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

استعمال مبدأ المسافة الدائرية لاكتشاف القيم الشاذة في أنموذج الانحدار الدائري

مهدي وهاب نعمة نصر الله	امال محمد جواد عبد الكاظم
mehdi.wahab@uokerbala.edu.iq	hibafarisnaji@gmail.com
قسم الإحصاء- كلية الإدارة والاقتصاد- جامعة كربلاء، كربلاء، العراق.	
تماضر كفاح حسن	
tmadrkfahhsnalhsnawy@gmail.com	
قسم الإحصاء- كلية الإدارة والاقتصاد- جامعة كربلاء، كربلاء، العراق.	

المستخلص

في هذا البحث تم استعمال مبدأ المسافة الدائرية لإنتاج إحصاءة COVA RATIO وذلك بعد حذف البيانات الشاذة من مجموعة البيانات الكاملة بعد الاخذ بنظر الاعتبار محدد مصفوفة التغاير (التباين المشترك) لغرض استعمالها في الكشف عن القيم الشاذة في البيانات الدائرية التي تمثل بيانات واقعية عن العين وهي زاوية العين التي تسبب الحول لغرض تقدير انموذج الانحدار الدائري للبيانات. وتم التوصل الى ان تطبيق احصاءة نسبة التباين قد ساعد في اكتشاف القيم الشاذة والمؤثرة على العين بشكل كامل والتي يختلف نسقها عن باقي القياسات للمرضى، حيث اعطى انموذج الانحدار الدائري ملائمة جيدة للبيانات الحقيقية في ظل المشاهدات الشاذة.

معلومات البحث

تواريخ البحث
تاريخ تقديم البحث: 2022/12/20
تاريخ قبول البحث: 2023/3/3
تاريخ رفع البحث على الموقع: 2023/12/31

الكلمات المفتاحية
قيم شاذة، انموذج الانحدار الدائري، احصاءة COVA RATIO، تقدير.

للمراسلة:
مهدي وهاب نعمة نصر الله
mehdi.wahab@uokerbala.edu.iq

https://doi.org/10.55562/jrucs.v54i1.592

1. مقدمة

يعد اول ظهور لأنموذج الانحدار الدائري على يد الباحث (Goud, 1969) وتم بعد اقتراح العديد من نماذج الانحدار الدائري. فقد طرح مفهوم الانحدار الدائري لبيانات عبارة عن اتجاهات تقاس كزوايا في مستوى مع الإشارة إلى إحساس ثابت بالدوران واتجاه ثابت للصفر, حيث ناقشا بعض الطرائق التجريبية والاختبارات المقاربة لتحديد درجة معادلة الانحدار وطورا بعض الخوارزميات العددية لايجاد معاملات الانحدار الدائري. تلتهم بعد ذلك الكثير من الدراسات والبحوث التي تناولت موضع الانحدار الدائري ودراسته من مختلف الجوانب منها الاخذ بنظر الاعتبار وجود القيم الشاذة في البيانات الدائرية ففي عام (2013) اقترح (A.H. Abuzaid) وآخرون إحصاءه عددية جديدة سميت متوسط الخطأ الدائري (Mean Circular error) وعرفت باستعمال المسافة الدائرية لغرض تحديد واكتشاف القيم الشاذة المحتملة في نماذج الإنحدار الدائري وبجانب نهج خذف الصف (Row deletion) تم الحصول على الدالة لغرض تقدير إنموذج الإنحدار الدائري وباستعمال تجارب محاكاة مكثفة تم التحقق من قوة اداء الطريقة المقترحة . واستعمل (A. Ibrahim) وآخرون طريقة المربعات الصغرى (OLS) لتقدير معلمات انموذج الانحدار الدائري واعتمدوا على إحصاءة COVARATIO وطريقة حذف الصف لتحديد واكتشاف القيم الشاذة في أنموذج الانحدار الدائري، وفي عام (2015) استعمل (Rambli) وآخرون إحصاءه COVARATIO لتحديد واكتشاف القيم الشاذة في أنموذج الانحدار الدائري الذي اقترحه كل من (Downs & Mardia, 2002) وتم تقييم اداء الطريقة عن طريق تجارب محاكاة مونت- كاروا .اما في عام (2017) اقترح (A. Mahmood) وآخرون احصاءة جديدة لتحديد القيم الشاذة المتعددة في أنموذج الانحدار الدائري البسيط تعتمد على حساب المسافة الدائرية الحصينة بين البواقي الدائرية ومعلمة الموقع الدائرية وتم تقييم اداء الطريقة المقترحة باستعمال المحاكاة على احجام عينات

مختلفة بنسب تلوث 10% و 20% من خلال نسب القيم الشاذة ومعدل الاخفاء والاغراق. في هذا البحث تم استعمال مبدأ المسافة الدائرية لانتاج إحصاءة COVA RATIO وذلك بعد حذف البيانات الشاذة من مجموعة البيانات الكاملة بعد الاخذ بنظر الاعتبار محدد مصفوفة التغاير (التباين المشترك) لغرض استعمالها في الكشف عن القيم الشاذة في البيانات الدائرية التي تمثل بيانات واقعية عن العين وهي زواية العين التي تسبب الحول لغرض تقدير انموذج الانحدار الدائري للبيانات.

2. الإحصاء الدائري (Circular Statistics) [8][9]

الأصل. في التحليل الدائري، يتم التركيز على الاتجاه فقط. ومن الأمثلة على البيانات الدائرية الاتجاهات المقاسة باستخدام أدوات مثل البوصلة أو المنقلة أو ريشة الطقس او الفرجال. degrees and radians فالمطلوب هو تحديد موقع النقطة والاتجاه وليس كما في البيانات على الخط الحقيقي ؛ فالقيم الموجودة على يسار نقطة الأصل (0) تكون سالبة والقيم الموجودة على اليمين تكون موجبة. وبالنسبة للبيانات الدائرية، تحدد كل زاوية تعرف كنقطة على محيط دائرة الوحدة، تمامًا كما تحدد كل قيمة لمتغير خطي كنقطة على الخط الحقيقي. ومع زيادة القيمة المطلقة للمتغير الخطي ، نبتعد عن نقطة الأصل.

3. القيمة الشاذة Outliers value [2] [10]

تعتبر القيمة الشاذة قيمة خارجة عن النسق المميز لمجموعة أو تركيبة معينة بحيث تشذ عن نمط باقي البيانات. وانها نقاط بيانات تكون مبتعدة عن غالبية نقاط البيانات الاخرى اي انها مشاهدات لا تنسجم مع بقية بيانات المجموعة لأي متغير من المتغيرات لظاهرة معينة أو لمجموعة من الظواهر، فقد تكون قيمة هذه المُشاهدة كبيرة أو قد تكون صغيرة واقعة في أحد طرفي مجموعة المُشاهدات المرتبة تصاعدياً أو تنازلياً، وأن شذوذها قد يكون في حالات كثيرة مسألة طبيعية ملازمة لبعض المتغيرات. وتعرف المشاهدة الشاذة (الملوثة) إحصائياً بأنها المُشاهدة المتأتية من مجتمع مختلف عن المجتمع قيد البحث. أي أن المجتمع الأصلي تم تلويثه بمشاهدات من مجتمع أخر. والقيم الشاذة مشاهدات تنحرف كثيراً عن المشاهدات الاخرى وهي مولدة بطريقة مختلفة عن طريقة توليد المشاهدات الاصلية) .

4. احصاءة نسبة التباين COVARATIO [7][8][9]

بيلسلي وآخرون. (1980) استخدم منهجية حذف الصف للتأكد في تأثير حذف صف واحد في كل مرة على المعاملات المقدرة والقيم المجهزة والبواقي ومصفوفة التغاير لنماذج الانحدار الخطي. حيث اهتموا بالتأثيرات المقابلة على مصفوفة التغاير لنموذج الانحدار الدائري. وذلك للاخذ بنظر الاعتبار تأثير حذف صف واحد من بيانات الانحدار الدائري على نسبة مصفوفة التغاير المقدرة باستخدام جميع المشاهدات المتاحة عبر مصفوفة التغاير المقدرة عند حذف المشاهدة j. "COV" "COV (-j)" j :COVARATIO

$$\text{COVARATIO} = \frac{|\text{COV}|}{|\text{COV}_{-j}|} \quad (1)$$

اذ ان بهذه الاحصاءة يتم استعمال المسافة بين البيانات الدائرية للكشف عن القيم الشاذة في البيانات الدائرية فاذا كان $|\text{COVARATIO}_{(-j)} - 1|$, فأي مشاهدة بـ $|\text{COVARATIO}_{(-j)} - 1|$ سوف يتم تجاوزها وتعيين بانها قيمة شاذة .

5. انموذج الانحدار الدائري The Circular Regression Model [1][3]

تم انموذج انحدار لمتغيرين عشوائيين دائريين U و V من حيث التوقع الشرطي لـ e^{iv} بمعلومية u وكالآتي:

$$E\left(e^{iv}|u\right) = \rho(u)e^{i\mu(u)} = g_1(u) + ig_2(u) \quad (2)$$

إذ أن:

$$e^{iv} = \cos v + i \sin v \quad (3)$$

و $\mu(u)$ يمثل المتوسط الشرطي الإتجاهي للـ v بمعلومية u, و $\rho(u)$ معلمة الكثافة الشرطية للدوال $g_1(u), ig_2(u)$ والتي تكتب كالآتي :

$$E(\cos v|u) = g_1(u) \quad (4)$$

$$E(\sin v|u) = g_2(u) \quad (5)$$

فان v بحيث أن:

$$\mu(u) = \hat{v} = arctan * \frac{g_2(u)}{g_1(u)} = \begin{cases} tan^{-1}\frac{g_2(u)}{g_1(u)} & if\ g_1(u) > 0 \\ \pi + tan^{-1}\frac{g_2(u)}{g_1(u)} & if\ g_1(u) \leq 0 \\ undefined & if\ g_1(u) = g_2(u) = 0 \end{cases} \quad (6)$$

ولصعوبة تقدير $g_1(u)$ و $g_2(u)$ يقودنا الى تقريب هذه الدوال باستعمال دوال مناسبة بالأخذ بنظر الاعتبار حقيقة دوريتها ضمن الفترة 2π. لذلك سيكون التقريب باستعمال متعددة حدود مثلثية مناسبة من الدرجة m وعلى الصيغة الآتية:

$$g_1(u) \approx \sum_{i=1}^{m}(A_k cos\ ku + B_k sin\ ku) \quad (7)$$

$$g_2(u) \approx \sum_{i=1}^{m}(C_k cos\ ku + D_k sin\ ku) \quad (8)$$

وبذلك نحصل على الإنموذج الآتي:

$$\cos v = \sum_{i=1}^{m}(A_k cos\ ku + B_k sin\ ku) + \varepsilon_1 \quad (9)$$

$$g_2(u) \approx \sum_{i=1}^{m}(C_k cos\ ku + D_k sin\ ku) + \varepsilon_2 \quad (10)$$

حيث أن $\varepsilon = (\varepsilon_1, \varepsilon_2)$ هو متجه الاخطاء العشوائية التابع للتوزيع الطبيعي بمتوسط صفر ومصفوفة تباين مجهولة Σ , وان (A_k, B_k, C_k, D_k) حيث أن k=0,1,2,…..,m هي الاخطاء القياسية .

6. تقدير انموذج الانحدار الدائري [1] Estimation of the Circular Regression Model

استعملت طريقة تقدير المربعات الصغرى العامة لنماذج الانحدار الدائري وكالآتي:

ليكن $(u_1, v_1) \ \dots \ (u_n, v_n)$ عينة عشوائية دائرية بحجم n , فمن معادلة (10) فان المعادلة المشاهدة يمكن ان تكتب كالآتي:

$$v_1 j = \cos v_j = \sum_{i=1}^{m}\left(C_k cos\ ku_j + D_k sin\ ku_j\right) + \varepsilon_{1j} \quad (11)$$

$$v_2 j = \sin v_j = \sum_{i=1}^{m}\left(C_k cos\ ku_j + D_k sin\ ku_j\right) + \varepsilon_{2j} \quad (12)$$

لـ $j = 1, \dots, n$

وأن:

$$\lambda^{(1)} = (A_0, A_1, \dots, A_m, B_1, \dots \dots B_m)'$$
$$\lambda^{(2)} = (C_0, C_1, \dots, C_m, D_1, \dots \dots D_m)'$$

فان تقدير المربعات الصغرى يكون :

$$\hat{\lambda}^{(1)} = (U'U)^{-1}U'^{V^{(1)}} \quad (13)$$

$$\hat{\lambda}^{(2)} = (U'U)^{-1}U'V^{(2)} \quad (14)$$

ومن ثم نحتاج الى تقدير مصفوفة التباين المشترك Σ باستعمال نظرية المربعات الصغرى , فليكن

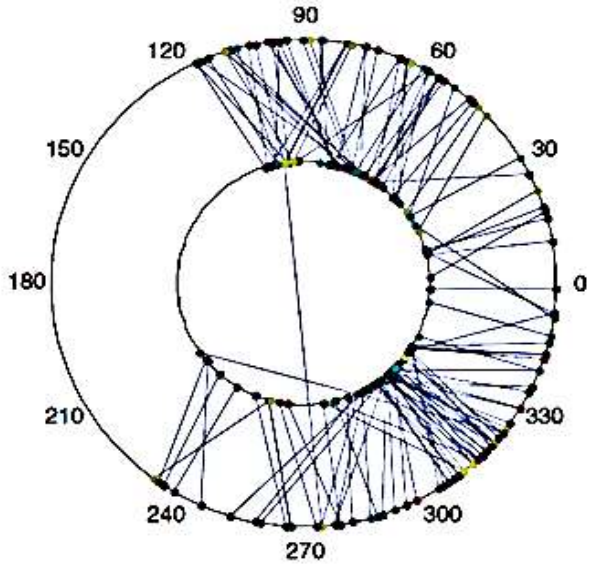
$$R_0(q,p) = V^{(p)'}V^{(q)} - V^{(p)'}(U'U)^{-1}U'V^{(q)} \quad (15)$$

حيث أن: $M = (U'U)^{-1}U'$ و $R_0 = (R_0(q,p))_{p,q=1,2}$ فان تقدير مصفوفة التباين يكون بالصيغة الآتية:

$$\widehat{\Sigma} = [n - 2(2m+1)]^{-1}\ R_0 \quad (16)$$

7. الجانب التطبيقي

تم اخذ بيانات عن العين لـ (130) مصابا بالحول والتي تمثل متغيرين دائريين مقاسان بالراديان، الأول هو زوايا انحناء العين (u) والثاني يمثل زاوية انحناء القزحية (v) والتي تم الحصول عليها من وحدة استشارية العيون في مستشفى الحسين التعليمي في محافظة كربلاء المقدسة. ويبين الشكل (1) والذي يمثل المخطط الدائري Spoke Plot للبيانات الحقيقية ان قيمة معامل الارتباط بين المتغيرين بلغت (0.8544) وهذه القيمة تشير الى ان هنالك ارتباطا قويا جدا وطرديا بين المتغيرين حيث نلاحظ ان هنالك فقط زوجين من المشاهدات يكونان على خط مستقيم ويعبران الدائرة الداخلية للمخطط الدائري بحيث يقطع احد الخطين الدائرة الداخلية بالكامل وهي المشاهدة رقم (38) بينما يطقع الآخر في وتر قصير لايعبر الدائرة الداخلية للمخطط الدائري وهي المشاهدة رقم (125) وهذا مؤشر اولي الى ان هاتين المشاهدتين من المشاهدات الفعالة والمؤثرة على الحالة المرضية للمريض.



شكل (1): المخطط الدائري Spoke plot للبيانات الحقيقية

والآن سيتم ملائمة انموذج الانحدار الدائري JS للبيانات الحقيقية، حيث ان لمعلمات المقدرة كانت كالآتي:

$$\hat{A}_0 = 0.0785$$

$\hat{A}_1 = 0.7685$
$\hat{B}_1 = -0.0877$
$\hat{C}_0 = -0.058$
$\hat{C}_1 = -0.1051$
$\hat{D}_1 = 0.8867$

وان الانموذج الملائم للداول $g_1(u)$ و $g_2(u)$ كالآتي:

$\hat{g}_1(u)$ =0.0785+0.7685cos(u)-0.0877sin(u)
$\hat{g}_2(u)$ =-0.058+0.1051cos(u)+0.8867sin(u)

وان:

$$\mu(u) = \hat{V}_j = arctan * \frac{-0.058 + 0.1051\cos(u) + 0.8867\sin(u}{0.0785 + 0.7685\cos(u) - 0.0877\sin(u)} \quad ; j = 1,2, \dots, n$$

$$\hat{\rho} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\hat{\rho}^2(u_j)} = \sqrt{\frac{1}{n}\left[{g^2}_1(u)\right] + {g^2}_2(u)} = 0.8544$$

الآن، نقوم بتطبيق إحصائية COVRATIO لاكتشاف أي من المشاهدات مؤثرة محتملة في بيانات العين.
ان محدد مصفوفة التغاير لمجموعة البيانات الكاملة | COV | =660.00
والقيم المقابلة لـ | 1− (−j) COVRATIO | تم احتسابها، اذ نجد أن المشاهدة (7) والمشاهدة رقم (38) مهمات للغاية. ويتضح ذلك من خلال الشكل (2)، الذي يوضح أن قيمة المشاهدة رقم (38) و (125) تختلفان عن القيم الأخرى.

8. مناقشة النتائج

ان تطبيق احصاءة نسبة التباين قد ساعد في اكتشاف القيم الشاذة والمؤثرة على العين بشكل كامل والتي يختلف نسقها عن باقي القياسات للمرضى، حيث اعطى انموذج الانحدار الدائري ملائمة جيدة للبيانات الحقيقية في ظل المشاهدات الشاذة.

المصادر

[1] Jammalamadaka, S. R., Sarma, Y. R. (1993). "Circular regression". In: Matsusita, K., ed. Statistical Science and Data Analysis. Utrecht: VSP, pp. 109–128.

[2] H. Abuzaid, A. G. Hussin & I. B. Mohamed , "Detection of outliers in simple circular regression models using the mean circular error statistic", Journal of Statistical Computation and Simulation Vol. 83, No. 2, February 2013, 269–277

[3] Alshqaq, Shokrya S., Abdullah A. Ahmadini , and Ali H. Abuzaid, (2021)(1) " Some New Robust Estimators for Circular Logistic Regression Model with Applications on Meteorological and Ecological Data", Hindawi Mathematical Problems in Engineering Volume 2021, Article ID 9944363, 15 pages https://doi.org/10.1155/2021/9944363.

[4] Pewsey, Arthur; Neuhuser, Markus; D. Ruxton, Graeme, (2013), Circular Statistics in R, Oxford University Press.

[5] Abuzaid, A. H., Hussin, A. G., Mohamed, I. B. (2008). "Identifying single outlier in linear circular regression model based on circular distance", Journal of Applied Probability and Statistics 3(1):107–117.

[6] Barrera , Azael, (2014), "Unit Circles and Inverse Trigonometric Functions, A method to determine all the inverse trigonometric functions directly from the unit circle", The Mathematics Teachers, Vol. 108, No. 2.

[7] Belsley, D. A., Kuh, E., Welsch, R. E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: John Wiley & Sons.

[8] Obikee Obikee, Adaku C., Ebuh; Godday U., Happiness; Obiora-Ilouno, (2014), "Comparison of Outlier Techniques Based on Simulated Data", Open Journal of Statistics, Vol. 4, No. 7.

[9] Ibrahim, S., 2013. Some Outlier Problems in a Circular Regression Model (Ph.D. thesis). Institute of Mathematical Sciences, University of Malaya.

Journal of AL-Rafidain University College for Sciences (2023); Issue 54; 211 – 215

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

Using the Circular Distance Principle to Detect Outliers in the Circular Regression Model

Mahdi W. Nea'ama Naserallah	Amal M. Abed Alkadhum
mehdi.wahab@uokerbala.edu.iq	hibafarisnaji@gmail.com
Statistics Department, Faculty of Administration and Economics, Karbala University, Karbala, Iraq	
Tamadur K. Hassan	
tmadrkfahhsnalhsnawy@gmail.com	
Statistics Department, Faculty of Administration and Economics, Karbala University, Karbala, Iraq	

Article Information

Article History:
Received: December, 20, 2022
Accepted: March, 3, 2023
Available Online: December, 31, 2023

Keywords:
Outliers, circular regression patterns, COVA RATIO statistics, estimation.

Correspondence:
Mahdi W. Nea'ama Naserallah
mehdi.wahab@uokerbala.edu.iq

Abstract

In order to detect abnormal values in the circular data that represent realistic data about the eye, which is an angle, the COVA RATIO statistic was calculated in this study using the principle of circular distance, after the abnormal data was removed from the entire data set and consideration of the covariance determinant of the covariance matrix. The amblyopia-causing eye in order to estimate the data's circular regression model. It was determined that the application of the contrast ratio statistics helped in discovering the abnormal values that affect the eye completely, and whose format differs from the rest of the measurements for patients, as the circular regression model provided a good fit for the real data in light of the abnormal observations.

https://doi.org/10.55562/jrucs.v54i1.592