AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

تقدير أنموذج انحدار بواسون المبتور صفريا بطريقة المربعات الصغرى معادة الوزن التكرارية مع التطبيق وبعض خصائص التوزيع

أ.م.د. ابتسام كريم عبد الله	محمد صبري الزبيدي
ekabdullah@coadec.uobaghdad.edu.iq	mohammed.sabri2101m@coadec.uobaghdad.edu.iq

قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد، بغداد، العراق

المستخلص

توزيع بواسون المبتور صفريا Zero Truncated (ZTPD) Poisson Distribution هو حالة خاصة من توزيع بواسون الاعتيادي يحدث عند خلو توزيع بواسون من القيمة الصفرية لبعض الظواهر مما يؤدي الى تغير في الدالته الاحتمالية و خصائصه الاحصائية .

يهدف هذا المبحث الى إيجاد نمذجة تلائم تاثير قيم البيانات الخالية من القيمة الصفرية لاي ظاهرة و توظيفها لبناء انموذج انحدار بواسون المبتور صفريا بطريقة المربعات الصغرى معادة الوزن التكرارية Iteratively re-weighted least squares (IRWLS) و توظيفها في تقدير أنموذج و معلمات إنحدار بواسون المبتور صفرياً -Zero Truncated Poisson Regression Model (ZTPRM) مع ذكر بعض خصائص التوزيع . اما في الجانب التطبيقي فقد تم سحب عينة من 68 مريض من مرضى الفشل الكلوي و الذين يتعالجون بالغسيل الكلوي في مستشفى الكرامة / مركز الحياة للغسيل الكلوي.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 22/2/2024

تاريخ قبول البحث: 12/4/2024

تاريخ رفع البحث على الموقع: 31/12/2024

الكلمات المفتاحية:

توزيع بواسون المبتور صفريا، طريقة المربعات الصغرى معادة الوزن التكرارية، مصفوفة فشر، العائلة الاسية.

للمراسلة:

محمد صبري الزبيدي

mohammed.sabri2101m@coadec.uobaghdad.edu.iq

DOI: https://doi.org/10.55562/jrucs.v56i1.23

1. المقدمة

قُدِمَ توزيع بواسون (عام 1837) [2] من قبل عالم الرياضيات و الفيزياء الفرنسي المشهور سيمون دنيز بواسون 1840-1781 (Simeon Denis Poisson) [5] . و هو توزيع إحتمالي متقطع يعبر عن إحتمالية حصول عدد من الحوادث نادرة الحدوث أو غير متوقعة الحدوث يكون فيها إحتمال النجاح ضعيفاً، ضمن مدة زمنية محددة و لعدد كبير من المحاولات، كحوادث القطارات و عدد الوحدات المعيبة في إنتاج واسع لمصنع معين و غيرها. اما بالنسبة لتوزيع بواسون المبتور صفريا فيُعد الباحثان (1952) David F. وJohnson N. أول من قدم توزيع ZTP [3] على أنه أنموذج لبيانات العد الخالية من القيمة الصفرية بالنسبة للمتغير العشوائي ($x_i = 1,2, \dots \dots \dots$) فكلما زاد القطع في أي مدة من مدد المتغير العشوائي زاد التوزيع تعقيدا [7]).

2. الدالة الاحتمالية و التجميعية لتوزيع بواسون المبتور صفرياً P.M.F. and C.D.F. of (ZTP)

معلوم ان دالة الكتلة الاحتمالية لتوزيع بواسون الاعتيادي هي[1] :

$$P_{(x)} = \frac{e^{-\lambda}\ \lambda^{x}}{x!} \quad , \ \lambda > 0 \ , \quad x_i = 0,1,2,3, \dots \dots \qquad (1)$$

ومنه يمكن اشتقاق دالة الكتلة الاحتمالية لتوزيع بواسون المبتور صفرياً [6] و هي[5] :

$$P_{(x)} = \frac{e^{-\lambda}\ \lambda^{x}}{x!\,(1-e^{-\lambda})} \quad , \lambda > 0 \quad , x_i = 1,2,3, \dots \dots \qquad (2)$$

وله الدالة التجميعية CDF هي:

$$F(x) = \frac{e^{-2\lambda}}{1-e^{-\lambda}} \left[\frac{\Gamma(x+1,\lambda)}{\Gamma(x+1)}\right] ، \; where \;\; \Gamma(x+1,\lambda) = \int_{\lambda}^{\infty} t^x \, e^{-t} \, dt \qquad (3)$$

3. بعض خصائص توزيع بواسون المبتور صفرياً: Some Properties of (ZTPD)

تم إيجاد بعض خصائص هذا التوزيع كما يأتي:

1. الوسط الحسابي لتوزيع ZTP :

$$E(x) = \frac{\lambda}{1 - e^{-\lambda}} \simeq \mu_i \qquad (4)$$

2. التباين لتوزيع ZTP :

$$Var(x) = \frac{\lambda}{(1 - e^{-\lambda})} \left[\frac{\lambda}{(1 - e^{-\lambda})}\right]^2 \simeq \sigma^2 \qquad (5)$$

3. الدالة المولدة للعزوم لتوزيع ZTP :

$$\mu_x^{(t)} = \frac{e^{\lambda(e^t-1)} - e^{-\lambda}}{1-e^{-\lambda}} \qquad (6)$$

4. الوسيط :

$$Median = \frac{e^{-\lambda}}{(1-e^{-\lambda})}\left[e^{-\lambda}\frac{\Gamma(x_{me}+1,\lambda)}{\Gamma(x_{me}+1)} - 1\right] \geq \frac{1}{2} \qquad (7)$$

5. الالتواء :

$$a = \frac{(\lambda^2+3\lambda+1)^2(1-e^{-\lambda})}{\lambda\ (\lambda+1)^3} \qquad (8)$$

6. التفرطح :

$$B = \frac{[\lambda^3+4\lambda^2+3\lambda+1]\big(1-e^{-\lambda}\big)}{\lambda\,(\lambda+1)^2} \qquad (9)$$

4. طريقة المربعات الصغرى معادة الوزن التكرارية[4]:

Iteratively re-weighted least)squares (IRWLS

تُعد هذه الطريقة جزء من طريقة الإمكان الأعظم التي يمكن إستعمالها لتقدير معلمات أنموذج إنحدار بواسون المبتور صفرياً من خلال توظيف خوارزمية درجة فشر (Fisher scoring) لإيجاد مقدرات معلمات الأنموذج. تعتمد هذه الطريقة على إعادة صياغة دالة الكتلة الإحتمالية لتوزيع ZTP لجعله ينتمي إلى العائلة الأُسية (Exponential family) حسب صيغة المعادلة (10)

$$P(y\,;\theta\,,\phi) = exp\left[\frac{\theta y - b(\theta)}{\alpha\,(\phi)} + C(y,\phi)\right] \qquad (10)$$

إذ إن :

y: متغير الاستجابة (response variable) الذي ينتمي للعائلة الأُسية. θ : معلمة الموقع (location parameter) أو المعلمة القانونية (canonical parameter). ϕ : معلمة القياس (scale parameter) و هذه المعلمة تظهر فقط في التوزيعات التي تحتوي على معلمتين أو أكثر ، و بما أن توزيع بواسون ضمن العائلة الأسية ذات المعلمة الواحدة إذن سوف لن تظهر هذه المعلمة في دالة الكثافة الإحتمالية الأسية لتوزيع بواسون. $b(\theta)$: دالة بدلالة معلمة الموقع. $\alpha(\phi)$: دالة بدلالة معلمة القياس. $C(y;\phi)$: دالة بدلالة المشاهدات و معلمة التشتت. و بأخذ اللوغاريتم الطبيعي للمعادلة (2-21) نستطيع الحصول على دالة العائلة الأُسية لتوزيع ZTP و كما يأتي:

$$log\,P(y) = y\,log\,\lambda - log\left[\frac{\left(1-e^{-\lambda}\right)}{e^{-\lambda}}\right] - lo\,g(y!) \qquad (11)$$

و بأخذ الـ exp لكلا الطرفين في المعادلة (11) نحصل على:

$$P(y,\theta,\phi) = exp\left\{y\,log\,\lambda - log\left[\frac{\left(1-e^{-\lambda}\right)}{e^{-\lambda}}\right] + [-log(y!)]\right\} \qquad (12)$$

نلاحظ أن دالة الكتلة الإحتمالية لتوزيع ZTP شبيهة بالصيغة العامة للعائلة الأُسية معادلة (13).

$$\therefore P(y,\theta,\phi) = exp\{y\,\theta - b(\theta) + C(y,\phi)\} \tag{13}$$

إذ إن:

$$\theta = log\,\lambda \quad \Rightarrow \quad \lambda = e^{\theta} \qquad \text{(دالة الربط)} \tag{14}$$

$$b(\theta) = log\,\left[\frac{\left(1-e^{-\lambda}\right)}{e^{-\lambda}}\right] \quad \Rightarrow \quad log\,\left[\frac{\left(1-e^{-e^{\theta}}\right)}{e^{-e^{\theta}}}\right]$$

$$\alpha\,(\phi) = 1$$

$$C\,(y,\phi) = -log(y!)$$

أصبح من الممكن إستخراج دالة التوقع بأخذ المشتقة الأولى لدالة معلمة الموقع $b'(\theta)$ أما التباين فيتم إستخراجه من خلال أخذ المشتقة الثانية لدالة معلمة الموقع $b''(\theta)$ و التان تمثلان التوقع و التباين لتوزيع بواسون المتور صفريا و على الترتيب:

من المعادلة (10) يمكن إيجاد دالة الإمكان اللوغاريتمية كما في المعادلة (15):

$$L(\theta;y,\phi) = \frac{y\,\theta - b(\theta)}{\alpha(\phi)} = y\,log\,\lambda - log\,\left[\frac{\left(1-e^{-\lambda}\right)}{e^{-\lambda}}\right] \tag{15}$$

و بإستعمال قانون السلسلة لإشتقاق المعادلة (15) كمشتقة أولى و كالآتي:

$$\frac{\partial L}{\partial \beta_j} = \left[\frac{\partial L}{\partial \lambda}\right]\left[\frac{\partial \lambda}{\partial \mu}\right]\left[\frac{\partial \mu}{\partial \eta}\right]\left[\frac{\partial \eta}{\partial \beta_j}\right] \tag{16}$$

إذ يمكن تمثيلها كالاتي:

$$\therefore \frac{\partial L}{\partial \beta_j} = [y-\mu]\,x_{ij} \tag{17}$$

المعادلة (17) تمثل المشتقة الأولى (Gradient) لمشاهدة واحدة فقط β_j ، أما n من المشاهدات فيمكن إيجادها حسب المعادلة (18) و التي يرمز لها بالرمز U في أدناه:

$$U = \frac{\partial L}{\partial \beta} = X^T A(y-\mu) \tag{18}$$

إذ إن :

$$A = W\left(\frac{\partial \eta}{\partial \mu}\right) \quad , \quad W = \frac{1}{V(y)}\left(\frac{\partial \mu}{\partial \eta}\right)^2$$

أما بالنسبة للمشتقة الثانية (Hessian) و التي يرمز لها بالرمز H فيمكن إيجادها من مصفوفة فيشر (Fisher matrix) أو مصفوفة المعلومات (The Information matrix) حسب المعادلة الآتية:

نستطيع إيجاد مقدرات المعلمة β_r عن طريق توظيف خوارزمية نيوتن رافسن لنحصل على:

$$H = -E\left[\frac{\partial^2 L}{\partial \beta_j \partial \beta_k}\right] = E\left[\left(\frac{\partial L}{\partial \beta_j}\right)\left(\frac{\partial L}{\partial \beta_k}\right)\right] = X^T W X \tag{19}$$

$$\beta_r = \beta_{r-1} + H^{-1}U \tag{20}$$

وبتعويض المعادلتين (18) و (19) في المعادلة (20) نحصل على:

$$\beta_r = \beta_{r-1} + (X^T W X)^{-1}\,X^T A(y-\mu)$$

بإستخراج معكوس المشتقة الثانية كعامل مشترك من الطرف الأيمن للمعادلة السابقة سنحصل على:

$$\beta_r = (X^T W X)^{-1}[X^T W X \beta_{r-1} + \ X^T A(y-\mu)]$$

$$X\beta = (\eta_1\,, \dots \dots, \eta_N)^T = \eta \qquad , \qquad A = W\left(\frac{\partial \eta}{\partial \mu}\right)$$

$$\beta_r = (X^T W X)^{-1}\,X^T\,W\,z. \tag{21}$$

$$z = \eta + \frac{\partial \eta}{\partial \mu}(y-\mu)$$

تدعى z بالتغير العامل "Working Variate" أو التغير التابع المعدل "Adjusted Dependent Variate"

5. الجانب التطبيقي

5.1 وصف بيانات البحث: Description of Research Data

تم أخذ عينة عشوائية مكونة من 68 مريض من مرضى العجز الكلوي في مستشفى الكرامة التعليمي / مركز الحياة لأمراض الكلى، و من هذه العينة أُخذت المعلومات الشخصية مباشرة من الملف الطبي لكل مريض و كذلك معلومات تحليل الدم (الذي يصف نوعين من التحاليل الأول يمثل تحاليل الكبد و الثاني يمثل تحليل الكلى) من السجلات المختبرية لنفس المرضى لشهر أيار لعام 2023. إذ تم تمثيل المتغيرات التوضيحية و المتغير المعتمد كما هو مبين في أدناه:

Y : يمثل المتغير المعتمد قيم حقيقية صحيحة هي عدد مرات الغسيل الكلوي بالشهر و تقاس (..... ,1,2).

X_1 : يمثل العمر و يقاس بالسنوات.

X_2 : يمثل مدة المرض و تقاس بالأيام.

X_3 : يمثل المركب (CLU) الذي يقيس نسبة الكلوكوز في مصل الدم و المسؤول عن خفض مستويات الكولسترول في الجسم لمقاومة بعض المشكلات الجلدية مثل الاكزما و الصدفية كذلك يساعدة الجسم على التخلص من السموم وإن تركيزه الطبيعي في جسم الانسان يتراوح ما بين (115 - 70) بوحدة قياس mg |dL

X_4 : يمثل إنزيم الكبد (GOT) أو (AST) هو مختصر لناقل امين الغلوماتيك للاكسالواسيتيك وظيفته هي الكشف عن وجود ضرر أو تلف في القلب أو الكبد و إن تركيزه الطبيعي في جسم الانسان يتراوح ما بين (40 - 2) بوحدة قياس U/L.

X_5 : يمثل نسبة البومين (ALB) و هو أحد إفرازات الكبد البروتينية الناقلة مثلا لمركبات الادوية وظيفته يحافظ على السوائل في مجرى الدم من التسرب لأنسجة الجسم الأُخرى و إن تركيزه الطبيعي في جسم الانسان يتراوح ما بين (5 - 3.5) بوحدة قياسg|dL.

X_6 : يمثل نسبة الفسفور (PHO) حيث يتحد الفوسفات مع الكالسيوم لبناء الاسنان و العظام و له أهمية كبيرة في إنتاج الطاقة وهو مهم جداً في تكوين الأعصاب و العضلات، و إن تركيزه الطبيعي في جسم الإنسان يتراوح ما بين (4.8 – 2.5) بوحدة قياس mg | dL.

n : تمثل حجم العينة أي 68 مريض .

ملاحظة : mg | dL تعني مليغرام| ديسيلتر، g | dL تعني غرام| ديسيلتر، U/L تعني وحدة|لتر.
الديسلتر : يمثل 100 مليلتر.

5.2 تقدير الأنموذج:

تم إستعمال طريقة IRWLS في تقدير أنموذج الإنحدار ، و قد كانت تقدير معلمات أنموذج إنحدار بواسون المبتور صفرياً كما في الجدول (1) في أدناه:

جدول (1): تقديرات معلمات أنموذج إنحدار ZTP بطريقة IRWLS

Parameters	Estimates
β_0	1.891876
β_1	-0.005213
β_2	0.000144
β_3	0.001394
β_4	0.001762
β_5	0.027897
β_6	-0.001459

وبناءً على هذه التقديرات تم تقدير معادلة الإنحدار و كالآتي:

$$\hat{Y}_i = exp\left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \hat{\beta}_5 X_{5i} + \hat{\beta}_6 X_{6i}\right)\ , i = 1, 2, , \dots n \qquad (22)$$

وبعد التعويض عن تقديرات المعلمات التي ذكرت في الجدول (1) في المعادلة (22) نحصل على معادلة الإنحدار الآتية:

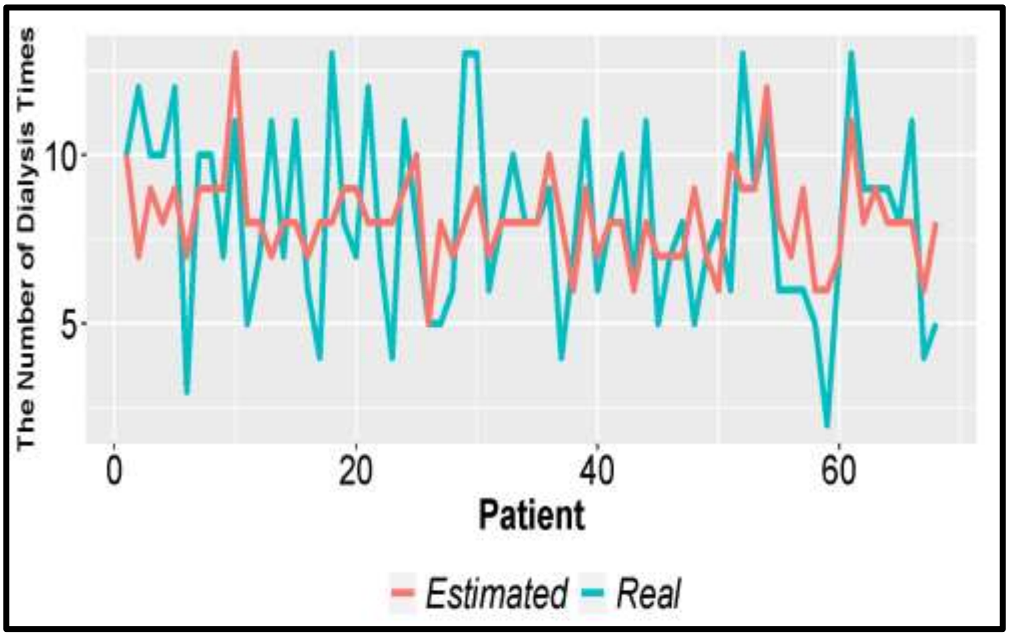
$$\begin{aligned}\hat{Y}_i = exp(&1.891876 - 0.005213X_{1i} + 0.000144X_{2i} + 0.001394X_{3i} + 0.001762X_{4i} \\ &+ 0.027897X_{5i} - 0.001459X_{6i})\end{aligned} \qquad (23)$$

و بناءً على هذه المعادلة تم تقدير مشاهدات المتغير التابع و وضعت النتائج الحقيقية مع التقديرية في الجدول (2)، و كما هو موضح في الشكل (1) الآتي:

جدول (2): القيم الحقيقية و القيم التقديرية للمتغير التابع

ت	Real	Estimated	ت	Real	Estimated	ت	Real	Estimated	ت	Real	Estimated
1	10	10	**18**	13	8	**35**	8	8	**52**	13	9
2	12	7	**19**	8	9	**36**	9	10	**53**	9	9
3	10	9	**20**	7	9	**37**	4	8	**54**	11	12
4	10	8	**21**	12	8	**38**	7	6	**55**	6	8
5	12	9	**22**	7	8	**39**	11	9	**56**	6	7

6	3	7	**23**	4	8	**40**	6	7	**57**	6	9
7	10	9	**24**	11	9	**41**	8	8	**58**	5	6
8	10	9	**25**	8	10	**42**	10	8	**59**	2	6
9	7	9	**26**	5	5	**43**	6	6	**60**	7	7
10	11	13	**27**	5	8	**44**	11	8	**61**	13	11
11	5	8	**28**	6	7	**45**	5	7	**62**	9	8
12	7	8	**29**	13	8	**46**	7	7	**63**	9	9
13	11	7	**30**	13	9	**47**	8	7	**64**	9	8
14	7	8	**31**	6	7	**48**	5	9	**65**	8	8
15	11	8	**32**	8	8	**49**	7	7	**66**	11	8
16	6	7	**33**	10	8	**50**	8	6	**67**	4	6
17	4	8	**34**	8	8	**51**	6	10	**68**	5	8



شكل (1): عدد مرات غسيل الكلية الحقيقي مع المقدر

6. الاستنتاجات:

تبين من أنموذج الانحدار المقدر المعادلة (23) أنه عند تقدم المرضى في السن بمقدار وحدة واحدة تقل قيمة Y (الذي يمثل عدد مرات الغسيل الكلوي للمرضى) بمقدار (0.005213) بثبوت باقي المتغيرات. أما عند الزيادة في نسبة الفسفور (PHO) في جسم المرضى بمقدار وحدة واحدة تقل قيمة Y بمقدار (0.001459) و بثبوت باقي المتغيرات.

في حالة طول مدة المرض بمقدار وحدة واحدة تزيد قيمة Y بمقدار (0.000144) بثبوت باقي المتغيرات. أما في حالة الزيادة في نسبة المركب (CLU) بمقدار وحدة واحدة تزيد قيمة Y بمقدار (0.001394) بثبوت باقي المتغيرات. في حالة الزيادة في إنزيم الكبد (GOT) أو (AST) بمقدار وحدة واحدة تزيد قيمة Y بمقدار (0.001762) بثبوت باقي المتغيرات. كذلك في حالة الزيادة في نسبة البومين (ALB) بمقدار وحدة واحدة تزيد قيمة Y بمقدار (0.027897) بثبوت باقي المتغيرات.

المصادر

[1] نعيمة، علي بندر و عبودي، عماد حازم ، "تقدير دوال الفشل للتوزيع الناتج من دمج توزيع بواسون ليندلي مع توزيعات أخرى"، أطروحة الدكتوراه فلسفة في الإحصاء، كلية الإدارة و الاقتصاد، جامعة بغداد,(2016).

[2] Chattamvelli, R. & Shanmugam R., “Discrete Distributions in Engineering and the Applied Sciences”, Morgan & Claypool,(2020).

[3] David F. & Johnson N., “The Truncated Poisson”, Biometrics 8(4), 275-285, (1952).

[4] Green P., “Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives”, Royal Statistical Society, Blackwell Publishing 46(2), 149 -192,(1984).

[5] Manjula D. & Uma G. et al., “Derivation of Zero –One Truncate Poisson Distribution”, International Journal of Applied Research 6(5), 253-255, (2020).
[6] Shamsur R. & Mohd Y., “Truncated Distributions and their Applications”, Dissertation submitted for the award of the degree of master philosophy in statistics in Department of statistics & Operations Research Aligarh Muslim University (India) (https://core.ac.uk/download/pdf/144520251.pdf), (2005).
[7] Taekyun K. & Kim D. et al., “A Study on Properties of Degenerate and Zero-Truncated Degenerate Poisson Random Variables”, Arxiv 2106. 13481v1 [math NT], (2021).

Journal of AL-Rafidain University College for Sciences (2024); Issue 56; pp. 258-264

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Estimation of the Zero-Truncated Poisson Regression Model using Iteratively Reweighted Least Squares Method with Application: and Some Distribution Properties

Mohammed S. Al-Zubaidi	Ebtisam K. Abdulah
mohammed.sabri2101m@coadec.uobaghdad.edu.iq	ekabdullah@coadec.uobaghdad.edu.iq

Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

Article Information

Article History:
Received: February, 22, 2024
Accepted: April, 12, 2024
Available Online: December, 31, 2024

Keywords:
Zero-Truncated Poisson Distribution, The Iterative Reweighted Least Squares method, Fisher matrix, Exponential family

Correspondence:
Mohammed S. Al-Zubaidi
mohammed.sabri2101m@coadec.uobaghdad.edu.iq

DOI: https://doi.org/10.55562/jrucs.v56i1.23

Abstract

The Zero Truncated Poisson Distribution (ZTPD) is a special case of the Poisson distribution that occurs when the Poisson distribution is devoid of zero values for certain phenomena, resulting in a change in its probability function and statistical properties. The aim of this study is to find a modeling approach that accommodates the impact of data values without the zero value for any phenomenon and to use it to build a Zero Truncated Poisson Regression Model (ZTPRM) using the Iteratively Reweighted Least Squares (IRWLS) method. This model is used to estimate the regression model and parameters for the Zero Truncated Poisson Distribution, with some distribution properties mentioned. In the applied side, a sample of 68 patients with renal failure undergoing renal dialysis at Al-Karama Hospital/Life Kidney Center was drawn.