$ , is observed only when  $Y_{2i}$  is positive. The regressors,  $X_{1i}$  and  $X_{2i}$ , are observed regardless of  $J_i$ .

The log-likelihood function for this model is

$$\ln L(\delta, \beta, \sigma, \rho_\epsilon) = \sum_{i=1}^n \{J_i[-\ln(\sigma) + \ln\phi(Z_i) + \ln\phi(W_i)] + (1 - J_i)\ln[1 - \phi(X_{2i}, \delta)]\} \quad (3)$$

Where  $Z_i = (Y_{1i} - X_{1i}\beta)/\sigma$ ,  $W_i = (X_{2i}\delta + \rho_\epsilon Z_i)/\sqrt{1 - \rho_\epsilon^2}$ , and where  $\rho_\epsilon$  is restricted to the open interval  $(-1,1)$ . This likelihood function is highly nonlinear, and a solution to the score equations is obtained by numerical methods. Unfortunately, the log-likelihood function is not globally concave. Gradient methods may converge to a local maximum likelihood estimator (MLE). One can only be assure of obtaining a global

MLE, assuming one exists, if the estimation processes is start in the neighborhood of the global maximum.

The two-stage method of Heckman (1976) and Lee (1976) is typically use to obtain starting values for numerical solution of the score equations. The small sample performance of this estimator can be erratic, particularly when the same regresses used in both equations. Zuehlke and Zeman (1991) show that under these conditions, the mean square error performance of the subsample OLS estimator of  $\beta$  is often superior to that of the Heckman-Lee estimator. Moreover, it is uncommon for the estimate of  $\rho_\epsilon$  to exceed one in absolute value. In an attempt to circumvent these problems, some authors have added quadratic terms to one or both equations. While the Heckman-Lee estimator is consistent, its use as starting values is not sufficient to insure convergence to a global MLE. There is a solution to this problem, however. Olsen (1982) shows that the log-likelihood function of the Type II Tobit model is globally concave conditional on  $\rho_\epsilon$ , He suggests that a grid search over the bounded parameter  $\rho_\epsilon$ , in conjunction with the corresponding conditional MLEs, may be used to trace the profile of the maximized value of  $\ln L(\delta, \beta, \sigma, \rho_\epsilon)$  over the space of  $\rho_\epsilon$ . The location of any local or global maxima is determined, and a simultaneous estimation procedure started in the neighborhood of the global maximum. Unfortunately, this algorithm is not available in current econometric software.

Olsen (1982) observes that with the Type II Tobit model the likelihood function is often flat with a local maximum (emphasis added) near  $\rho = 0$ . This raises a question about the practice, common in empirical work, of estimating a sample selection model as a robustness check for OLS estimates. In cases with multiple roots, tests based on the global root might lead to a different conclusion than tests based on the local root are not.

Now assume that the aim is to estimate parameters in a parametric model. Usually, this can be derived from the Maximum Likelihood (ML) method. As suggested by its name, this method obtains estimators by maximizing a likelihood function.

A model for latent variable  $y^*$ , which is only partially observed:

$$y^* = \beta_0 + \beta_1 x_i + \epsilon_i . \quad \epsilon_i \sim N(0, \sigma^2) \quad \dots(4)$$

The Likelihood function,  $L$ , for e the whole sample is:

$$L(\beta_0, \beta_1, \sigma) = \prod_{i=1}^n L_i = \prod_{i=1}^n \left[ \frac{1}{\sigma} \varphi \left( \frac{y_i - \beta_0 - \beta_1 x_i}{\sigma} \right) \right]^{D_i} \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right]^{1-D_i} \dots (5)$$

The values of  $\beta_0$ ,  $\beta_1$  and  $\sigma$  that maximize the likelihood function are the Tobit estimators of the parameters. As usual the  $\ln(L)$  is:

$$\ln l = \sum_{i=1}^n D_i \ln \left[ \frac{1}{\sigma} \varphi \left( \frac{y_i - \beta_0 - \beta_1 x_i}{\sigma} \right) \right] + (1 - D_i) \ln \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right]$$

$$= \frac{N}{2} [\ln(\sigma^2) + \ln(2\pi)] + \sum_{i=1}^n D_i \left[ -\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} + (1 - D_i) \ln \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right] \right] \dots (6)$$

The first –order partial derivatives of  $l$  with respect to  $\beta_0$  and  $\beta_1$  and equating them to zero are as follows:

$$\frac{\partial l}{\partial \beta_0} = \sum_{i=1}^n D_i \left[ \frac{(y_i - \beta_0 - \beta_1 x_i)}{\sigma^2} + \frac{(1 - D_i)}{\sigma \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right]} \right] = 0 \dots\dots\dots (7)$$

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^n D_i \left[ \frac{x_i (y_i - \beta_0 - \beta_1 x_i)}{\sigma^2} + \frac{(1 - D_i)(-x_i)}{\sigma \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right]} \right] = 0 \dots\dots\dots (8)$$

The normal equations (7) and (8) do not have explicit solution and they have to be solved numerically.

**Fisher information matrix**

The elements of the Fisher information matrix are obtained by taking the negative expectation of the second derivatives of the natural logarithm of the likelihood function with respect to  $\Theta$ .

Amemiya (1985) presents the following representation for the information matrix:

$$I(\Theta) = \begin{bmatrix} \sum_{i=1}^T a_i x_i \dot{x}_i & \sum_{i=1}^T b_i \dot{x}_i \\ \sum_{i=1}^T b_i \dot{x}_i & \sum_{i=1}^T c_i \end{bmatrix} \dots\dots\dots (9)$$

Where

$$z_i = \frac{\dot{x}_i \beta}{\sigma} \cdot a_i = \frac{-1}{\sigma^2} \left[ z_i f(z_i) - \frac{f(z_i)^2}{1 - F(z_i)} - F(z_i) \right] \cdot b_i$$

$$= \frac{1}{2\sigma^3} \left[ z_i^2 f(z_i) + f(z_i) - \frac{f(z_i)^2}{1 - F(z_i)} \right]$$

$$c_i = \frac{1}{4\sigma^4} \left[ z_i^3 f(z_i) + z_i f(z_i) - \frac{z_i^2 f(z_i)^2}{1 - F(z_i)} - 2F(z_i) \right]$$

The elements of the Fisher information matrix are obtained by taking the negative expectation of the second derivatives of the natural logarithm of the likelihood function as follows:

$$\frac{\partial l}{\partial \beta_0^2} = \sum_{i=1}^n D_i \left[ \frac{-1}{\sigma^2} - \frac{(1 - D_i)}{\sigma^2 \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right]^2} \right] \dots\dots\dots (10)$$

$$\frac{\partial l}{\partial \beta_0 \beta_1} = \sum_{i=1}^n D_i \left[ \frac{-x_i}{\sigma^2} - \frac{(1 - D_i)x_i}{\sigma^2 \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right]^2} \right] \dots\dots\dots (11)$$

$$\frac{\partial l}{\partial \beta_1^2} = \sum_{i=1}^n D_i \left[ \frac{-x_i^2}{\sigma^2} - \frac{(1 - D_i)(x_i)^2}{\sigma^2 \left[ 1 - \varphi \left( \frac{\beta_0 + \beta_1 x_i}{\sigma} \right) \right]^2} \right] \dots\dots\dots (12)$$

Under particular regularity conditions, the two-sided  $100(1 - \alpha)\%$ ,  $0 < \alpha < 1$ , asymptotic CIs (Asy-CIs) for the vector of unknown parameters  $\Theta$  can be obtained.

### Monte Carlo Results

The purpose of the Monte Carlo portion of this study is to analyze the performance of estimation methods, including MLE for one can use linear models and for, one can use Type II Tobit model, a Monte Carlo simulation study is employed, under non-ignorable missing data. For MLE, 1000 observation, number of replications, is generated from normal distribution for random error and covariate of independent variable  $X$  are generating from uniform distribution. The following assumptions are hold for Monte-Carlo simulation:

Coefficients  $(\beta_0, \beta_1)$  assumed to be:

$$(\beta_0 = -1, \beta_1 = 1) \text{ And } (\beta_0 = -0.5, \beta_1 = 0.5)$$

Random error of the proposed model is generated from normal distribution with mean zero and standard deviation  $(\sigma) 1$ .

Sample sizes of  $n = 25, 50, 100, 200, 500$  and  $1000$ .

Steps for simulation:

Step 1: Generate independent variable ( $X$ ) and error part ( $U$ ) as follows:

$$\begin{aligned} x_i &\sim U(0,2) \quad .i = 1,2, \dots, n \\ u_i &\sim N(0,1) \quad .i = 1,2, \dots, n \end{aligned}$$

Step 2: Compute dependent variable  $Y$  as follows:

$$y_i = \beta'x_i + u_i$$

Step 3: Convert dependent variable ( $Y$ ) to Tobit II model variable ( $Y^*$ ) as follows:

$$y_i^* = \begin{cases} y_i & \text{if } y_i \geq 0 \\ \text{ignorable} & \text{if } y_i < 0 \end{cases}$$

and define indicator variable  $d_i$  as:

$$d_i = \begin{cases} 1 & \text{if } y_i \geq 0 \\ 0 & \text{if } y_i < 0 \end{cases}$$

Step 4: Find estimates of  $\beta_0, \beta_1$  and  $\sigma$  from:

Traditional model:  $Y \sim \beta_0 + \beta_1 X$  as linear regression model

By solving likelihood equations of Tobit II model (3.7) and (3.8) to obtain maximum likelihood estimates of  $\beta_0, \beta_1$  and  $\sigma$  which denoted by:  $\widehat{\beta}_0, \widehat{\beta}_1$ , and  $\widehat{\sigma}$ .

Measures of AIC and BIC also computed as:

$$\begin{aligned} AIC(\widehat{\theta}) &= -2\log\text{likelihood}(\widehat{\theta}) + 2q \\ BIC(\widehat{\theta}) &= -2\log\text{likelihood}(\widehat{\theta}) + q \log n \end{aligned}$$

where  $q$  is the number of parameters and  $n$  is the proposed sample size.

Step 5: Repeat step 1 to step 4 number of times  $B = 1000$ .

Step 6: Compute the following statistical measures:

Mean square error (MSE)

$$MSE(\hat{\theta}) = \frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i - \theta)^2 \quad \dots\dots\dots (13)$$

Relative biases (RBias)

$$RBias(\hat{\theta}) = \frac{1}{B} \sum_{i=1}^B \frac{|\hat{\theta}_i - \theta|}{\theta} \times 100 \quad \dots\dots\dots (14)$$

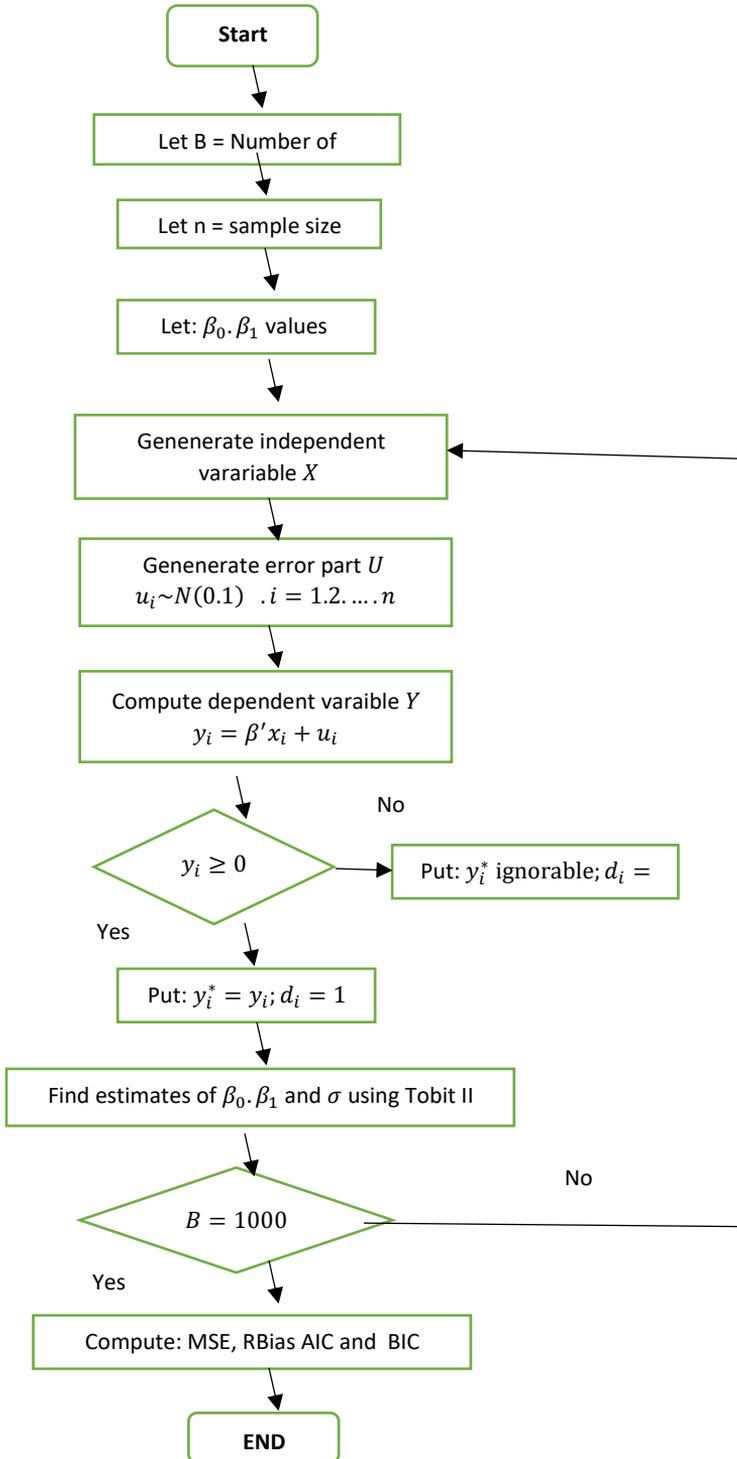
Based on generated data and assumed two cases for  $\beta_1$ , all statistical measures are computed and reported in Table 3.1 for the initial parameter of  $\beta_1 = 1$  and Table 3.2 for the initial parameter of  $\beta_1 = 0.5$ .

From the tabulated results, one can indicate that: With increasing in sample size  $n$ , MSEs and Rbiases are decreasing for all parameters in two different models AIC and BIC are decreasing in two models.

In comparison with two different proposed models namely; linear (traditional) models and Tobit type-II models, one can indicate that:

MSEs in linear models is smaller than MSEs in Tobit type-II models.

AIC and BIC in linear model is greater than in Tobit type-II models which indicate that Tobit type-II model is more fitted than linear models.



**Table (1): Estimated values, MSEs, RBias (in %)and model criteria of the linear model and Tobit Type-II model for different sample size  $n$  and initial parameters:  $\beta_0 = -1$ ,  $\beta_1 = 1$ , and  $\sigma = 1$**

$n$	Parm →	Linear Model			Tobit Type-II Model		
		$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\sigma}$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\sigma}$
25	MSE	0.16725	0.12277	0.02303	0.36533	0.20101	0.04845
	RBias	1.57	0.96	4.68	0.63	0.02	5.32
	AIC	73.96026			53.63765		
	BIC	77.61688			57.29428		
50	MSE	0.08383	0.06734	0.01130	0.17190	0.10168	0.02678
	RBias	0.64	0.02	3.03	1.63	1.17	2.87
	AIC	144.26289			103.84047		
	BIC	149.99896			109.57654		
100	MSE	0.04247	0.03079	0.00508	0.08215	0.04698	0.01238
	RBias	0.77	0.65	1.41	1.39	0.82	1.36
	AIC	286.45275			204.75444		
	BIC	294.26826			212.56995		
200	MSE	0.01943	0.01573	0.00255	0.03826	0.02327	0.00618
	RBias	0.41	0.38	0.57	1.14	0.74	0.33
	AIC	570.77785			406.50665		
	BIC	580.67281			416.40161		
500	MSE	0.00748	0.00585	0.00109	0.01481	0.00887	0.00270
	RBias	0.16	0.01	0.29	0.39	0.18	0.39
	AIC	1421.45016			1012.75682		
	BIC	1434.09399			1025.40065		
1000	MSE	0.00407	0.00292	0.00051	0.00780	0.00431	0.00129
	RBias	0.04	0.18	0.17	0.49	0.45	0.09
	AIC	2839.95861			2022.42687		
	BIC	2854.68187			2037.15013		

**Table (2): Estimated values, MSEs, RBias (in %) and model criteria of the linear model and Tobit Type-II model for different sample size  $n$  and initial parameters:  $\beta_0 = -0.5$ ,  $\beta_1 = 0.5$ , and  $\sigma = 1$**

$n$	Parm →	Linear Model			Tobit Type-II Model		
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}$
25	MSE	0.16050	0.12106	0.02382	0.29825	0.18340	0.05621
	RBias	0.26	0.67	5.76	5.18	1.25	5.99
	AIC	73.38586			54.58556		
	BIC	77.04249			58.24218		
50	MSE	0.08745	0.06276	0.00970	0.13281	0.08102	0.02441
	RBias	0.26	0.56	2.57	3.83	1.08	2.30
	AIC	144.81639			106.98088		
	BIC	150.55246			112.71695		
100	MSE	0.03877	0.02847	0.00516	0.06300	0.03811	0.01245
	RBias	0.08	0.25	1.26	2.42	1.19	0.69
	AIC	286.74406			211.91490		
	BIC	294.55957			219.73041		
200	MSE	0.01924	0.01468	0.00260	0.03134	0.01902	0.00670
	RBias	2.49	2.18	0.78	2.70	2.20	0.89
	AIC	569.90682			419.07701		
	BIC	579.80178			428.97196		
500	MSE	0.00818	0.00614	0.00098	0.01265	0.00783	0.00240
	RBias	0.10	0.16	0.30	0.92	0.46	0.15
	AIC	1421.46692			1045.79543		
	BIC	1434.11075			1058.43926		
1000	MSE	0.00400	0.00295	0.00053	0.00630	0.00398	0.00128
	RBias	0.21	0.29	0.19	0.51	0.40	0.11
	AIC	2839.56523			2088.65979		
	BIC	2854.28850			2103.38306		

### Applications

A private datasets with different characteristics was employed in the process of empirical model evaluation. The data is studied from two way, one for independent variable and second for four independent variables. This data set is related to the loan completion process for customers details provided while filling out the online application form. These details are gender, marital statues, education, number of dependents, income, loan amount, credit history and others. The data set was taken from the online website (<http://www.Kaggle.com>).

**Case I: One independent variable**

Define variables of the proposed model from dataset:

**Dependent variable:** Co-applicant income ( $y_i$ ), thus,

$$y_i^* = \begin{cases} y_i & \text{if } y_i > 0 \\ \text{ignorable} & \text{if } y_i = 0 \end{cases}$$

and define indicatro variable  $d_i$  as:

$$d_i = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{if } y_i = 0 \end{cases}$$

Where.  $i = 1.2. \dots 614$ . Thus, we have 273 ignorable observations ( $y_i = 0$ ) and 341 un-censored observations ( $y_i^* = y_i[y_i > 0]$ ).

**Independent variable:** Applicant income ( $x$ )

In Table (3), maximum likelihood estimates of  $\beta_0, \beta_1, \sigma$  are obtained from the given real data set for two different models, (Linear and Tobit type-II). Note that, in linear model we have an estimate of the standard error ( $\hat{\sigma}$ ) but in Tobit type-II model we have an estimate for parameter  $\sigma$ . From tabulated values of real data set, we notice that the measures of fitting (AIC and BIC) in Tobit type-II model is less than those values in linear model which indicate that the proposed Tobit type-II model is better than linear models.

**Table (3): Estimated values, standard errors (St.Er), and model criteria of the linear model and Tobit Type-II model for given real data set of loan prediction: Case I.**

Model	Parameter	Estimate	St.Er	AIC	BIC
Linear Model	$\widehat{\beta}_0$	1923.0502	156.7689	11540.30	11553.56
	$\widehat{\beta}_1$	-0.0559	0.0192		
	$\hat{\sigma}$	2908.66	----		
Tobit Type-II Model	$\widehat{\beta}_0$	1570.7325	325.8952	7047.868	7061.128
	$\widehat{\beta}_1$	-0.3139	0.0576		
	$\hat{\sigma}$	4394.941	0.0410		

**Case II: Four independent variables**

Define variables of the proposed model from dataset:

**Dependent variable:** same as in case I.

**Independent variables:**

- Applicant income ( $x_1$ )
- Loan amount in thousands ( $x_2$ )
- Term of loan in months ( $x_3$ )
- Credit History ( $x_4$ )

In Table (4), maximum likelihood estimates of coefficients:  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \sigma$  are obtained from the given real data set for two different models, (Linear and Tobit type-II). Note that, in linear model we have an estimate of the standard error ( $\hat{\sigma}$ ) but in Tobit type-II model we have an estimate for parameter  $\sigma$ . From tabulated values of real data set, we notice that the measures of fitting (AIC and BIC) in Tobit type-II model is less than those values in linear model which indicate that the proposed Tobit type-II model is better than linear models.

**Table (4): Estimated values, standard errors (St.Er) and model criteria of the linear model and Tobit Type-II model for given real data set of loan prediction: Case II.**

Model	Parameter	Estimate	St.Er	AIC	BIC
Linear Model	$\widehat{\beta}_0$	1187.7616	640.2544	9748.179	9773.805
	$\widehat{\beta}_1$	-0.1261	0.0200		
	$\widehat{\beta}_2$	10.2707	1.5238		
	$\widehat{\beta}_3$	-1.0995	1.6259		
	$\widehat{\beta}_4$	-84.896	294.3358		
	$\hat{\sigma}$	2412.087	---		
Tobit Type-II Model	$\widehat{\beta}_0$	-0.9765	1054.4152	5893.004	5918.63
	$\widehat{\beta}_1$	-0.5111	0.0611		
	$\widehat{\beta}_2$	22.7578	2.8029		
	$\widehat{\beta}_3$	-1.8414	2.7001		
	$\widehat{\beta}_4$	113.1238	486.3130		
	$\hat{\sigma}$	3569.4791	0.0444		

## Conclusions

In this paper Type II Tobit (sample selection) model studied statistically point of view depending on maximum likelyhood. A Monte Carlo simulation study is introduced to examine the behaviour of the suggested methods: using linear model in case of ignoring missing and Type II Tobit model in case of non-ignoring missing data. Results show that, strange behavior that has never been reported before for the Type II Tobit MLE. In addition, a real data set is studied from two way, one for independent variable and second for four independent variables. The results show that the measures of fitting (AIC and BIC) in Tobit type-II model is less than those values in linear model which indicate that the proposed Tobit type-II model is better than linear models.

## References

1. **Abdou, H. A. H. (2009)**. Credit scoring models for Egyptian banks: Neural nets and genetic programming versus conventional techniques (Ph.D. Thesis). The University of Plymouth, UK.
2. **Abbod, M. F and Ala'raj, M. (2016)**. Classifiers consensus system approach for credit scoring. *Knowledge-Based Systems*, 104 , 89–105
3. **Amemiya, T., & AMEMIYA, T. A. (1985)**. *Advanced econometrics*. Harvard university press.
4. **Bailey, M. 2001**. *Credit scoring: the principles and practicalities*. Kingswood, Bristol: White Box Publishing.
5. **Chuang, C., Lin, R. 2009**. Constructing a reassigning credit scoring model. *Expert Systems with Applications* 36 (2/1): 1685-1694.
6. **Gup, B. E., Kolari, J. W. 2005**. *Commercial Banking: The management of risk*. Alabama: John Wiley & Sons, Inc.
7. **Harris, T. (2015)**. Credit scoring using the clustered support vector machine. *Expert Systems with Applications*, 42 , 741–750
8. **Lewis, E. M. 1992**. *An Introduction to Credit Scoring*. California: Fair, Isaac & Co., Inc.
9. **Kozodoi, Nikita, et al.** "Shallow self-learning for reject inference in credit scoring." *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part III*. Springer International Publishing, 2020.
10. **Malhotra, R., Malhotra, D. K. 2003**. Evaluating consumer loans using Neural Networks. *Omega the International Journal of Management Science* 31 (2): 83-96.
11. **Mays, E. 2001**. *Handbook of Credit Scoring*. Chicago: Glenlake Publishing Company, Ltd.
12. Olsen, Randall J., "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator," *International Economic Review*, 1982, 23, 223–240.
13. **Owen, A. B. (2001)**. *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton.
14. **Siddiqi, N. 2006**. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. New Jersey: John Wiley & Sons, Inc.

15. **Sustersic, M., Mramor, D., Zupan J. 2009.** Consumer credit scoring models with limited data. *Expert Systems with Applications* 36 (3): 4736-4744.
16. **Thomas, L. C., Edelman, D. B., Crook, J. N. 2004.** *Readings in Credit Scoring: recent developments, advances, and aims.* New York: Oxford University Press
17. **Thompson, Paul (1998).** "Bank lending and the environment: policies and opportunities." *International Journal of Bank Marketing* 16.6 : 243-252.

## A Five State Progression Model: Estimation and Application

**Ayat Ahmed\*, Nahed Helmy\*, Mervat Elgohary\***

\*Al- Azhar University Faculty of Commerce – Girls’ Branch Department of Statistics.

Dependence, like occurrence of disease changing the risk of death. It can also model paired data. It is useful for recurrent events, but has limitations.

تاريخ استلام البحث: 2023/05/14

تاريخ قبول البحوث: 2023/06/22

نشر البحث في العدد الواحد والعشرين: ديسمبر / كانون اول 2023

2522-64X/519.2

رمز التصنيف ديوي / النسخة الالكترونية (Online):

2519-948X/519.2

رمز التصنيف ديوي / النسخة الورقية (Print):

## A Five State Progression Model: Estimation and Application

**Ayat Ahmed\*, Nahed Helmy\*, Mervat Elgohary\***

\*Al- Azhar University Faculty of Commerce – Girls' Branch Department of Statistics. dependence, like occurrence of disease changing the risk of death. It can also model paired data. It is useful for recurrent events, but has limitations.

### Abstract

Multi-state models (MSMs) are an extension of classical survival analysis which allows adjustment to the prediction of survival duration of the patient in the course of time by incorporating new information regarding the progression of the medical history and to better understand how prognostic factors influence the different phases of the disease/recovery process. In recent years, a wide range of medical situations have been modelled using MSMs such as problems following lung transplantation, problems following heart transplantation, hepatic cancer, HIV infection and AIDS. Disease progression model is needed for understanding the progression of disease and important in retrospective cohort analyses. In these paper five states progression model suggested. Transition intensities of the suggested model estimated by maximum likelihood method. The likelihood function constructed using transition probabilities, therefore they are derived first. Confidence intervals of estimated parameters obtained. Some extractor quantities of the model derived. An application using dataset-containing histories of bronchiolitis obliterans syndrome (BOS) from lung transplant recipients applied using the suggested model. The BOS data set provided in the R package msm.

**Keywords:** Markov processes, Interval-censored, Staged progression model, Transition rates, Transition probabilities.

### المستخلص

النماذج متعددة الحالات (MSMs) هي امتداد لتحليل البقاء على قيد الحياة الكلاسيكي الذي يسمح بالتكيف مع التنبؤ بمدة بقاء المريض على قيد الحياة بمرور الوقت من خلال دمج معلومات جديدة تتعلق بتطور التاريخ الطبي وفهم أفضل لكيفية تأثير العوامل النذير على المراحل المختلفة للمرض / عملية الشفاء. في السنوات الأخيرة، تم نمذجة مجموعة واسعة من الحالات الطبية باستخدام MSMs مثل المشاكل التالية لزراعة الرئة، والمشاكل التالية لزراعة القلب، وسرطان الكبد، وعدوى فيروس نقص المناعة البشرية والإيدز. هناك حاجة إلى نموذج تطور المرض لفهم تطور المرض ومهم في التحليلات الأترابية بأثر رجعي. في هذه الورقة يقترح نموذج تقدم خمس حالات. يتم تقدير شدة الانتقال للنموذج المقترح بطريقة الاحتمال الأقصى. يتم إنشاء دالة الاحتمال باستخدام احتمالات الانتقال، وبالتالي يتم اشتقاقها أولاً. فترات الثقة من الأسمية المقطرة عن طريق طريقة الاحتمال الأقصى. يتم إنشاء دالة الاحتمال باستخدام احتمالات الانتقال، وبالتالي يتم اشتقاقها أولاً. يتم الحصول على فترات الثقة للمعلومات المقطرة. يتم اشتقاق بعض كميات المستخرج من النموذج. يتم تطبيق تطبيق باستخدام مجموعة بيانات تحتوي على تاريخ متلازمة التهاب القصيبات (BOS) من متلقي زراعة الرئة باستخدام النموذج المقترح. يتم توفير مجموعة بيانات BOS في حزمة R msm.

الكلمات المفتاحية: عمليات ماركوف، الفاصل الزمني الخاضع للرقابة، نموذج التقدم المرحلي، معدلات الانتقال، احتمالات الانتقال

## Introduction

In epidemiology and medical statistics, the models used to describe health-related processes over time, where states usually defined by the severity of a disease or condition of subjects in a follow up study. Standard survival analysis can analyze separately different endpoints but fails to give insight into what happens to a patient after a first event.

Multi-state models are an extremely flexible approach that can model almost any kind of longitudinal failure time data. This is particularly relevant for modeling different events, which have an event-related

Disease progression models help in understanding the expected burden of the disease. Moreover, the results of studying these models will be useful for public health policymakers to implement the policies and treatment plans to improve the survival of the patients.

Kalbfleisch and Lawless (1985) introduced the MSMs. They developed a general procedure for obtaining maximum likelihood estimates of the model parameters. Kay (1986) proposed a similar method that calculates the first and second order derivatives of some particular multi-state processes and provided methods for hypothesis testing and model diagnostics. Satten and Longini (1996) developed a method for fitting these models when states are subject to measurement errors. Commenge (1999) discussed some assumptions for multi-state models in epidemiology and considered different inference approaches. Andersen and Keiding (2002) discussed event history analysis through multi-state models and introduced statistical model specification via transition intensities and likelihood inference. Machado *et al* (2009) introduced modelling approaches for multi-state models, and focused on the estimation of quantities such as the transition probabilities and survival probabilities. In addition, they discussed the existing software available to fit the various models. Hinchliffe *et al* (2013) presented a parametric survival model for transition hazard rates and provided some programs that used to estimate transition hazards and probabilities within an illness-death model framework using flexible parametric survival model. Then, they generated confidence intervals and transition hazard functions for Rotler dam Lambert breast cancer data. Machado *et al* (2018) discussed flexible modelling of time-dependency using penalised maximum likelihood to estimate the models. They proposed an automatic and efficient method to estimate multistate models with splines in the presence of interval censoring. Machado *et al* (2021) presented a new and efficient method to estimate multi-state models with splines using automatic estimation of penalty parameters. They showed that using splines with penalty parameters can improve model fit. Lintu *et al.* (2022) estimated covariate effects on the bidirectional transition rates for a continuous time homogeneous multi-state Markov model with three transient states, and an absorbing state (death) for kidney disease progression.

In this paper, staged disease progression model with four transient states and one absorbing state is proposed. The suggested model studied in the case of continuous time homogeneous multistate model based on Markov processes. The parameters of the suggested model estimated by ML method. The likelihood function constructed using transition probabilities, therefore they are derived first. The probability of survival for patients in different states of disease for some time in the future is predicted.

This paper is organized as follows: Section (2) introduces multi-state models and discusses the framework of Markov processes. In section (3) inference of multi-state models is presented. Section (4) discusses some extractor quantities. Section (5) is devoted to the suggested model and the estimation of its parameters. In section (6) an application using BOS data set is applied to represent the performance of the suggested model. Section (7) contains conclusions.

### Multi-state models

MSMs are the most commonly used models for describing the development for longitudinal data. MSMs are models for a stochastic process, which at any time point occupies one of a set of discrete states. In medicine, the states can describe conditions like healthy, diseased, diseased with complications, and death. A change of state is call a transition. This corresponds to outbreak of disease, occurrence of complications and death.

MSMs can be illustrated using diagrams with boxes representing the states and with arrows between the states representing the possible transitions. A state is call an absorbing state if transitions cannot occur from that state. A transient state meaning that at least one transition is possible from that state. The complexity of a MSM depends on the number of states and on the possible transitions.

MSMs for interval-censored data are commonly formulated in a Markov processes framework [Kalbfleisch and Lawless (1985)]. The Markov property states that the future of the process only depends on the current state. There are two types of Markov process in the literature; one is discrete-time Markov process, and the continuous-time Markov process. This paper is concerned with continuous-time multi-state processes.

Continuous-time Markov process expresses the condition that the state space,  $S$ , is discrete and the time  $T$  is continuous. Given the time points  $t_1, t_2, \dots, t_n$ , it is of interest to examine the joint distribution  $Y_1, Y_2, \dots, Y_n$ , where  $Y_j = Y(t_j)$  for  $j = 1, 2, \dots, n$ . Commonly,  $(Y(t) | t \in T)$  is assume a Markov process, which means that the future state of the process only depends on the current state. Thus, a continuous-time Markov process on the discrete states  $D$  is defined through a set of probabilities,  $p_{rs}(t)$ , such that,

$$p_{rs}(t, u) = p(Y(u + t) = s | Y(u) = r) \text{ for } u \geq 0 \ t \geq 0, \quad (2.1)$$

which represents probability of being in state  $s$  at a specified time  $u + t$  in the future. [Van den Hout (2017).

In applications, models are specified through the transition rates (transition intensities),  $q_{rs}(t)$ , which represent the rate of transition to state  $s$  for someone currently in state  $r$  over a small time interval. In time-homogeneous multi-state models all transition intensities,  $q_{rs}$ , are assumed to be constant as functions of time and specified as follows:

$$q_{rs}(t) = q_{rs} = \lim_{\Delta t \rightarrow 0} \frac{p(Y(t + \Delta t) = s | Y(t) = r)}{\Delta t}, r \neq s, \quad (2.2)$$

where  $q_{rs}(t)$  is the instantaneous risk of moving from state  $r$  to state  $s$

Transition intensity matrix (the generator matrix),  $\mathbf{Q}$ , given by:

$$\mathbf{Q}(t) = \begin{bmatrix} q_{11} = -\sum_{s \neq 1} q_{1s} & q_{12} & q_{13} & \dots & q_{1n} \\ q_{21} & q_{22} = -\sum_{s \neq 2} q_{2s} & q_{23} & \dots & q_{2n} \\ \vdots & & q_{32} & \ddots & \dots & q_{3n} \end{bmatrix}, \quad (2.3)$$

Where  $\mathbf{Q}(t)$ , is a matrix with off-diagonal entries  $q_{rs}$  and diagonal entries  $q_{rr} = -\sum_{s \neq r} q_{rs}$ . If  $q_{rr} = 0$  the state  $r$  is called absorbing.

The generator matrix satisfies:

- $q_{rs} \geq 0$  for  $r \neq s$
- $\sum_s q_{rs} = 0$

The Markov process  $(Y(t) | t \in T)$  is time homogeneous if the probability (2.1) only depends on the initial state as follows:

$$p_{rs}(t) = p(Y(t) = s | Y(0) = r), \quad (2.4)$$

The probabilities in (2.4) satisfy:

$$1) \quad 0 \leq p_{rs}(t) \leq 1, \quad (2.5)$$

$$2) \quad p_{rk}(t) = \sum_s p_{rs}(u) p_{rk}(t - u), \quad t > u \quad (2.6)$$

$$3) \quad \sum_s p_{rs}(t) = 1, \quad (2.7)$$

Eq. (2.6) is the Chapman-Kolmogorov equation for a time homogeneous Markov process. Then, the matrix  $\mathbf{P}(t)$  which contains these probabilities is called the *transition probability matrix*,  $\mathbf{P}(t)$ , and is defined as

$$\mathbf{P}(t) = \mathbf{P}(u)\mathbf{P}(t - u) \quad \text{with } \mathbf{p}(0) = \mathbf{I}, \quad (2.8)$$

For a given generator matrix,  $\mathbf{Q}$ , a Markov process is defined. The link between a generator matrix and its probability matrix is established by the forward and backward equations as follows:

$$\dot{\mathbf{P}}(t) = \mathbf{P}(t) \mathbf{Q}, \quad (2.9)$$

$$\dot{\mathbf{P}}(t) = \mathbf{Q} \mathbf{P}(t), \quad (2.10)$$

Given the initial condition  $\mathbf{P}(0) = \mathbf{I}$ , the unique solution of both forward and backward equations in (2.9) and (2.10) is

$$\mathbf{P}(t) = \exp(\mathbf{Q}t) = \sum_{k=0}^{\infty} \frac{t^k \mathbf{Q}^k}{k!}, \quad (2.11)$$

Where  $\exp$  is the matrix exponential. [cox and Miller (2017)].

Using the eigen-decomposition of  $\mathbf{Q}$ , the transition probability matrix  $\mathbf{P}(t)$  is computed. If  $\mathbf{Q}$  has an eigen-decomposition, then the matrix exponential can

be expressed in a simple form. Let  $\mathbf{B}$  be a diagonal matrix of eigenvalues of  $\mathbf{Q}$  and let  $\mathbf{A}$  be a corresponding matrix of eigenvectors, so that  $\mathbf{Q} = \mathbf{A}\mathbf{B}\mathbf{A}^{-1}$  and  $\mathbf{Q}t = \mathbf{A}\mathbf{B}t\mathbf{A}^{-1}$ .

It follows that

$$\exp(\mathbf{Q}t) = \mathbf{A} \exp(\mathbf{B}t) \mathbf{A}^{-1}, \quad (2.12)$$

This is much simpler, since  $\exp(\mathbf{B}t)$  is a diagonal matrix whose  $(i, i)$ th element is simply  $e^{b_{ii}t}$ .

Thus, the formulas for  $p_{rs}(t)$  in terms of  $q_{rs}$  can be derived by carrying out the following steps:

1. Form  $\mathbf{Q}$  with algebraic elements of  $q_{rs}$ .
2. Derive formulas for the elements of  $\mathbf{B}$  by solving the characteristic equation  $\det(\mathbf{Q} - b\mathbf{I}) = 0$ .
3. Derive formulas for the elements of  $\mathbf{A}$  by solving  $\mathbf{Q}\mathbf{x} = b\mathbf{x}$ , where  $\mathbf{x}$  is an  $n \times 1$  vector.
4. Derive formulas for the elements of  $\mathbf{A}^{-1}$  by any of several standard methods.
5. Derive formulas for the elements of  $\mathbf{P}(t) = \mathbf{A} \exp(\mathbf{B}t)\mathbf{A}^{-1}$  by using the rules of matrix multiplication. [Jones *et al.* (2017)].

The transition probability matrix  $\mathbf{P}(t)$  can be used for predicting the probability of survival for patients in all states of disease, for some time  $t$  in the future by plotting the expected probability of survival against time, from each transient state.

## Inference of Multi-State Models:

### Maximum likelihood estimation

The likelihood function is formed with the transition probabilities. Suppose that each of a random sample of  $n$  individuals is observed at times  $t_0, t_1, \dots, t_m$ . If  $n_{rsl}$  denotes the number of individuals in state  $r$  at  $t_{l-1}$ , and  $s$  at  $t_l$ , then the likelihood function for  $\theta$  is given by:

$$L(\theta) = \prod_{l=1}^m \left\{ \prod_{r,s=1}^D p_{rs}(t_{l-1}, t_l)^{n_{rsl}} \right\}, \quad (3.1)$$

In the time-homogeneous  $w = t_l - t_{l-1}$ ,  $l = (1, \dots, m)$  the likelihood (3.1) can be written as

$$L(\theta) = \prod_{r,s=1}^D p_{rs}(w)^{n_{rs}} \quad (3.2)$$

where  $n_{rs} = \sum_{l=1}^m n_{rsl}$  is the total number of recorded transitions from  $r$  to  $s$ . [Kalbfleisch and Lawless (1985)]

Since, for each  $r$ , the likelihood (3.2) is of multinomial form,

$$L_r = \frac{n_r!}{\prod_{s=1}^D n_{rs}!} \prod_{s=1}^D p_{rs}(w)^{n_{rs}}, \quad s = 1, 2, \dots, D \quad (3.3)$$

Then, the likelihood function for the entire group is

$$L(\boldsymbol{\theta}) = \prod_{r=1}^D L_r, \quad (3.4)$$

The natural logarithm of the likelihood function is of the form:

$$\ln L(\boldsymbol{\theta}) = \sum_{rs} n_{rs} \ln p_{rs}(w) \quad (3.5)$$

The maximum likelihood equations are:

$$U_i(\boldsymbol{\theta}) = \frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_i} = \sum_{rs} \frac{n_{rs}}{p_{rs}(w)} \frac{\partial p_{rs}(w)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, d \quad (3.6)$$

The maximum likelihood estimates of  $U_i(\boldsymbol{\theta})$  are derived using numerical methods.

### Confidence Intervals for the Parameters

The ML estimators of the parameters are asymptotically normal, asymptotically unbiased and have asymptotic variance–covariance matrix given by the inverse of the Fisher information matrix. The elements of the fisher information matrix are obtained by taking the negative expectation of the second derivatives of the natural logarithm of the likelihood function,  $L(\boldsymbol{\theta})$ , with respect to  $\boldsymbol{\theta}$  as follows:

$$I(\boldsymbol{\theta}) = -E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_v} \right), \quad (3.7)$$

Then the second derivatives of the natural logarithm of the likelihood function is given by:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_v} = \sum_{r,s} n_{rs} \times \left\{ \frac{\frac{\partial^2 p_{rs}(w)}{\partial \theta_i \partial \theta_v}}{p_{rs}(w)} - \frac{\frac{\partial p_{rs}(w)}{\partial \theta_i} \frac{\partial p_{rs}(w)}{\partial \theta_v}}{p_{rs}^2(w)} \right\}, \quad (3.8)$$

Unfortunately, the exact mathematical expressions for the above expectations are very difficult to obtain.

Therefore, the asymptotic Fisher information matrix is

$$\hat{I}_{ij}(\boldsymbol{\theta}) = - \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_v} \right) = \sum_{r,s} \frac{n_{rs}}{p_{rs}(w)} \frac{\partial p_{rs}(w)}{\partial \theta_i} \frac{\partial p_{rs}(w)}{\partial \theta_v}, \quad (3.9)$$

which is obtained by dropping the expectation operator  $E$  and using the fact that  $\frac{\partial^2 p_{rs}(w)}{\partial \theta_i \partial \theta_v} = 0$ .

The approximate asymptotic variance –covariance matrix for the ML estimators is the inverse of asymptotic Fisher information matrix. It is useful for computing the standard error of ML estimation.

$$I(\boldsymbol{\theta})^{-1} = \frac{1}{|I(\boldsymbol{\theta})|} \text{adj } I(\boldsymbol{\theta}) \quad (3.10)$$

where  $\text{adj } I(\boldsymbol{\theta})$  is the ad joint of  $I(\boldsymbol{\theta})$

For large sample size, the ML estimators under regularity conditions are consistent and asymptotically unbiased as well as asymptotically normally

distributed; hence the asymptotic two sided confidence intervals of the parameters for the ML estimators are obtained by:

$$p \left[ -z < \frac{\hat{\theta}_{iML} - \theta_i}{\sigma_{\hat{\theta}_{iML}}} < z \right] = 1 - \alpha, \tag{3.11}$$

where  $z$  is the  $100(1 - \frac{\tau}{2})$  th standard normal percentile. The two sided approximate  $100(1 - \frac{\tau}{2})100\%$  confidence intervals are given by:

$$L_{\theta} = \hat{\theta}_{iML} - Z_{\frac{\tau}{2}} \hat{\sigma}_{\hat{\theta}_{iML}}, \quad \text{and} \quad U_{\theta} = \hat{\theta}_{iML} + Z_{\frac{\tau}{2}} \hat{\sigma}_{\hat{\theta}_{iML}}, \quad i = 1, 2, \dots, b \tag{3.12}$$

where  $\hat{\sigma}_{\hat{\theta}_{iML}}$  is the standard deviation of the parameters  $\hat{\theta}_i$  and  $\hat{\theta}_{iML}$  is the ML estimates of the parameters  $\hat{\theta}_i$ .

**Some extractor quantities**

A set of quantities can be used to extract interesting features of the fitted multi-state model including expected duration in each state and the population size of states.

**Expected duration time in each state:**

It represents the average duration that an individual is expected to stay in each state with a time period of length  $t$  depending on the initial state. The expected duration of stay in state  $s$  between times  $t_0$  and  $t$  for an individual in state  $r$  at time  $t_0$ , is defined by the integral from  $t_0$  to  $t$  of the  $r, s$  entry of the transition probability matrix,  $\mathbf{P}(t)$ .

For person in state  $r$  at current time  $t_0$ , let

$e_{rs}(t)$  is the expected duration of stay in state  $s$  in the interval  $(t_0, t)$ ,  $s = 1, 2, \dots, D$

Let an individual be in state  $r$  at time  $t_0$  and, for each  $u$ ,  $t_0 \leq u \leq t$ , define indicator function  $I_{rs}(u)$  so that

$$I_{rs}(u) = \begin{cases} 1 & \text{if the individual is in state } r \text{ at time } t \\ 0 & \text{otherwise} \end{cases}, \tag{4.1}$$

With the expectation

$$E[I_{rs}(u)] = p_{rs}(u) \tag{4.2}$$

Then,

$$e_{rs}(t) = E \int_{t_0}^t I_{rs}(u) du, \tag{4.3}$$

Interchanging the expectation and the integral sign gives

$$e_{rs}(t) = \int_{t_0}^t p_{rs}(u) du, \quad t_0 \leq u \leq t, \quad r, s = 1, 2, \dots, D \tag{4.4}$$

The sum of the expected durations of stay over all states is equal to the entire length of the interval,

$$e_{r1}(t) + e_{rs}(t) + \dots + e_{rs}(t) = t, \quad r = 1, 2, \dots, D \tag{4.5}$$

[ Chiang (1968)].

**The population size of states:**

It provides a description of population size in all states over a period of time. An individual in state  $S_r$  at time 0 must be either in the same state or move to

another state at time  $t$ ; consequently the corresponding transition probabilities is given by equation (2.7).

At time  $t = 0$ , let there be  $x_1(0)$  individuals in state  $S_1$  and  $x_2(0)$  individuals in state  $S_2$ , and so on  $x_r(0)$  individuals in states  $S_r$ , where  $r = 1, 2, \dots, D$ . Thus the sum  $x(0) = x_1(0) + x_2(0) + \dots + x_D(0)$  be the initial size of the population.

Suppose that the  $x(0)$  individuals travel independently from one state to another, and that at the end of the interval  $(0, t)$  individuals are in states  $S_1, S_2, \dots, S_D$ . Obviously

$$x(t) = X_1(t) + X_2(t) + \dots + X_D(t) \tag{4.6}$$

Each  $r$  of the random variables on the right side of (4.6) have a multinomial distribution with probability generating function given by:

$$E \left[ z_1^{X_{r1}(t)} z_2^{X_{r2}(t)} \dots z_D^{X_{rs}(t)} \mid x_r(0) \right] = [p_{r1}(t) z_1 + p_{r2}(t) z_2 + \dots + p_{rD}(t) z_D]^{x_r(0)} \tag{4.7}$$

Therefore, the probability generating function of the joint probability distribution for the population sizes of all the states at time  $t$  is

$$E \left[ z_1^{X_{r1}(t)} z_2^{X_{r2}(t)} \dots z_D^{X_{rs}(t)} \mid x_1(0), x_2(0), \dots, x_r(0) \right] = \prod_r [p_{r1}(t) z_1 + p_{r2}(t) z_2 + \dots + p_{rD}(t) z_D]^{x_r(0)} \tag{4.8}$$

and the joint probabilities are

$$\begin{aligned} p(X_1(t) = x_1, X_2(t) = x_2, \dots, X_s(t) = x_s \mid x_1(0), x_2(0), \dots, x_r(0)) \\ = \sum_{r_s} \prod_{r=1}^D \frac{x_r(0)!}{x_{r1}! x_{r2}! \dots x_{rD}!} p_{r1}(t)^{x_{r1}} p_{r2}(t)^{x_{r2}} \dots p_{rD}(t)^{x_{rD}} \quad r, s \\ = 1, 2, \dots, D \end{aligned} \tag{4.9}$$

The expected number of individuals in state  $S_\beta$  at time  $t$  is given by:

$$E[X_\beta(t) \mid x_1(0), x_2(0), \dots, x_r(0)] = x_1(0) p_{1\beta}(t) + x_2(0) p_{2\beta}(t) + \dots + x_r(0) p_{r\beta}(t) \tag{4.10}$$

Chiang (1968)].

### The Suggested Model

The model contains four transient states and one absorbing state. The four transient states are well state  $S_1$ , mild disease state  $S_2$ , moderate disease state  $S_3$  and the severe disease state  $S_4$ . An absorbing state is the death state  $S_5$ . The considered disease is so dangerous that recovery from it is not allowed. Also, there is no transitions between disease states. The possible transitions are shown in the following figure:

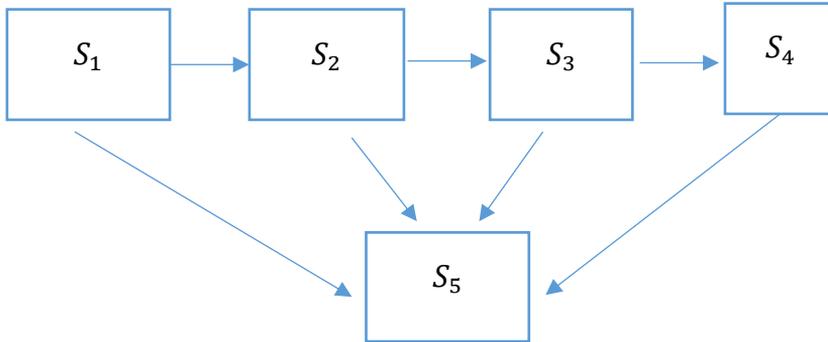


Figure 5.1: The proposed 5-state model.

The computation of the transition probability matrix,  $P(t)$ , following steps in section 2.4 is as follows:

The transition intensity matrix,  $Q$ , is as follows:

$$Q = \begin{bmatrix} q_{11} & q_{12} & 0 & 0 & q_{15} \\ 0 & q_{22} & q_{23} & 0 & q_{25} \\ 0 & 0 & q_{33} & q_{34} & q_{35} \\ 0 & 0 & 0 & q_{44} & q_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{5.1}$$

where  $q_{11} = -(q_{12} + q_{15})$ ,  $q_{22} = -(q_{23} + q_{25})$ ,  $q_{33} = -(q_{34} + q_{35})$ ,  $q_{44} = -q_{45}$ .

The matrix of eigenvalue,  $B$ , is as follows:

$$B = \begin{bmatrix} q_{11} & 0 & 0 & 0 & 0 \\ 0 & q_{22} & 0 & 0 & 0 \\ 0 & 0 & q_{33} & 0 & 0 \\ 0 & 0 & 0 & q_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{5.2}$$

The matrix of eigenvector,  $A$ , is as follows:

$$A = \begin{bmatrix} 1 & 1 & \frac{-q_{12}}{q_{11}-q_{33}} & \frac{-q_{12} q_{23}}{(q_{11}-q_{44})(q_{22}-q_{44})} & 1 \\ 0 & \frac{q_{11}-q_{22}}{-q_{12}} & 1 & \frac{-q_{23}}{q_{22}-q_{44}} & 1 \\ 0 & 0 & \frac{q_{22}-q_{33}}{-q_{23}} & 1 & 1 \\ 0 & 0 & 0 & \frac{q_{33}-q_{44}}{-q_{34}} & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5.3}$$

The inverse of the matrix,  $A$  is given by:

$$A^{-1} = \begin{bmatrix} 1 & \frac{-q_{12}}{q_{11}-q_{22}} & \frac{q_{12}q_{23}}{(q_{11}-q_{33})(q_{11}-q_{22})} & \frac{-f}{D} & \frac{h}{D} \\ 0 & \frac{-q_{12}}{q_{11}-q_{22}} & \frac{-q_{12}q_{23}}{(q_{11}-q_{22})(q_{22}-q_{33})} & \frac{-q_{12}q_{23}q_{34}}{(q_{22}-q_{44})(q_{11}-q_{22})(q_{22}-q_{33})} & \frac{-l}{D} \\ 0 & 0 & \frac{-q_{23}}{q_{22}-q_{33}} & \frac{-q_{23}q_{34}}{(q_{33}-q_{44})(q_{22}-q_{33})} & \frac{-L}{D} \\ 0 & 0 & 0 & \frac{-q_{34}}{q_{33}-q_{44}} & \frac{q_{34}}{q_{33}-q_{44}} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (5.4)$$

where

$$f = (q_{22} - q_{33}) \left[ \frac{(q_{22} - q_{44})(q_{11} - q_{44}) - (q_{11} - q_{33})(2q_{11} - q_{44} - q_{22})}{(q_{11} - q_{33})(q_{22} - q_{44})(q_{11} - q_{44})} \right]$$

$$D = \frac{(q_{11} - q_{22})(q_{22} - q_{33})(q_{33} - q_{44})}{-q_{12} q_{23} q_{34}}$$

$$h = (q_{22} - q_{33}) \left[ \frac{q_{34} + q_{33} - q_{44}}{q_{34}(q_{11} - q_{33})} + \frac{q_{12} + q_{11} - q_{22}}{q_{12}} - \frac{2 q_{11} - q_{44} - q_{22}}{(q_{22} - q_{44})(q_{11} - q_{44})} \right]$$

$$l = (q_{33} - q_{44}) \left[ \frac{q_{34}q_{23} + (q_{22} - q_{44})(q_{23} + q_{22} + q_{33})}{q_{34} q_{23} (q_{22} - q_{44})} \right]$$

$$L = \frac{(q_{11} - q_{22})(q_{34} + q_{33} + q_{44})}{-q_{12} q_{34}}$$

Then, the transition probability matrix is as follows:

$$P(t) = \begin{bmatrix} e^{q_{11}t} & M & N & R & T \\ 0 & e^{q_{22}t} & P & U & V \\ 0 & 0 & e^{q_{33}t} & W & Y \\ 0 & 0 & 0 & e^{q_{44}t} & 1 - e^{q_{44}t} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (5.5)$$

where

$$M = \frac{q_{12}(e^{q_{11}t} - e^{q_{22}t})}{q_{11} - q_{22}}$$

$$N = \frac{q_{12} q_{23} [(q_{22} - q_{33})e^{q_{11}t} - (q_{11} - q_{33})e^{q_{22}t} + (q_{11} - q_{22})e^{q_{33}t}]}{(q_{11} - q_{33})(q_{22} - q_{33})(q_{11} - q_{22})}$$

$$R = \left[ \frac{1}{q_{11}-q_{33}} - \frac{2q_{11}-q_{44}-q_{22}}{(q_{22}-q_{44})(q_{11}-q_{44})} \right] \left[ \frac{q_{12}q_{23}q_{34} e^{q_{11}t}}{(q_{11}-q_{22})(q_{33}-q_{44})} \right] - \frac{q_{12}q_{23}q_{34} e^{q_{22}t}}{(q_{11}-q_{22})(q_{22}-q_{33})(q_{22}-q_{44})} + \frac{q_{12}q_{23}q_{34} e^{q_{33}t}}{(q_{11}-q_{33})(q_{22}-q_{33})(q_{33}-q_{44})} + \frac{q_{12}q_{23}q_{34} e^{q_{44}t}}{(q_{33}-q_{44})(q_{11}-q_{44})(q_{22}-q_{44})}$$

$$T = \frac{1}{D} \left[ h e^{q_{11}t} - l e^{q_{22}t} - q_{12} \frac{L e^{q_{33}t}}{q_{11} - q_{33}} \right] - \frac{q_{12}q_{23}q_{34} e^{q_{44}t}}{(q_{11} - q_{44})(q_{22} - q_{44})(q_{33} - q_{44})} + 1$$

$$P = \frac{q_{23}}{(q_{22} - q_{33})} [e^{q_{22}t} - e^{q_{33}t}]$$

$$U = q_{23}q_{34} \left[ \frac{e^{q_{22}t}}{(q_{22} - q_{44})(q_{22} - q_{33})} - \frac{e^{q_{33}t}}{(q_{33} - q_{44})(q_{22} - q_{33})} - \frac{e^{q_{44}t}}{(q_{33} - q_{44})(q_{22} - q_{44})} \right]$$

$$V = \frac{1}{D} \left[ \frac{q_{11} - q_{22}}{q_{12}} I e^{q_{22}t} + L e^{q_{33}t} \right] - \frac{q_{23}q_{34} e^{q_{44}t}}{(q_{22} - q_{44})(q_{33} - q_{44})} + 1$$

$$W = \frac{q_{34}}{(q_{33} - q_{44})} [ e^{q_{33}t} - e^{q_{44}t} ]$$

$$Y = \frac{-L}{D} \left( \frac{(q_{22} - q_{33})}{q_{23}} \right) e^{q_{33}t} + \frac{q_{34}}{q_{33} - q_{44}} e^{q_{44}t} + 1$$

Using equation (3.3) to obtain the likelihood function for each row:

For the first row ( $r = 1$ )

$$L_{r=1} = \frac{n_1!}{n_{11}! n_{12}! n_{15}!} p_{11}(w)^{n_{11}} p_{12}(w)^{n_{12}} p_{15}(w)^{n_{15}}, \quad (5.6)$$

For the second row ( $r = 2$ )

$$L_{r=2} = \frac{n_2!}{n_{22}! n_{23}! n_{25}!} p_{22}(w)^{n_{22}} p_{23}(w)^{n_{23}} p_{25}(w)^{n_{25}}, \quad (5.7)$$

For the third row ( $r = 3$ )

$$L_{r=3} = \frac{n_3!}{n_{33}! n_{34}! n_{35}!} p_{33}(w)^{n_{33}} p_{34}(w)^{n_{34}} p_{35}(w)^{n_{35}}, \quad (5.8)$$

For the fourth row ( $r = 4$ )

$$L_{r=4} = \frac{n_4!}{n_{44}! n_{45}!} p_{44}(w)^{n_{44}} p_{45}(w)^{n_{45}}, \quad (5.9)$$

Then, the likelihood function for the entire group is:

$$(\theta) = \prod_{r=1}^4 L_r = L_{r=1} * L_{r=2} * L_{r=3} * L_{r=4}, \quad (5.10)$$

The natural logarithm of the entire group is:

$$\begin{aligned} \ln L(\theta) \propto & n_{11} \ln p_{11}(w) \\ & + n_{12} \ln p_{12}(w) \\ & + n_{15} \ln p_{15}(w) + n_{22} \ln p_{22}(w) + n_{23} \ln p_{23}(w) \\ & + n_{25} \ln p_{25}(w) \\ & + n_{33} \ln p_{33}(w) + n_{34} \ln p_{34}(w) + n_{35} \ln p_{35}(w) \\ & + n_{44} \ln p_{44}(w) \\ & + n_{45} \ln p_{45}(w), \end{aligned} \quad (5.11)$$

Substituting with probabilities in (5.11) , differentiating with respect to  $\theta$  and equating to zero yields the score equations. These equations don't have closed form solution. They can be solved numerically to obtain the estimator of the elements of the transition intensities matrix

$\hat{q}_{11}, \hat{q}_{12}, \hat{q}_{15}, \hat{q}_{22}, \hat{q}_{23}, \hat{q}_{25}, \hat{q}_{33}, \hat{q}_{34}, \hat{q}_{35}, \hat{q}_{44}$  and  $\hat{q}_{45}$  .

Using equation (4.4) the expected duration of stay in  $S_1, S_2, S_3, S_4$  and  $S_5$  in the interval  $(0, t)$  can be derived respectively as follows:

- For an individual in state  $S_1$  at time 0 are:

$$e_{11}(t) = \frac{e^{q_{11}t} - 1}{q_{11}}, \quad (5.12)$$

$$e_{12}(t) = \frac{q_{12} \left[ \frac{(e^{q_{11}t} - 1)}{q_{11}} - \frac{(e^{q_{22}t} - 1)}{q_{22}} \right]}{q_{11} - q_{22}}, \quad (5.13)$$

$$e_{13}(t) = \frac{q_{12} q_{23} \left[ (q_{22} - q_{33}) \frac{(e^{q_{11}t} - 1)}{q_{11}} - (q_{11} - q_{33}) \frac{(e^{q_{22}t} - 1)}{q_{22}} + (q_{11} - q_{22}) \frac{(e^{q_{33}t} - 1)}{q_{33}} \right]}{(q_{11} - q_{33})(q_{22} - q_{33})(q_{11} - q_{22})}, \quad (5.14)$$

$$e_{14}(t) = \left[ \frac{1}{q_{11} - q_{33}} - \frac{2q_{11} - q_{44} - q_{22}}{(q_{22} - q_{44})(q_{11} - q_{44})} \right] \left[ \frac{q_{12} q_{23} q_{34} \frac{(e^{q_{11}t} - 1)}{q_{11}}}{(q_{11} - q_{22})(q_{33} - q_{44})} \right] -$$

$$\frac{q_{12} q_{23} q_{34} \frac{(e^{q_{22}t} - 1)}{q_{22}}}{(q_{11} - q_{22})(q_{22} - q_{33})(q_{22} - q_{44})} + \frac{q_{12} q_{23} q_{34} \frac{(e^{q_{33}t} - 1)}{q_{33}}}{(q_{11} - q_{33})(q_{22} - q_{33})(q_{33} - q_{44})} +$$

$$\frac{q_{12} q_{23} q_{34} \frac{(e^{q_{44}t} - 1)}{q_{44}}}{(q_{33} - q_{44})(q_{11} - q_{44})(q_{22} - q_{44})}, \quad (5.15)$$

$$e_{15}(t) = \frac{1}{D} \left[ h \frac{(e^{q_{11}t} - 1)}{q_{11}} - \frac{I(e^{q_{22}t} - 1)}{q_{22}} - q_{12} L \frac{(e^{q_{33}t} - 1)}{q_{11} - q_{33}} \right]$$

$$- \frac{q_{12} q_{23} q_{34} \frac{(e^{q_{44}t} - 1)}{q_{44}}}{(q_{11} - q_{44})(q_{22} - q_{44})(q_{33} - q_{44})} + t \quad (5.16)$$

- For an individual in state  $S_2$  at time 0 are:

$$e_{22}(t) = \frac{e^{q_{22}t} - 1}{q_{22}}, \quad (5.17)$$

$$e_{23}(t) = \frac{q_{23}}{(q_{22} - q_{33})} \left[ \frac{(e^{q_{22}t} - 1)}{q_{22}} - \frac{(e^{q_{33}t} - 1)}{q_{33}} \right], \quad (5.18)$$

$$e_{24}(t) = q_{23} q_{34} \left[ \frac{e^{q_{22}t} - 1}{q_{22}(q_{22} - q_{44})(q_{22} - q_{33})} - \frac{e^{q_{33}t} - 1}{q_{33}(q_{33} - q_{44})(q_{22} - q_{33})} \right. \\ \left. - \frac{e^{q_{44}t} - 1}{q_{44}(q_{33} - q_{44})(q_{22} - q_{44})} \right], \quad (5.19)$$

$$e_{25}(t) = \frac{1}{D} \left[ \frac{(q_{11} - q_{22})(e^{q_{22}t} - 1)}{q_{12}q_{22}} I + \frac{L(e^{q_{33}t} - 1)}{q_{33}t} \right] - \frac{q_{23}q_{34}(e^{q_{44}t} - 1)}{q_{44}(q_{22} - q_{44})(q_{33} - q_{44})} + t, \tag{5.20}$$

- For an individual in state  $S_3$  at time 0 are:

$$e_{33}(t) = \frac{e^{q_{33}t} - 1}{q_{33}}, \tag{5.21}$$

$$e_{34}(t) = \frac{q_{34}}{(q_{33} - q_{44})} \left[ \frac{e^{q_{33}t} - 1}{q_{33}} - \frac{e^{q_{44}t} - 1}{q_{44}} \right], \tag{5.22}$$

$$e_{35}(t) = \frac{L}{D} \left( \frac{(q_{22} - q_{33})(1 - e^{q_{33}t})}{q_{23}q_{33}} \right) + \frac{q_{34}(e^{q_{44}t} - 1)}{q_{44}(q_{33} - q_{44})} e^{q_{44}t} + t, \tag{5.23}$$

For an individual in state  $S_4$  at time 0 are:

$$e_{44}(t) = \frac{e^{q_{44}t} - 1}{q_{44}}, \tag{5.24}$$

$$e_{45}(t) = t + \frac{1 - e^{q_{44}t}}{q_{44}}, \tag{5.25}$$

The distribution of individuals in all states at time  $t$  according to initial state at time 0 are summarized in Table (5.1) .

Table 5.1: The distribution of individuals in all states at time  $t$  according to initial state at time 0.

State at time $t$ \ State at time 0	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	Initial population Sizes
$S_1$	$X_{11}(t)$	$X_{12}(t)$	$X_{13}(t)$	$X_{14}(t)$	$X_{15}(t)$	$x_1(0)$
$S_2$		$X_{22}(t)$	$X_{23}(t)$	$X_{24}(t)$	$X_{25}(t)$	$x_2(0)$
$S_3$			$X_{33}(t)$	$X_{34}(t)$	$X_{35}(t)$	$x_3(0)$
$S_4$				$X_{44}(t)$	$X_{45}(t)$	$x_4(0)$
population sizes at time $t$	$X_1(t)$	$X_2(t)$	$X_3(t)$	$X_4(t)$	$X_5(t)$	$x(0)$

### Application

The suggested model is illustrated using an application of BOS data set. The dataset containing histories of bronchiolitis obliterans syndrome (BOS) from lung transplant recipients. BOS is a chronic decline in lung function, often

observed after lung transplantation. The BOS data set is provided in the R package `msm`.

### Data description

The data come from Papworth Hospital U.K and are available in the `msm` package. It contains 638 rows, grouped by patient, including histories of 204 patients. The first observation for each patient is defined to be in state 1 after six months from transplant. Subsequent observations denote the entry times into stages 2, 3, 4, representing mild, moderate and severe BOS respectively, and stage 5, representing death. BOS data includes `ptnum` representing patient identification number, `time` representing months after transplant and `state` representing BOS state entered at this time. The entry time of each patient into each stage of BOS was estimated by clinicians, based on their history of lung function measurements and acute rejection and infection episodes. BOS is only assumed to occur beyond six months after transplant. In the first six months the function of each patient's new lung is stable. Subsequently BOS is diagnosed by comparing the lung function against the baseline value.

### Results

A five-state continuous time homogeneous multi-state Markov model is fitted for BOS data as defined in Figure (5.1).

The state table for BOS data is summarized by the frequencies in Table (6.1).

Table 6.1: The state table for BOS data.

To state \ From state	1	2	3	4	5
1	72	103	0	0	29
2	0	15	77	0	11
3	0	0	9	52	16
4	0	0	0	9	41

Table 6.1 shows that there were 29 deaths from state 1, 11 deaths from state 2, 16 deaths from state 3 and 41 deaths from state 4. The number of transitions from stage 4 to stage 5 is highest in comparison to transition from other stages to stage 5. There are no observations of mild, moderate and severe BOS followed by an observation of no BOS (state 1).

The estimated transition probability matrix at  $t = 30$  months is given in Table (6.2).

Table 6.2 : The estimated transition probability matrix at  $t = 30$  months.

From state \ To state	1	2	3	4	5
1	0.4578	0.2248	0.0827	0.1293	0.1054
2	0	0.1762	0.1032	0.3529	0.3677
3	0	0	0.0119	0.3565	0.6316
4	0	0	0	0.3196	0.6803
5	0	0	0	0	1

Table 6.2 shows that

- A person in state 1, has a probability of 0.1054 of being dead after thirty months, a probability of 0.4578 being still in state 1, a probability of 0.2248 of being alive with mild BOS and probabilities of 0.0827, 0.1293 of being alive with moderate/ or severe BOS, respectively.
- A person in state 2, has a probability of 0.3677 of being dead after thirty months, a probability of 0.1762 being still in state 2 and probabilities of 0.1032, 0.3529 of being alive with moderate/ or severe BOS, respectively.
- A person in state 3, has a probability of 0.6316 of being dead after thirty months, a probability of 0.0119 being still in state 3 and a probability of 0.3565 of being alive with severe BOS.
- A person in state 4, has a probability of 0.6803 of being dead after thirty months and a probability of 0.3196 being still in state 4 .

The likelihood estimates of transition intensities (estimated hazards) between various stages, standard error and 95% confidence intervals are shown in Table (6.3).

Table 6.3: The likelihood estimates of transition intensities (estimated hazards), standard error and 95% confidence intervals.

Transition	Transition intensities (estimates)	SE	Confidence intervals		Length
			L	U	
$q_{12}$	0.0254	0.0027	0.0211	0.0306	0.0095
$q_{15}$	0.0006	0.0027	0.0000	0.0057	0.0057
$q_{23}$	0.0564	0.0087	0.0435	0.0733	0.0298
$q_{25}$	0.0014	0.3137	0.0000	0.6163	0.6163
$q_{34}$	0.1271	0.1561	0.0903	0.1789	0.0885
$q_{35}$	0.0206	0.0574	0.0046	0.0920	0.0874
$q_{45}$	0.0380	0.0059	0.0292	0.0495	0.0203

Table 6.3 shows that transition intensity rate of moving from one state to the next is greater than the transition intensity rate of moving to death state. It is

noticed that the SE of the transition intensities are small, the estimated transition intensities lie in 95% confidence intervals and the lengths of the confidence intervals are small.

The expected duration of stay in each state is given in Table (6.4).

Table 6.4: The expected duration of stay in each state.

Transient state Initial state	State 1	State 2	State 3	State 4
	State 1	38	17	6
State 2	-	17	7	22
State 3	-	-	7	22
State 4	-	-	-	26

Table 6.4 shows that

- a person in state 1 is forecasted to spend 38 months in state 1, 17 months in state 2, 6 months in state 3 and finally 22 months in state 4. These results show that a person is expected to spend more time in state 1 and state 4 compared to the time spent in other states.
- a person in state 2 is forecasted to spend 17 months in state 2, 7 months in state 3, and finally 22 months in state 4. These results show that a person is expected to spend more time in state 4 compared to the time spent in other states.
- a person in state 3 is forecasted to spend 7 months in state 3, 22 months in state 4.
- a person in state 4 is forecasted to spend 26 months in state 4.

The distribution of individuals in all states at time  $t = 30$  months according to initial state at time  $t = 6$  months is given in Table (6.5).

Table 6.5 : The distribution of individuals in all states at time  $t = 30$  months according to initial state at time  $t = 6$  months.

State at Time $t = 6$	State at Time $t = 30$					Initial population sizes
	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	
$S_1$	93	46	17	26	22	204

Table 6.5 shows the distribution of individuals in all states at time  $t = 30$  months according to initial state at time  $t = 6$  months. It is clear that State 1 have the highest number of individuals. Since BOS assumed to occur beyond six months after transplant, all individual start from state 1 at time  $t = 6$  months. A plot of the expected probability of survival function for each transient state given in Figure (6.1).

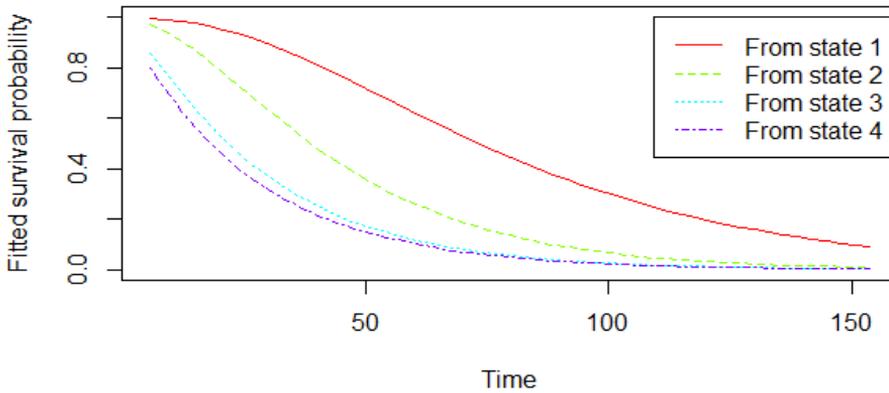


Figure 6.1 A plot of the expected probability of survival against time from each transient state.

Figure (6.1) shows that the survival probability curve of state 1 is the highest and the survival probability curve of state 2 is somewhat high which means that the probability of survival for individuals in state 1 and state 2 is high. The survival probability curves of state 3 and state 4 are lower than the survival probabilities in state 1 and 2 which reflects the rapid deterioration of individuals in state 3 and 4.

## Conclusions

The multi-state model has been used to understand the progression of several chronic diseases involving transitions across different intermediate states indicating the severity of the disease in continuous time. The model provides more insight into the complex event pattern and thus it can be used as an effective tool to study the effectiveness of treatments.

In this paper, five states progression model is suggested. The parameters of the suggested model are estimated by maximum likelihood method. The transition probabilities are derived. Confidence intervals of the estimated parameters are obtained. Some extractor quantities of the model are derived. The probability of survival for patients in different states of disease for some time  $t$  in the future is predicted.

The analysis in the current paper used the BOS data set provided in the R package `msm`. By analyzing this data, we found that transition intensity rate of moving from one state to the next is greater than the transition intensity rate of moving to death state. The probabilities of transition from state 3 to state 5 and from state 4 to state 5 are high. The expected duration of stay in state 1 is the highest compared to the expected duration of stay in other states. Survival probability of severe and moderate BOS decreases over time.

## References

1. **Andersen, P. K. and Keiding, N.** (2002). Multi-state Models for Event History Analysis. *Statistical Methods in Medical Research*, Vol. 11, pp. 91–115.
2. **Chiang, C. (1968).** *Introduction to Stochastic Processes in Biostatistics*. John Wiley & Sons, USA.
3. **Commenges, D.** (1999). Multi-state Models in Epidemiology. *Lifetime data analysis*, Vol. 5, pp. 315-327.
4. **Cox, D. and Miller H.** (2017). *The Theory of Stochastic Processes*. CRC/Chapman & Hall. USA.
5. **Hinchliffe, S., Sco, D. and Lambert, P.** (2013). Flexible Parametric Illness-Death models. *The Stata Journal*, Vol.13, No. 4, pp. 759-775.
6. **Jones, E. Epstein, D. and Garcí'a-Mocho'n, L.** (2017). A Procedure for Deriving Formulas to Convert Transition Rates to Probabilities for Multistate Markov Models. *Medical Decision Making*, Vol.37, pp.779–789.
7. **Kalbfleisch, J. and Lawless, J. F.** (1985). The Analysis of Panel Data under a Markov Assumption. *Journal of the American Statistical Association*, Vol. 80, pp. 863–871.
8. **Kay, R.** (1986). A Markov model for Analysing Cancer Markers and Disease States in Survival Studies. *Biometrics*, Vol. 80, pp.855–865.
9. **Lintu, M.K. Shreyas, K.M. and Kamath, A.** (2022). A Multi-State Model for Kidney Disease Progression. *Clinical Epidemiology and Global Health*, Vol. 13.
10. **Machado, L., Jacobo, D. and Andersen, P.** (2009). Multi-state Models for the Analysis of Time-to-Event Data. *statistical methods in medical research*, Vol.18, No. 2, pp.195-222.
11. **Machado, R. J. and van den Hout, A.** (2021). Flexible Multistate Models for Interval Censored Data: Specification, Estimation, and an Application to Ageing research. *Statistics in Medicine*, Vol. 37, No.10, pp.1636–1649.
12. **Satten, G. A. and Longini, I. M.** (1996). Markov Chains with Measurement Error: Estimating the 'True' Course of a Marker of the Progression of Human Immunodeficiency Virus Disease. *Applied Statistics*, Vol. 45, pp. 275–309.
13. **Van den Hout, A.** (2017). *Multi-state Survival Models for Interval-Censored Data*. CRC/Chapman & Hall.UK.

## Effect of COVID 19 on women and childcare in Iraq

Dr. Arjwan Marawan Shaaban 1, Dr.Eman A. Al-Tae 2, Dr.Raouf Tariq Raouf 3.

1. Family physician, Director of department of specialized centers and hospitals at MOH.
2. Family physician at Nineveh health directorate.
3. Director of the health statistical division at MOH.

تاريخ استلام البحث: 2023/05/29

تاريخ قبول البحث: 2023/06/22

نشر البحث في العدد الواحد والعشرين: ديسمبر / كانون اول 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/614.5

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/614.5

## Effect of COVID 19 on women and childcare in Iraq

Dr. Arjwan Marawan Shaaban 1, Dr.Eman A. Al-Taee 2, Dr.Raouf Tariq Raouf 3.

1. Family physician, Director of department of specialized centers and hospitals at MOH.
2. Family physician at Nineveh health directorate.
3. Director of the health statistical division at MOH.

### Abstract:

**Introduction:** Coronavirus disease (COVID-19) is a new mutation of coronavirus (SARS-CoV) that cause a pandemic in 2019. During COVID-19 pandemic, women and children were affected by the disruption of health services, specifically in developing countries. **Patients and methods:** The study was done in Iraq ,Descriptive study was done to identify the effect of Covid19 on women and child care in many aspects including: site of births, primary health care services and breastfeeding for three years (2019,2020,2021).

**Results:** The total births decreased in year 2020 compared to 2019, while in 2021 it is more than 2019, a decrease in births within governmental hospitals with the onset of the pandemic, and then return to the rise. Increased births in the private hospitals during the years 2020, 2021.

**Discussion:** Decrease of governmental hospital birth and antenatal care visits because the women afraid from getting infection with COVID-19 in health centers.

**Keywords:** COVID 19, antenatal care, breastfeeding, birth.

### مستخلص

مرض فيروس كورونا (COVID-19) هو طفرة جديدة لفيروس كورونا (SARS-CoV) تسبب جائحة في عام 2019. خلال جائحة كوفيد-19، تأثرت النساء والأطفال بتعطيل الخدمات الصحية، وتحديدًا في البلدان النامية. المرضى والأساليب: أجريت الدراسة في العراق، وتم إجراء دراسة وصفية لتحديد تأثير Covid19 على رعاية النساء والأطفال في العديد من الجوانب بما في ذلك: موقع الولادة وخدمات الرعاية الصحية الأولية والرضاعة الطبيعية لمدة ثلاث سنوات (2019، 2020، 2021). النتائج: انخفض إجمالي المواليد في عام 2020 مقارنة بعام 2019، بينما في عام 2021 كان أكثر من عام 2019، وانخفض عدد المواليد داخل المستشفيات الحكومية مع بداية الجائحة، ثم عاد إلى الارتفاع. زيادة الولادات في المستشفيات الخاصة خلال عامي 2020، 2021. المناقشة: انخفاض زيارات الولادة الحكومية والرعاية السابقة للولادة في المستشفيات لأن النساء يخشين الإصابة بـ COVID-19 في المراكز الصحية.

الكلمات المفتاحية: COVID 19، رعاية ما قبل الولادة، الرضاعة الطبيعية، الولادة.

**Introduction:**

Coronavirus disease (COVID-19) is a new mutation of coronavirus (SARS-CoV) that cause a pandemic in 2019, which overspread internationally. <sup>(1)</sup> According to WHO, during pandemics all supports directed to overcome the emergencies, so, the routine and basic health care services usually neglected. <sup>(2)</sup> During epidemics of Ebola at 2014, The maternal and neonatal mortality caused by Ebola virus itself is less than that caused by indirect effect of health care services destruction due to epidemic shift of services. <sup>(3)</sup> During COVID-19 pandemic, women and children were affected by the disruption of health services, specifically in developing countries. <sup>(4)</sup> Pregnant women and children in low resources countries are likely to face big impact during pandemics. <sup>(5)</sup> For example , In 2020, the maternal mortality ratio (MMR) was 211 per 100,000 live births while there were around 5.3 million under five deaths. <sup>(6)</sup> Restriction for visiting hospitals to patients relatives during pandemics ,and depending on tele-consultations, together with fear of infection, make confusion to pregnant ladies and patients about seeking health care from health centers or not. <sup>(7-9)</sup> in order to not overburden the hospitals, during epidemic many patients think that their health problems are not important enough to go to hospitals which are already overcrowded with patients. <sup>(10)</sup> Many countries adopt lockdown and low person-person contact due to unavailability of vaccine at the beginning and deficiency of treatment. <sup>(11-13)</sup> about breast-feeding and covid19, theoretically newborns do not have immunity against disease. <sup>(14)</sup> This put breastfeeding among infected ladies understudy during pandemic, <sup>(15)</sup> also the mother afraid from transmitting the virus to their babies. <sup>(16)</sup> Finally two reviews in 2020 favored continue lactation even if the mother is infected. <sup>(17,18)</sup>

**Patients and methods:**

**Study setting:** The study was done in Iraq, the data from the monthly reports sent from the primary health care centers, hospitals and the statistics of the Ministry of health and environment.

**Study design:** Descriptive study was done to identify the effect of Covid19 on women and child care in many aspects including: site of births, primary health care services and breastfeeding for three years (2019,2020,2021).

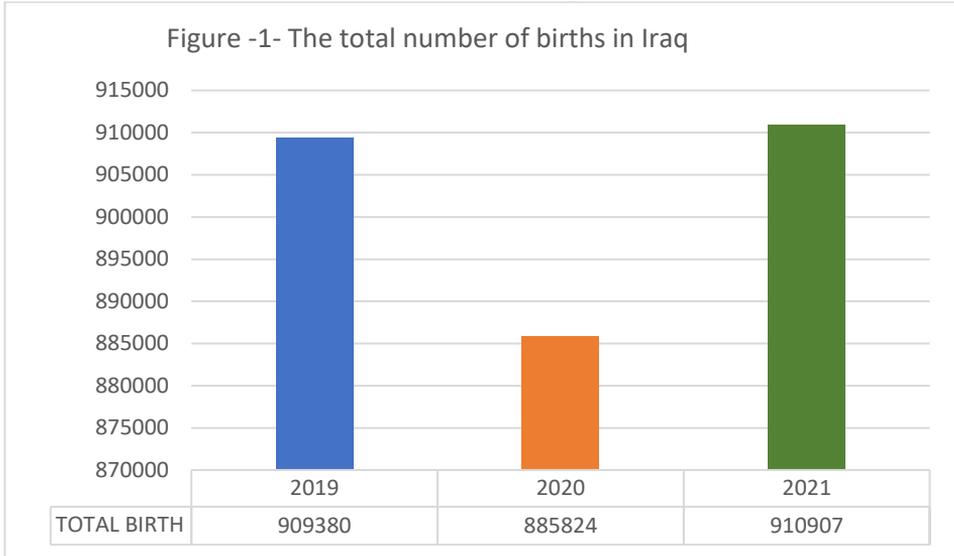
**Data analysis:** Computer feeding and statistical Analysis was done by Acer computer and Minitab (18th edition), many statistical diagrams was used to explain the relations.

**Results:**

**I. Site of births:**

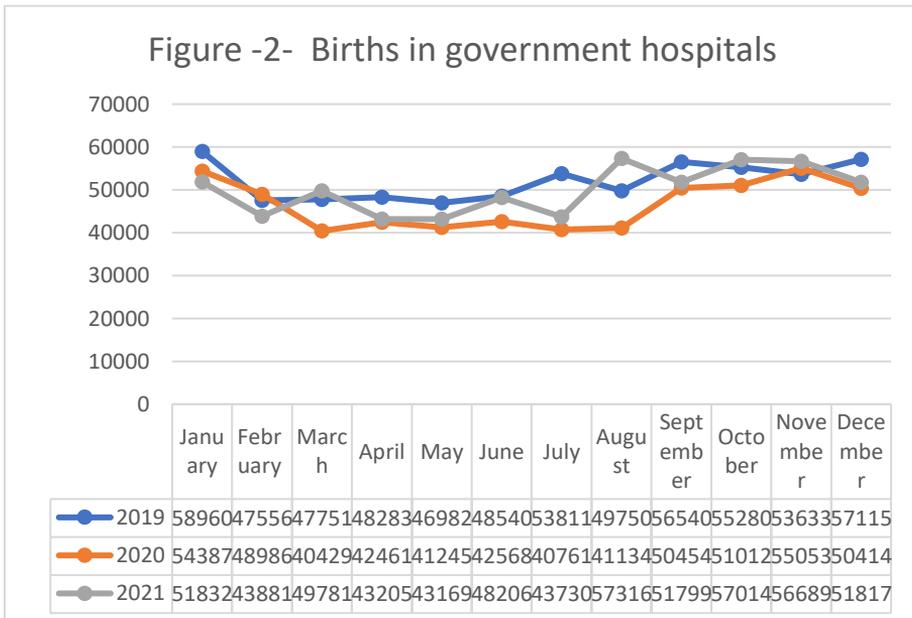
**I.1 Total births:**

The total births in Iraq decreased in year 2020 compared to 2019, while in 2021 it is more than 2019 as shown in figure -1-



**I.2 Births in governmental hospitals:**

we note a decrease in births within governmental hospitals with the onset of the pandemic (2020) and then return to the rise by 2021,



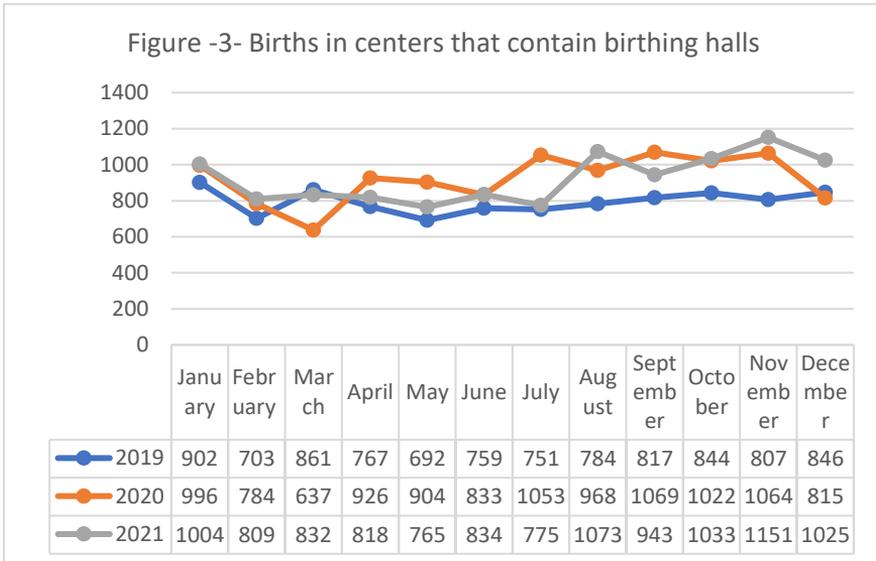
especially in the months after the first quarter to the end of the year as shown in figure -2 - with non-significant p-value(0.336).

**I.3 Births in centers that contain birthing halls:**

The results show fluctuating increase and decrease during the three years of the study as shown in figure -3- with non-significant p-value(0.0154).

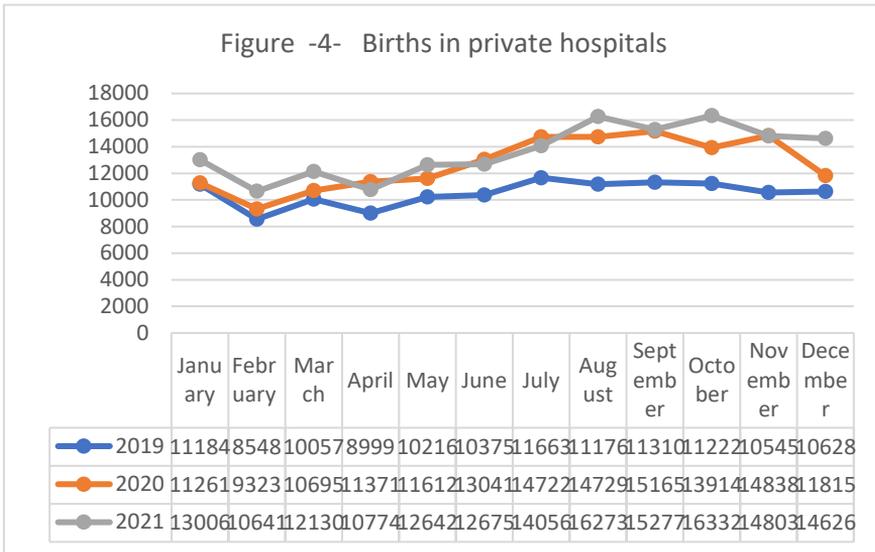
**I.4 Births in private hospitals:**

Results showed increased births in private hospitals during the years

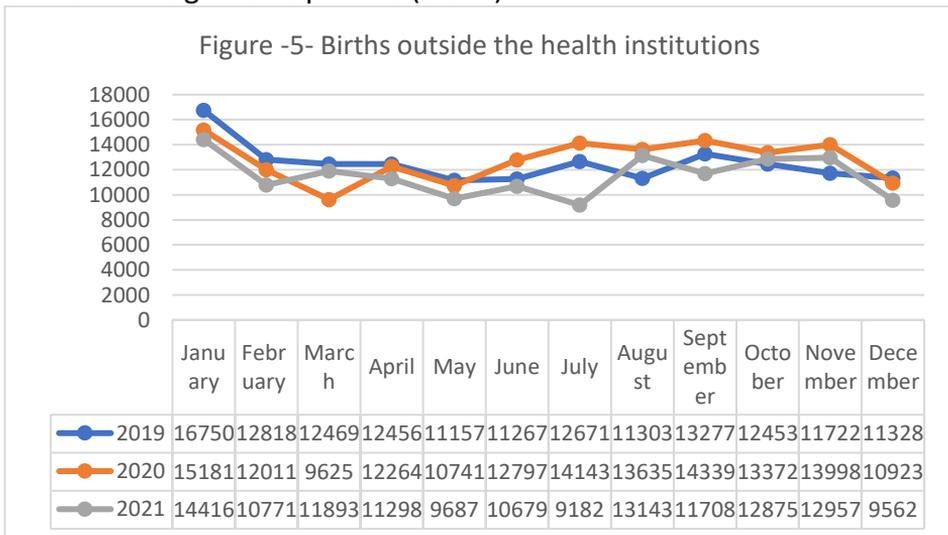


2020,2021 in comparison to year of 2019 as shown in figure -4- with highly significant p-value (0.007).

**I.5 Births outside the health institutions:**



Results shows increase of births outside the health institutions in second half of the years 2020, 2021 in comparison to 2019 as showed in figure - 5- with non-significant p-value (0.511).



**II. Primary health care services**

Comparisons were made between the first and fourth antenatal care visits for pregnant women in Iraq on monthly basis for the three years of the study. In addition, the results shows that the first and forth visits percentages were declined in 2020, 2021 in comparison to 2019 for the

months (January to April), while in months (May to August) there is increase of visits in comparison to first third of the year. From September to December, there is fluctuation in the percentage of visits (figures 6-17). In addition, the results shows that Anbar directorate recorded the lowest visits (first and fourth) in comparison to other health directorates.

Figure -6- Periodic medical examination for pregnant women and mothers for the month of January

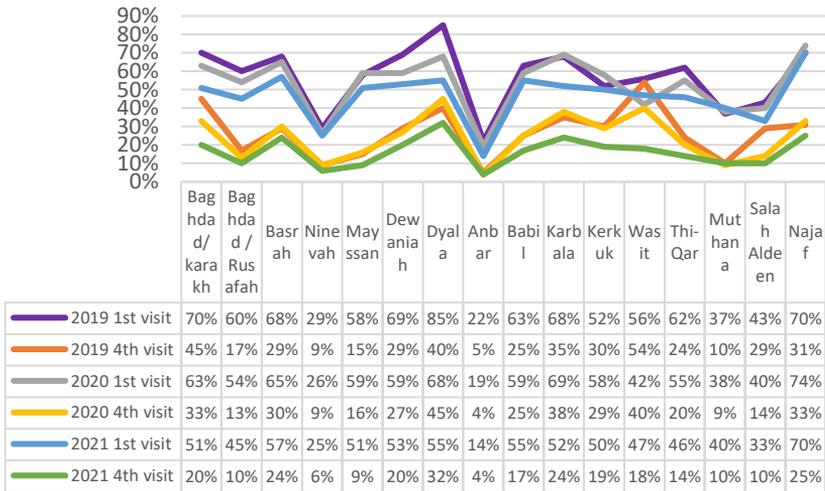


Figure -7- Periodic medical examination for pregnant women and mothers for the month of February

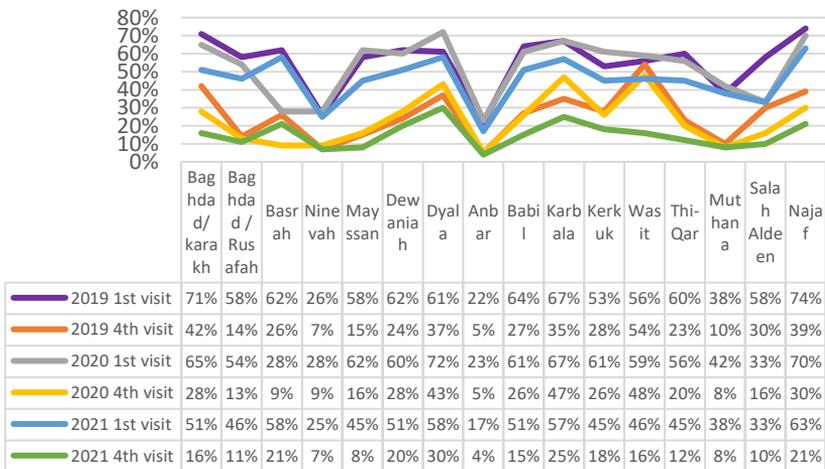


Figure -8- Periodic medical examination for pregnant women and mothers for the month of March

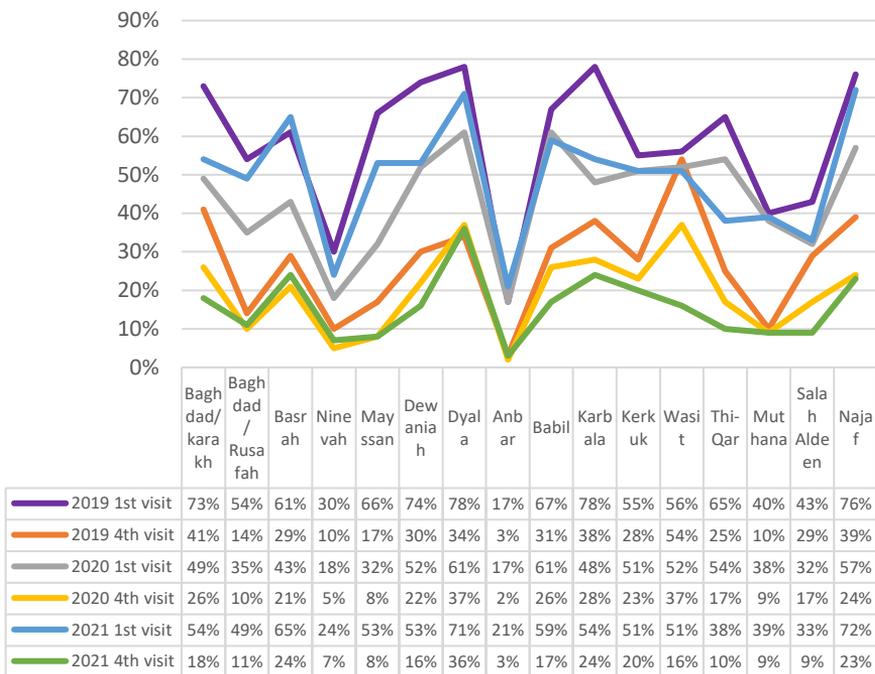


Figure -9- Periodic medical examination for pregnant women and mothers for the month of April

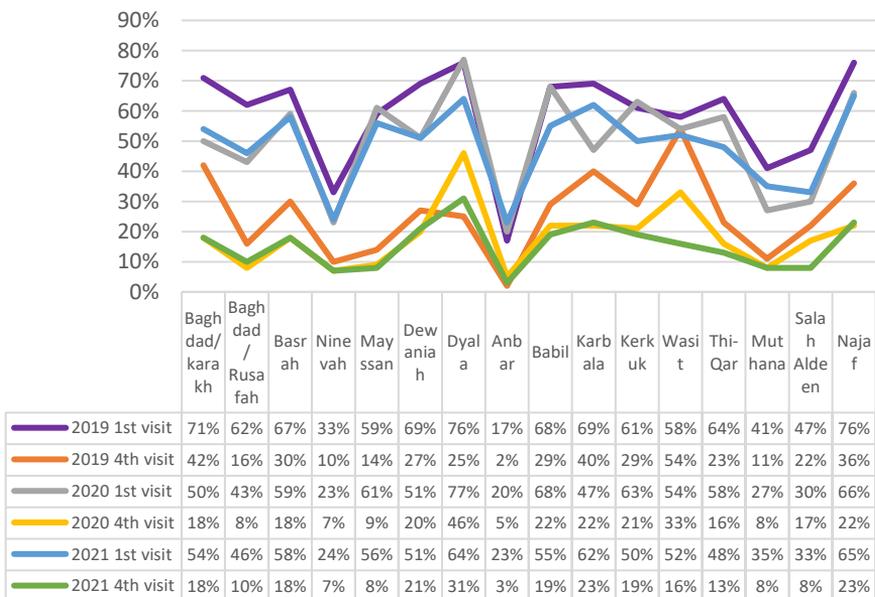


Figure -10- Periodic medical examination for pregnant women and mothers for the month of May

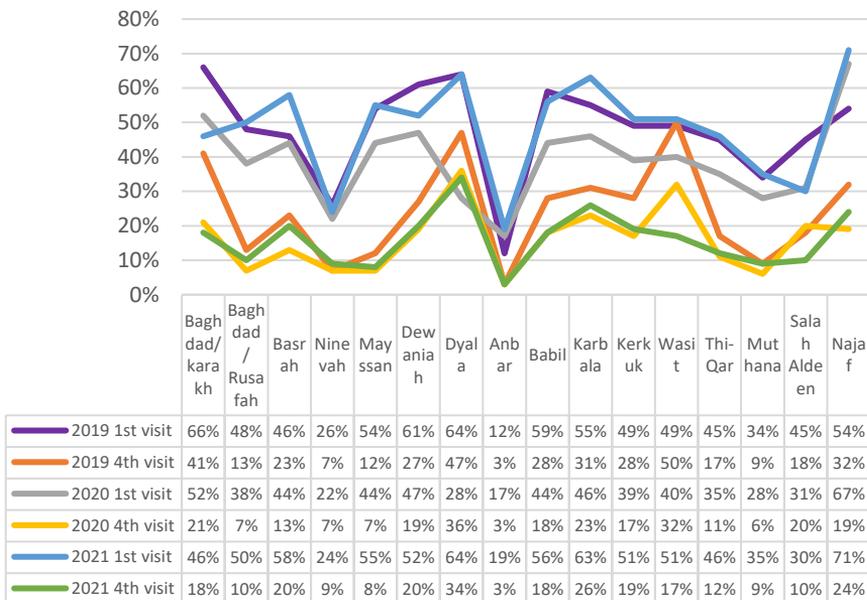


Figure -11- for the periodic medical examination of pregnant women and mothers for the month of June

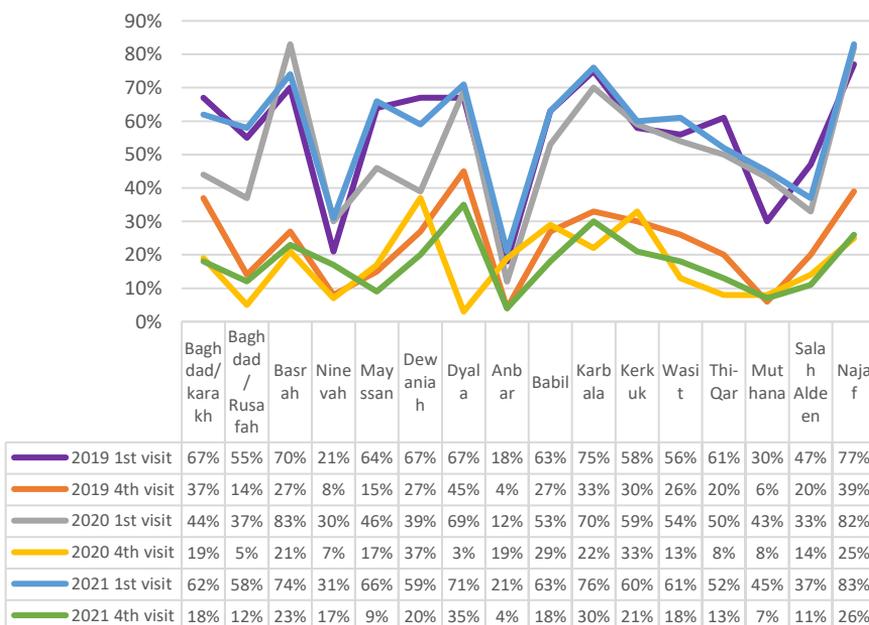
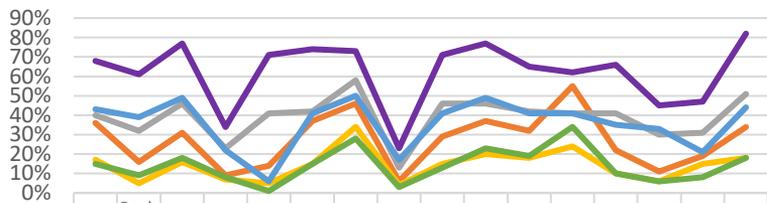
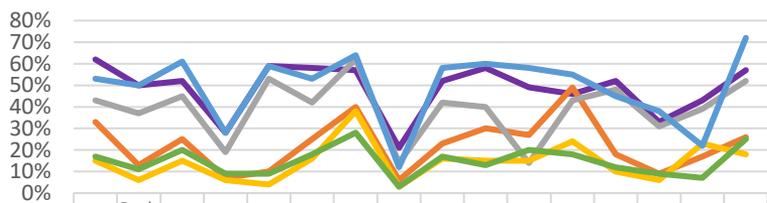


Figure -12- Periodic medical examination for pregnant women and mothers for the month of July



	Baghdad/karakh	Baghdad/Rusafah	Basrah	Ninevah	Mayssan	Dewaniyah	Dyalah	Anbar	Babil	Karbala	Kerkuk	Wasit	Thi-Qar	Mutahana	Salah Aldeen	Najaf
2019 1st visit	68%	61%	77%	34%	71%	74%	73%	23%	71%	77%	65%	62%	66%	45%	47%	82%
2019 4th visit	36%	16%	31%	9%	14%	37%	46%	6%	29%	37%	32%	55%	22%	11%	19%	34%
2020 1st visit	40%	32%	46%	23%	41%	42%	58%	13%	46%	46%	42%	41%	41%	30%	31%	51%
2020 4th visit	17%	5%	16%	7%	5%	15%	34%	4%	15%	20%	18%	24%	10%	6%	15%	18%
2021 1st visit	43%	39%	49%	22%	6%	41%	50%	17%	41%	49%	41%	41%	35%	33%	21%	44%
2021 4th visit	15%	9%	18%	8%	1%	15%	28%	3%	13%	23%	19%	34%	10%	6%	8%	18%

Figure -13- Periodic medical examination for pregnant women and mothers for the month of August



	Baghdad/karakh	Baghdad/Rusafah	Basrah	Ninevah	Mayssan	Dewaniyah	Dyalah	Anbar	Babil	Karbala	Kerkuk	Wasit	Thi-Qar	Mutahana	Salah Aldeen	Najaf
2019 1st visit	62%	50%	52%	28%	59%	58%	57%	21%	52%	58%	49%	46%	52%	33%	43%	57%
2019 4th visit	33%	13%	25%	7%	10%	25%	40%	6%	23%	30%	27%	49%	18%	9%	17%	26%
2020 1st visit	43%	37%	45%	19%	53%	42%	62%	14%	42%	40%	14%	43%	48%	31%	39%	52%
2020 4th visit	15%	6%	15%	6%	4%	16%	38%	3%	16%	15%	15%	24%	10%	6%	23%	18%
2021 1st visit	53%	50%	61%	28%	59%	53%	64%	12%	58%	60%	58%	55%	45%	38%	22%	72%
2021 4th visit	17%	11%	20%	9%	9%	18%	28%	3%	17%	13%	20%	18%	12%	9%	7%	25%

Figure -14- Periodic medical examination for pregnant women and mothers for the month of September

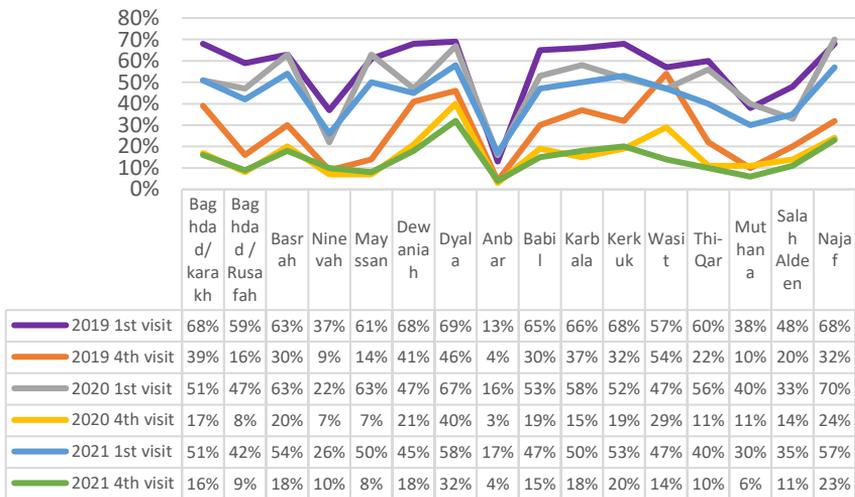


Figure -15- A periodic medical examination for pregnant women and mothers for the month of October

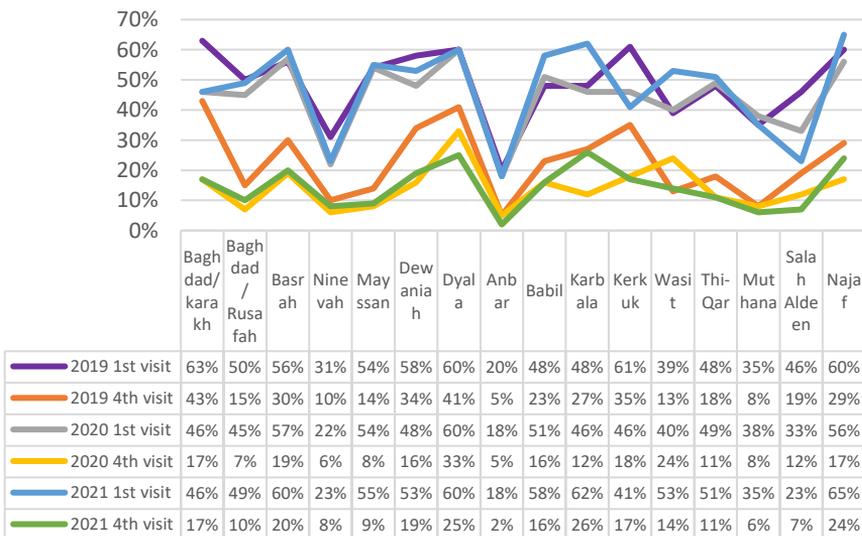


Figure -16- Periodic medical examination for pregnant women and mothers for the month of November

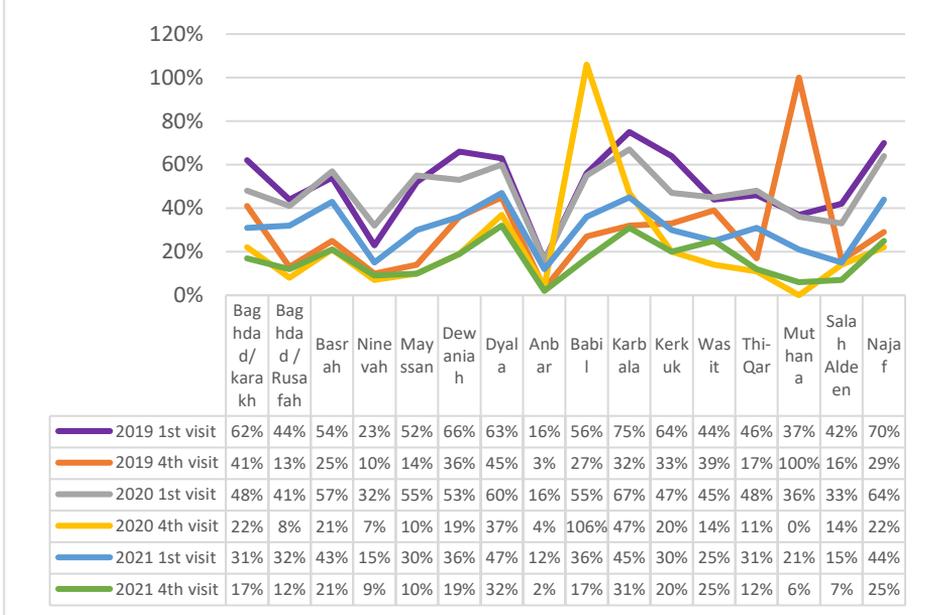
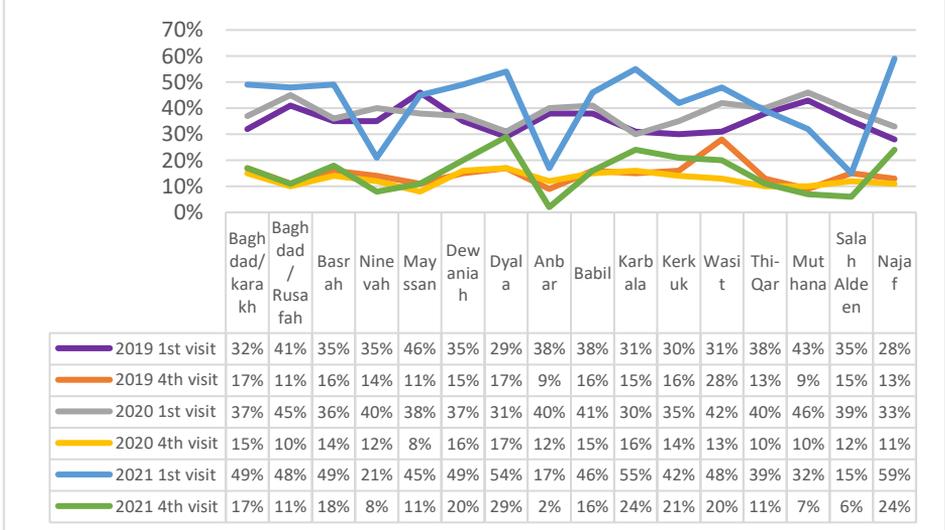


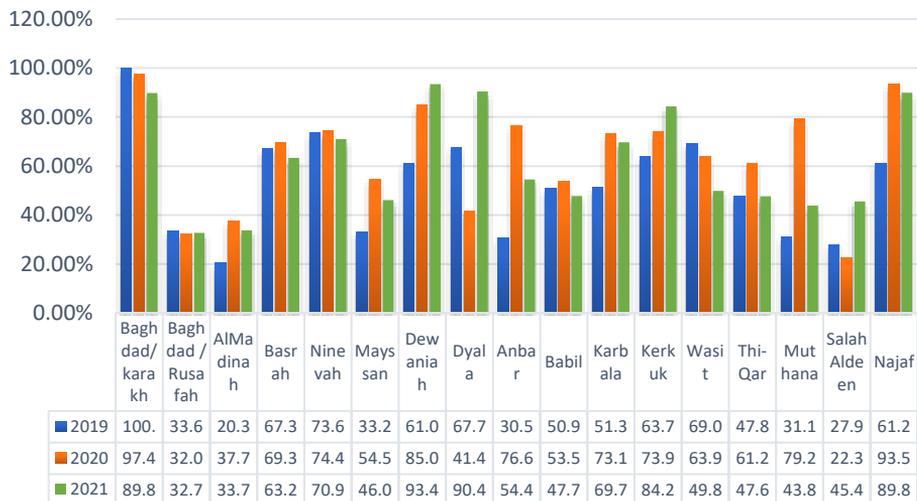
Figure -17- Periodic medical examination for pregnant women and mothers for the month of December



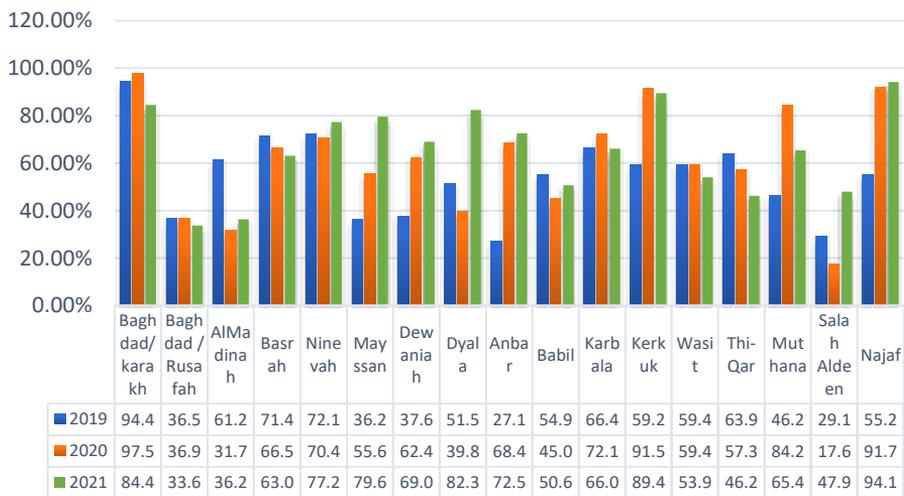
**Breast-feeding:**

The results show increase percentage of breast feeding in most of Iraqi cities for the years of 2020, 2021 in comparison to 2019 (figures 18-29).

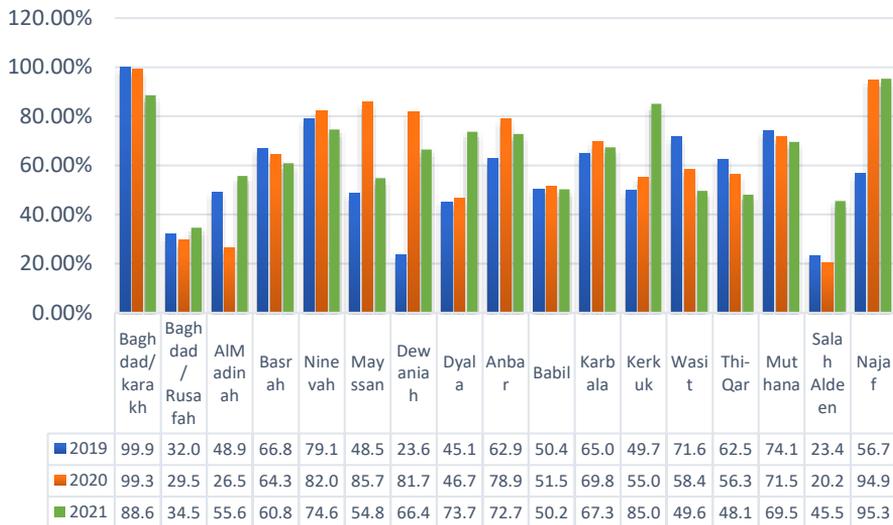
**Figure -18- Percentage of newborns who were breastfed for the month of January**



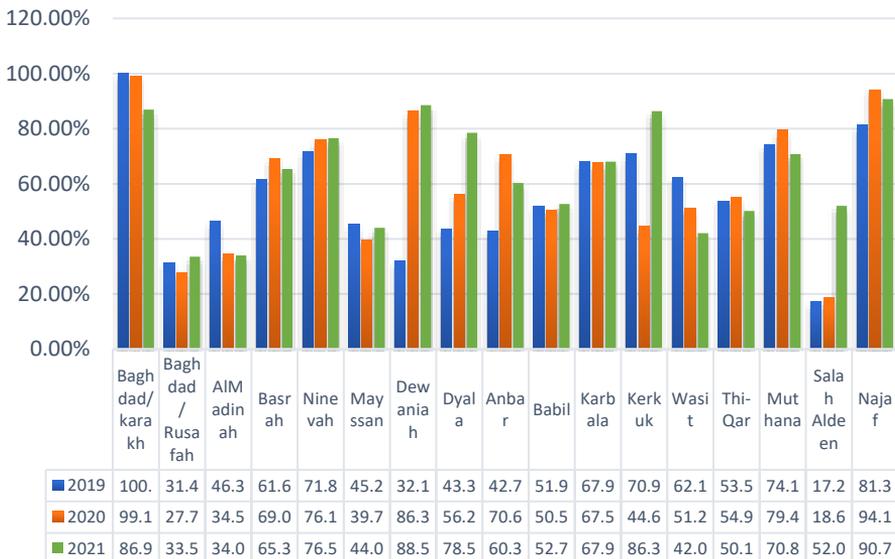
**Figure -19- Percentage of newborns who were breastfed for the month of February**

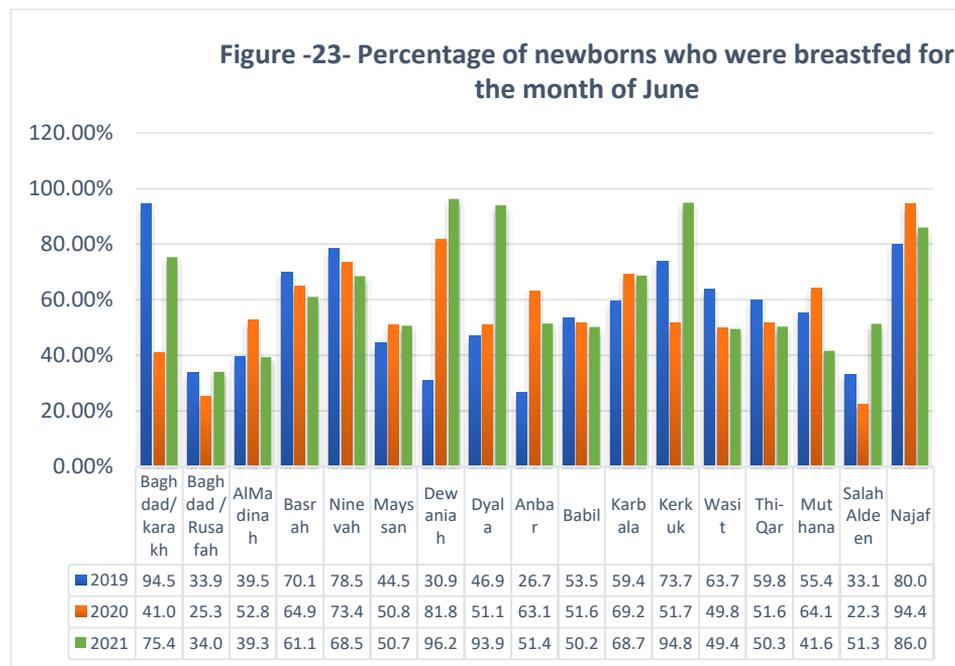
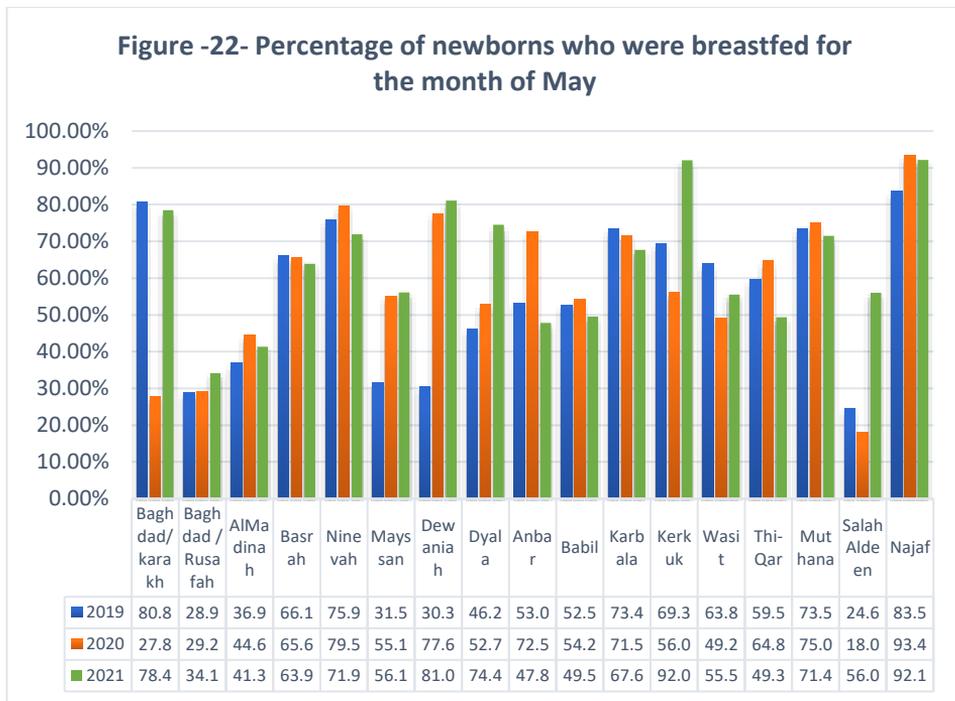


**Figure -20- Percentage of those who were breastfed for the month of March**

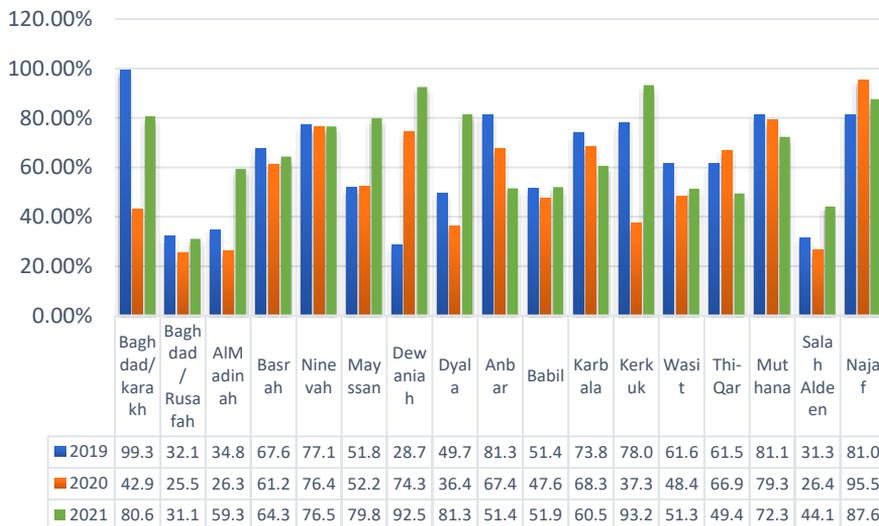


**Figure -21- Percentage of those who were breastfed for the month of April**

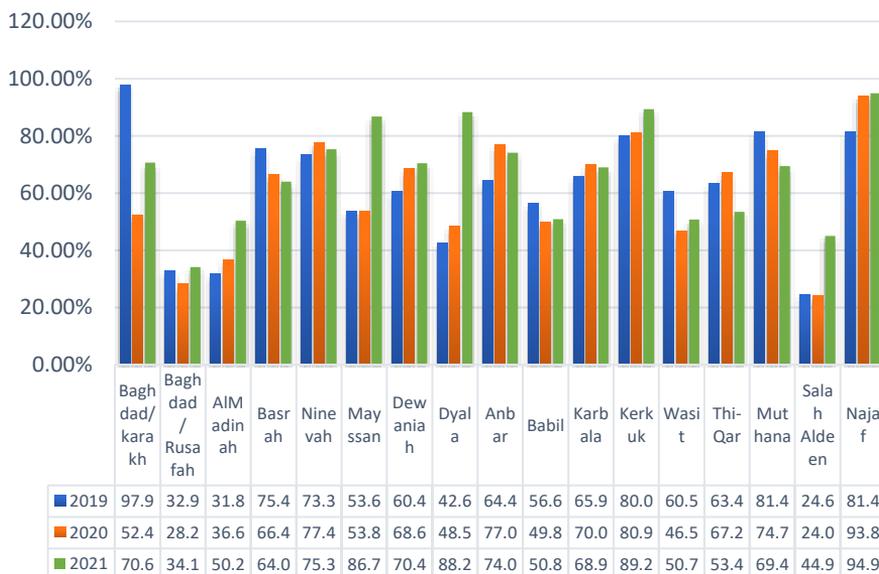




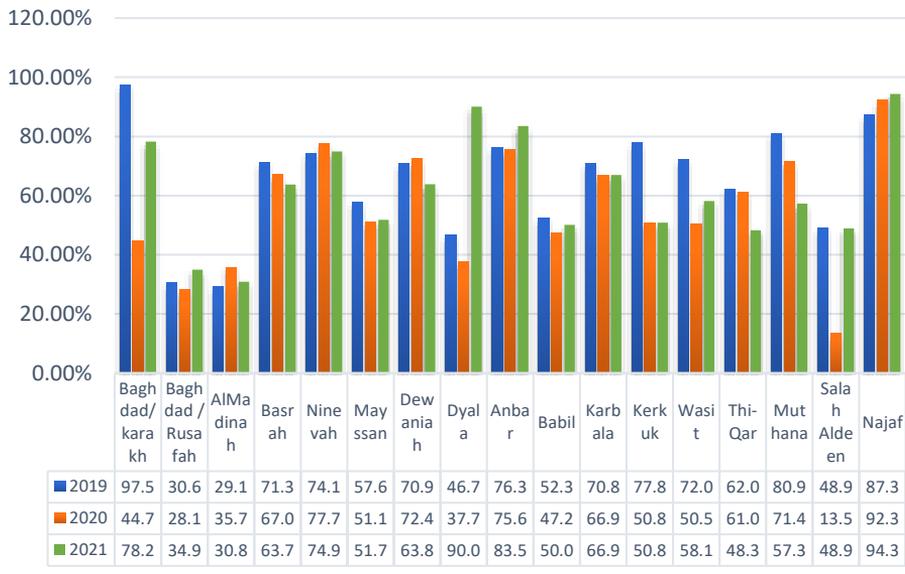
**Figure-24- Percentage of newborns who were breastfed for the month of July**



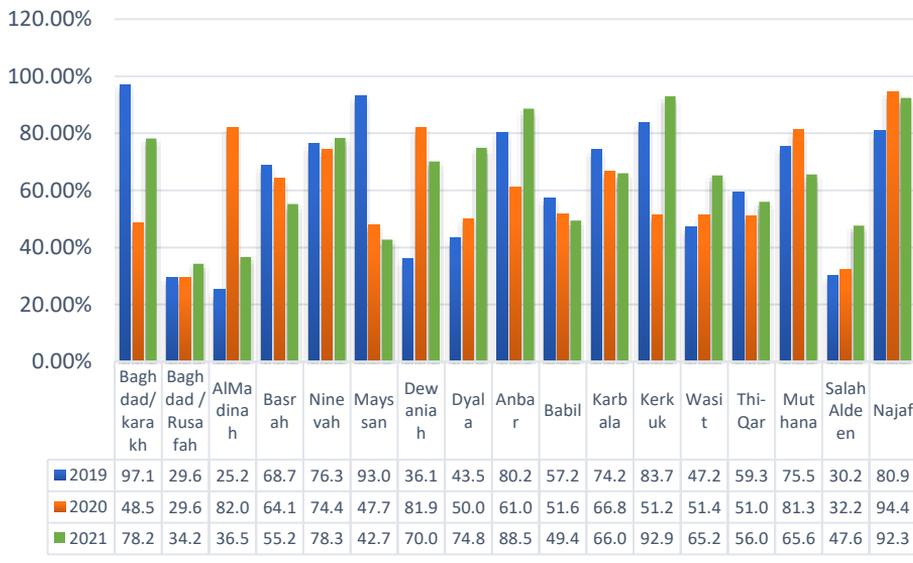
**Figure -25- Percentage of newborns who were breastfed for the month of August**



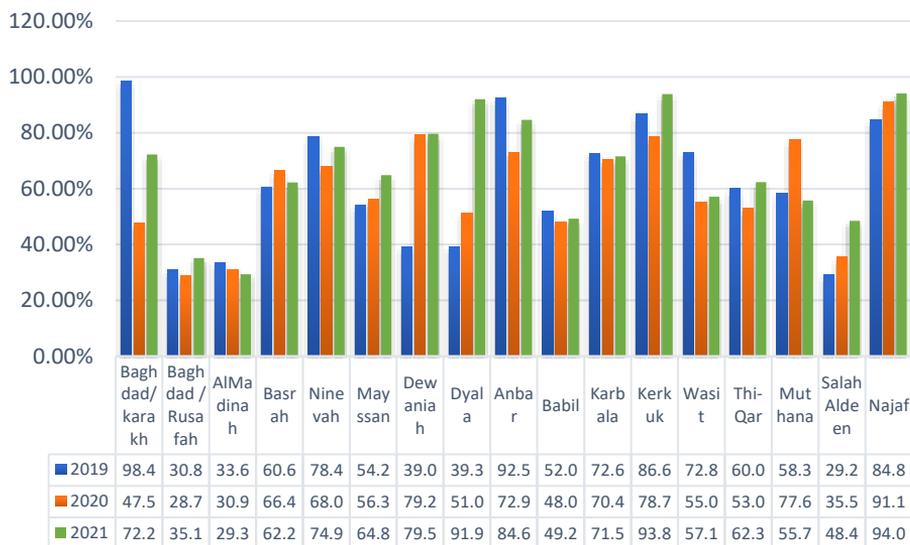
**Figure -26- Percentage of newborns who were breastfed for the month of September**



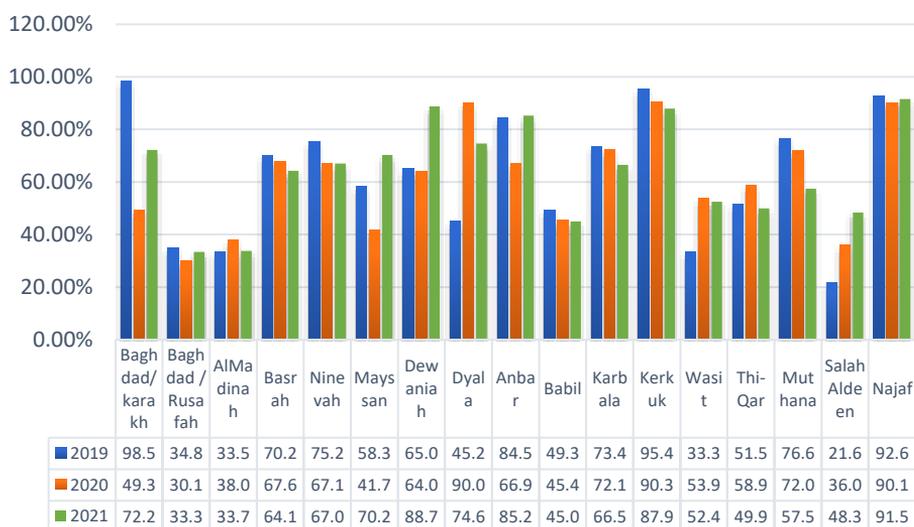
**Figure -27- Percentage of newborns who were breastfed for the month of October**



**Figure -28- Percentage of newborns who were breastfed for the month of November**



**Figure -29- Percentage of newborns who were breastfed for the month of December**



**Discussion:****Site of births****Total births**

This study show decreased total births in 2020, this may be due to fear of having congenital anomalies in newborns if getting the infection, while in 2021 it is more than 2019 and this might be due to increase in vaccination rate and the people want to complete their families after the decrease in births in the previous year.

**Births in governmental hospitals:**

This study show decrease in deliveries in these hospitals in 2020, like result of Anil K. Singh et al, Binyam Tilahun et al results <sup>(19,20)</sup> this may be due to the fact that the ladies afraid from getting infection with covid 19 in hospitals, the percentage begin to rise in 2021 after increase in vaccination rate and increase in preventive care strategies, on the contrary, Shweta Bankar et al study shows that deliveries in hospitals are increased during pandemics, because women think that this will let them able to receive better medication for themselves and their newborn. <sup>(21)</sup>

**Births in centers that contain birthing halls:**

The results show fluctuation in percentages of births in these centers with no direct relationship with the pandemic, may be because these centers are the only place for labor in rural areas.

**Births in private hospitals:**

The study show Increase in deliveries in private hospitals because the governmental hospitals was used to treat covid 19 patients mainly and the people thinks that the isolation and protective care equipment and sanitization techniques are better in private hospitals than other places, the result is same like Shweta Bankar et al results. <sup>(21)</sup>

**Births outside health institutes:**

The results show increase in deliveries outside health centers during the pandemic like Binyam Tilahun et al results, <sup>(20)</sup> in which 60-70% of mothers delivered at home, authors suspect this is due to fear from being infected with covid 19.

**Primary health care services:**

The results show decrease visits in the winter, like Binyam Tilahun et al Shweta Bankar et al, Mar Requena-Mullor et al, Faran Emmanuel et al results, <sup>(20-23)</sup> in May to august there is increase in visits, fluctuating percentages of visits from September to December. The decrease in visits belong to many reasons like: fear from getting infected with covid 19 virus, <sup>(24)</sup> some women think that their complains are not so important

and don't want to increase the burden on health workers and may be due to the availability of tele-consultations programs which don't need health centers visits.

### III. Breast feeding:

This study show increase breast feeding rate during the pandemic may be due to decrease job times of the working mothers because many countries depend on the lockdown to flatten the curve of infections which gave the women more time for lactation, also they know that breast feeding support the immunity of their infants, and forbidding many private jobs with crowded nature during the pandemic lead to decrease the income of many families so they depend on breast feeding as a cheap and available method of feeding, on the contrary, Mar Requena-Mullor etal results shows breastfeeding of newborns decreased during the pandemic (76.0% before versus 70.6% during the pandemic), <sup>(22)</sup> and this may be due to geographic and social differences between the studies.

Nowadays, many articles recommend breastfeeding regardless of infection, arguing that transmission of the virus by this route is impossible and that the benefits outweigh the dangers. <sup>(25-29)</sup>

### References:

1. Ministerio de Sanidad. Centro de Coordination de Alerts y Emergencies Sanitarias, 2021. Actualización N 377. Enfermedad por el Coronavirus (COVID-19) a 18 de Mayo de 2021.
2. World Health Organization. Managing Epidemics: Key Facts about Major Deadly Diseases. World Health Organization; 2018.
3. Sochas L, Channon AA, Nam S. Counting indirect crisis-related deaths in the context of a low-resilience health system: the case of maternal and neonatal health during the Ebola epidemic in Sierra Leone. Health Policy Plan. 2017 Nov 1; 32(suppl\_3):iii32–9
4. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. Lancet Infect Dis. 2020; 20:773. Doi:10.1016/S1473-3099(20)30195-X
5. Robertson T, Carter ED, Chou VB, Stegmuller AR, Jackson BD, Tam Y, et al. Early estimates of the indirect effects of the COVID-19 pandemic on maternal and child mortality in low-income and middle-income countries: a modelling study. Lancet Glob Health. 2020 Jul; 8:901–8.
6. World Health Organization. World Health Statistics 2016: Monitoring Health for the SDGs Sustainable Development Goals. World Health Organization; 2016.

7. Jardine, J.; Relph, S.; Magee, L.; Dadelszen, P.; Morris, E.; Ross-Davie, M.; Draycott, T.; Khalil, A. Maternity services in the UK during the coronavirus disease 2019 pandemic: A national survey of modifications to standard care. *BJOG Int. J. Obstet. Gynaecol.* 2021, 128, 880–889.
8. Royal College of Obstetricians and Gynaecologists. Guidance for Healthcare Professionals on Coronavirus (COVID-19) Infection in Pregnancy, 11th ed.; Royal College of Obstetricians and Gynaecologists: London, UK, 2020.
9. Ravaldi, C.; Wilson, A.; Ricca, V.; Homer, C.; Vannacci, A. Pregnant women voice their concerns and birth expectations during the COVID-19 pandemic in Italy. *Women Birth* 2021, 34, 335–343.
10. Karavadra, B.; Stockl, A.; Prosser-Snelling, E.; Simpson, P.; Morris, E. Women's perceptions of COVID-19 and their healthcare experiences: A qualitative thematic analysis of a national survey of pregnant women in the United Kingdom. *BMC Pregnancy Childbirth* 2020, 20, 600.
11. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C et al. A global database of COVID-19 vaccinations. *Nat Hum Behav.* 2021 Jul;5(7):947–53.
12. Rome BN, Avorn J. Drug evaluation during the Covid-19 pandemic. *N Engl J Med.* 2020 Jun 11; 382:2280–2.
13. Atalan A. Is the lockdown important to prevent the COVID-9 pandemic? Effects on psychology, environment and economy-perspective. *Ann Med Surg.* 2020 Aug; 56:38–42.
14. Biasucci G, Cannalire G, Raymond A, Capra ME, Benenati B, Vadacca G, et al. Safe perinatal management of neonates born to SARS-CoV-2 positive mothers at the epicenter of the Italian epidemic. *Front Pediatr.* 2020; 8:565522.
15. World Health Organization. Q&A on COVID-19 and breastfeeding. <https://www.who.int/news-room/q-a-detail/q-a-on-covid-19-and-breastfeeding> Accessed 4 Jun 2021.
16. Davanzo R. Breast feeding at the time of COVID-19: do not forget expressed mother's milk, please. *Arch Dis Child Fetal Neonatal Ed.* 2020; 105(4):455.
17. Davanzo R, Moro G, Sandri F, Agosti M, Moretti C, Mosca F. Breastfeeding and coronavirus disease-2019: ad interim indications of the Italian Society of Neonatology endorsed by the Union of European Neonatal & perinatal societies. *Matern Child Nutr.* 2020; 16(3):e13010.
18. Walker KF, O'Donoghue K, Grace N, Dorling J, Comeau JL, Li W, et al. Maternal transmission of SARS-COV-2 to the neonate, and possible routes for such transmission: a systematic review and critical analysis. *BJOG.* 2020; 127(11):1324–36.
19. Singh, A. K., Jain, P. K., Singh, N. P., Kumar, S., Bajpai, P. K., Singh, S., & Jha, M. (2021). Impact of COVID-19 pandemic on maternal and child health services in Uttar Pradesh, India. *Journal of family medicine and primary care*, 10(1), 509–513.

20. Tilahun B, Nigusie A, Zelalem M, Mekonnen ZA. Effect of COVID-19 Pandemic on Maternal and Child Health Services and Strategies for Effective Service Implementation in Ethiopia. *J Multidiscip Healthc.* 2022; 15:2781-2795.
21. Bankar, S., Ghosh, D. Accessing Antenatal Care (ANC) services during the COVID-19 first wave: insights into decision-making in rural India. *Reprod Health* 19, 158 (2022).
22. Requena-Mullor, M., García-González, J., Wei, R., Romero-Del Rey, R., & Alarcón-Rodríguez, R. (2022). The Impact of COVID-19 on the Monitoring of Pregnancy and Delivery of Pregnant Women in the Dominican Republic. *Healthcare (Basel, Switzerland)*, 10(11), 2266.
23. Faran Emmanuel, 1 Ahsan Ahmad, 1 Tahira Reza et al. Indirect effects of COVID-19 pandemic on reproductive, maternal, newborn and child health services in Pakistan. *Eastern Mediterranean Health Journal* 2022, 28(4).
24. Wu, H.; Sun, W.; Huang, X.; Yu, S.; Wang, H.; Bi, X.; Sheng, J.; Chen, S.; Akinwunmi, B.; Zhang, C.J.P.; et al. Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges. *J. Med. Internet Res.* 2020, 22, e19916.
25. Lubbe,W.; Botha, E.; Niela-Vilen, H.; Reimers, P. Breastfeeding during the COVID-19 pandemic—a literature review for clinical practice. *Int. Breastfeed. J.* 2020, 15, 82.
26. Royal College of Obstetricians and Gynaecologists. *Guidance for Healthcare Professionals on Coronavirus (COVID-19) Infection in Pregnancy*, 11th ed.; Royal College of Obstetricians and Gynaecologists: London, UK, 2020.
27. Davanzo, R.; Merewood, A.; Manzoni, P. Skin-to-Skin Contact at Birth in the COVID-19 Era: In Need have Help! *Am. J. Perinatol.* 2020, 37, S1–S4.
28. Lebrao CW, Cruz MN, da Silva MH, Dutra LV, Cristiani C, Affonso Fonseca FL, et al. Early identification of IgA anti-SARSCoV-2 in milk of mother with COVID-19 infection. *J Hum Lact.* 2020; 36(4):609–13.
29. Pace RM, Williams JE, Jarvinen KM, Belfort MB, Pace CDW, Lackey KA, et al. Characterization of SARS-CoV-2 RNA, antibodies, and neutralizing capacity in milk produced by women with COVID-19. *mBio.* 2021; 12(1):e03192–20.

**Publication Prerequisites and terms**

- 1- The journal publishes scientific research and studies in statistics and informatics written in Arabic, English and French, to make it clear that research submitted for publication has been published or submitted for publication in magazines or other periodicals or presented and published in periodicals for conferences or seminars.
- 2- Send electronic copies (word & PDF) of the research and studies to the editor should include the name of the researcher or researchers and their scientific titles and places of work with the address of the correspondence, the numbers of telephones and e-mail. The research to be published should be sent electronically in accordance with the specifications below:
  - a. To be printed on A4 paper and be in the form of a single column and use the Type simplified Arabic and Times New Roman for English and French and with a font size (12). Using Microsoft Word and on one face of the paper.
  - b. The margin is 2.5 cm for all sides of the paper.
  - c. The researcher will attach a summary of his research in Arabic, English, or French in no more than one page.
  - d. Place references at the end of the paper and separated page. It is recommended to use the Harvard system of referencing, which (author's name, year of publication, source address, publishing house, country).
  - e. Numbered tables, illustrations, and others as they are received in the research, documents as aliases of the original sources.
  - f. The number of search or study pages should not exceed (25) pages.
- 3- Authors will notified of receiving their research within two working days from the date of receipt of the research.
- 4- Referees will evaluate all submitted research, the Authors will informed of the proposed evaluation and modifications if any within two weeks of receipt of the research.
- 5- The editorial board of the Journal has the right to accept or reject the research and has the right to make any modification or partial redrafting of the material submitted for publication in accordance with the format adopted in its publication after the approval of the researcher.
- 6- Published research becomes the property of the Journal and may not republished elsewhere.
- 7- The articles published in the magazine reflect the opinions of the authors, and do not necessarily reflect the view of the Journal or the Arab Institute for Training and Research in Statistics.
- 8- The research is sent to the magazine's e-mail address:  
[journal@aitrs.org](mailto:journal@aitrs.org) or [Info@aitrs.org](mailto:Info@aitrs.org)

# Journal of Statistical Sciences

Scientific Referred Journal

## Editorial Board

### Editor-in Chief

Mr. Hedi Saidi

### Editorial Secretary

Dr. Bachioua Lahcene

## Editorial Board Members

Prof. Dr. Faisal Al-Sharabi  
Dr. Salwa Mahmoud Assar  
Dr. Hassan Abuhassan

Prof. Dr. Abed Khaliq Tohami  
Prof Dr. Ahmed Shaker Almutwali  
Dr. Hamid Bouzida

Prof. Dr. Mukhtar Al-Kouki  
Prof. Dr. Issa Masarweh

## Scientific Consulting Committee

Dr. Qassim Al-Zoubi  
Dr. Diaa Awad  
Dr. luay shabaneh

Dr. Nabeel M. Shams  
Dr. Khalifa Al-Barwani  
Prof. Dr. Ghazi Raho  
Dr. Ola Awad

Prof. Dr. Awad Haje Ali  
Prof. Dr. Maytham Elaibi Ismael  
Dr. Mohammed Husain Ali Al-Janabi

Listed in Ulrich's website

[www.ulrichsweb.com](http://www.ulrichsweb.com)

Classified in The Arab Citation & Impact Factor (Arcif)

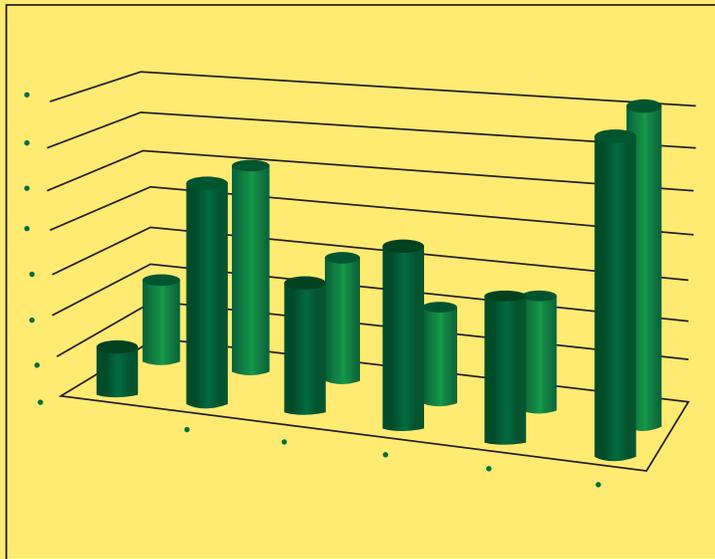
[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)



Arab Institute for Training and Research in Statistics

# Journal of Statistical Sciences



**Issue No. 21**

**Scientific Peer-reviewed Journal issued by  
Arab Institute for Training and Research in Statistics**

Listed in Ulrich's website  
[www.ulrichsweb.com](http://www.ulrichsweb.com)

Classified in The Arab Citation & Impact Factor (Arcif)  
[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)