Journal of Education and Science (ISSN 1812-125X), Vol: 31, No: 02, 2022 (29-41)

Gauss-Hermite Cubature Method to estimate parameters of a multivariate GLMM

Adnan M. H. Al-Sinjary[1*], Dr. Auday Taha Raheem [2]

[1*,2] Statistics Department, College of Management and Economics, Mustansiriyah University

E-mail: [1*]adnan.m.alsinjary@uomustansiriyah.edu.iq, [2]uday_adm@uomustansiriyah.edu.iq

(Received February 14, 2022; Accepted March 24, 2022; Available online June 01, 2022)

DOI: 10.33899/edusj.2022.133041.1216,

Abstract
In this paper, the multivariate generalized linear mixed model (GLMM) was studied when there are three response variables, distributed as Normal, Bernoulli, and Poisson. And because there is a multiple integration in the likelihood function for the model under study, it is necessary to use mathematical methods to solve this integration, and because it is not possible to obtain the result of this integration by the well-known methods of integration, numerical methods have been used, the Gauss-Hermite Cubature (GHC) algorithm, which is one of the most common numerical integration methods. Then the estimates were obtained by maximizing the resulting likelihood function with respect to the parameters, and thus, estimates were obtained for the parameters. On the practical side, we have used real data representing the effect of potassium as a fixed effect, and referring patients were considered as a random effect on three response variables: calcium, creatinine, and urea as they follow Normal, Bernoulli, and Poisson distributions taken from the records of the Vajin Hospital in Dohuk city. The regression coefficients showed that the effect of potassium is positive on both calcium and urea because the values of the coefficients are positive, while its effect is negative on the creatinine because the value of the coefficient is negative. Based on the results, the researchers recommended several recommendations, including estimating the standard error of the estimates in their light, in order to construct hypotheses for a significance test about regression coefficients.

Keywords: Numerical Integration, Likelihood Function, Mixed Models.

طريقة گاوس–هرمت التكعيبية لتقدير معلمات النموذج المختلط الخطي المعمم متعدد المتغيرات

عدنان مصطفى حسين السنجاري [1*]، د. عدي طه رحيم [2]

[1*، 2] قسم الإحصاء، كلية الإدارة والاقتصاد، الجامعة المستنصرية

الخلاصة

درس في هذا البحث النموذج المختلط الخطي المعمم متعدد المتغيرات بوجود ثلاثة متغيرات استجابة، هذه المتغيرات تتبع التوزيع الطبيعي وبرنولي وبواسون، ولوجود تكامل متعدد في دالة الإمكان بالنسبة للنموذج محل الدراسة وجب استعمال طرائق رياضية

للحصول على ناتج هذا التكامل، وبسبب عدم إمكانية الحصول على ناتج هذا التكامل بطرائق التكامل المعروفة فقد تم اللجوء إلى طرائق عددية، إذ استعملت خوارزمية گاوس–هرمت التكعيبية، وهي واحدة من أكثر طرائق التكامل العددي شيوعاً. بعد ذلك تحصل على التقديرات بتعظيم دالة الإمكان الناتجة نسبة إلى المعلمات وبذلك تم الحصول على تقديرات للمعلمات. أما في الجانب التطبيقي فقد طبق على بيانات حقيقية تمثل تأثير البوتاسيوم كتأثير ثابت ومراجعة المريض تم اعتبارها تأثير عشوائي على ثلاث متغيرات استجابة هي الكالسيوم والكرياتنين واليوريا باعتبارها تتبع التوزيع الطبيعي وتوزيع برنولي وتوزيع بواسون على التوالي مأخوذة من سجلات مستشفى ئازين الأهلية في مدينة دهوك. وقد أظهرت معاملات الانحدار أن تأثير البوتاسيوم موجب على كل من الكالسيوم واليوريا لأن قيمة المعامل لها موجب في حين أن تأثيره سالب على الكرياتنين بسبب كون قيمة المعامل سالبة، وبناءً على النتائج أوصى الباحثان عدة توصيات، منها تقدير الخطأ القياسي للتقديرات ليتم بناء اختبار فرضيات على ضوئها والتأكد من معنوية التأثير.

الكلمات المفتاحية: التكامل العددي، دالة الإمكان، النماذج المختلطة.

المقدمة Introduction

يطلق لفظ النموذج الخطي العام (General Linear Model) على النماذج التي يكون حد الخطأ فيها يتبع التوزيع الطبيعي متعدد المتغيرات، مما يؤدي إلى جعل متغير الاستجابة يتبع التوزيع الطبيعي متعدد المتغيرات، إلا أنه في كثير من الحالات لا يتحقق هذا الشرط كأن يتبع متغير الاستجابة توزيع برنولي أو توزيع ثنائي الحدين أو ثنائي الحدين السالب أو متعدد الحدود أو بواسون أو غيرها، في هذه الحالة يعمم النموذج الخطي ليشمل، بالإضافة إلى التوزيع الطبيعي، هذه التوزيعات، حينها يسمى النموذج بالنموذج الخطي المعمم (Generalized Linear Model).

في حين يمتاز النموذج الخطي العام بأنه مكون بالإضافة إلى تركيبة الخطأ العشوائي من حد ثابت (Fixed)، فأنه قد يحصل أن يكون مضافاً له حد آخر غير ثابت، أي عشوائي (Random)، وهذه النماذج التي تحتوي على الحد الثابت والحد العشوائي تسمى بالنماذج المختلطة العامة (General Linear Mixed Model)، ولفظ مختلط هنا يعني وجود التأثير الثابت والعشوائي معاً. أما النموذج المختلط الذي لا يتبع متغير الاستجابة فيه التوزيع الطبيعي متعدد المتغيرات بل يتبع إحدى توزيعات العائلة الأسية فإنه يسمى النموذج المختلط الخطي المعمم (Generalized Linear Mixed Model).

كل ما سبق هو في حالة وجود متغير استجابة واحد، إلا أنه في بعض الأحيان، وبالنسبة لجميع هذه النماذج، قد يكون هناك أكثر من متغير استجابة في نفس الوقت، حينها نحتاج إلى ما يسمى بالنماذج متعددة المتغيرات. كلما زادت متغيرات الاستجابة، بالطبع، ستزيد عدد المعلمات المراد تقديرها وبذلك يصبح من الصعب عملية التقدير هذه. تكمن صعوبة إيجاد تقديرات الإمكان الأعظم بشكل اعتيادي ليس فقط في وجود أكثر من متغير استجابة وهذه المتغيرات مختلفة، بل الصعوبة الأكبر بسبب وجود التأثيرات العشوائية والتي تجعل من دالة الإمكان محتوية على تكامل وهذا التكامل لا يمكن حله بشكل بسيط لذا يتم الذهاب إلى طرائق عددية لحل هذا التكامل بل قد يتعدى ذلك إلى طرائق محاكاة، في هذا البحث تستعمل طريقة تكامل عددي متمثلة بطريقة گاوس–هرمت التكعيبية (GHC).

ولتوضيح هذه النماذج وطريقة تقدير معلماتها وكذلك تطبيقها فقد تم تقسيم البحث إلى قسمين رئيسين تمثل القسم الأول المواد وطرائق العمل الذي يشمل نبذة عن النموذج المختلط الخطي المعمم أحادي متغير الاستجابة وكذلك نبذة عن النموذج في حالة تعدد متغيرات الاستجابة بالإضافة إلى توضيح طريقة گاوس هرمت التكعيبية (GHC) للحصول على تكاملات متعددة الأبعاد وتطبيقها على

نموذج البحث. أما القسم الثاني والذي يمثل النتائج والمناقشة فقد تضمن نبذة عن الكلية ووظائفها ووصف البيانات محل الدراسة وكذلك تقدير معلمات النموذج لهذه البيانات بناءً على طريقة گاوس–هرمت التكعيبية (GHC) بالإضافة إلى الاستنتاجات والتوصيات.

المواد وطرائق العمل Martials and Methods

1. النموذج المختلط الخطي المعمم

يعتبر النموذج الخطي المعمم (GLM) Generalized Linear Models الذي اقترح من قبل كل من (Nelder & Wedderburn) سنة 1972 امتداداً للنموذج الخطي العام والذي يمثل نموذج الانحدار الخطي المعروف والذي يمتاز أن متغير الاستجابة فيه يتبع التوزيع الطبيعي. هذا التوسيع يشمل جميع التوزيعات الاحتمالية التي تنتمي إلى العائلة الأسية ومن أهم التوزيعات المستعملة في هذه النماذج، التوزيع الطبيعي، توزيع برنولي، توزيع ثنائي الحدين، توزيع بواسون، وتوزيع گاما. ويكمن الفرق بين النموذج الخطي المعمم والنموذج المختلط الخطي المعمم (GLMM) Generalized Linear Mixed Models في الجزء الخطي النظامي من المركبات التي يتكون منها النموذج المعمم، وهذه المركبات هي [1]:

1. مركبة الجزء العشوائي (Random Component) والمتمثلة بمتغير الاستجابة الذي له توزيع احتمالي ينتمي إلى العائلة الأسية.
2. مركبة الجزء الخطي النظامي (Linear Systematic Component) والتي تحتوي على مشاهدات المتغيرات التوضيحية (X) ومتجه المعالم ($\underline{\beta}$)، إذ إن الجزء النظامي في النموذج الخطي المعمم والذي يسمى المتنبئ الخطي هو $\eta_i = \underline{x}_i^T \underline{\beta}$ بينما في النموذج المختلط الخطي المعمم هو $\eta_{ij} = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T \underline{b}_i$.

حيث إن: i=1,2,…,n يمثل المؤشر الخاص بالمفردات (المشاهدات) و$j=1,2,\ldots,r_i$ يمثل المؤشر الخاص بالتكرارات وإن n يمثل حجم العينة وr_i يمثل عدد مرات تكرار المفردة i.

η_{ij}: يمثل المتنبئ الخطي.

$\underline{x}_{ij}^T$: يمثل الصف j من المصفوفة X_i. و X هي مصفوفة التصميم للتأثيرات الثابتة.

$\underline{\beta}$: يمثل متجه معلمات التأثير الثابت بالبعد (p×1).

$\underline{z}_{ij}^T$: يمثل الصف j من المصفوفة Z_i. و Z هي مصفوفة التصميم للتأثيرات العشوائية.

$\underline{b}_i$: متجه التأثير العشوائي للمفردة i بالبعد (q×1).

3. دالة الربط Link Function القابلة للاشتقاق $g(.)$، والتي تربط بين الجزء العشوائي والجزء النظامي، والتي تحقق $g(.) = \underline{x}_{ij}^T \underline{\beta} + \underline{z}_{ij}^T \underline{b}_i$.

وسيتم في هذه الفقرة تحديد دالة الإمكان المشروطة للمفردة i في حالة النموذج المختلط الخطي المعمم (GLMM) لكل من التوزيع الطبيعي وتوزيع برنولي وتوزيع بواسون لكونها التوزيعات محل اهتمام البحث وكما يأتي [2]:

أولاً: التوزيع الطبيعي:

$$L_{i1}\left(\underline{\beta}|\underline{b}_i;\underline{y}_i\right) = exp\left[\frac{1}{\sigma^2}\sum_{j=1}^{r_i}\left(y_{ij}\left(\underline{x}_{ij}^T\underline{\beta} + \underline{z}_{ij}^T\underline{b}_i\right) - \left(\underline{x}_{ij}^T\underline{\beta} + \underline{z}_{ij}^T\underline{b}_i\right)^2/2\right) - \frac{1}{2}\sum_{j=1}^{r_i}\left\{\frac{y_{ij}^2}{\sigma^2} + ln(2\pi\sigma^2)\right\}\right] \qquad (1)$$

ثانياً: توزيع برنولي:

$$L_{i2}\left(\underline{\beta}|\underline{b}_i;\underline{y}_i\right) = exp\left[\sum_{j=1}^{r_i}\left(y_{ji}\left(\underline{x}_{ji}^T\underline{\beta}+\underline{z}_{ji}^T\underline{b}_i\right)-ln\left\{1+exp\left(\underline{x}_{ji}^T\underline{\beta}+\underline{z}_{ji}^T\underline{b}_i\right)\right\}\right)\right] \quad (2)$$

ثالثاً: توزيع بواسون:

$$L_{i3}\left(\underline{\beta}|\underline{b}_i;\underline{y}_i\right) = exp\left[\sum_{j=1}^{r_i}\left(y_{ij}\left(\underline{x}_{ij}^T\underline{\beta}+\underline{z}_{ij}^T\underline{b}_i\right)-exp\left\{\underline{x}_{ij}^T\underline{\beta}+\underline{z}_{ij}^T\underline{b}_i\right\}\right) - ln(y_{ij}!)\right] \quad (3)$$

2. نبذة تاريخية عن النموذج المختلط الخطي المعمم متعدد المتغيرات

بدأ النموذج المختلط الخطي المعمم متعدد المتغيرات(Multivariate Generalized Linear Mixed Model (MGLMM عام 1999 من قبل الباحثة Gueorguieva وبقيت الدراسات المتعلقة به نادرة على المستوى العالمي إلى منتصف العقد الثاني من هذا القرن أما في العراق فلم تقع في يد الباحثين أية دراسة بهذا الخصوص.

ويمكن تلخيص الدراسات السابقة بدراسة Gueorguieva عام 1999 [3] والتي اقترحت النموذج المختلط الخطي المعمم متعدد المتغيرات إلا أنها عملت على متغيرين، الأول طبيعي والثاني برنولي، ودراسة Gueorguieva عام 2001 [4] والتي قامت بالتطبيق على نموذجين مختلفين الأول في حالة كون المتغير الأول يتبع توزيع گاما والثاني يتبع توزيع ذي الحدين السالب، والنموذج الثاني فيه المتغير الأول طبيعي والثاني برنولي، ودراسة Gebregziabher وآخرون عام 2013 [5] الذين قدموا نموذج من ثلاثة متغيرات يتكون من توزيعي گاما وتوزيع طبيعي لوغاريتمي، ودراسة Karl وآخرون عام 2014 [6] الذين قاموا بالعمل على نموذج ثنائي المتغيرات يتكون من توزيع بواسون وتوزيع برنولي، ودراسة Chen و Wehrly عام 2014 أيضاً [7] الذين استعملا نماذج ذات توزيعات مختلفة مثل نموذج ثنائي من توزيع بواسون وتوزيع گاما ونموذج ثلاثي من توزيع بواسون والتوزيع الأسي والتوزيع الطبيعي. بالإضافة إلى دراسة Mikulich-Gilbertson وآخرين سنة 2015 [8] التي استعملت نموذج ثنائي بمتغير يتوزع توزيع برنولي والآخر ذي الحدين السالب، ودراسة Jaffa وآخرون عام 2015 [9] التي عملت على نموذج متكون من ثلاثة متغيرات جميعها يتبع التوزيع الطبيعي اللوغاريتمي، ودراسة Jaffa وآخرون عام 2016 [10] على نموذج رباعي المتغيرات الثلاث الأولى تتبع التوزيع الطبيعي اللوغاريتمي والرابع يتبع توزيع برنولي، ودراسة [11] لثلاثة نماذج هي نموذج ثنائي المتغيرات من توزيعي بواسون وگاما ونموذج ثنائي آخر من التوزيع الطبيعي وتوزيع ثنائي الحدين السالب، ونموذج ثلاثي من توزيع بواسون وتوزيع گاما والتوزيع الطبيعي. وكذلك ودراسة Crowther عام 2017 [12] لنموذج ثلاثي متكون من توزيع ويبل وتوزيعين طبيعيين، ودراسة Broatch و Kar عام 2018 [13] التي شملت على نموذجين ثنائيين، النموذج الأول هو نموذج مكون من التوزيع الطبيعي وتوزيع برنولي والثاني هو نموذج من توزيع بواسون وتوزيع برنولي، ودراسة Mikulich-Gilbertson وآخرين عام 2018 [14] لنموذج يتكون من متغيرين الأول يتبع توزيع برنولي والثاني يتبع توزيع ثنائي الحدين السالب، ودراسة Chauvet وآخرين عام 2019 [15] لنموذج مكون من توزيع بواسون وتوزيع برنولي، ودراسة Pelck و Labouria عام 2020 [16] لنموذج ذي متغيرين الأول ثنائي الحدين والثاني بواسون، ودراسة Achana وآخرين عام 2021 [17] لنموذج من متغيرين كل منهما يتبع توزيع گاما.

3. وصف النموذج المختلط الخطي المعمم متعدد المتغيرات

النموذج الذي قدمته Gueorguieva [3] كان عبارة عن متغيري استجابة وقد عمم الباحثان في هذا البحث النموذج إلى m من المتغيرات وكما يأتي:

لتكن لدينا المشاهدة y_{ikj}، حيث i=1,2,…,n تمثل المفردات وk=1,2,…,m تمثل متغيرات الاستجابة، و $j=1,2,\ldots,r_{ik}$ تمثل تكرار كل مفردة، عندئذٍ يمكننا التعبير عن كل مفردة كما يأتي:

$$\underline{y}_i = \begin{bmatrix} \underline{y}_{i1} \\ \underline{y}_{i2} \\ \vdots \\ \underline{y}_{im} \end{bmatrix} \quad for\ i = 1,2,\ldots,n$$

إذ:

$$\underline{y}_{ik} = \begin{bmatrix} y_{i11} \\ y_{i12} \\ \vdots \\ y_{i1r_{ik}} \end{bmatrix} \quad for\ i = 1,2,\ldots,n \;\; ; \;\; k = 1,2,\ldots,m$$

و $\underline{b}_i$ تمثل التأثير العشوائي والمعرف كما يأتي:

$$\underline{b}_i = \begin{bmatrix} \underline{b}_{i1} \\ \underline{b}_{i2} \\ \vdots \\ \underline{b}_{im} \end{bmatrix} \quad for\ i = 1,2,\ldots,n$$

حيث إن $\underline{b}_{ik}$ يمثل التأثير العشوائي المرافق للمتغير k.

ولتكن y_{i1j} مستقلة شرطياً باستبعاد $\underline{b}_{i1}$ لكل $j=1,2,\ldots,r_{i1}$ بدالة كثافة $f_1(.)$ ولتكن y_{i2j} مستقلة شرطياً باستبعاد $\underline{b}_{i2}$ لكل $j=1,2,\ldots,r_{i2}$ بدالة كثافة $f_2(.)$ وهكذا بالنسبة لجميع متغيرات الاستجابة التي عددها m، إذ إن $f_1(.)$ و $f_2(.)$ إلى $f_m(.)$ هي دوال منتمية إلى العائلة الأسية، كما أن $\underline{y}_{ik}$ مستقلة شرطياً باستبعاد $\underline{b}_i$ لكل k=1,2,…,m، وكذلك جميع المفردات مستقلة عن بعضها الآخر.

الآن وباعتبار أن $g_1(.)$ و $g_2(.)$ إلى $g_m(.)$ هي دوال ربط لكل من $f_1(.)$ و $f_2(.)$ إلى $f_m(.)$ على التوالي، في حين المتوسط الشرطي لكل من y_{i1j} و y_{i2j} إلى y_{imj} هو μ_{i1j} و μ_{i2j} إلى μ_{imj} على التوالي. مع الأخذ بنظر الاعتبار أن التوقع والتباين الهامشيين لكل واحد من المتغيرات يكونان كما يأتي:

$$E\left(\underline{y}_{ik}\right) = E\left[\underline{\mu}_{ik}\right]; \quad Var\left(\underline{y}_{ik}\right) = E\left[Var\left(\underline{\mu}_{ik}\right)\right] + Var\left[\underline{\mu}_{ik}\right]$$

حيث إن:

$$\underline{\mu}_{ik} = \begin{bmatrix} \mu_{ik1} \\ \mu_{ik2} \\ \vdots \\ \mu_{ikr_{ik}} \end{bmatrix} \quad for\ i = 1,2,\ldots,n \;\; ; \;\; k = 1,2,\ldots,m$$

وبفرض أن:

$$g_k(\underline{\mu}_{ik}) = X_{ik}\underline{\beta}_k + Z_{ik}\underline{b}_{ik}$$
$$\eta_{ikj} = \underline{x}_{ikj}^T\underline{\beta}_k + \underline{z}_{ikj}^T\underline{b}_{ik} \qquad for\ k = 1,2,\ldots,m \qquad (4)$$

إذ:

$g_k\left(\underline{\mu}_{ik}\right)$: دالة الربط للمتغير k.

η_{ikj}: المتنبئ الخطي للمتغير k.

X_{ik}: مصفوفة التصميم للتأثير الثابت بالنسبة للمفردة i والمتغير k بالبعد $(r_{ik} \times p_k)$، وتكون معرفة كما يأتي:

$$X_{ik} = \begin{bmatrix} x_{ik1}^{(1)} & x_{ik1}^{(2)} & \cdots & x_{ik1}^{(p_k)} \\ x_{ik2}^{(1)} & x_{ik2}^{(2)} & \cdots & x_{ik2}^{(p_k)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{ikr_{ik}}^{(1)} & x_{ikr_{ik}}^{(2)} & \cdots & x_{ikr_{ik}}^{(p_k)} \end{bmatrix}$$

$\underline{x}_{ikj}^T$: الصف j من مصفوفة التصميم للتأثير الثابت للمتغير k.

$\underline{\beta}_k$: متجه التأثير الثابت المرافق للمتغير k بالبعد $(p_k \times 1)$، ويكون معرفاً لكل k=1,2,…,m كما يأتي:

$$\underline{\beta}_k = \begin{bmatrix} \beta_{k1} \\ \beta_{k2} \\ \vdots \\ \beta_{kp_k} \end{bmatrix}$$

Z_{ik}: مصفوفة التصميم للتأثير العشوائي بالنسبة للمفردة i والمتغير k بالبعد $(r_{ik} \times q_k)$، وتكون معرفة كما يأتي:

$$Z_{ik} = \begin{bmatrix} z_{ik1}^{(1)} & z_{ik1}^{(2)} & \cdots & z_{ik1}^{(q_k)} \\ z_{ik2}^{(1)} & z_{ik2}^{(2)} & \cdots & z_{ik2}^{(q_k)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{ikr_{ik}}^{(1)} & z_{ikr_{ik}}^{(2)} & \cdots & z_{ikr_{ik}}^{(q_k)} \end{bmatrix}$$

$\underline{z}_{ikj}^T$: الصف j من مصفوفة التصميم للتأثير العشوائي للمتغير k.

$\underline{b}_{ik}$: متجه التأثير العشوائي المرافق للمفردة i والمتغير k بالبعد $(q_k \times 1)$، ويكون معرفاً لكل k=1,2,…,m كما يأتي:

$$\underline{b}_{ik} = \begin{bmatrix} b_{ik1} \\ b_{ik2} \\ \vdots \\ b_{ikq_k} \end{bmatrix}$$

ويتبع عادة التوزيع الطبيعي متعدد المتغيرات، ويعبر عنه وصفياً كما يأتي:

$$\underline{b}_{ik} \sim N_{q_k}\left(\underline{0}, \Sigma_{kk}\right)$$

علماً أن المصفوفة Σ_{kk} تكون قطرية لكل k=1,2,…,m.

فيما المتجه $\underline{b}_i$ يتوزع توزيعاً طبيعياً متعدد المتغيرات أيضاً، والذي يعبر عنه وصفياً كما يأتي:

$$\underline{b}_i = \begin{bmatrix} \underline{b}_{i1} \\ \underline{b}_{i2} \\ \vdots \\ \underline{b}_{im} \end{bmatrix} \sim N_q\left(\underline{0}, \Sigma\right);\ where\ q = \sum_{k=1}^{m} q_k$$

إذ:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \cdots & \Sigma_{mm} \end{bmatrix}$$

عندما تكون التباينات المشتركة مساوية للصفر فأننا نستطيع أن نتعامل مع النموذج متعدد المتغيرات على أنه عبارة عن نماذج منفصلة، لكل متغير نموذجه الخاص. بالنسبة لدراستنا هذه سيتم العمل على النموذج الذي يحتوي على ثلاثة متغيرات بمعنى m=3.

وبالنسبة لدالة الإمكان الهامشية للمفردة i فهي موضحة في الصيغة الآتية:

$$L_i\left(\underline{\beta},\Sigma\right)=\int\int\int L_{i1}L_{i2}L_{i3}f\left(\underline{b}_i;\Sigma\right)d\underline{b}_{i1}d\underline{b}_{i2}d\underline{b}_{i3} \qquad (5)$$

حيث إن L_{i1} و L_{i2} و L_{i3} موضحة في المعادلات $(1-3)$ على التوالي و $f\left(\underline{b}_i;\Sigma\right)$ هي دالة الكثافة الاحتمالية للتوزيع الطبيعي وموضحة كما يأتي:

$$f\left(\underline{b}_i;\Sigma\right)=\frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma|^{\frac{1}{2}}}\,e^{-\frac{1}{2}\underline{b}_i^T\Sigma^{-1}\underline{b}_i}$$

ونلاحظ من المعادلة (5) أنها تحتوي على تكامل، وهذا التكامل لا يمكن الحصول عليه بطرائق تحليلية لذا يتم اللجوء إما إلى طرائق محاكاة أو طرائق عددية مثل طريقة كاوس–هرمت التكعيبية (GHC).

3.4 خوارزمية كاوس–هرمت التكعيبية

في التحليل العددي Numerical Analysis، الطريقة التي يمكن بها إيجاد التكامل المحدد بشكل عددي للدوال التي لا يمكن إيجاد تكاملها تحليلياً تسمى التكامل العددي Numerical Integration أو التربيع Quadrature، وكلمة Quadrature، ويشير مصطلح التربيع هنا إلى إيجاد المربع الذي مساحته نفس المساحة تحت منحنى الدالة [18]. ويهتم التكامل العددي بإيجاد حلول تقريبية لتكاملات الدوال التي يصعب إيجادها بالطرائق التحليلية، وكما يأتي [19]:

$$Q(f)\approx\sum_{k=0}^{v}w_kf(x_k) \qquad (6)$$

حيث إن x_s تمثل العقدة التربيعية ذات الترتيب s، و w_s: الوزن ذي الترتيب s عندما s=0,1,…,v. وفي العادة تكون الأوزان معتمدة على العقد والتي ينبغي أن تكون مستقلة عن الدالة $f(x)$. مع ملاحظة أن $\sum_{s=0}^{v}w_s=b-a$.

تعد طريقة كاوس التربيعية Gaussian Quadrature واحدة من أهم الطرائق التربيعية والتي توصف بأنها أداة ممتازة لتقريب التكاملات، كما تمتاز بأنها تقوم باختيار العقد بطريقة تجعل المعادلة دقيقة لأعلى درجة ممكنة من متعددة الحدود [20]. تبنى طريقة كاوس التربيعية على دوال موزونة، فعلى سبيل المثال لو أن لدينا دالة الوزن w معرفة على الفترة [a,b] وبرتبة تساوي v، حيث إن $x_1,x_2,\dots,x_v$ تمثل العقد التربيعية و $w_1,w_2,\dots,w_v$ تمثل الأوزان التربيعية فإن التكامل يقرب كما يأتي [21]:

$$\int_a^b f(x)w(x)dx\approx\sum_{i=1}^{v}w_if(x_i) \qquad (7)$$

لدالة الوزن $w(x)$ عدة أشكال شائعة وتختلف معها حدود التكامل وهذه الأوزان مع الفترات والدالة الأكثر شيوعاً والتي تستعمل في حالة كون حدود التكامل $(-\infty,\infty)$ والمسماة هرمت Hermite، حيث دالة الوزن هي e^{-x^2} وطريقة كاوس التي تستعمل هذه الدالة تسمى طريقة كاوس–هرمت التربيعية (GHQ) Gauss-Hermite Quadrature. ويمكن الحصول على متعددة حدود هرمت التي تستعمل في إيجاد قيم x_i كما يأتي [22]:

$$H_v(x)=\sum_{s=0}^{v/2}\frac{(-1)^s v!}{s!\,(v-2s)!}(2x)^{v-2s}\,,\quad v=0,1,2,\dots \qquad (8)$$

الآن، وبالعودة إلى المعادلة (7) والتعويض عن دالة الوزن بما تساويها في حالة هرمت نحصل على الآتي:

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}\,dx \approx \sum_{i=1}^{v} w_i f(x_i) \qquad (9)$$

حيث إن العقد التربيعية x_i هي الجذر ذي الترتيب i لمتعددة حدود هرمت $H_v(x)$، وإن الأوزان التربيعية w_i يتم الحصول عليها من الصيغة الآتية [23]:

$$w_i = \frac{2^{v-1} v! \sqrt{\pi}}{v^2 [H_{v-1}(x_i)]^2} \qquad (10)$$

فيما سبق درس التكامل العددي في حالة وجود تكامل أحادي، ولكن في الكثير من الأحيان تكون هناك عدة تكاملات كأن يكون تكامل ثنائي أو ثلاثي، في هذه الحالة تسمى الطرائق العددية بالطرائق التكعيبية Cubature [24]. ليكن لدينا m من المتغيرات (تكامل من البعد m)، بمعنى $\underline{x} = [x_1, \ldots, x_n]^T$ بدالة $f(\underline{x})$ معرفة على خط الأعداد الحقيقية ودالة وزن $w(\underline{x})$ معرفة على خط الأعداد الحقيقية الموجبة، فأن المعادلة (7) يمكن إعادة صياغتها كما يأتي [25]:

$$\int_{R^m} f(\underline{x}) w(\underline{x}) d\underline{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_m) w(x_1, \ldots, x_m) dx_1 \ldots x_m$$

وعلى شرط أن تكون دالة الوزن معطاة بالشكل الآتي:

$$w(\underline{x}) = e^{-\underline{x}^T \underline{x}} = e^{-x_1^2} \ldots e^{-x_m^2}$$

عندئذٍ يمكن استعمال طريقة گاوس–هرمت التربيعية أحادية البعد لكل متغير من المتغيرات، مستعملين لذلك v_r من العقد في المتغير ذات الترتيب r، حيث r=1,…,m، وبذلك ينتج:

$$\int_{R^m} f(\underline{x}) w(\underline{x}) d\underline{x} \approx \sum_{i_1=1}^{v_1} w_{i_1}^{(1)} \ldots \sum_{i_m=1}^{v_m} w_{i_m}^{(m)} f\left(x_{i_1}^{(1)}, \ldots, x_{i_m}^{(m)}\right) \qquad (11)$$

ويمكن للباحثين الاستفادة من المعادلة (11) لحل التكامل (5) وكما يأتي:

ليكن $\underline{b}_i = \sqrt{2}\Delta \underline{u}_i$، حيث $\Sigma = \Delta\Delta^T$ وإن Δ هي معامل تشوليسكي المثلثي السفلي Lower Triangular Cholesky Factor للمصفوفة Σ، ستصبح كما يأتي بعد تحويل $\underline{b}_i$ كما يأتي:

$$L_i(\underline{\psi}) = \frac{1}{\pi^{\frac{q}{2}}} \int_{R^q} \gamma(\underline{u}_i)\, exp(-\underline{u}_i^T \underline{u}_i) d\underline{u}_i$$

إذ إن:

$\underline{\psi}$: يمثل متجه المعلمات.

γ: هي دالة بدلالة المتغير $\underline{u}_i$ وتكون معطاة كما في الصيغة الآتية:

$$\gamma(\underline{u}_i) = \prod_{j=1}^{r_i} L_1\left(\underline{y}_{i1j} \middle| \underline{u}_{i1}; \underline{\beta}_1\right) L_2\left(\underline{y}_{i2j} \middle| \underline{u}_{i2}; \underline{\beta}_2\right) L_3\left(\underline{y}_{i3j} \middle| \underline{u}_{i3}; \underline{\beta}_3\right)$$

وكل تكامل يمكن تقريبه كما يأتي:

$$L_i^{GQ} = \sum_{s_1=1}^{v} \omega_{s_1}^{(1)} \ldots \sum_{s_q=1}^{v} \omega_{s_q}^{(q)} \gamma(\underline{u}^{(s)}) \qquad (12)$$

باعتبار أن عدد العقد والأوزان متساو لكل بعد. مع ملاحظة أن $\underline{u}^{(s)} = \sqrt{2}\Delta \underline{d}^{(s)}$ وإن $\underline{d}^{(s)}$ تمثل العقد التربيعية أحادية المتغير من الرتبة v، حيث $s=(s_1,s_2,\ldots,s_q)$ تمثل دليل متعدد وإن:

$$\omega_{s_l}^{(l)} = \frac{1}{\pi^{\frac{q}{2}}} w_{s_l}^{(l)}, l = 1,\ldots,q$$

باعتبار أن $w_{s_l}^{(l)}$ تمثل الأوزان التربيعية للبعد l والترتيب s_l.

إلى الآن نكون قد حصلنا على قيم دالة الإمكان الهامشية بعد أن تم الحصول على قيم عددية للتكامل، وفيما بعد ولأن الهدف بالأصل هو تعظيم دالة الإمكان بالنسبة لمتجه المعلمات $\underline{\psi}$ فأنه سيتم تعظيم هذه الدالة باتباع الخوارزمية الآتية:

1. وضع t=0، واختيار قيم ابتدائية لمتجه المعلمات $\underline{\hat{\psi}}^{(0)}$ و v.
2. جعل t=t+1. ومن ثم لكل مفردة (i=1,2,…,n) يتم تقريب التكامل $L_i\left(\underline{\hat{\psi}}^{(t-1)}\right)$ بواسطة $L_i^{GQ}\left(\underline{\hat{\psi}}^{(t-1)}\right)$ باستعمال v من النقاط التربيعية.
3. تعظيم المعادلة (21) بالنسبة لمتجه المعلمات $\underline{\psi}$ والحصول على $\underline{\hat{\psi}}^{(t)}$.
4. تكرار الخطوتين الثانية والثالثة لحين حصول التقارب.

النتائج والمناقشة Results and Discussion

1. وصف البيانات

تعد الكلية Kidney من أهم الأعضاء التي يتكون منها الجهاز البولي في جسم الإنسان والتي تقوم بوظيفة تخليص الجسم من الفضلات السائلة فضلاً عن حفظ التوزان المائي والأيوني داخل الجسم. تعد الفحوصات والتحاليل المختبرية هي الوسيلة الوحيدة للكشف عن مدى صحة الكلى وقدرتها على أداء فعالياتها الحيوية، ومن أشهر هذه التحاليل، اختبار اليوريا Urea Test واختبار الكرياتنين Creatinine Test واختبار الكالسيوم Calcium Test واختبار البوتاسيوم Potassium Test [26].

تم الحصول على البيانات محل الدراسة والخاصة بثلاثين مريض (n=30) لمراجعتين متتاليتين (كل مراجعة تمثل تكرار) ($r_i=r=2$) ولثلاثة متغيرات استجابة (m=3) هي نتيجة اختبار اليوريا (Y_1) واختبار الكرياتنين (Y_2) واختبار الكالسيوم (Y_3)، بالإضافة إلى التأثير الثابت المتمثل بالبوتاسيوم (X_1) وكذلك حد المقطع، حيث إن ($p_1=p_2=p_3=2$)، والتأثير العشوائي المتمثل بالتكرار (Z_1) حيث تساوي قيمته صفراً في المراجعة الأولى والقيمة واحد في المراجعة الثانية، وأن ($q_1=q_2=q_3=1$). وبالنسبة للمتنبئ الخطي المذكور في المعادلة (4) فأنه يكون كالآتي:

$$\eta_{ikj} = \beta_{k0}x_{ikj}^{(0)} + \beta_{k1}x_{ikj}^{(1)} + b_{ik1}z_{ikj}^{(1)}$$

حيث إن $i = 1,2,\ldots,n$ و $k = 1,2,3$ و $j = 1,2$، وإن:

β_{k0}: معلمة حد المقطع لانحدار المتغير k على المتغيرات التوضيحية.

β_{k1}: معامل تأثير البوتاسيوم على متغير الاستجابة k.

b_{ik1}: التأثير العشوائي.

وبالنسبة لمتجه المعلمات $\underline{\psi}$ فيكون كما يأتي:

$$\underline{\psi} = [\beta_{10} \quad \beta_{11} \quad \beta_{20} \quad \beta_{21} \quad \beta_{30} \quad \beta_{31} \quad \sigma^2 \quad \sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \quad \sigma_{12} \quad \sigma_{13} \quad \sigma_{23}]^T$$

يتضح أن المتجه $\underline{\psi}$ يحتوي على 13 معلمة (6 معاملات انحدار و 6 مكونات تباين بالإضافة إلى تباين متغير الاستجابة الأول σ^2).

الآن يتم اختبار حسن مطابقة المتغيرات للتوزيعات المشار إليها، باستعمال اختبار مربع كاي لحسن المطابقة، والمعادلة الآتية تمثل صيغة إحصاءة هذا الاختبار:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

حيث إن k تمثل عدد الفئات أو المستويات، وإن:

χ^2: إحصاءة مربع كاي.

O_i: التكرار المشاهد، والذي نحصل عليه من العينة.

E_i: التكرار المتوقع، والذي نحصل عليه من حجم العينة مضروب بالاحتمالية عند التوزيع محل الاهتمام.

والجدول الآتي يوضح نتائج هذا الاختبار والمتمثل بقيم P-Value المرافقة لإحصاءة اختبار مربع كاي:

جدول (1): نتائج P-Value لاختبار مربع كاي

المتغير	المراجعة الأولى	المراجعة الثانية
كالسيوم	0.86	0.94
كرياتنين	0.70	
يوريا	0.18	0.38

يتبين من الجدول أعلاه أن المتغير الأول المتمثل بالكالسيوم يتبع التوزيع الطبيعي والمتغير الثاني المتمثل بالكرياتنين يتبع توزيع برنولي بينما يتبع المتغير الثالث المتمثل باليوريا توزيع بواسون، لأن قيم P-Value جميعها أكبر من 0.05.

2. مقدرات الإمكان للبيانات المدروسة

يتم في هذه الفقرة تقدير معلمات النموذج المختلط الخطي المعمم متعدد المتغيرات باستعمال طريقة الإمكان الأعظم بأسلوب مونتي كارلو تعظيم التوقيع (MCEM) لبيانات الكلية.

قد اعتمدت تقديرات نموذج الانحدار الخطي لمعلمات الانحدار بالنسبة للمتغير الذي يتبع التوزيع الطبيعي ونتائج نموذج الانحدار اللوجستي بالنسبة للمتغير الذي يتبع توزيع برنولي، ونتائج انحدار بواسون بالنسبة للمتغير الذي يتبع توزيع بواسون، وبالنسبة لمعلمة التوزيع الطبيعي σ^2 فقد اعتبرت القيمة الأولية لها التباين للمتغير الذي يتبع التوزيع الطبيعي، أما مركبات التباين فتم الاعتماد على مصفوفة التباين والتباين المشترك لمتغيرات الدراسة الثلاثة، وجميع هذه النتائج تم الحصول عليها بواسطة البرنامج الإحصائي الجاهز Minitab 20 بناءً على بيانات المراجعة الأولى. أما القيم التقديرية فتم الحصول عليها باستعمال خوارزمية GHC بواسطة برنامج تم برمجته من قبل الباحثين بلغة R لهذا الغرض.

استعمل عدد نقاط تربيعية تساوي 10 لطريقة گاوس–هرمت التكعيبية (GHC) وقد حصل التقارب عند التكرار 38. والجدول الآتي يوضح القيم الأولية والتقديرية وفق هذه الطريقة:

جدول (2): القيم الأولية والمقدرة لمعلمات النموذج بطريقة GHC

ت	المعلمة	القيمة الأولية	GHC
1	β_{10}	6.95	5.308
2	β_{11}	0.213	0.583
3	β_{20}	20.61	11.912
4	β_{21}	-5.1	-2.883
5	β_{30}	-1.44	-1.308
6	β_{31}	0.797	0.401
7	σ^2	1	0.997
8	σ_1^2	1	0.055
9	σ_2^2	0.1851	21.882
10	σ_3^2	2.4367	0.382
11	σ_{12}	-0.1325	0.751
12	σ_{13}	0.2782	-0.105
13	σ_{23}	-0.4253	-2.885

نلاحظ من الجدول أن قيمة معامل الانحدار لتأثير البوتاسيوم على الكالسيوم ($\beta_{11} = 0.583$) ذي إشارة موجبة وهذا يعني أن تأثير البوتاسيوم على الكالسيوم موجب. كما نلاحظ أن قيمة معامل الانحدار لتأثير البوتاسيوم على الكرياتنين ($\beta_{21} = -2.883$) ذي إشارة سالبة بمعنى أن تأثير البوتاسيوم على الكرياتنين سلبي. في حين يكون معامل تأثير البوتاسيوم على اليوريا ($\beta_{31} = 0.401$) ذا إشارة موجبة وهذا يدل على التأثير الموجب للبوتاسيوم على اليوريا.

3. الاستنتاجات

بناءً على الجانبين النظري والتطبيقي تم التوصل إلى:

1. يستعمل النموذج المختلط الخطي المعمم متعدد المتغيرات في حالة وجود علاقة معنوية بين متغيرات الاستجابة.
2. عدم وجود أية دراسات داخل العراق لنماذج مختلطة خطية معممة متعددة المتغيرات، مع وجود القليل في الدراسات العالمية.
3. من معاملات الانحدار يظهر أن تأثير البوتاسيوم موجب على الكالسيوم واليوريا ولكن سالب على الكرياتنين.
4. التوازن في استهلاك البوتاسيوم يساعد في الحفاظ على صحة الكلية لأنه يؤثر على أغلب العناصر المؤثرة في صحة الكلى.

4. التوصيات

بناءً على الجانبين النظري والتطبيقي إضافة إلى الاستنتاجات نوصي بما يأتي:

1. بناء نموذج متعدد متغيرات يتكون من توزيع طبيعي وتوزيعي گاما باعتبار الكالسيوم يتبع التوزيع الطبيعي وكل من اليوريا والكرياتنين يتبعان توزيع گاما.
2. استعمال طرائق أخرى لإيجاد التكامل مثل طريقة تكامل مونتي كارلو.
3. تقدير الخطأ القياسي للتقديرات ليتم بناء اختبار فرضيات على ضوئها.
4. استعمال النموذج في دراسات في مجالات أخرى.

الشكر والتقدير Acknowledgment

يود الباحثان أن يشكرا إدارة مستشفى ڤەژين الأهلية في مدينة دهوك لتوفيرها البيانات الخاصة بوظائف الكلى والتي استعملت في الجانب التطبيقي من هذا البحث.

المراجع

[1] A. Agresti, An introduction to categorical data analysis, John Wiley & Sons, 2018.

[2] J. &. N. T. Jiang, Linear and generalized linear mixed models and their applications, New York: Springer, 2021.

[3] R. V. Gueorguieva, Models for repeated measures of a multivariate response, PhD Disseration, University of Florida, 1999.

[4] R. Gueorguieva, "A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family," *Statistical modelling 1(3),* pp. 177-193, 2001.

[5] M. Z. Y. D. C. E. A. N. H. K. J .& .E. L. E. Gebregziabher ،"Joint modeling of multiple longitudinal cost outcomes using multivariate generalized linear mixed models "،*Health Services and Outcomes Research Methodology, 13 ،(1)*pp .2013 ،57-39 .

[6] A. T. Y. Y. &. L. S. L. Karl, "Computation of maximum likelihood estimates for multiresponse generalized linear mixed models with non-nested, correlated random effects," *Computational Statistics & Data Analysis,* pp. 73, 146-162, 2014.

[7] H. C. &. W. T. E. Chen, "Assessing correlation of clustered mixed outcomes from a multivariate generalized linear mixed model," *Statistics in Medicine, 34(4),* pp. 704-720.

[8] S. K. W. B. D. R. P. D. &. Z. G. O. Mikulich-Gilbertson, "On estimating and testing associations between random coefficients from multivariate generalized linear mixed models of longitudinal outcomes," *Statistical Methods in Medical Research 26(3),* pp. 1130-1145, 2015.

[9] M. A. G. M .& .J. A. A. Jaffa ،"Analysis of multivariate longitudinal kidney function outcomes using generalized linear mixed models "،*Journal of translational medicine, 13 ،(1)*pp. 1-12 .2015 ،

[10] A. A. Jaffa, "Multivariate generalized linear mixed models with random intercepts to analyze cardiovascular risk markers in type-1 diabetic patients," *Journal of applied statistics, 43(8),* pp. 1447-1464, 2016.

[11] H. C. &. W. T. E. Chen, "Approximate uniform shrinkage prior for a multivariate generalized linear mixed model," *Journal of Multivariate Analysis, 145,* pp. 148-161, 2016.

[12] M. J. Crowther, "Extended multivariate generalised linear and non-linear mixed effects models," *arXiv preprint arXiv:1710.02223,* 2017.

[13] J. E. &. K. A. T. Broatch, "Multivariate generalized linear mixed models for joint estimation of sporting outcomes," *Italian Journal of Applied Statistics, 30(2),* pp. 189-211, 2018.

[14] S. K. W. B. D. G. G. K. R. P. D. &. Z. G. O. Mikulich-Gilbertson, "Using empirical Bayes predictors from generalized linear mixed models to test and visualize associations among longitudinal outcomes," *Statistical Methods in Medical Research, 28(5),* pp. 1399-1411, 2018.

[15] J. T. C. &. B. X. Chauvet, "Component-Based Regularization of Multivariate Generalized Linear Mixed Models," *Journal of Computational and Graphical Statistics, 28(4),* pp. 909-920, 2019.

[16] J. S. &. L. R. Pelck, "Using multivariate generalised linear mixed models for studying roots development: An example based on minirhizotron observations," *arXiv preprint arXiv:2011.00546,* 2020.

[17] F. G. D. O. R. K. S. P. S. M. J. &. C. M. Achana, "Multivariate Generalized Linear Mixed-Effects Models for the Analysis of Clinical Trial–Based Cost-Effectiveness Data," *Medical Decision Making, 41(6),* pp. 667-684, 2021.

[18] B. S. &. S. A. Everitt, The Cambridge dictionary of statistics, Cambridge University Press, 2010.

[19] T. R. L. G. &. X. F. Heister, Numerical Analysis: An Introduction, Walter de Gruyter GmbH & Co KG, 2019.

[20] A. C. Faul, A Concise introduction to numerical analysis, Chapman and Hall/CRC, 2018.

[21] J. F. Epperson, An introduction to numerical methods and analysis, John Wiley & Sons, 2021.

[22] B. S. P. S. K. &. N. V. K. Koranga, Special Functions and their Applications, River Publishers, 2021.

[23] M. &. S. I. A. Abramowitz, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards Applied Mathematics Series 55. Tenth Printing, 1972.

[24] C. F. &. X. Y. Dunkl, Orthogonal polynomials of several variables, Cambridge University Press, 2014.

[25] L. T. G. Fahrmeir, Multivariate statistical modelling based on generalized linear models, Springer-Verlag, 2001.

[26] S. C. &. N. S. Litin, Mayo Clinic Family Health Book, Mayo Clinic, 2018.