Vol 06 / Issue 02 / Juin 2024 / pp 121 – 135
ISSN : 2676-234X / EISSN : 2716-9006

Social Empowerment Journal
University Amar Telidji of Laghouat-Algeria

Methods for detecting and treating outliers in simple linear regression

Mohammed Doua
Faculty of Economics, Management Sciences and Commercial Sciences, University of Laghouat, Algeria, e-mail: m.doua@lagh-univ.dz
Economic development studies laboratory

ORCID: https://orcid.org/0009-0008-7559-3214

Received: 09/01/2024; **Accepted**: 25/05/2024, **Published:** 30/06/2024

Keywords

Box plot, Sorting method, scatter Z, T statistic, Cook's distance.

Abstract

In data analysis, one or more values may appear that are sometimes far from a group of other values. These values appear irrationally compared to the rest of the data, and they can be very small or very large compared to the rest of the data. They are usually undesirable because they cause imbalances, and they are called These are outliers, in this article we discussed a concept, the reasons for their appearance, and the most important methods used to detect them in the simple linear regression technique, and in the end, we suggested some methods to treat them.

Corresponding Author : Mohammed Doua
DOI 10.34118/sej.v6i2.3924

مجلة التمكين الاجتماعي
جامعة عمار ثليجي الأغواط – الجزائر

المجلد 06 / العدد 02 / جوان 2024 / ص ص 121 – 135
ISSN : 2676-234X / EISSN : 2716-9006

طرق كشف ومعالجة القيم المتطرفة في الانحدار الخطي البسيط

محمد دوة

كلية العلوم الاقتصادية وعلوم التسيير والعلوم التجارية - جامعة الأغواط – الجزائر، البريد الإلكتروني:
m.doua@lagh-univ.dz

مخبر دراسات التنمية الاقتصادية

ORCID: https://orcid.org/0009-0008-7559-3214

تاريخ الاستلام: 2024/01/09 - **تاريخ القبول:** 2024/05/25 - **تاريخ النشر:** 2024/06/30

الكلمات المفتاحية

مخطط صندوق، طريقة الفرز، انتشار Z، احصاءة t، مسافة كوك.

الملخص

في تحليل البيانات قد تظهر قيمة واحدة أو أكثر تكون أحيانا بعيدة عن مجموعة القيم الأخرى، تظهر هذه القيم بشكل غير منطقي مقارنة ببقية البيانات، ويمكن أن تكون صغيرة جدًا أو كبيرة جدًا مقارنة ببقية البيانات، وعادة ما تكون غير مرغوب فيها لأنها تسبب اختلالات، وتسمى هذه بالقيم المتطرفة، في هذا المقال ناقشنا مفهوما، وأسباب ظهورها، وأهم الطرق المستخدمة للكشف عنها في تقنية الانحدار الخطي البسيط، و في الأخير اقترحنا بعض طرق معالجتها.

المؤلف المرسل: محمد دوة

DOI 10.34118/sej.v6i2.3924

- مقدمة:

في بعض مجموعات البيانات، توجد قيم (نقاط بيانات ملحوظة) تسمى القيم المتطرفة، القيم المتطرفة هي نقاط بيانات ملحوظة بعيدة عن خط المربعات الصغرى تحتوي على أخطاء كبيرة، حيث لا يكون الخطأ أو المتبقي قريبا جدا من أفضل خط تسوية مناسب.

يجب فحص القيم المتطرفة عن كثب في بعض الأحيان، حيث لا ينبغي تضمينها في تحليل البيانات مثل ما إذا كان من الممكن أن يكون الانحراف نتيجة لبيانات غير صحيحة، في أوقات أخرى، قد تحتوي القيم المتطرفة على معلومات قيمة حول المجتمع قيد الدراسة ويجب أن يظل مدرجا في البيانات.

إلى جانب القيم المتطرفة، قد تحتوي العينة على نقطة أو بضع نقاط تسمى النقاط المؤثرة، النقاط المؤثرة هي نقاط بيانات ملحوظة بعيدة عن نقاط البيانات الأخرى الملحوظة في الاتجاه الأفقي، قد يكون لهذه النقاط تأثير كبير على ميل خط الانحدار، للبدء في تحديد نقطة مؤثرة يمكننا إزالتها من مجموعة البيانات وتحديد ما إذا كان ميل خط الانحدار قد تغير بشكل ملحوظ.

2- تعريف القيم المتطرفة:

هناك العديد من التعريفات التي تخص القيم المتطرفة نذكر منها:

2-1- تعريف 1:

هي الانحراف هو المشاهدة التي تقع على مسافة غير طبيعية من القيم الأخرى في عينة عشوائية . (nist.gov، 2023)

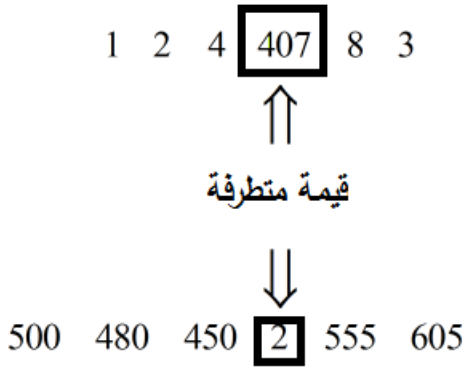
2-2- تعريف 2:

القيمة المتطرفة (outlier) هي عنصر شاذ وخارج عن النسق المميز لمجموعة أو تركيبة معينة، ففي عموم الدراسات في الإحصاء، الرياضياتيون أنجزوا خوارزميات قادرة على التخفيف من تأثير القيم الشاذة، أو إلغائها، وحتى حذفها، مستخدمين طرق الإحصاء المتين، غير أنه في بعض الأحيان يكون وجودها مفيدا لمعرفة سلوك تركيبة، أو منظومة.

أما القيمة المستحيلة (anomaly) فتعتبر قراءة خاطئة لأنها تدل على ظاهرة مستحيل حدوثها. (Matthias Klusch, 2008, p. 121)

ونورد المثال التالي عن القيمة المتطرفة:

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en



3- أسباب ظهور القيم المتطرفة: (حسينة، 2023، صفحة 55)

قد تظهر القيم المتطرفة في مجموعة البيانات لأسباب عدة نذكر منها:

1) الأخطاء التي يقع فيها الباحث عند رصد القياسات أو قد تحدث نتيجة إلى وجود خلل في أجهزة القياس وخاصة في التجارب المختبرية، أو نتيجة لأخطاء في الحسابات مما يؤدي إلى ظهور القيم المتطرفة.

2) قد تأتي البيانات من نوعين من التوزيعات أحدهما التوزيع الأساسي والذي يولد مشاهدات جيدة، بينما الأخر يسمى التوزيع الملوث والذي يولد قيما متطرفة.

3) قد تكون هذه البيانات حقيقية ناتجة عن ظروف غير عادية، فمثلا حدوث كوارث طبيعية كالزلازل، الأعاصير، الأمطار الغزيرة تؤثر على مستويات الإنتاج الزراعي، الحيواني والصناعي، إضراب عمال في مؤسسة ما يؤثر على إنتاجها، الحروب بين الدول تؤثر على اقتصاد هذه الدول...الخ. (حسينة، 2023)

3-1- الرسم البياني للقيم المتطرفة:

باستخدام البرامج الاحصائية يمكننا تحديد القيم المتطرفة بيانيا، حيث يمكننا قياس المسافة العمودية من أي نقطة بيانات إلى النقطة المقابلة على الخط الأفضل ملاءمة (الأفضل تسوية). (learn.saylor، 2023)

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en



شكل 1. أفضل تسوية

المصدر:(saylor, 2023)

جدول 1

مثال 1

	x	y	x^2	y^2	xy
	195	525	38025	275625	102375
	201	399	40401	159201	80199
	213	555	45369	308025	118215
	213	489	45369	239121	104157
	198	378	39204	142884	74844
	225	594	50625	352836	133650
	201	459	40401	210681	92259
	210	489	44100	239121	102690
	213	453	45369	205209	96489
	180	477	32400	227529	85860
المتوسط	204,9	481,8	42126,3	236023,2	99073,8

المصدر: من إعداد الباحث

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

$$\hat{\beta} = \frac{Cov(x,y)}{V(x)}$$

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x} \cdot \bar{y} = 99073{,}8 - (204{,}9)(481{,}8) = 352{,}98$$

$$V(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 = 142{,}29$$

$$V(y) = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - (\bar{y})^2 = 3891{,}96$$

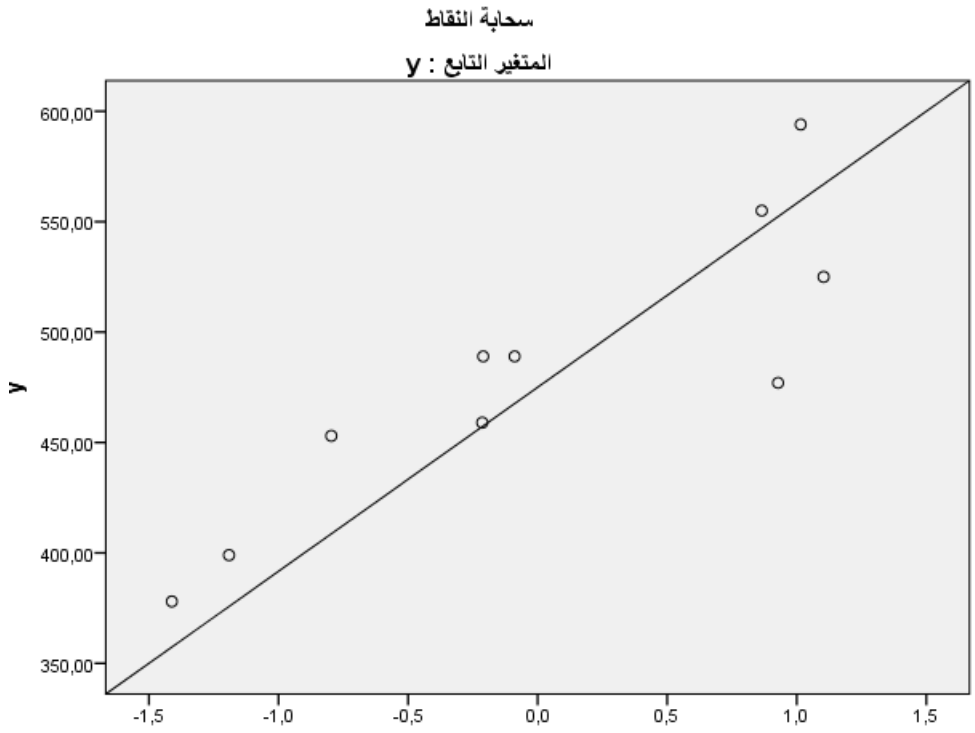
$$\hat{\beta} = \frac{Cov(x,y)}{V(x)} = \frac{352{,}98}{142{,}29} = 2.48$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 481{,}8 - (2.48 \times 204{,}9) = -26.352$$

$$\hat{y}_i = -26.352 + 2.48 x_i$$

x	y	$\hat{y}_i$	$y - \hat{y}_i$
195	525	457,248	67,752
201	399	472,128	-73,128
213	555	501,888	53,112
213	489	501,888	-12,888
198	378	464,688	-86,688
225	594	531,648	62,352
201	459	472,128	-13,128
210	489	494,448	-5,448
213	453	501,888	-48,888
180	477	420,048	56,952

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en



شكل 2. سحابة النقط

المصدر: مخرجات برنامج spss

2-3- طرق الكشف عن القيم المتطرفة:

أدخل هنا محتوى العنوان الفرعي الثاني، أدخل هنا محتوى العنوان الفرعي الثاني، أدخل هنا محتوى العنوان الفرعي الثاني، أدخل هنا محتوى العنوان الفرعي الثاني، أدخل هنا محتوى العنوان الفرعي الثاني، أدخل هنا محتوى العنوان الفرعي الثاني، أدخل هنا محتوى العنوان الفرعي الثاني.

هناك أربع طرق لحساب القيم المتطرفة، يمكننا الاختيار من بين عدة طرق لاكتشاف القيم المتطرفة حسب الوقت والموارد، نذكر منها:

أولا: طريقة الفرز :Sorting method

يمكننا فرز المتغيرات الكمية من الصغيرة إلى العالية والبحث عن قيم صغيرة للغاية أو عالية للغاية، نضع علامة على أي قيم متطرفة نجدها.(Prabhakar Gupta, 2007, p. 184)

هذه طريقة بسيطة للتحقق مما إذا كنا بحاجة إلى التحقق من نقاط بيانات معينة قبل استخدام طرق أكثر تعقيدا.

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

ثانيا: استخدام التصوير البياني Using visualizations

يمكننا استخدام برنامج لتصور بياناتنا باستخدام مخطط الصندوق، حتى نتمكن من رؤية توزيع البيانات في لمح البصر، يبرز هذا النوع من المخططات القيم الدنيا والقصوى(النطاق) والوسيط والمدى الربيعي لبياناتنا.

تبرز العديد من برامج الكمبيوتر قيمة خارجية على الرسم البياني بعلامة النجمة، وستقع خارج حدود الرسم البياني.

ثالثا: الكشف الإحصائي للقيم المتطرفة Statistical outlier detection

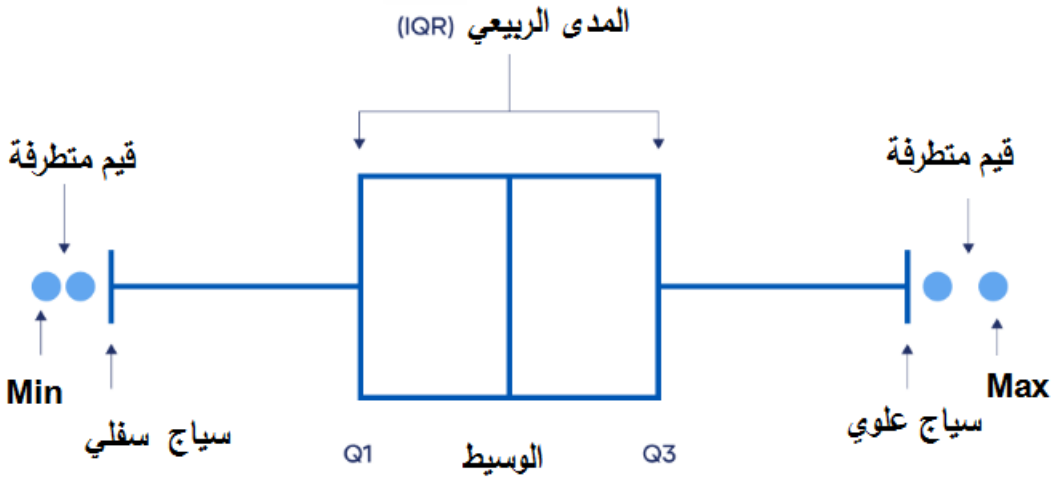
يتضمن الكشف للقيم المتطرفة تطبيق اختبارات أو إجراءات إحصائية لتحديد القيم المتطرفة، يمكننا تحويل نقاط البيانات القصوى إلى درجات z التي تخبرنا بعدد الانحرافات المعيارية التي تبعدها عن المتوسط .

إذا كانت القيمة تحتوي على درجة z عالية بما يكفي أو منخفضة بما يكفي، فيمكن اعتبارها متقطعة، كقاعدة عامة غالبا ما يتم تحديد القيم التي تحتوي على درجة z أكبر من 3 أو أقل من -3 على أنها قيم متطرفة.

استخدام المدى الربيعي:Using the interquartile range

يخبرنا المدى الربيعي (IQR) بمدى النصف الأوسط (الوسيط) من مجموعة البيانات الخاصة بنا.

نوضح مخطط الصندوق[1] كالتالي:



شكل 3. مخطط الصندوق

[1]-ابتكرت هذه الطريقة سنة 1977على يد الرياضياتي والاحصائي الأمريكي جون توكي John Tukey .

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

المصدر: مخرجات برنامج spss

يعرض مخطط الصندوق، ويسمى أيضا مخطط whisker ملخصا مكونا من خمسة أرقام لمجموعة من البيانات (الحد الأدنى، والربيع الأول، والربيع الثاني (الوسيط)، والربيع الثالث، والحد الأقصى. (بداوي، 2017، صفحة 57)

- مثال:

في محل لبيع الأجهزة الكهرو منزلية سجل خلال يوم بيع 100 جهاز، حيث تراوحت الأسعار ما بين 200 و1200 وحدة نقدية، حيث كانت المعطيات ممثلة في الجدول التكراري الاتي:

جدول 2

مثال2

الأسعار (و.نقدية)	400–200	600–400	800–600	1000–800	1200–1000
الوحدات المباعة	15	25	35	10	15

المصدر: من إعداد الباحث

- المطلوب :إيجاد الربيعات وقدمهما في علبة (Boite à moustaches) whiskers، ثم أحسب الانحراف الربيعي؟

- حل المثال:

- إيجاد الربيعات:

– تحديد قيمة الربيع الأول: $(100\times 0.25 = 25)$ ، إذن فئة الربيع الأول هي $]400-600[$ ، ومن صيغة الربيع الأول نحدد المقادير الاتية:

$$L_1 : 400 \text{ ،} n_{1/4} : 25 \text{ ،} N_{01} : 15 \text{ ،} c : 200$$

فقيمة الربيع الأول هي كما يلي: $Q_1 = L_1 + \frac{\frac{n}{4} - N_{01}}{n_{1/4}} \times c = 400 + \frac{25-15}{25} \times 200 = 480$

– تحديد قيمة الربيع الثالث: $(100\times 0.75 = 75)$ ، إذن فئة الربيع الثالث هي $]600-800[$ ، ومن صيغة الربيع الثالث نحدد المقادير الاتية:

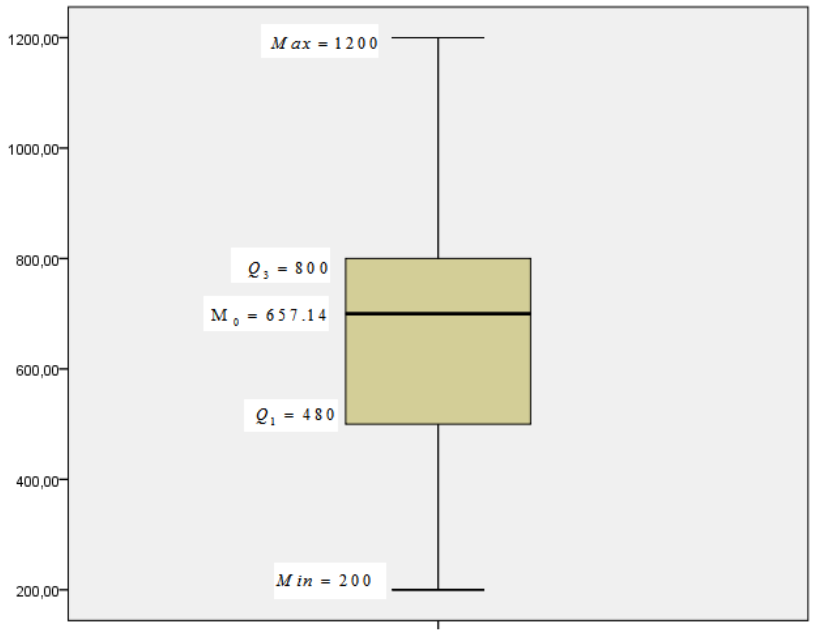
$$L_1 : 600 \text{ ،} n_{3/4} : 35 \text{ ،} N_{03} : 40 \text{ ،} c : 200$$

فقيمة الربيع الثالث هي كما يلي: $Q_3 = L_1 + \frac{\frac{3n}{4} - N_{03}}{n_{3/4}} \times c = 600 + \frac{75-40}{35} \times 200 = 800$

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

إذن قيمة الانحراف الربيعي هي كما يلي: $ET = Q_3 - Q_1 = 800 - 480 = 320$

تمثيل الربيعات في مخطط صندوق Box plot (Boite à moustaches) يكون كما يلي:



شكل 4. مخطط الصندوق

المصدر: مخرجات برنامج spss

4- استخدام انتشار Z_i Using scatter:

نستخدم المعادلة التالية:

$$Z_i = Y_i - \hat{\beta} X_i$$

من خلال تسوية النموذج بطريقة المربعات الصغرى، وحساب قيم Z_i و ترتيبها تصاعديا مقابل رتبها، تبرز القيم المتطرفة:

- مثال:

نستخدم بيانات المثال الأول:

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

جدول 3

مثال3

x	y	$Z_i = Y_i - \widehat{\beta} X_i$
195	525	433,35
201	399	304,53
213	555	454,89
213	489	388,89
198	378	284,94
225	594	488,25
201	459	364,53
210	489	390,3
213	453	352,89
180	477	392,4

المصدر: من إعداد الباحث

$$\beta = 2.48$$

x	y	$Z_i = Y_i - \widehat{\beta} X_i$
195	525	41,4
201	399	-99,48
213	555	26,76
213	489	-39,24
198	378	-113,04
225	594	36
201	459	-39,48
210	489	-31,8
213	453	-75,24
180	477	30,6

اظهار الرتب مقابل رتبة كل مشاهدة:

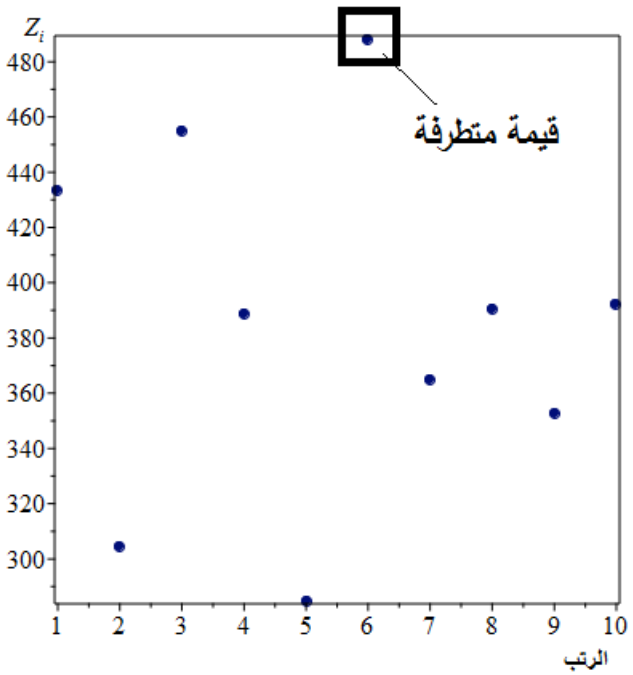
This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

جدول 4

مثال4

الرتب	Z_i
5	36
2	-99,48
9	-75,24
7	-39,48
10	30,6
4	-39,24
8	-31,8
1	41,4
3	26,76
6	36

باستخدام برنامج Maple نقوم بتمثيل الجدول الأخير:

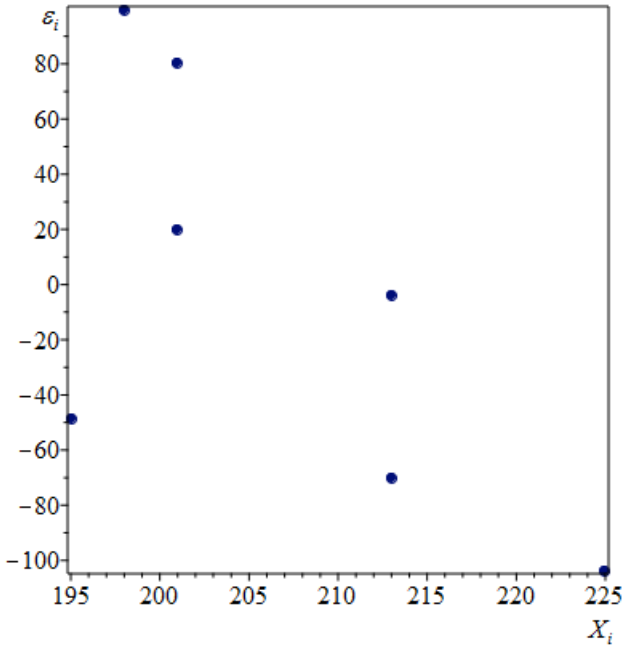


شكل 5. انتشار Z

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

المصدر: مخرجات برنامج spss

يمكننا اكتشاف القيم المتطرفة كذلك عن طريق تمثيل الأخطاء (البواقي) $\varepsilon_i = Y_i - \hat{Y}_i$ مع قيم X_i، باستخدام برنامج Maple نقوم بتمثيلها:



شكل 6. الأخطاء (البواقي)

المصدر: مخرجات برنامج spss

5- استخدام احصاءة t:

في حالة الانحدار الخطي البسيط تعطى كما يلي:

$$t = Max\left|\frac{\varepsilon_i}{s_i}\right|$$

$$\frac{\varepsilon_i}{V(\varepsilon_i)} = \frac{\varepsilon_i}{\frac{SS_E}{n-2}} = \frac{\varepsilon_i}{s_i}$$

$$\therefore SS_E = SS_T - \hat{\beta}\ S_{xy} \quad , \quad SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$\therefore S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)$$

لمعرفة ما إذا كانت المشاهدة عبارة عن قيمة متطرفة يلزم حساب احصاءة t^* التي تعطى كما يلي:

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

$$t^* = Max\frac{|\varepsilon_i / s_i|}{\sqrt{\frac{(n-2)V(\varepsilon_i)}{n}}} = Max\frac{|\varepsilon_i / s_i|}{\sqrt{\frac{\sum_{i=1}^{n}\varepsilon_i^2}{n}}}$$

إذا كان $t \square t^*$ فأن الزوج (x_i, y_i) يحقق قيمة متطرفة.

إذا كان $t \neq t^*$ فأن الزوج (x_i, y_i) لا يحقق قيمة متطرفة.

بالرجوع إلى المثال (3) نجد: $t = 0,0274878$ التي تقابل الزوج $(225\ ,\ 594)$، وكذلك $(x_i, y_i) = (225\ ,\ 594)$

$t^* = 1,72916495$ التي تقابل الزوج $(x_i, y_i) = (225\ ,\ 594)$.

نلاحظ أن $t \square t^*$ فأن الزوج $(x_i, y_i) = (225\ ,\ 594)$ يحقق قيمة متطرفة.

-ملاحظة:

في حالة الانحدار الخطي المتعدد نستخدم الاحصاءة التالية:

$$t_i = \frac{\varepsilon_i}{\sqrt{MSE(1-h_{ij})}}$$

حيث: MSE: متوسط مربع الخطأ، h_{ij}: عناصر قطر مصفوفة القبعة Hat matrix

$H = X(X'X)^{-1}X'$.

6- مسافة كوك Cook's distance:

في الإحصاء ، تعد مسافة Cook's أو Cook's D تقديرا شائعا لتأثير نقطة البيانات عند إجراء تحليل انحدار المربعات الصغرى، في تحليل المربعات الصغرى الاعتيادية ، يمكن استخدام مسافة كوك بعدة طرق: للإشارة إلى نقاط البيانات المؤثرة التي تستحق بالتحديد التحقق من صحتها ، تم تسميته على اسم الإحصائي الأمريكي آر دينيس كوك ، الذي قدم المفهوم في عام1977 (statistical-methods, 2023) ، تعطى هذه الاحصاءة كما يلي:

$$D_i = \frac{r_i^2}{p} \cdot \frac{h_{ij}}{(1-h_{ij})} \quad , \quad i = 1,2,3,...$$

r_i هي i^{th} عمود للمصفوفة $R = (X'X)^{-1}X'$ ، p: عدد معلمات نموذج الانحدار.

إذا كانت $D_i > 1$ فنعتبر أن المشاهدة متطرفة. (statistical-methods، 2023)

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

7- نقاط التأثير العالي High-leverage points :

إذا كان لمتغير تفسيري معين قيمة واحدة أو أكثر أكبر بكثير من المشاهدات الأخرى، فإن هذه المشاهدات يمكن أن تؤثر بقوة على نتائج الانحدار، تتمتع نقاط التأثير العالي بإمكانية كبيرة للتأثير على نتائج التحليل إذا كانت تتوافق مع الملاحظات التي لا تتبع النموذج الخطي، ولكن المشكلة الناتجة قد لا تكون واضحة في فحص القيم المتطرفة لذلك من المهم تحديد نقاط التأثير العالي، يتم حسابها كما يلي (Alain Zuur, 2007, p. 64):

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_j (x_i - \bar{x})^2}$$

في حالة الانحدار الخطي البسيط $\sum_{i=1}^{n} h_i = 2$ ، أما في حالة الانحدار الخطي المتعدد نقاط التأثير هي مجموع قطر مصفوفة القبعة.

8- الحلول المقترحة لمعالجة القيم المتطرفة:

نذكر بعضها:

8-1- معالجة القيم المتطرفة للخطأ: في هذه الطريقة، يجب إما إزالة كل خطأ أو تصحيحه، إذا كانت هناك بيانات متوفرة ضمن قيم مهمة، فإن الإدخال الأصلي لنقاط البيانات لتجنب فقدان المعلومات بشكل كبير من خلال الحذف، إذا كانت البيانات تحتوي على بعض الأخطاء، فإن أفضل طريقة هي إزالة الإدخالات. (Dash, 2023)

معالجة القيم المتطرفة غير الخطأ: هناك ثلاث طرق تُستخدم للتعامل مع القيم المتطرفة غير الخطأ: الاحتفاظ والحذف والتسجيل:

1- عند الاحتفاظ بالقيم المتطرفة، نكون على دراية بأنها يمكن أن تشوه نتائج مهمتنا الفعلية.
2- الحذف: الخيار الأكثر مباشرة هو حذف أي ملاحظة خارجية.
3- إعادة الترميز- تجنب فقدان كمية كبيرة من البيانات باستخدام طريقة winsorizing[2] للتعامل مع القيم المتطرفة.

9- الخاتمة:

[2] winsorizing هو تحويل الإحصائيات عن طريق الحد من القيم المتطرفة في البيانات الإحصائية لتقليل تأثير القيم المتطرفة التي قد تكون زائفة، سميت على اسم المهندس الذي تحول إلى الإحصائي الحيوي تشارلز ب. وينسور Charles P. Winsor (1895–1951). التأثير هو نفس القص في معالجة الإشارة.

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en

تطرقنا في هذا المقال تسليط الضوء على موضوع القيم المتطرفة في الانحدار الخطي البسيط، نظرا للأهمية الكبيرة في البحوث الإحصائية، إن القيم المتطرفة ميدانا واسعا للبحث، لذلك حاولنا تبيان طرق كشف ومعالجة القيم المتطرفة، ودعمنا ذلك بأمثلة تطبيقية.

10- قائمة المراجع:

حسينة، ع. ا .(2023) .طرق كشف ومعالجة القيم الشاذة في الانحدار الخطي البسيط .مستغانم: جامعة مستغانم. الجزائر.

بداوي، محمد. (2017). الاحصاء الوصفي. الجزائر: دار هومة.

Alain, Z. N. (2007). Analyzing Ecological Data. New York: Springer.

Dash, C. S. (2023). An outlier's detection and elimination framework in classification task of data. Elsevier, 1-8.

learn. Saylor. (2023). https://learn.saylor.org/mod/book/view.php?id=55086&chapterid=40788

Matthias, K. P. (2008). Cooperative Information Agents XII : 12th International Workshop. Prague: Springer.

nist.gov. (2023). Product and Process Comparisons. Récupéré sur https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm

Othor, C. S. (2023). An outliers detection and elimination framework in classification task of data mining, Volume 6, March 2023, 100164. Elsevier, 1-8.

Prabhakar, G. A. (2007). Data Structure Using C. Firewall.

Saylor. (2023). Récupéré sur https://learn.saylor.org/mod/book/view.php?id=55086&chapterid=4078

Scribbr. (2023). https://www.scribbr.com/statistics/outliers/

Statistical-methods. (2023). Récupéré sur https://medium.com/analytics-vidhya/statistical-methods-for-identifying-outliers-regression-analysis-approach-partii-977b035b65a0

- Arabic references in English:

Hasina, P. A. (2023). Methods for detecting and processing anomalous values in simple linear regression. Mostaganem: University of Mostaganem. Algeria.

Badaoui, M. (2017). Descriptive statistics. Algeria: Dar Houma.

***Citation**: Doua, M. **Methods for detecting and treating outliers in simple linear regression**. Social Empowerment Journal. 2024; 6(2): pp. 121-135. https://doi.org/10.34118/sej.v6i2.3924*

***Publisher's Note:** SEJ stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.*

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
https://creativecommons.org/licenses/by-nc/4.0/deed.en