مجلة

كلية التـــراث الجامعة

مجلة علمية محكمة

متعددة التخصصات نصف سنوية

العدد السابع والثلاثون

15 حزيران 2023

ISSN 2074-5621

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم (ب 3059/4) والمؤرخ في (4/7 /2014)

مقارنة بين أنموذج انحدار بواسون وأنموذج ثنائي الحدين السالب مع تطبيق عملي

الباحثة سارة عدنان ردم أ.م.د. لمياء محمد علي حميد

المهندس جاسم محمد كاظم الزبيدي

جامعة بغداد – كلية الإدارة والاقتصاد

الملخص:
يعد انموذج انحدار بواسون من النماذج اللوغاريتمية الخطية الذي يتعامل مع بيانات تكون اما معدودة او معدلات وايضاً نادرة الحدوث كعدد الرجال المصابين بسرطان الثدي او معدل الرجال المصابين بسرطان الثدي. وأيضا يعد انموذج انحدار ذي الحدين السالب هو احد النماذج العددية والذي يستعمل للتمثيل عن الحالات والظواهر التي لا يمكن التعبير عنها بالنماذج الاعتيادية ، حيث تم استعمال طريقة الامكان الاعظم لتقدير معلمات النموذجين وذلك بهدف الوصول الى افضل انموذج ، اذ سميت عينة عشوائية بسيطة حجمها (100) حاله من دائرة مدينة الطب وقد بينت النتائج من خلال استعمال معيار متوسط مربعات الخطأ (MSE) ومعيار (AIC).

Abstract:
The Poisson regression model is a linear logarithmic model that deals with data that are either numerical or average and also rare, such as the number of men with breast cancer or the rate of men with breast cancer. Also, the negative binomial regression model is one of the numerical models that is used to represent cases and phenomena that cannot be expressed in the usual models, where the method of greatest possibility was used to estimate the parameters of the two models in order to reach the best model, as it was called a simple random sample of size (100). A case from the Medical City Department, and the results were shown through the use of the mean square error (MSE) standard and the (AIC) standard

المقدمة :
بعد أرتفاع نسبة الأصابة بسرطان الثدي في العراق وأزدياد الحالات للنساء وأنتشار المرض بكثرة وجب وضع دراسة تقف على أهم مسببات المرض والحد منه لذلك تم أستعمال نموذجين أحصائيين للمقارنة بينهما وهما أنموذج إنحدار بواسون وإنموذج إنحدار ثنائي الحدين السالب ومع أستعمال طريقة الإمكان الأعظم لتقدير معلمات كلا النموذجين وتطبيق معيار متوسط مربعات الخطأ (MSE) وإيضاً معيار الاكاكي (AIC) لمعرفة الأنموذج الأفضل بينهما ومن الدراسات التي تناولت هذه النماذج :

- في عام (2013) م قدم (صبري) [3] بحثاً حول مقارنة طرائق تقدير معلمات انموذج انحدار بواسون في حال وجود مشكلة التعدد الخطي للبيانات حيث قدم طريقة انحدار الحرف وطريقة مقدرات ليو ,وايضاً قدم طريقتين مقترحة حيث استعمل المحاكاة لتوليد البيانات باستعمال اسلوب مونتي – كارلو حيث اعتمد على متوسط مربعات الخطأ كمعيار للمقارنة بين الطرائق المستخدمة لتقدير معلمات الانموذج واظهرت نتائج المحاكاة تفوق الطريقتين المقترحتين للتقدير وتم تطبيق الطريقة المقترحة الاولى على البيانات الحقيقية .
- في عام (2017) م قدمت (العوادي) [1] بحثاً حول تقدير معلمات أنموذج انحدار بواسون الهرمي الجزئي على حالة وفيات الأمهات وكانت الدراسة تهدف الى أهم العوامل المؤثرة في زيادة نسبة الوفيات للأمهات بعد الولادة إذ تم استعمال انحدار بواسون و انحدار بواسون الهرمي لتقدير المعلمات وكانت طرائق التقدير المتاحة هي طريقة الإمكان الأعظم بالنسبة الى انحدار بواسون أما بالنسبة الى انحدار بواسون الهرمي ثم استخدام طريقة الإمكان الأعظم الكاملة وطريقة بيز , وأظهرت النتائج أن طريقة الإمكان الأعظم الكاملة هي الفضلى في حالة استعمال أنموذج أنحدار بواسون الهرمي الجزئي حيث يمكن

تطبيقه على العينات الصغيرة وأن تسلسل الحمل يؤثر بشكل عكسي على عدد الوفيات أما باقي المتغيرات فكان تأثيرها إيجابي على نسبة الوفاه وأن أكثر المؤشرات خطراً على زيادة نسبة الوفاة هو النزف ما قبل الولادة و أيضاً النزف الدماغي.

- في عام (2018) م قدم (عدنان طعمه) [2] بحثاً حول تقدير أفضل أنموذج لتوزيع ثنائي الحدين السالب وذلك لمعرفة التشوهات الخلقية متعددة الأنواع والشدة وتم أستعمال أربع طرائق لتقدير معلمات الإنموذج (طريقة الإمكان الأعظم ,طريقة الإمكان الأعظم الموزونة ,طريقة المربعات الصغرى الموزونة , طريقة المربعات الصغرى الموزونة المكررة) ومن خلال المقارنة بين نتائج هذه الطرائق أثبتت أن طريقة المربعات الصغرى الموزونة المكررة هي الفضلى بين الطرائق المستعملة إذ تمتلك أقل متوسط مربعات للخطأ وأعلى معامل تحديد وجاءت التوصيات باعتماد هذه الطريقة في حالة البيانات الطبية والبيانات المعدودة .
- في عام (2018) م قدمت (Avci) [14] بحثاً حول استخدام نماذج العد لتحديد العوامل التي تؤثر على عدد الأشخاص المصابين بالانفصام والذي يدعى (Schizophrenia) حيث تناول البحث ثلاثة نماذج عددية وهي انموذج انحدار بواسون وانموذج انحدار ثنائي الحدين السالب وانموذج انحدار كونواي ماكسويل – بواسون وباستخدام معيارين للمقارنة أولاً معيار (Log-likelihood) وثانياً معيار (Akakie) ولعينة مكونة من (205) مشاهدة مأخوذ من عام 2011 ولغاية (2014) وقد كانت النتائج الفضلى لنموذج انحدار ثنائي الحدين السالب إذ كان الأفضل من خلال معياري المقارنة .
- في عام (2019) م قدمت (غادة يوسف) [4] بحثاً حول اختبار تغيرات في أنموذج انحدار بواسون وكان هدف البحث الى استعمال خوارزمية الأعشاب الضارة ومقارنتها مع عدت طرق في تحديد المتغيرات في أنموذج انحدار بواسون وذلك باستعمال المحاكاة والبيانات الحقيقة التي تم أخذها من مرضى مصابين في العجز الكلوي المزمن والذين يتعالجون بالغسيل المستمر إذ تم استعمال طريقة مونت – كارلو في المحاكاة لتوليد البيانات التي تتبع أنموذج انحدار بواسون وأظهرت النتائج أن استعمال خوارزمية الأعشاب الضارة هي أكثر دقة من الطرائق الأخرى .

أنموذج انحدار بواسون [3] [1] :

يعد أنموذج أنحدار بواسون من أهم نماذج الأنحدار اللوغارتمية الخطية وهو يعتبر أحد الوسائل المهمة لنمذجة المتغير المعتمد عندما تكون قيم المتغير قابلة للعد حيث يفترض هذا النموذج أن المتغير المعتمد (y_i) هو متغير استجابة يتبع توزيع بواسون بمعلمة مقدارها (θ) وأيضاً تتبع الأخطاء العشوائية في أنموذج توزيع بواسون معلمة مقدارها (θ) [7] [6].

حيث يتم الأعتماد عليه بشكل كبير وفي هذا الأطار فأن الدالة الأحتمالية ستكون معرفة حسب الصيغة الأتية [10] [8] [7] :

$$Y=e^{x\underline{\beta}+u}$$

إذ أن :

y : موجه المتغير المعتمد ذي درجة $n\times1$.

X : مصفوفة المتغيرات المستقلة ذات الدرجة $n\times(p+1)$.

$\underline{\beta}$: موجه المعلمات للنموذج ذي الدرجة ($n\times1$).

u: موجه الأخطاء العشوائية ذات الدرجة ($n\times1$) .

أفتراضات أنموذج أنحدار بواسون :

يعتمد أنموذج أنحدار بواسون على ثلاثة أفتراضات أساسية يتم الأعماد عليها لبناء هذا الإنموذج :

<u>الأفتراض الأول :</u>

أن الدالة الأحتمالية الشرطية للمتغير المعتمد (Y) تتبع توزيع بواسون عندما تكون معلمة التوزيع (θ) تتبع توزيع بواسون بمعلمة مقدارها (θ) [8] [7] .

<u>الأفتراض الثاني :</u>

إن معلمة التوزيع للمتغير المعتمد (y) تكون مساوية إلى [10] [9] [8] [7] :

$$\mu_i = e^{x_i`\underline{\beta}}$$

إذ أن :

$x_i`$: تمثل الصف (i) لمصفوفة المتغيرات المستقلة X .

<u>الافتراض الثالث :</u>

إن أزواج المتغيرين (yi,xi) تكون بينها استقلالية [6] أن مع اعتماد هذه الافتراضات الثلاثة بالإضافة إلى خصائص توزيع بواسون سوف يكون الوسط الحسابي والتباين كما يلي :

$$E\,(y_i/x) = \theta_i = e^{\acute{x}_i \underline{\beta}}$$
$$var\,(y_i/x) = \theta_i = e^{\acute{x}_i \underline{\beta}}$$
$$E\,(y_i/x) = var(y_i/x) = \theta_i = e^{\acute{x}_i \underline{\beta}}$$

انحدار ثنائي الحدين السالب [13] [12] [11] : (Negative Binomial Regression)

يعد أنموذج انحدار ثنائي الحدين السالب هو أحد النماذج العددية وهنا يأخذ المتغير المعتمد (y) مشاهدات عددية صحيحة غير سالبة $(0,1,2,\dots)$.

أما في توزيع ثنائي الحدين السالب يعتمد على معلمتين هما معلمة النجاح (P) ومعلمة الفشل (q) وغالباً ما يكون التباين أكبر من المتوسط لهذا التوزيع .حيث إن الدالة الاحتمالية لتوزيع ثنائي الحدين السالب ستكون كما يلي [11]:

$$y \sim NB(p, r)$$
$$p(x = y) = \binom{x + r - 1}{y} p^y q^\alpha$$

ويمكن الحصول على توزيع ثنائي الحدين السالب من التوزيع المركب بين توزيع كاما وتوزيع بواسون وذلك بعملية خلط التوزيعات, إذ أن المتغير y يتبع توزيع كاما بمعلمات (r, β) [13] [12] :

$$y \sim Gam(r, \beta)$$

وستكون الدالة الاحتمالية كما يلي :

$$f(y) = \frac{y^{r-1}\beta^r e^{-\beta y}}{\Gamma(r)}$$

وبما وان :

$\Gamma(r) = (r-1)!$ فأن $r \in N$ وأن $\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} d_y$

وعند استعمال توزيع كاما مع توزيع بواسون نحصل على :

$$p(x = y) = \int_0^\infty \frac{e^{-r}\mu^y}{y!} \frac{\mu^{r-1}\beta^r e^{-\beta\mu}}{\Gamma(r)}\, d_\mu$$

$$= \frac{\beta^r}{y!\ \Gamma(r)} \int_0^\infty e^{-\mu(1+\beta)} \mu^{y+r-1} d_m$$
$$= \frac{\beta^r}{y!\,(r-1)!}\ (y + \beta - 1)!\, \frac{1}{(1+\beta)^{y+r}}$$
$$= \binom{y + r - 1}{y}\ (\frac{1}{1+\beta})^y\ (\frac{\beta}{1+\beta})^y$$

أذ إننا سنحصل على توزيع ثنائي الحدين السالب بمعلمات $\left(r, \frac{1}{1+r}\right)$

حيث أن :

$$E(y) = \frac{r}{\beta} \qquad var(y) = \frac{(1+\beta)^r}{\beta^2}$$

لذلك نلجأ الى توزيع خليط (بواسون _ كاما)وهذه الصيغة هي الأكثر شيوعاً له لأنها تسمح بنمذجة عدم تجانس بواسون باستعمال توزيع كاما وكما يلي [13] :

$$f(Y = y/\mu, r) = \frac{\Gamma\,(y + \beta)}{\Gamma\,(y+1)\Gamma(\beta)}\ \ (\frac{\beta}{\beta + \mu})^\beta\ (\frac{\mu}{\beta + \mu})^y$$

حيث إن :

$$\beta = \frac{1}{r}$$

$r > 0$ وهي معلمة التشتت
$\mu > 0$ الوسط الحسابي ل (y)
حيث إن :

$E(y)=\mu$
$V(y)=\mu+\frac{\mu^2}{\beta}$

التقدير (Estimation) :

- أنموذج انحدار بواسون :

طريقة الإمكان الأعظم : (Maximum Likelinood Method)

من أجل عملية تقدير معلمات إنموذج انحدار بواسون والتي تم ذكرها في المعادلة (2-7) واعتمادا على الفرضيات التي تم توضيحها سابقاً فأن حد الخطأ العشوائي يتبع توزيع بواسون لمعلمة مقداره (μ_i)وأن المتغير المعتمد (y_i)يتبع أيضاً توزيع بواسون لمعلمة مقدارها (μ_i)فسوف تكون دالة التوزيع للمتغير المعتمد (y_i)بالشكل التالي [10] :

$$P(y=y_i)=\frac{e^{-\mu}\theta i^{yi}}{yi!}$$

إذ سيكون متوسط مربعات الخطأ لتقدير معلمات انموذج انحدار بواسون وفق طريقة الإمكان الأعظم (MLE) كما يلي [9] [10] :

$$MSE(\hat{\beta}pML)= E(\hat{\beta}pML-\beta)`(\hat{\beta}pML-\beta)$$

- أنموذج انحدار ثنائي الحدين السالب :

طريقة الإمكان الأعظم [13] [5] : (Maximum Likelinood Method)

$$L=\sum_{i=1}^{n}[\mathrm{Ln}\,(y_i+{}^{1}/_{\alpha})]-\mathrm{Ln}\left[\left(\frac{1}{\alpha}\right)\right]-\mathrm{Ln}$$
$$[\Gamma\,(y_i+1)]-\frac{1}{\alpha}\,\mathrm{Ln}\,(1+\alpha\mu_i\ )-y_i\mathrm{Ln}\,(1+\alpha\mu_i)+y_i\,\mathrm{Ln}(\alpha)$$
$$+y_i\,\mathrm{Ln}\,(\mu_i)$$

وبأخذ اللوغاريتم الطبيعي للصيغة أعلاه والاشتقاق بالنسبة إلى موجة المعلمات (β) (α) ومساوة المشتقة للصفر نحصل على :

$$\sum_{i=1}^{n}\left[\alpha^{-2}\left(\mathrm{Ln}(1+\alpha\mu_i)-\sum_{j=0}^{yi-1}\frac{1}{j+\alpha^{-1}}\right)+\frac{y_i-\mu_i}{\alpha(1+\alpha\mu_i)}\right]$$

التطبيق العملي :

تم استحصال البيانات الحقيقة الخاصة بمرضى سرطان الثدي للنساء وبما يخص هذه الدراسة من مستشفى الأورام التعليمي ومستشفى الأمل التخصصي وعلى مدى خمس سنوات من (2018) م لغاية (2022) م وصف بيانات البحث:
بعد زيارة وزارة الصحة والدوائر المتعلقة بها وكذلك مركز الأبحاث السرطانية ودائرة مدينة الطب تم الحصول على البيانات المطلوبة للدراسة (سرطان الثدي) في بغداد وقد تم الحصول على عينة مكونة من (100) مشاهدة حيث سيتم التعرف بشكل أدق على كل متغير من المتغيرات التي تم العمل عليها في الأنموذج .

Y : يمثل عدد الغدد اللمفاوية المصابة و المستحصلة لكل مريضة خلال فترة زمنية معينة .
β_0:معلمة الحد الثابت .
$\beta_6, \beta_5, \beta_4, \beta_3, \beta_2, \beta_1$:معلمات الميل الحدي في الأنموذج .
x_1: يمثل نوع الرضاعة (طبيعي , غير طبيعي) .
x_2: يمثل الأجهاض (وجود أجهاض , عدم وجود أجهاض) .
x_3: يمثل العمر عند البلوغ .
x_4: يمثل العامل الوراثي للعائلة أي وجود أحد أفراد العائلة مصابين بمرض سرطان الثدي (مصاب , غير مصاب) .

x_5: يمثل العامل الوراثي للعائلة أي وجود أحد أفراد العائلة مصابين بأنواع أخرى من السرطانات (مصاب , غير مصاب).

x_6:: يمثل مرحلة المرض (أولى , ثانية , ثالثة , رابعة) .

اختبار شابيرو ويلك (Shapiro-Wilk Test) :

هو اختبار احصائي يستخدم للكشف عن البيانات التي تحقق فرضية التوزيع المطلوب حيث ستكون الفرضية للتوزيعين كما يلي :

الفرضية الاولى :(توزيع بواسون)

H_0: البيانات تتبع توزيع بواسون .

H_1: البيانات لا تتبع توزيع بواسون .

الفرضية الثانية : (توزيع ثنائي الحدين السالب)

H_0: البيانات تتبع توزيع ثنائي الحدين السالب.

H_1: البيانات لا تتبع توزيع ثنائي الحدين السالب.

ويتم اعتماد الفرضية المناسبة من خلال احصاءه الاختبار $w < w(\alpha, n)$ حيث ان (w) هو قيمة الاختبار المستخرجة وكما يلي :

$$w = \frac{(\sum_{i=1}^{\frac{n}{2}} a_i(x(n-i+1)X(i))^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

حيث ان :

a_i : قيمة ثابتة تستخرج من جداول خاصة بالاختبار .

x_i : هي قيمة المتغير x مرتبة تصاعدياً .

$\bar{x}$: هو الوسط الحسابي .

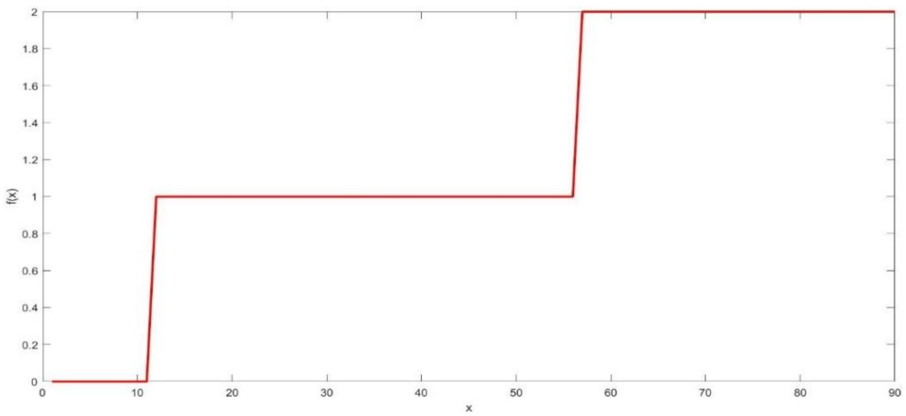
اما $w(\alpha, n)$ هي القيمة الجدولية الخاصة بالاختبار ففي حالة $w < w(\alpha, n)$ ترفض فرضية العدم .

اما الطريقة الثانية من خلال القيمة الاحتمالية (P-Value) فاذا كانت اصغر من $(\alpha = 0.05)$ فأننا نرفض فرضية العدم اما اذا كانت قيمة (P-Value) اكبر من $(\alpha = 0.05)$ فأننا لا نرفض فرضية العدم .

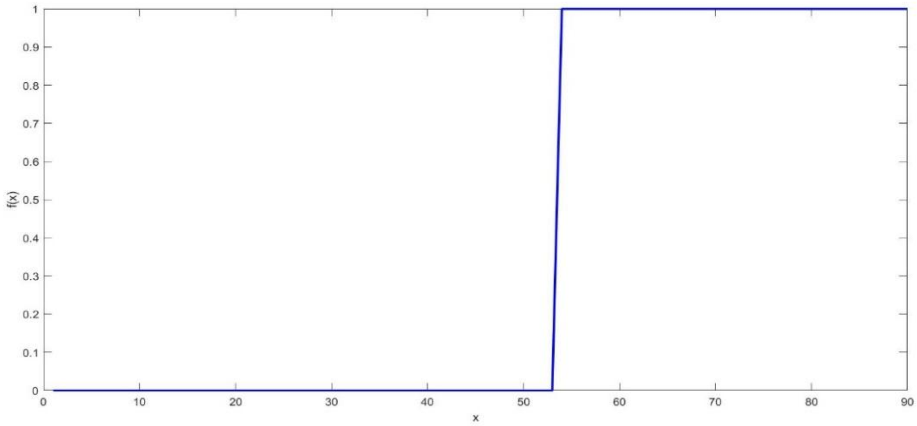
P-value	قيمة الاختبار	Shapiro-Wilk
0.7807	0.7810	Poisson
0.1312	0.6248	Negative Binom

الجدول يمثل اختبار البيانات للنموذجين

أظهرت نتائج الاختبار في جدول (4-1) ان قيمة الـP-value واختبار Shapiro-Wilk اكبر من مستوى المعنوية (0.05) مما يؤدي الى قبول فرضية H_0 .



الشكل يوضح منحى تطابق Poisson للبيانات الحقيقية



الشكل يوضح منحى تطابق Negative Binom للبيانات الحقيقية

Distribution	poisson		Negative Binom	
	poisson	Negative Binom	poisson	Negative Binom
β_0	0.181232	-1.71051	0.11324	-4.64415
β_1	-0.00702	-0.01257	-0.0154	-0.0254
β_2	0.149203	1.59037	0.187822	4.51062
β_3	0.025251	0.172825	0.03056	0.294437
β_4	0.178588	0.833882	0.204865	0.99215
β_5	-0.00738	0	-0.02213	7.07E-06
β_6	0.133081	-0.12299	0.153017	-0.28343
MSE	0.098879	0.29323	1.402559	48.66927
AIC	23.53852	18.64313	22.45498	26.07083

من خلال الجدول المذكور انفاً نلاحظ ان اقل MSE كان لا نموذج (Poisson)للبيانات التي تتبع توزيع (Poisson) اما بالنسبة للبيانات التي تتبع توزيع (Negative Binom) فقد كان اقل MSE لا نموذج (Poisson) وقد حقق معيار (AIC) اقل قيمة لا نموذج (Negative Binom) للبيانات التي تتبع توزيع (Poisson) وان البيانات التي تتبع توزيع(Negative Binom) كان اقل (AIC) لا انموذج (Poisson) وبذلك يكون (Poisson) قد حقق تطابق مع البيانات الحقيقة لمعيار (AIC) في النموذجين . وان افضل انموذج هو انموذج بواسون وتحت توزيع (Negative Binom) وذلك لتحقيقه اقل (MSE) واقل (AIC).

الاستنتاجات :

1) ان افضل انموذج هو (Poisson) من خلال معيار (MSE) فقد حقق اقل متوسط مربعات الخطأ (MSE) تحت توزيع (Negative Binom) (Poisson) .
2) ان اقل اكايكي (AIC) كان لا نموذج (Pisson) وايضاً تحت توزيع (Pisson,Negative Binom) فقد حقق تطابق واضح مع البيانات الحقيقة للنموذجين.
3) يوصى بتطبيق توزيع وانموذج بواسون على البيانات التي تعد من بيانات العد وذلك لجودته .
4) توسيع نطاق البحث بأخذ متغيرات توضيحية اخرى مثل (التدخين).

المصادر :

اولاً المصادر العربية :

1) العوادي، ايثار حسين جواد،(2017) , " مقارنة بعض طرائق تقدير معلمات انموذج انحدار بواسون الهرمي الجزئي مع تطبيق عملي " ،رسللة ماجستير ، قسم الاحصاء، كلية الادارة والاقتصاد،جامعة بغداد.
2) طعمة، عدنان فاضل، (2018),"اختيار افضل تقدير انموذج انحدار لتوزيع ثنائي الحدين السالب مع تطبيق عملي" ، رسالة ماجستير، قسم الاحصاء، كلية الادارة والاقتصاد ،جامعة كربلاء.
3) صبري ، حسام موفق،(2013)," مقارنة طرائق تقدير معلمات انموذج انحدار بواسون في ظل وجود مشكلة التعدد الخطي مع تطبيق عملي، اطروحة دكتوراه، قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.
4) عبدالله ، غادة يوسف اسماعيل،(2019), " اختيار المتغيرات في انموذج انحدار بواسون باستخدام خوارزمية الاعشاب الضارة" ، رسالة ماجستير، قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.
5) محمد , نور اياد ,(2017) , " تقدير معلمات توزيع بواسون مع تطبيق عملي",رسالة ماجستير,كلية الادارة والاقتصاد , جامعة بغداد .

ثانياً : المصادر الاجنبية :

6) Long, J . S (1997) "Regression Models for Categorical and Limited Independent Variables " ,SAGE Publicayion Inc, USA.
7) Winkelmann, R (2008) "Economic Analysis of Count Data ", 5 th ed., Springer, Verlag Berlin Heidelberg Germany.
8) Cameron , A .C & Trivedi, P .K (1998) "Regression Analysis of Count Data ", Cambridge University Press, New York USA.
9) Mansson, K & Kebria, B . M & Sjolander , P & Shukur, G (2012) New Liu Estimates for the Poisson Regression Model : Methods and App ;ication.
10) Mansson, K & Shukur , G (2011) A Poisson Ridge Regression Estimator " , Economic Modeling, Vol. 28, Issue 4 ,pp. 1475-1491 .
11) Emmmanual , C . "Modelling and data analysis , Acomparison of Poisson or Negative Binomial Regression and Lee-Carter Models of Forecasting Norwegian Male mortality , Master thesis for faculty of Mathematics and Natural sciences , University of Oslo (2015).
12) Hilbe, J .M ., log negative binomial regression as a generalized linear model , technical report COS 93/94-5-26 ,department of sociology , Arizona state university (1993)
13) Hilbe , J . M ."Negative Binomial Regression , second edition , Jet Propulsion Laboratory , California Institute of Technology and Arizona State University , Cambridge University Press(2011).
14) Avci , Esin , "Using Count Regression Models to Determine the Factors Which Effects the Hospitalization Number of people with Schizophrenia " ,Giresun University , (2018)