مجلة العلوم الرياضية والمرتبطة
المجلد الرابع / العدد الحادي والثلاثون – ديسمبر 2021 ISSN 2518-5470

تحويلة القوة لبوكس-كوكس وكيفية تنفيذها فيMATLAB
Box-Cox power transformation and how to implement it in MATLAB

د. البهلول عمر علي شلابي
الايميل : b.shalabi@uot.edu.ly
ا. سعاد محمد أحمد البرقاوي
الايميل :s.elbargawi@uot.edu.ly

تاريخ الاستلام / 2021/7/12
تاريخ القبول /2021/10/28

الكلمات المفتاحية : لبوكس- كوكس – طريقة ماتلاب

الملخص البحث

منذ أن قام كل من بوكس وكوكس بكتابة ورقتهما الأصلية خلال سنة 1964 والتي عرضت نوع من تحويلات القوة التي تهدف إلى تحويل البيانات التي لا تتبع التوزيع الطبيعي إلى بيانات تتبع التوزيع الطبيعي، هذا النوع من تحويلات القوة كان ذو أهمية كبيرة لدى الباحثين، سواء في البحوث النظرية أو في التطبيقات العملية (Box and Cox, 1964).

هذه الورقة تطرقت إلى الموضوعات التالية المتعلقة بتحويلة القوة لبوكس-كوكس:

- التعريف بتحويلة القوة لبوكس-كوكس.
- أستعراض مختصر لبعض الدراسات السابقة حول التحويلة.
- شرح لكيفية تقدير معلمة تحويلة القوة لبوكس-كوكس بطريقتين مختلفتين وذلك بالإستعانة بالبرمجية ماتلاب (MATLAB Software).
- عرض لبعض طرق الكشف عن مدى نجاح تحويلة القوة لبوكس-كوكس في تحويل البيانات التي لا تتبع التوزيع الطبيعي إلى بيانات تتبع التوزيع الطبيعي.

مقدمة

يشترط عادة أن تكون مشاهدات عينة الدراسة قد جاءت من مجتمع إحصائي يتبع التوزيع الطبيعي، وذلك كشرط لاستخدام عدد من أدوات التحليل الإحصائي المعلمية، مثل اختبارات t وتحليل التباين (ANOVA) وغيرها من الأدوات. عندما لا يكون توزيع مشاهدات العينة هو التوزيع الطبيعي، فإنه في هذه الحالة يجب معالجة مشاهدات العينة بأحد الإجراءات العلاجية المناسبة.

تعد تحويلة القوة لـ بوكس-كوكس (Box-Cox power transformation) أحد هذه الإجراءات العلاجية التي قد تساعد في جعل مشاهدات العينة تتبع التوزيع الطبيعي. فمن خلال فهم الباحث لمفهوم التحويل وطريقة التحويل لـ بوكس-كوكس سيكون لديه المقدرة على التعامل مع البيانات التي لا تتبع التوزيع الطبيعي.

منذ سنة 1964 عندما قام كلاً من بوكس وكوكس بنشر ورقتهما البحثية حول تحويلات القوة (Box and Cox, 1964) تولد قدراً كبيراً من الاهتمام بالجانب النظري وكذلك بالتطبيقات العملية لتحويلات القوة لـ بوكس-كوكس.

تهدف هذه الورقة إلى ما يلي:

1. التعريف بتحويلة القوة لـ بوكس-كوكس.
2. عرض لأهم الدراسات السابقة لتطوير التحويلة.
3. شرح لكيفية تقدير معلمة تحويلة القوة لـ بوكس-كوكس باستخدام برنامج ماتلاب (MATLAB Software).
4. عرض لبعض طرق الكشف عن مدى نجاح تحويلة القوة لبوكس-كوكس في تحويل البيانات التي لا تتبع التوزيع الطبيعي إلى بيانات تتبع التوزيع الطبيعي.

تحويلة القوة لـ بوكس-كوكس

إذا كان $y_1, y_2, \dots, y_n$ تمثل مشاهدات العينة التي حجمها n والتي جاءت من مجتمع إحصائي لا يتبع التوزيع الطبيعي، الشكل الأصلي لتحويلة القوة لـ بوكس-كوكس، كما ظهرت في بحثهم الصادر عام (Box 1964 and Cox, 1964)، يتخذ الصورة التالية لجميع $y_i > 0$:

$$y_i(\lambda) = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda}, & if\ \lambda \neq 0; \\ \log y_i, & if\ \lambda = 0. \end{cases} \quad \text{...............(1)}$$

حيث $y_1(\lambda), y_2(\lambda), \ldots, y_n(\lambda)$ هي مشاهدات العينة المحولة بواسطة تحويلة القوة لـ بوكس-كوكس بمعلمة λ.

في نفس الورقة، اقترح كل من بوكس وكوكس أيضاً نموذجاً موسعاً يمكن أن يستوعب قيم مشاهدات العينة السالبة:

$$y_i(\lambda) = \begin{cases} \dfrac{(y_i + c)^{\lambda_1} - 1}{\lambda}, & if\ \lambda_1 \neq 0; \\ \log(y_i + c), & if\ \lambda = 0. \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ (2)$$

عملياً، يمكننا أن نختار c بحيث يكون $y_i + c > 0$ لجميع قيم مشاهدات العينة.

الهدف من تحويلة القوة لـ بوكس-كوكس هو ضمان أن تكون البيانات المحولة تتبع التوزيع الطبيعي. أحياناً لا تعمل تحويلة القوة لـ بوكس-كوكس جيداً، لذا يجب فحص مشاهدات العينة المحولة باستخدام شكل **P-P Normal Plot** أو شكل **Q-Q Normal Plot** أو استخدام أحد اختبارات الكشف عن ما إذا كانت مشاهدات العينة تتبع التوزيع الطبيعي أم لا (**Normality Tests**).

عرض مختصر لأهم الدراسات السابقة

منذ سنة 1964 عندما قام بوكس وكوكس بنشر ورقتهما حول تحويلة القوة (**Box and Cox, 1964**)، تم اقتراح العديد من التعديلات المتعلقة بالتحويلة، ففي سنة 1971 اقترح مانلي (**Manly, 1976**) التحويلة الأسية التالية:

$$y_i(\lambda) = \begin{cases} \dfrac{e^{\lambda y_i} - 1}{\lambda}, & if\ \lambda \neq 0; \\ y_i, & if\ \lambda = 0. \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ (3)$$

هذه التحويلة تصلح لمشاهدات العينة التي تحتوي على قيم سالبة. هذه التحويلة كانت ناجحة في تحويل مشاهدات العينات التي لها توزيع ملتوي أحادي المنوال إلى التوزيع الطبيعي، ولكنها لم تكن ناجحة بدرجة كافية عندما يكون توزيع مشاهدات العينة ثنائي المنوال أو على شكل حرف U.

خلال عام 1981 قام كلا من بيكل ودوكسوم (**Bickel and Doksum, 1981**) بإجراء التعديل الطفيف التالي في نموذج تحويلة القوة لـ بوكس-كوكس:

$$y_i(\lambda)=\frac{|y_i|^{\lambda}\,Sign(y_i)-1}{\lambda},\quad for\ \lambda>0, \qquad \text{……………. (4)}$$

حيث

$$Sign(y_i)=\begin{cases}1, & if\ y_i\ge 0;\\ -1, & if\ y_i<0.\end{cases} \qquad \text{……………. (5)}$$

خلال سنة 1992 قدم ساقية (Sakia, 1992) عرض وافي للأعمال المتعلقة بتحويلة القوة لـ بوكس-كوكس.

اقترح في عام 2000 كلا من يو وجونسون (Yeo and Johnson, 2000) التحويلة التالية:

$$y_i(\lambda)=\begin{cases}\dfrac{(y_i+1)^{\lambda}-1}{\lambda}, & if\ \lambda\ne 0, y_i\ge 0;\\ \log(y_i+1), & if\ \lambda=0, y_i\ge 0;\\ \dfrac{(1-y_i)^{2-\lambda}-1}{\lambda-2}, & if\ \lambda\ne 2, y_i<0;\\ -\log(1-y_i), & if\ \lambda=2, y_i<0.\end{cases} \qquad \text{……………. (6)}$$

لمزيد من الإطلاع حول تحويلة القوة لـ بوكس-كوكس والتعديلات التي تم اقتراحها عليها من قبل العديد من الباحثين، أنظر (Riani, and Corbellini, Atkinson, 2020).

بعض طرق تقدير معلمة تحويلة القوة لبوكس-كوكس

أولاً: طريقة الاحتمال الأعظم

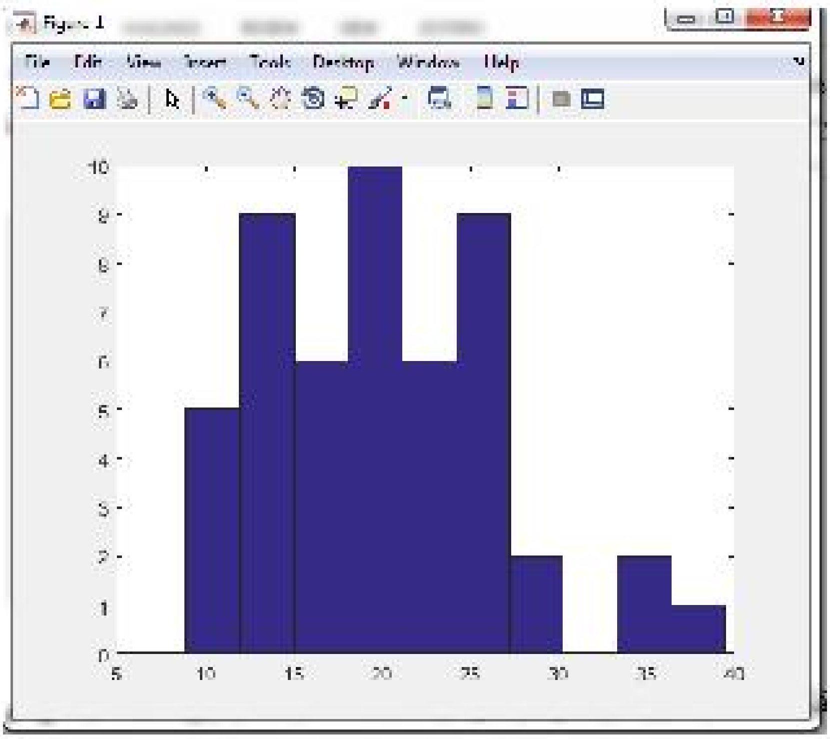
الهدف الرئيسي من تحويلة القوة لبوكس-كوكس هو الاستدلال على المعلمة λ، معلمة تحوية القوة لبوكس-كوكس، حيث استخدم (Box and Cox, 1964) طريقة ا لاحتمال الأعظم (Maximum Likelihood Method)، والتي تُستخدم بشكل شائع في عمليات التقدير.

دالة ماتلاب boxcox تستخدم طريقة الاحتمال الأعظم لتقدير المعلمة λ. خطوات استخدام هذه الدالة لتقدير المعلمة λ لبيانات عينة عشوائية حجمها 50 مفردة تم توليدها من توزيع مربع كاي بدرجات حرية 20 موضحة في الشكل التالي:

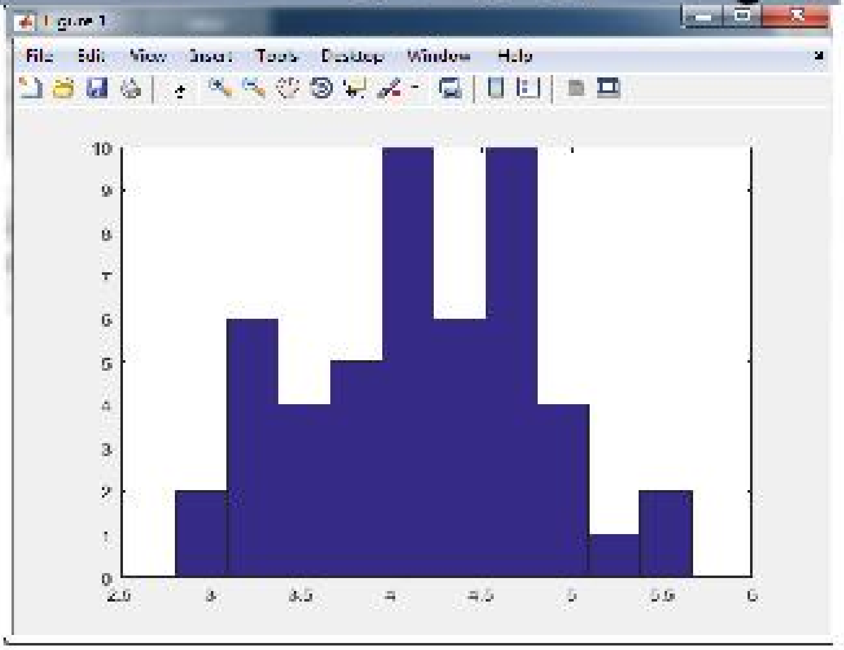
 ISSN 2518-5470

```
Command Window
>> y=chi2rnd(20,[50,1]);
>> [ylambda, lambda]=boxcox(y);
>> lambda
lambda =
            0.220375
>>
```

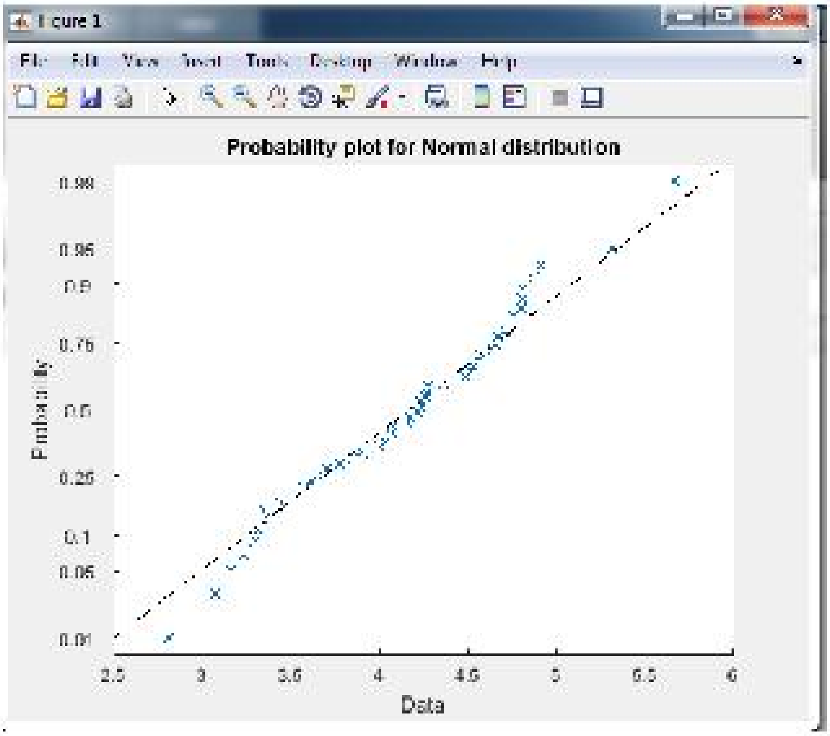
مدرج التكرار لمشاهدات العينة قبل استخدام تحويلة القوة لبوكس-كوكس، والذي تم الحصول عليه باستخدام الأمر hist(y)، موضح في الشكل التالي:



كما أن مدرج التكرار لمشاهدات العينة بعد استخدام تحويلة بوكس-كوكس، وذلك باستخدام الأمر hist(ylambda)، موضح في الشكل التالي:



يمكننا أيضاً التحقق من أن مشاهدات العينة المحولة تتبع التوزيع الطبيعي وذلك عن طريق رسم شكل البياني **p-p Normality plot** عن طريق ستخدام الأمر **probplot(ylambda)** فنتحصل على المطلوب كما هو مبين في الشكل التالي:



يمكننا أيضاً حساب قيمة معامل الالتواء للكشف عن مدى قرب توزيع مشاهدات العينة المحولة من التماثل وذلك باستخدام الأمر **S = skewness(ylambda)** ، فنتحصل على قيمة معامل الالتواء كما هو مبين في الشكل التالي:

```
Command Window
>> S = skewness(ylambda)
S =
     -0.0123458681669921
>>
```

ملاحظة هامة: لاستخدام دالة ماتلاب **boxcox** لتقدير معلمة تحويلة القوة لبوكس-كوكس يشترط أن أن تكون جميع مشاهدات العينة التي يراد تحويلها موجبة. في حالة ما إذا كانت مشاهدات العينة غير موجبة يجب إضافة عدد ثابت مناسب، **c**، إلى كل مشاهدة من مشاهدات العينة لتصبح جميع المشاهدات موجبة.

الشكل التالي يوضح أوامر ماتلاب للكشف عن وجود قيم سالبة في مشاهدات العينة، وفي حالة وجودها، يتم تحويلها إلى مشاهدات موجبة قبل استخدام دالة ماتلاب **boxcox** بهدف الحصول عل تقدير لمعلمة تحويلة القوة لبوكس-كوكس.

```
Command Window
fx >> %y = the non-normal data on which the Cox Box transformation is to be applied
   %lambda = the Box-Cox power transformation parameter
   n=length(y);% the size of the non-normal data, y.
   c=0;
   minx=min(y);
   if minx<0
     kk=find(y<zeros(n,1));
     Ny=y(kk);
     minNy=min(Ny);
      c=abs(minNy);
      y=y+c;
      jj=find(y==zeros(n,1));
      y(jj)=y(jj)+0.000001;
    end
```

أوامر ماتلاب للكشف عن وجود قيم سالبة في مشاهدات العينة يمكن الحصول عليها من خلال الرابط التالي:

https://docs.google.com/document/d/1K2rH3IrZDvkzCH4woshm8R0l9xW6naoP/edit?usp=sharing&ouid=103075293608624046456&rtpof=true&sd=true

 ISSN 2518-5470

للحصول على مشاهدات العينة المحولة باستخدام معلمة تحويلة القوة لبوكس-كوكس، **lambda**، لأي مشاهدات عينة موجبة كانت أم سالبة تمت كتابة دالة ماتلاب التي تحمل اسم **BoxCoxtransformation** لتنفيذ ذلك، والمبينة في الشكل التالي:

```
function [lambda, y_lambda,c] = BoxCoxtransformation(y, lambda);
%y = the non-normal data on which the Box-Cox transformation is to be applied.
%lambda = the Box-Cox power transformation parameter.
n=length(y);% the size of the non-normal data, y.
c=0;
minx = min(y);
if minx<0
  kk=find(y<zeros(n,1));
  Ny = y(kk);    c = abs(minNy);

  minNy=min(Ny);
   y=y+c;
   jj=find(y==zeros(n,1));
   y(jj) = y(jj)+0.000001;
 end
if lambda ==0
   y_lambda=log(y);
else
   y_lambda=((y.^lambda)-1)/lambda;
end
%y_lambda = the Cox-Box transformed data using the estimated lambda
```

يمكن الحصول على دالة ماتلاب **BoxCoxtransformation** من خلال الرابط التالي:

https://docs.google.com/document/d/1EKmzYPVLyhBBpgnOYXQArVfYyO69Qsci/edit?usp=sharing&ouid=106822803012039915729&rtpof=true&sd=true

ثانياً: طريقة فحص قيمة الاحتمال لاختبار شابيرو-ويلك (Shapiro-Wilk Test)

يعتبر اختبار شابيرو-ويلك أحد الاختبارات المشهورة للكشف عما إذا كانت مشاهدات عينة عشوائية قد جاءت من مجتمع يتبع التوزيع الطبيعي أم لا. يصلح هذا الاختبار حتى للعينات الصغيرة الحجم. تم نشره في عام 1965 من قبل صمونيل سانفورد شابيرو ومارتن ويلك (Shapiro, S. S. ad Wilk, M. B., 1965).

صياغة فرض العدم لهذا الاختبار هي أن مشاهدات العينة العشوائية قد جاءت من مجتمع يتبع التوزيع الطبيعي، وبالتالي إذا كانت قيمة الاحتمال (p-value) أقل من مستوى ألمعنوية المرغوب، α، فسيتم رفض فرض العدم ويكون هناك دليل كافي على أن مشاهدات العينة المختبرة لم تأتي من مجتمع يتبع التوزيع

الطبيعي. من ناحية أخرى، إذا كانت قيمة الاحتمال أكبر من مستوى المعنوية المرغوب، α، ، فلا يمكن رفض فرض العدم بأن مشاهدات العينة جاءت من مجتمع يتبع التوزيع الطبيعي.

خلال سنة 2015 قدم كل من Vélez JI وآخرون ورقة تتضمن مقترح لمنهجية جديدة لتقدير قيمة λ ، معلمة تحويلة القوة لبوكس-كوكس Vélez JI, Correa JC and Marmolejo-Ramos F (2015). تتلخص هذه الطريقة في الخطوات التالية:

1- بافتراض أن لدينا سلسلة من قيم λ تكون مرتبة تصاعدياً على النحو التالي:

$$\lambda_{(1)} < \lambda_{(2)} < \lambda_{(3)} < \cdots \lambda_{(k)}$$

هنا، $\lambda_{(1)}$ و $\lambda_{(k)}$ هما، على التوالي، الحدود الدنيا والعليا لهذه السلسلة التي تحتوي على عدد محدود من قيم معلمة تحويلة القوة لبوكس-كوكس، λ.

2- باستخدام قيم λ في الخطوة (1) السابقة نقوم باتباع الخطوات التالية بهدف تقدير معلمة تحويلة بوكس-كوكس، λ بالطريقة المقترحة:

- يتم تطبيق المعادلة (1)، تحويلة القوة لبوكس-كوكس، على مشاهدات العينة العشوائية التي لا تتبع التوزيع الطبيعي، $y_1, y_2, \cdots, y_n$ ، مع كل قيمة من قيم λ الموجودة في السلسلة

$$\lambda_{(1)} < \lambda_{(2)} < \lambda_{(3)} < \cdots \lambda_{(k)}.$$

- عند الحصول على مشاهدات العينة المحولة باستخدام معلمة بوكس-كوكس، $\lambda = \lambda_{(j)}; j = 1,2,3,\cdots,k$ ، نقوم باستخدام اختبار شابيرو-ويلك لاختبار ما إذا كانت مشاهدات العينة المحولة بالمعلمة λ تتبع التوزيع الطبيعي أم لا وذلك بعد حساب قيمة الاحتمال (p-value) للاختبار بعد مقارنتها بمستوى المعنوية للاختبار المرغوبة، مثلاً $\alpha = 0.05$.

- بعد الإنتهاء من الخطوة السابقة تتحصل على أزواج مرتبة من قيم الاحتمال لاختبار شابيرو-ويلك، P، وقيم معالم تحويلة القوة لبوكس-كوكس التي تم تكوينها في الخطوة (1)، والتي تكون على الصورة التالية:

$$(\lambda_{(1)}, P_{(1)}), (\lambda_{(2)}, P_{(2)}), \cdots, (\lambda_{(k)}, P_{(k)})$$

- نبحث عن أكبر قيمة لقيمة الاحتمال لاختبار شابيرو-ويلك، P^* ، والتي تكون مناظرة لقيمة معلمة تحويلة القوة لبوكس-كوكس، λ^*، ضمن الأزواج المرتبة التي تم الحصول عليها في الخطوة السابقة.

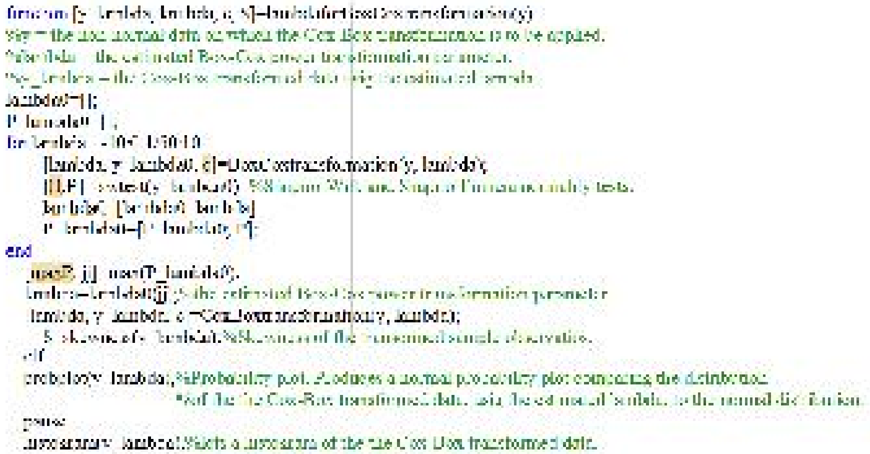
لاحظ أن قيمة λ^* ، تجعل قيمة الاحتمال P^* لاختبار مدي كون مشاهدات العينة العشوائية تتبع التوزيع الطبيعي هي الأعلى لجميع قيم λ ، وذلك عند تطبيق اختبار شابيرو-ويلك، ومع ذلك، قد لا يكون الأمر كذلك، لنفس قيمة λ^* عند استخدام اختبارات أخرى غير اختبار شابيرو-ويلك.

مجلة العلوم الرياضية والمرتبطة
المجلد الرابع / العدد الحادي والثلاثون – ديسمبر 2021 ISSN 2518-5470

لتنفيذ هذه الطريقة باستخدام برنامج ماتلاب تم في البداية البحث عن دالة ماتلاب تقوم بتنفيذ اختبار شابيرو- ويلك للكشف عن كون مشاهدات عينة عشوائية تتبع التوزيع الطبيعي أم لا. تم استخدام دالة ماتلاب التي تحمل الاسم **swtest** والتي تم كتابتها بواسطة أحمد بن سعيدة (**Ahmed Ben Saida**). يمكن الحصول على هذه الدالة من خلال الرابط التالي:

https://docs.google.com/document/d/1qghw6lCCna9Cy-Xm1iqO8swXXZHbqywM/edit?usp=sharing&ouid=103075293608624046456&rtpof=true&sd=true

بعد ذلك تم كتابة دالة ماتلاب تحمل الاسم **lambdaforBoxCoxtransformation** لتنفيذ هذه الطريقة والمبينة في الشكل التالي:



```
function [y_lambda, lambda, c, S]=lambdaforBoxCoxtransformation(y)
%y = the non-normal data on which the Box-Cox transformation is to be applied.
%lambda = the estimated Box-Cox power transformation parameter.
%y_lambda = the Box-Cox transformed data using the estimated lambda.
lambda0=[];
P_lambda0=[];
for lambda = -10:0.1:10
      [lambda, y_lambda0, c]=BoxCoxtransformation(y, lambda);
      [H,P] = swtest(y_lambda0); %Shapiro-Wilk and Shapiro-Francia normality tests.
      lambda0=[lambda0 lambda];
      P_lambda0=[P_lambda0, P];
end
   [maxP, jj]=max(P_lambda0);
   lambda=lambda0(jj); %the estimated Box-Cox power transformation parameter
   [lambda, y_lambda, c]=BoxCoxtransformation(y, lambda);
     S=skewness(y_lambda);%Skewness of the transformed sample observation.
   clf
   probplot(y_lambda);%Probability plot. Produces a normal probability plot comparing the distribution
                      %of the the Box-Cox transformed data using the estimated lambda to the normal distribution.
   pause
   histogram(y_lambda);%Plots a histogram of the the Box-Cox transformed data.
```

يمكن الحصول على الدالة **lambdaforBoxCoxtransformation** من خلال الرابط التالي:

https://docs.google.com/document/d/10EfKk4mdZXQB2hkWAIG7Ks6Zmw9lOLth/edit?usp=sharing&ouid=106822803012039915729&rtpof=true&sd=true

لغرض مقارنة نتائج هذه الطريقة بطريقة الاحتمال الإعظم التي سبق شرحها، سنستخدم مشاهدات العينة العشوائية التي تم توليدها من توزيع مربع كاي بحجم 50 ودرجات حرية 20 عند استخدامنا لطريقة الاحتمال الأعظم ونتبع الخطوات المبينة في الشكل التالي:

ISSN 2518-5470

```
Command Window
>> [y_lambda, lambda, c, S]=lambdaforCoxBoxtransformation(y)
y_lambda
         4.3278461144372
        4.42818090744936
        5.00095295952495
         3.8114349254544
        3.39931125726828
        4.85609557932748

        4.22350792132768
        3.69309891484714
        4.38105502145083
        4.18888012446265
        3.54418869582094
        4.53331988095082
          4.659112143962
        5.55425881998658
        5.00921422209153
        4.14428158311545
lambda
       0.242000000000001
c
     0
S
      0.0037877040982765
fx >>
```

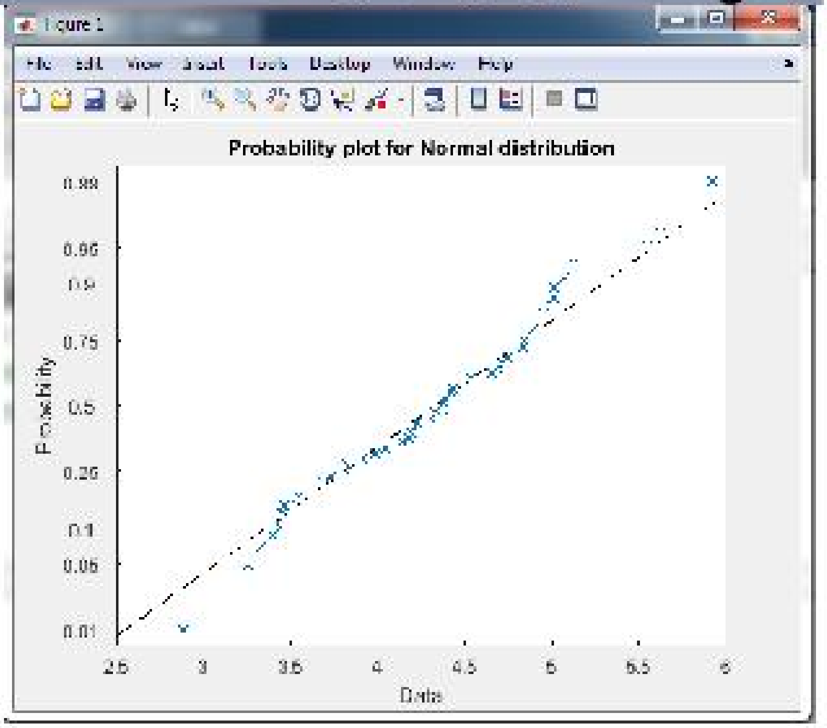


Figure 1
File Edit View Insert Tools Desktop Window Help
Probability plot for Normal distribution
Probability
Data
0.99
0.95
0.9
0.75
0.5
0.25
0.1
0.05
0.01
2.5
3
3.5
4
4.5
5
5.5
6

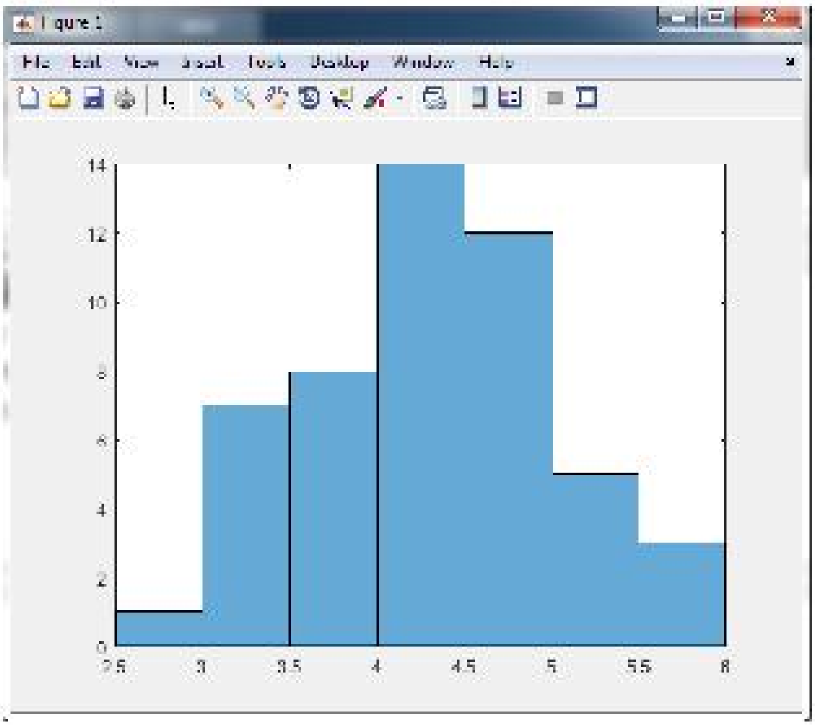


Figure 1
File Edit View Insert Tools Desktop Window Help
14
12
10
8
6
4
2
0
2.5
3
3.5
4
4.5
5
5.5
6

مجلة العلوم الرياضية والمرتبطة
المجلد الرابع / العدد الحادي والثلاثون – ديسمبر 2021 ISSN 2518-5470

بملاحظة قيمة λ وقيمة معامل الالتواء وشكل **P-P Normal Plot** وشكل مدرج التكرار للبيانات المحولة بالطريقة الثانية، نجد نتائج الطريقة الثانية تعطي نتائج مقاربة جداً لطريقة الاحتمال الأعظم. دوال ماتلاب التي تم استخدامها في هذه الطريقة تم عرضها في ملحق بنهاية هذه الورقة.

ثالثاً: طريقة استخدام قيم مقربة لمعلمة تحويلة بوكس-كوكس

هناك الكثير من البحاث يفضلون استخدام قيمة مقربة لمعلمة تحويلة القوة لبوكس-كوكس، λ، بدلاً من القيمة المثلى المقدرة، وذلك كما هو مبين في الجدول التالي:

قيمة λ	تحويلة القوة لكوكس-بوكس، $y(\lambda)$	اسم التحويلة
-3	$y(-3)=\frac{1}{y^3}$	**المقلوب التكعيبي (Inverse Cube)**
-2	$y(-2)=\frac{1}{y^2}$	**المقلوب التربيعي (Inverse Square)**
-1	$y(-1)=\frac{1}{y}$	**المقلوب (Inverse)**
-0.5	$y(-0.5)=\frac{1}{\sqrt{y}}$	**مقلوب الجذر التربيعي (Inverse Square Root)**
0	$y(0)=\ln(y)$	**اللوغاريتمية (Logarithmic)**
0.5	$y(0.5)=\sqrt{y}$	**الجذر التربيعي (Square Root)**
1	$y(1)=y$	**لا تحويلة (N0 Transformation)**
2	$y(2)=y^2$	**التربيعية (Square)**
3	$y(3)=y^3$	**التكعيبية (Cube)**

يجب مرعاة شرط أنه عند استخدام التحويلات المبينة في الجدول أعلاه يجب أن تكون جميع مشاهدات العينة موجبة.

الاستنتاج

تبحث طريقة بوكس-كوكس عن الأس الأفضل لتحويل مشاهدات العينات العشوائية التي لا تتبع التوزيع الطبيعي إلى مشاهدات تتبع التوزيع الطبيعي، حيث يتم رفع كل مشاهدة من مشاهدات العينة لهذا الأس، هذا الأس يطلق عليه اسم معلمة تحويلة القوة لبوكس-كوكس ويرمز له عادة بالرمز λ. توجد عدة طرق لتقدير قيمة λ. تم في هذه الورقة التعرض لشرح طرقتين منها: (1) طريقة الاحتمال الأعظم و(2) طريقة فحص قيمة الاحتمال لاختبار شابيرو-ويلك (Shapiro-Wilk Test). تم تنفيذ الطريقتين باستخدام دوال البرمجية ماتلاب علي مشاهدات عينة عشوائية حجمها 50 مشاهدة تم توليدها من توزيع مربع كاي بدرجات 20 حيث تمت مقارنة نتائج الطريقتين وتبين بأنهما تعطيان نتائج متقاربة جداً مع نجاحهما في تحويل مشاهدات العينة إلى مشاهدات تتبع التوزيع الطبيعي.

ملحق

دوال ماتلاب المستخدمة في طريقة فحص قيمة الاحتمال لاختبار شابيرو-ويلك

```
function [y_lambda, lambda, c, S]=lambdaforBoxCoxtransformation(y)
%y = the non-normal data on which the Cox-Box transformation is to be applied.
%lambda = the estimated Box-Cox power transformation parameter.
%y_lambda = the Cox-Box transformed data usig the estimated lambda.
lambda0=[];
P_lambda0=[];
for lambda =-10:0.1/50:10
[lambda, y_lambda0, c]=BoxCoxtransformation(y, lambda);
[H,P] =swtest(y_lambda0); %Shapiro-Wilk and Shapiro-Francia normality tests.
lambda0=[lambda0; lambda];
P_lambda0=[P_lambda0; P];
end
[maxP, jj]=max(P_lambda0);
```

lambda=lambda0(jj);%the estimated Box-Cox power transformation parameter.

[lambda, y_lambda, c]=BoxCoxtransformation(y, lambda);

S=skewness(y_lambda);%Skewness of the transormed sample observatios.

clf

probplot(y_lambda),

%Probability plot. Produces a normal probability plot comparing the distribution

%of the Cox-Box transformed data, using the estimated lambda, to the normal %distribution.

pause

histogram(y_lambda),%lots a histogram of the the Cox-Box transformed data.

function [lambda, y_lambda,c]=BoxCoxtransformation(y, lambda);

%y = the non-normal data on which the Cox-Box transformation is to be applied.

%lambda = the Box-Cox power transformation parameter.

n=length(y);% the size of the non-normal data, y.

c=0;

minx=min(y);

if minx<0

kk=find(y<zeros(n,1));

Ny=y(kk); c=abs(minNy);

minNy=min(Ny);

y=y+c;

jj=find(y==zeros(n,1));

```
y(jj)=y(jj)+0.000001;
end
if lambda ==0
y_lambda=log(y);
else
y_lambda=((y.^lambda)-1)/lambda;
end
%y_lambda = the Cox-Box transformed data using the estimated lambda.
```

```
function [H, pValue, W] = swtest(x, alpha)
%SWTEST Shapiro-Wilk parametric hypothesis test of composite normality.
%SWTEST Shapiro-Wilk parametric hypothesis test of composite normality.
% by Ahmed BenSaïda. https://uk.mathworks.com/matlabcentral/fileexchange/13964-%shapiro-wilk-and-%shapiro-francia-%normality-tests
if numel(x) == length(x)
x = x(:); % Ensure a column vector.
else
error(' Input sample "X" must be a vector.');
end
x = x(~isnan(x));
if length(x) < 3
error(' Sample vector "X" must have at least 3 valid observations.');
end
if length(x) > 5000
```

```
warning('Shapiro-Wilk test might be inaccurate due to large sample size ( > 5000).');
end
if (nargin >= 2) && ~isempty(alpha)
if ~isscalar(alpha)
error(' Significance level ''Alpha'' must be a scalar.');
end
if (alpha <= 0 || alpha >= 1)
error(' Significance level ''Alpha'' must be between 0 and 1.');
end
else
alpha = 0.05;
end
x = sort(x); % Sort the vector X in ascending order.
n = length(x);
mtilde = norminv(((1:n)' - 3/8) / (n + 1/4));
weights = zeros(n,1); % Preallocate the weights.
if kurtosis(x) > 3
% The Shapiro-Francia test is better for leptokurtic samples.
weights = 1/sqrt(mtilde'*mtilde) * mtilde;
W = (weights' * x)^2 / ((x - mean(x))' * (x - mean(x)));
nu = log(n);
u1 = log(nu) - nu;
u2 = log(nu) + 2/nu;
mu = -1.2725 + (1.0521 * u1);
sigma = 1.0308 - (0.26758 * u2);
newSFstatistic = log(1 - W);
```

```
NormalSFstatistic = (newSFstatistic - mu) / sigma;
pValue = 1 - normcdf(NormalSFstatistic, 0, 1);
else
% The Shapiro-Wilk test is better for platykurtic samples.
c = 1/sqrt(mtilde'*mtilde) * mtilde;
u = 1/sqrt(n);
PolyCoef_1 = [-2.706056 , 4.434685 , -2.071190 , -0.147981 , 0.221157 , c(n)];
PolyCoef_2 = [-3.582633 , 5.682633 , -1.752461 , -0.293762 , 0.042981 , c(n-1)];
PolyCoef_3 = [-0.0006714 , 0.0250540 , -0.39978 , 0.54400];
PolyCoef_4 = [-0.0020322 , 0.0627670 , -0.77857 , 1.38220];
PolyCoef_5 = [0.00389150 , -0.083751 , -0.31082 , -1.5861];
PolyCoef_6 = [0.00303020 , -0.082676 , -0.48030];
PolyCoef_7 = [0.459 , -2.273];
weights(n) = polyval(PolyCoef_1 , u);
weights(1) = -weights(n);
if n > 5
weights(n-1) = polyval(PolyCoef_2 , u);
weights(2) = -weights(n-1);
count = 3;
phi = (mtilde'*mtilde - 2 * mtilde(n)^2 - 2 * mtilde(n-1)^2) / ...
(1 - 2 * weights(n)^2 - 2 * weights(n-1)^2);
else
count = 2;
phi = (mtilde'*mtilde - 2 * mtilde(n)^2) / ...
(1 - 2 * weights(n)^2);
end
```

```
if n == 3
weights(1) = 1/sqrt(2);
weights(n) = -weights(1);
phi = 1;
end
weights(count : n-count+1) = mtilde(count : n-count+1) / sqrt(phi);
W = (weights' * x) ^2 / ((x - mean(x))' * (x - mean(x)));
newn = log(n);
if (n >= 4) && (n <= 11)
mu = polyval(PolyCoef_3 , n);
sigma = exp(polyval(PolyCoef_4 , n));
gam = polyval(PolyCoef_7 , n);
newSWstatistic = -log(gam-log(1-W));
elseif n > 11
mu = polyval(PolyCoef_5 , newn);
sigma = exp(polyval(PolyCoef_6 , newn));
newSWstatistic = log(1 - W);
elseif n == 3
mu = 0;
sigma = 1;
newSWstatistic = 0;
end
NormalSWstatistic = (newSWstatistic - mu) / sigma;
pValue = 1 - normcdf(NormalSWstatistic, 0, 1);
% Special attention when n = 3 (this is a special case).
if n == 3
```

```
pValue = 6/pi * (asin(sqrt(W)) - asin(sqrt(3/4)));
end
end
H = (alpha >= pValue);
```

المراجع

Box, G. E. P., and D. R. Cox, (1964). An Analysis of Transformations, *Journal of the Royal Statistical Society*, Ser. B, 26 (Apr. 1964), 211-243.

Manly, B. F., (1976). Exponential data transformation, *The Statistician*, 25, 37-42.

Yeo, I.K. and Johnson, R.A., (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), pp.954-959.

Atkinson, A., Riani, M., and Corbellini, A. (2020). *The Box–Cox Transformation: Review and Extensions*, *Statistical Science*, pp. ISSN 0883–ISSN 4237

Shapiro, S. S.and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples)". Biometrika. 52 (3–4): 591–611.

Vélez JI, Correa JC and Marmolejo-Ramos F (2015). A new approach to the Box–Cox transformation. *Front. Appl. Math. Stat.* 1:12. doi: 10.3389/fams.2015.00012.