استخدام المحاكاة لمقارنة طرق حساب فترات الثقة لتوزيع Lognormal Distribution

م.صفاء كريم كاظم/ قسم الرياضيات وتطبيقات الحاسوب/ جامعة المثنى/كلية العلوم

الخلاصة:
هناك الكثير من الحالات في الواقع الحياتي تمتلك بيانات غالبا" ما تكون ذات التواء موجب positive skewed ،مثل هذه البيانات تتم نمذجتها باستخدام التوزيع Lognormal Distribution. وبما ان الاستدلال الإحصائي يعتمد بيانات لها توزيع الطبيعي كما هو معروف لذا تستخدم بعض التحويلات لتحويل البيانات التي تمتلك توزيعات غير طبيعية إلى بيانات تتوزع توزيعا" طبيعيا". ويعتبر التحويل اللوغارتمي Log-Transformation واحدا من أهم التحويلات التي يتم من خلالها تحويل البيانات التي تمتلك التوزيع Lognormal Distribution إلى بيانات تمتلك توزيعا" طبيعيا".
وكما معروف يعتبر الوسيط median احد مقاييس النزعة المركزية الأفضل في حالة البيانات المتناثرة أو المتباعدة المستخدمة في وصف التوزيعات ذات الالتواء الموجب(التوزيع اللوغاريتم الطبيعي).
أن الهدف من هذا البحث هو تركيب وتقدير فترات ثقة لبيانات تتوزع توزيع لوغاريتم طبيعي وذلك بتقدير المتوسط θ بدلا" من الوسيط بعد تحويل البيانات للتوزيع الطبيعي حيث يعتبر المتوسط أفضل من الوسيط في تقدير معالم المجتمع ،حيث تم أجراء دراسة محاكاة لحساب ومقارنة فترات الثقة لبيانات تتوزع Lognormal Distribution باستخدام عدة طرائق لحساب فترات ثقة.

المقدمة:
أن الكثير من الظواهر الطبيعية تتصف بياناتها بميزة الالتواء الموجب Positive Skewed مثل هذه البيانات تمتلك توزيع Lognormal Distribution ، وبما أن الإحصائيين يعتمدون بالدرجة الأساس على التوزيع الطبيعي في عملية الاستدلال لذا يتم تحويل هذه البيانات باستخدام التحويل الوغارتمي لتتوزع توزيعا طبيعيا ".

لو افترضنا ان المتغير العشوائي X يمتلك توزيع الطبيعي اللوغارتمي Lognormal Distribution بمتوسط μ وتباين σ^2 ، وإذا فرضنا إن :

$$Y=\log(x) \ldots\ldots\ldots (1)$$

فان المتغير y يمتلك توزيعا" طبيعيا بمتوسط μ وتباين σ^2 .

وكما هو معروف أذا كان x~lognormal فان[٣]:

$$Median(y) =\log ((median(x)) \ldots\ldots\ldots (2)$$

في هذا البحث سيتم استخدام المتوسط الحسابي بدلا من الوسيط ،حيث سيتم الاستدلال حول ذلك من خلال استخدام الطرق المستخدمة لتركيب ومقارنة فترات الثقة قيد الدراسة.

المتوسط الحسابي θ =E(x) للمتغير x يحسب كالأتي :

$$\theta = e^{(\mu+\frac{1}{2}\sigma^2)} \ldots\ldots\ldots (2)$$

هذا يعني أن قيمة المتوسط للمتغير x لا تساوي لوغاريتم قيمة المتوسط للمتغير y .

أن قيمة المقدر للوسط الحسابي $\hat{\theta}$ المستخدم في تركيب فترات الثقة لـــ Lognormal Distribution تحسب من بيانات العينة التي تمتلك توزيعا" طبيعيا" وفق الصيغة التالية :

$$\hat{\log(\theta)} = \left(\bar{y} + \frac{S^2}{2} \right) \text{.....(3)}$$

أما المقدر للتباين فيحسب وفق الصيغة التالية:

$$\hat{\text{var}}\left(\hat{\log(\theta)} \right) = \frac{S^2}{n} + \frac{S^4}{2(n-1)} \text{.....(4)}$$

طرق حساب فترات الثقة للمتوسط الحسابي $E(x)=\theta$: [٣]،[٤]

هناك العديد من الطرق المستخدمة في تركيب فترات الثقة ، حيث سيتم اعتماد بعض منها لحساب فترات ثقة الوسط الحسابي $E(x)=\theta$ هي كالأتي :

١- طريقة naïve:-

يتم في هذه الطريقة تركيب فترة ثقة للمتوسط μ ، أن فترة الثقة لمتوسط بيانات التحويل اللوغارتمي

يتم حسابها كالأتي :

$$\bar{y} \pm Z_{\left(1-\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}} \text{.....(5)}$$

حيث أن z قيمة جدوليه.

٢- طريقةCox :-

تقترح هذه الطريقة بان فترة الثقة للمتوسط الحسابي $\log(\theta)$ يمكن أن تحسب وفق الصيغة الآتية :

$$\bar{Y} + \frac{S^2}{2} \pm Z\sqrt{\frac{S^2}{n} + \frac{S^4}{2(n-1)}} \text{.....(6)}$$

حيث أن حدود فترة الثقة لهذه الطريقة تحول بشكل خلفي للحصول على فترة ثقة لـ θأي أن :

$$[e^{\text{Lower}}, e^{\text{Upper}}]$$

٣- طريقة Cox المعدلة:-

تعتبر هذه الطريقة تعديلا للطريقة Cox السابقة، حيث يتم في هذه الطريقة الاعتماد على t بدلا" من z

بدرجة حرية معينة تعتمد على درجة حرية تقدير σ^2:

$$\bar{Y} + \frac{S^2}{2} \pm t\sqrt{\frac{S^2}{n} + \frac{S^4}{2(n-1)}} \text{.....(7)}$$

٤- طريقة تعتمد على نظرية المعاينة الكبيرة Large-sample approach

تنص هذه الطريقة على انه بدلا" من حساب فترة الثقة اعتمادا" على البيانات المحولة y حيث أن y=log(x) يمكن أن تحسب مباشرة" من متوسط وتباين العينة $\overline{X}$ و S_x^2 على التوالي وبدون استخدام أي تحويلات.

وحسب نظرية الغاية المركزية Central Limit Theorem أن توزيع متوسط العينة $\overline{X}$ يمكن أن يقرب للتوزيع الطبيعي عندما يكون حجم العينة n كبيرا"، أي أن فترة الثقة وفق هذا الأسلوب تحسب كالأتي:

$$\overline{X} \pm Z\sqrt{\frac{S_X^2}{n}} \quad \text{.....(8)}$$

دراسة محاكاة:

تم إجراء التطبيق العملي باستخدام المحاكاة لمقارنة فترات الثقة المحسوبة باستخدام الطرائق الأربعة التي تم ذكرها سابقا" حيث تم كتابة برنامج باستخدام برنامج S-Plus خاص بالتجربة حيث تم تكرارها ١٠٠٠ مرة، أذا تم بناء التجربة وفق الخوارزمية أدناه:

خوارزمية التجربة

١- تولد عينة بحجم (n) تتوزع توزيع lognormal distribution.
٢- حساب $\bar{y}$ وs^2 للعينة المحسوبة في الخطوة رقم (١) .
٣- حساب فترة الثقة بالاعتماد على $\bar{y}$ وs^2 في الخطوة رقم (٢) وبالطرائق الأربع التي تم استعراضها وتحت مستوى ثقة معين .
٤- المقارنة بين فترات الثقة على أساس معيار طول الفترة length criterion. أو أضيق فترة محسوبة لتمثل القيمة الأقرب أو الأمثل لمعلمة المجتمع .

تحليل النتائج

الجدول رقم (١) يبين النتائج التي تم الحصول عليها من خلال إجراء تجربة محاكاة كررت ١٠٠٠ مرة ولإحجام عينة مختلفة (n=5, 15, 30, 50, 100, 250)، حيث تم سحب العينات من بيانات تتوزع توزيع Lognormal Distribution بالمعلمة μ=5 وσ^2 =1 ولكافة المجتمعات التي سحبت منها العينات .

جدول رقم (١) يبين النتائج تجربة محاكاة لحساب فترة الثقة باستخدام أربع طرائق لبيانات تتوزع Lognormal Distribution بالمعلمة μ=5 وσ^2 =1 باستخدام أربعة طرائق ولإحجام عينات مختلفة وتحت مستوى ثقة ٩٥%

method	Naïve method	Cox method	Cox modify method	Large-sample approach
n	Length conf.intrv.	Length conf.intrv.	Length conf.intrv.	Length conf.intrv.
5	110.05	219.58	226.7	192.32

15	176.13	217.33	223.45	163.85
30	188.01	211.57	221.79	155.74
50	202.22	201.89	218.07	144.43
100	213.13	199.77	212.66	139.45
250	234.44	198.78	205.38	133.33

نلاحظ الجدول أعلاه أن طريقة Naïve method تكون غير فعالة في العينات الكبيرة، حيث كلما زاد حجم العينة زاد طول الفترة.على العكس من طريقة Cox method حيث نلاحظ استقرار ملحوظ في طول الفترة ولكافة إحجام العينة قيد الدراسة، وكذلك الحال بالنسبة لطريقة Cox modify method حيث نلاحظ استقرار في طول فترة الثقة ولكافة أحجام العينة، حيث كلما زاد حجم العينة يقابله ضيق في طول الفترة. أما بالنسبة لطريقة Large-sample approach نلاحظ انخفاض شديد في طول الفترة وهذا يفسر نظرية الغاية المركزية التي تعتمد عليها هذه الطريقة بحساب فترة الثقة انه كلما زاد حجم العينة تكون البيانات اقرب للتوزيع الطبيعي الذي يعكس أمثلية هذه الفترات لتمثل التقدير الأقرب لمتوسط المجتمع المسحوبة منه العينة.

الاستنتاجات

١- بعد تحليل النتائج نلاحظ من الجدول رقم (١) أن طريقة Large-sample approach هي الأفضل بشكل مطلق عند زيادة حجم العينة، على العكس من طريقة Naïve method غير جيدة كلما زاد حجم العينة.

٢- تعتبر طريقة Cox method وCox modify method هي الأفضل من ناحية طول الفترة بعد طريقة Large-sample approach وبالأخص طريقة Cox modify method كون هذه الطريقة تعمد في تركيبها على t على عكس من طريقة Cox method التي تعتمد على z .

المصادر

1-land, C.E., 1971."Confidence intervals for linear functions of the normal mean and variance".Ann.Math.Stat. V (42):1187-1205.

2-Olsson, U.," Confidence intervals for the mean of a Lognormal Distribution", Jou.of Stas .Edu.V (13), No.1, 2005.

3- Weerahndi, S.,"Exact statistical methods for data analysis", New York: spring-verlag, 1995.

4- Zhou, X-H., and Gao, S.," Confidence intervals for the lognormal mean", statistics in medicine, no.16, pp 783-790, 1997.