المجلة العراقية للعلوم الاحصائية (16) 2009
ص ص [185–202]

توصيف التوقع الشرطي بنموذج انحدار خطي بسيط لتوزيع ثنائي الحدين المشروط بحجم عينة متغير

هيثم طه اليوسف* فداء نوئيل عنائي**

الملخص

يتناول هذا البحث دراسة مسالة التوقع الشرطي وعلاقته بنموذج الانحدار من خلال توصيف معادلة التوقع الشرطي ذات الطابع الخطي بنموذج انحدار خطي بسيط ، واستخدام تحويل القوى للمتغير التوضيحي للحصول على أفضل مطابقة ممكنة للنموذج المشتق من توزيع ثنائي الحدين المشروط بحجـم عينـة متغير . تم تقدير معلمات النموذج الموصوف بالاعتماد على بيانات تجريبية تم توليدها باستخدام قيم مختلفة لمعلمات التوزيع الاحتمالي المشروط . تم الحصول على أحسن مقدر لمعلمة القوى في النموذج الموصوف بطريقتي الترجيح الأعظم و Draper & Smith والتي أعطت أفضل مطابقة للنموذج المقترح وأفضل مقدر للتوقع الشرطي للتوزيع الاحتمالي المشروط وقد تم اثبات ان منهجية التطبيق المقترحة اكفأ تجريبيا من الطرائق التقليدية .

لقد كان للزيادة في احتمال النجاح p تأثير كبير على أفضلية مطابقة النموذج ومن ثم كان له تأثير على مقدر التوقع الشرطي في توزيع ثنائي الحدين المشروط وكانت هذه النتيجة واضحة بسبب تناقص معامل التحديد R^2 في طريقة Draper & Smith وكذلك تزايد معدل مربعات البواقي بزيادة p.

Specification of the Conditional Expectation by Simple Linear Regression Model For Binomial Distribution Conditioned with Varying Sample Size.

Abstract

In this research, we consider the study of conditional expectation and it's relationship with regression model. The

---

*استاذ مساعد / كلية دهوك الجامعة
**مدرس مساعد/ كلية علوم الحاسبات والرياضيات

تاريخ التسلم :2009/6/25 ———— تاريخ القبول : 12 /10 /2009

conditional expectation has a linear form which is specified as a simple linear regression model. The power transformation was used on the predictor variable which gave the best possible fit for the model which was derived from the binomial distribution conditioned with varying sampl size.

The parameters of specified models were estimated by depending on emprical data which were simulated with different values for the parameter of conditional probability distribution. The best estimator for the power parameter was found in two specified models by the maximum liklihood and Draper & Smith methods. These estimators gave the best fit to the suggested model and best estimator to the conditional expectations of conditional probability distribution and it was concluded that the suggested method was better than the ordinary method.

The increments in the probability of success (p) had a great effect on the best fitted model also the estimated conditional expectation of conditional binomial distribution was affected. This result was clear because of decreasing the coefficient of determination (R^2) in Draper & Smith and the mean square of residuals in maximum liklihood method with increase in (p).

1. المقدمـة:

إن المنحني الذي يمثل دالة التوقع الشرطي بين متغيرين في مجتمع التوزيعات الثنائية التي تضم زوجا من المتغيرات العشوائية - بدالة توزيع مشتركة ومعامل ارتباط يعبر عن اعتماد خطي غير تام بين المتغيرين - يسمى منحني الانحدار الحقيقي للمتغير الشرطي نسبة إلى المتغير الآخر. إذا كان التوقع الشرطي خطيا فان مخطط الانحدار سيكون عندئذٍ عبارة عن خط مستقيم يعبر عنه بدالة الانحدار الخطي البسيط . تستخدم هذه الدالة في وصف بيانات التوزيع وتقدير المعلمات المجهولة التي تصف تلك البيانات ويستدل منها على أهمية العلاقة بين المتغيرين وقوة هذه العلاقة واتجاهها ، كذلك يمكن استخدام معادلة الانحدار في تقدير الاستجابة والتنبؤ بما يفيد في عملية اتخاذ القرار.

تواجه الباحثين مشكلة تحقيق مطابقة معنوية للتوزيعات المقترحة على البيانات المفترضة او المتيسرة للظاهرة المدروسة عموما وتزداد هذه المشكلة تعقيدا عندما يتعلق الامر بالتوزيع الشرطي – لان وجود المتغير المستقل او التوضيحي يمثل قيدا مضافا في اصل الدالة الشرطية . يهتم هذا البحث بمسالة تحقق المطابقة من عدمها لتوزيع ثنائي الحدين المشروط بحجم عينة متغير على البيانات ذات العلاقة.

تضمن البحث عرضا للإطار النظري لتوزيع ثنائي الحدين المشروط بحجم عينة متغير والإطار النظري لمنهجية التطبيق المقترحة لبناء نموذج التوقع الشرطي للتوزيع من خلال استخدام تحويل القوى Power Transformation للمتغير التوضيحي لضمان تحقيق افضل مطابقة ممكنة للبيانات على التوزيع المفترض ، كما تم استخدام طريقتين لتقدير قيمة معلمة القوى هما طريقة الترجيح الأعظم Maximum Likelihood وطريقة Draper & Smith التكرارية.

هـدف البحـث:

يهدف البحث إلى اقتراح طريقة لبناء نموذج للتوقع الشرطي لتوزيع ثنائي الحدين المشروط بحجم عينة متغير من خلال توصيف العلاقة الشرطية بنموذج انحدار خطي بسيط يحقق أعلى متطلبات المطابقة (Fitting) باستخدام تحويل القوى .

2. اطار نظري لنماذج تقدير التوقع الشـرطي ومنهجية التطبيق المقترحة :

يمكن التعبير عن العلاقة الخطية بين المتغيرين العشوائيين X و Y وتحت افتراض ان الاول يمثل المتغير التوضيحي والثاني يمثل متغير الاستجابة بنموذج الانحدار الخطي البسيط على وفق المعادلة الآتية:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \qquad ....(1)$$

وكما هو معروف فان ε_i يمثل الحد الذي يعبر عن الخطأ العشوائي، أما α و β فتمثلان معلمات النموذج اللتين تصفان موقع خط الانحدار وميله .

من المعروف ان نموذج الانحدار الخطي البسيط يمثل حالة خاصة من النموذج الخطي العام (General Linear Model GLM)، وان شروط تحليل هذا النموذج واختباره تخضع لفرضيتين معروفتين [2] الفرضية الأولى هي ان

الخطأ يتوزع طبيعيا بمتوسط (صفر) وتباين σ^2، وهذه الفرضية ذات شقين ، الأول يتعلق بالتوقع $E(\varepsilon_i)=0$ ، ويمثل الجزء الذي يحدد ملاءمة النموذج (Appropriate Model)، والشق الثاني يمثل التباين $Var(\varepsilon_i)=\sigma^2$، وهو الجزء الذي يثبت تباين مشاهدات متغير الاستجابة دون أن يرتبط هذا التباين بقيمة التوقع التي تعتمد على المتغيرات التوضيحية. اما الفرضية الثانية فتتعلق بالأخطاء المتعاقبة التي يجب ان تكون غير مرتبطة مع بعضها .

إن الهدف الرئيس من تحليل الانحدار هو تقدير متوسط القيم لمتغير الاستجابة في مدى القيم المعطاة للمتغير التوضيحي، وان دالة الانحدار تصف الاعتمادية لكمية متغير الاستجابة على المتغير التوضيحي، كذلك فان التنبؤ المعقول A Plausible Prediction لقيم متغير الاستجابة سيكون عبارة عن التوقع الشرطي المنسوب إلى المتغير التوضيحي، أي أن توقع متغير الاستجابة عبارة عن دالة بدلالة المتغير التوضيحي، وهذا ما يسمى بانحدار المربعات الصغرى Least Squears Regression. فإذا كانت $f(x,y)$ تمثل دالة الكثافة الاحتمالية المشتركة Joint Probability density function لـ X وY و $f_X(x)$ تمثل دالة الكثافة الاحتمالية Probability density function للمتغير X، وΩ تمثل مجال معلمات التوزيع (Parameters Space) ، فان التوقع الشرطي لـY نسبة الى X=x ويكون دالة لـ x وΩ ولتكن φ(x) اما التوقع الشرطي لـ X نسبة إلى Y=y فسيكون دالة لـ y فقط ولتكن $\psi(y)$. ففي حالة كون φ(x) عبارة عن دالة خطية من x كأن تكون $\varphi(x)=a+bx$ حيث ان a و b تقديران للمعلمات الحقيقية لنموذج العلاقة فيمكن القول ان التوقع الشرطي لـY خطيا في x وان لـY توقع شرطي خطي وهو:

$$E(Y|x)=\frac{\int yf(y,x)dy}{f_X(x)}=a+bx$$

وان:

$$\int yf(y,x)dy=(a+bx)f_X(x) \qquad ....(2)$$

وباجراء التكامل لطرفي المعادلة (2) بالنسبة للمتغير x فان الناتج يكون:

$$E(Y) = a + bE(X)$$

وبالطريقة نفسها يمكن ايجاد $\varphi(x)$ في الحالة المتقطعة.

إن دالة التوقع الشرطي $E(Y|X = x)$ تسمى دالة الانحدار، وهذه الدالة توضح حقيقة أن متوسط التوزيع للمتغير Y نسبة إلى x له علاقة دالية تعود إلى المتغير X، التي يمكن تمثيلها بمعادلة الانحدار الخطي البسيط لـ Y نسبة إلى x . شرط تحقيق مطابقة جيدة وتمثيل مقبول للبيانات . ولتحقيق أعلى مطابقة ممكنة لنموذج الانحدار ومن ثم تحقيق أفضل تقدير لدالة التوقع الشرطي فان منهجية التطبيق التي يقترحها الباحثان تستند إلى تحويل المتغير التوضيحي باستخدام معلمة قوى تغطي انحناءات الدالة – تلك الانحناءات التي تفشل أية محاولة لتوصيف العلاقة بين المتغيرين بنموذج خطي – وكالاتي :

$$(Y|X = x) = \begin{cases} \alpha + \beta x_i^{\lambda} + \varepsilon_i & ; \quad \lambda \neq 0 \\ \alpha + \beta \ln x_i + \varepsilon_i & ; \quad \lambda = 0 \end{cases} \qquad ....(3)$$

وبذلك يمكن تقدير نموذج التوقع الشرطي على وفق الآتي:

$$\Psi(x;\Omega) = \begin{cases} \alpha + \beta x_i^{\lambda} + \varepsilon_i & ; \quad \lambda \neq 0 \\ \alpha + \beta \ln x_i + \varepsilon_i & ; \quad \lambda = 0 \end{cases} \qquad ....(4)$$

وتتيسر للباحثين طرائق عديدة لتقدير معلمات النموذج (5) ومعلمة القوى ومنها طريقتا الترجيح الأعظم والمربعات الصغرى الشائعتان وطريقة Draper & Smith وطريقة Box & Cox [4] والخوارزميات الاتية تمثل خلاصة طريقتي الترجيح الأعظم و Draper & Smith:

(أولا): <u>خوارزميـة تقديـر المعلمـات باستخـدام طريقـة الترجيح الأعظـم</u> :

1) يتم اختيار قيمة λ من المدى المختار، الذي يحدد عادة في المدى (2,2-) وربما أقل من ذلك في البداية، من ثم يتم القيام بتوسيع المدى عند الحاجة فيما بعد.

2) تحويل المتغير التوضيحي $x_i \to \psi(x_i)$.

3) تقدير معلمات نموذج الانحدار $Y_i = \alpha + \beta\psi(x_i)$ بطريقة المربعات الصغرى.

4) يتم تقدير دالة الترجيح الاعظم $L_{max}(\lambda)$ لقيم λ المختارة على وفق المعادلة الترجيح الآتية:

$$L_{\max}(\lambda) = -0.5n\ln\sigma_e^2 - 0.5n\ln(2\pi) - \frac{1}{2\sigma_e^2}\sum(e_i^2) \quad ....(5)$$

إذ أن n تمثل العدد الكلي للمشاهدات وان:

$$\hat{\sigma}_e^2 = \frac{\sum e_i^2}{n-2} \quad ....(6)$$

وان $e_i = y_i - \hat{\alpha} - \hat{\beta}X^{\hat{\lambda}}$.

5) تقدير دالة الترجيح الأعظم $L_{\max}(\lambda)$ للعديد من قيم λ في المدى المختار ورسم المنحني الخاص

بـ $L_{\max}(\lambda)$ ضد λ، إذ يتم البحث عن قيمة λ التي تعظم $L_{\max}(\lambda)$، وفي كل الأحوال يتم

اختيار λ من الباحث بعد أن يتم اختبار الحسابات وإكمالها.

ثانيا): خوارزميـة تقديـر المعلمـات باستخـدام طريقـة Draper & Smith:

1) يتم اختيار قيمة λ من المدى المفترض.

2) تحويل المتغير التوضيحي $x_i \rightarrow \psi(x_i)$.

3) تقدير معلمات نموذج الانحدار $Y_i = \alpha + \beta\psi(x_i)$ بطريقة المربعات الصغرى الاعتيادية.

4) تقدير قيمة معيار القرار R^2 أو MSE.

5) اتخاذ القرار.

إن عدم الحاجة الى تحويل تكافئ بناء نموذج للتوقع الشرطي للمتغير Y مشروطا بقيمة المتغير X أي $E(Y|X = x)$، أما اذا استدعت الضرورة اجراء التحويل على المتغير التوضيحي X فان الحالة تكون مكافئة لبناء نموذج للتوقع الشرطي للمتغير Y مشروطا بقيمة دالة المتغير التوضيحي $\psi(x;\Omega)$ المعرف على وفق المعادلة (4). إن اهتمام هذه الدراسة ببناء نموذج للتوقع الشرطي $E(Y|X = x)$ بوصفه حالةً خاصة او التوقع الشرطي $E(Y|\psi(x))$ بوصفه

حالةً عامة لا يؤثر في الفرضيات المطلوبة لبناء نموذج الانحدار الخطي البسيط بوصفه حالة خاصة من نموذج الانحدار الخطي العام، مادام التوقع المطلوب للمتغير Y من دون أي تحويل يمكن ان يتعارض مع مفهوم هذا المتغير وتعريفه في الواقع العملي أو في إطار الظاهرة تحت البحث.

4 توزيع ثنائي الحدين المشـروط بحجم عينة متغير:

يقال للمتغير العشوائي X انه يتوزع على وفق دالة توزيع ثنائي الحدين إذا كانت دالة الكتلة الاحتمالية لهذا المتغير تأخذ الشكل الآتي:

$$f_x(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0,1,2,...,n \quad ....(7)$$

إذ أن n وp تمثلان معلمتي التوزيع وان n عدد موجب صحيح وp تمثل احتمال نجاح المحاولة (x) وان $0 < p < 1$. ومن الواضح أن الدالة الاحتمالية (7) تمثل دالة لمتغير واحد وهو X . أما عند افتراض أن n متغير عشوائي يأخذ قيماً مختلفة ، وان $x \le n$ فان دالة توزيع ثنائي الحدين ستكون بالشكل الآتي بعد التعبير عن الدالة الاحتمالية $f_x(x;n,p)$ بالدالة $f_{x/n}(x|n)$ الآتية:

$$f_{x/n}(x|n) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0,1,2, ... ,n \quad ....(8)$$

$$n = 1,2, ...$$

وفي هذه الحالة فان الدالة (8) تعبر عن أية ظاهرة يمكن تبويب بياناتها على وفق الجدول (1) الآتي:

الجدول (1)

جـدول يوضـح كيفيـة تبـويب البيانـات المولـدة

n	x=0	x=1	x=2	x=3	x=4	x=5	x=6	x=7	x=8	x=9	x=10	Sum
1	$f_{1.1}$	$f_{1.2}$	$f_{1.3}$	$f_{1.4}$	$f_{1.5}$	$f_{1.6}$	$f_{1.7}$	$f_{1.8}$	$f_{1.9}$	$f_{1.10}$	$f_{1.11}$	$f_{1.j}$
2	$f_{2.1}$	$f_{2.2}$	$f_{2.3}$	$f_{2.4}$	$f_{2.5}$	$f_{2.6}$	$f_{2.7}$	$f_{2.8}$	$f_{2.9}$	$f_{2.10}$	$f_{2.11}$	$f_{2.j}$
3	$f_{3.1}$	$f_{3.2}$	$f_{3.3}$	$f_{3.4}$	$f_{3.5}$	$f_{3.6}$	$f_{3.7}$	$f_{3.8}$	$f_{3.9}$	$F_{3.10}$	$f_{3.11}$	$f_{3.j}$
4	$f_{4.1}$	$f_{4.2}$	$f_{4.3}$	$f_{4.4}$	$f_{4.5}$	$f_{4.6}$	$f_{4.7}$	$f_{4.8}$	$f_{4.9}$	$f_{4.10}$	$f_{4.11}$	$f_{4.j}$
5	$f_{5.1}$	$f_{5.2}$	$f_{5.3}$	$f_{5.4}$	$f_{5.5}$	$f_{5.6}$	$f_{5.7}$	$f_{5.8}$	$f_{5.9}$	$f_{5.10}$	$f_{5.11}$	$f_{5.j}$
6	$f_{6.1}$	$f_{6.2}$	$f_{6.3}$	$f_{6.4}$	$f_{6.5}$	$f_{6.6}$	$f_{6.7}$	$f_{6.8}$	$f_{6.9}$	$f_{6.10}$	$f_{6.11}$	$f_{6.j}$
7	$f_{7.1}$	$f_{7.2}$	$f_{7.3}$	$f_{7.4}$	$f_{7.5}$	$f_{7.6}$	$f_{7.7}$	$f_{7.8}$	$f_{7.9}$	$f_{7.10}$	$f_{7.11}$	$f_{7.j}$
8	$f_{8.1}$	$f_{8.2}$	$f_{8.3}$	$f_{8.4}$	$f_{8.5}$	$f_{8.6}$	$f_{8.7}$	$f_{8.8}$	$f_{8.9}$	$f_{8.10}$	$f_{8.11}$	$f_{8.j}$
9	$f_{9.1}$	$f_{9.2}$	$f_{9.3}$	$f_{9.4}$	$f_{9.5}$	$f_{9.6}$	$f_{9.7}$	$f_{9.8}$	$f_{9.9}$	$f_{9.10}$	$f_{9.11}$	$f_{9.j}$
10	$f_{10.1}$	$f_{10.2}$	$f_{10.3}$	$f_{10.4}$	$f_{10.5}$	$f_{10.6}$	$f_{10.7}$	$f_{10.8}$	$f_{10.9}$	$f_{10.10}$	$f_{10.11}$	$f_{10.j}$
sum	$f_{i.1}$	$f_{i.2}$	$f_{i.3}$	$f_{i.4}$	$f_{i.5}$	$f_{i.6}$	$f_{i.7}$	$f_{i.8}$	$f_{i.9}$	$f_{i.10}$	$f_{i.11}$	$f_{i.j}$

إذ أن $f_{i,j}$ تمثل التكرار في الخلية (i , j) في جدول البيانات.
وباستخدام طريقة الترجيح الأعظم لتقدير معلمة التوزيع p فان:

$$\hat{p} = \bar{x}/n \qquad n = 1,2,\dots \qquad \dots\dots\dots\dots(9)$$

ان البحث يهتم ببناء نموذج انحدار خطي بين المقدر $\hat{p}$ والمتغير $\Psi(n)$ الذي يمثل البيانات المحولة للمتغير n من خلال استخدام تحويل القوى ثم مساواة النموذج المقدر بمعادلة التوقع الشرطي لتحقيق اعلى ما يمكن تحقيقه من مطابقة للبيانات موضوع البحث على توزيع ثنائي الحدين المشروط بحجم عينة متغير .
وباستخدام نموذج تحويل المتغيرات التوضيحية الآتي:

$$\psi(n) = \begin{cases} n^{\lambda} & , \quad \lambda \neq 0 \\ \ln n & , \quad \lambda = 0 \end{cases} \qquad \dots\dots\dots\dots\dots\dots(10)$$

وبتوظيف هذا النموذج لتوصيف العلاقة بين n و $\hat{p}$ خطيا فان النموذج الآتي يمكن ان يعبر عن هذه العلاقة:

$$\hat{p} = \begin{cases} \alpha + \beta n^{\lambda} + \varepsilon & ; \quad \lambda \neq 0 \\ \alpha + \beta \ln n + \varepsilon & ; \quad \lambda = 0 \end{cases} \qquad \dots\dots\dots\dots\dots\dots\dots(11)$$

وباستخدام طريقة المربعات الصغرى الاعتيادية وبتعويض النموذج المقدر في المعادلة (4) يمكن الحصول على نموذج التوقع الشرطي الآتي:

$$\mathrm{E}(X|n) = \begin{cases} n(\hat{\alpha} + \hat{\beta} n^{\hat{\lambda}}) & ; \quad \lambda \neq 0 \\ n(\hat{\alpha} + \hat{\beta} \ln n) & ; \quad \lambda = 0 \end{cases} \qquad \dots\dots\dots\dots\dots\dots(12)$$

وان التباين الشرطي يكون:

$$V(X|n) = \begin{cases} n(\hat{\alpha} + \hat{\beta} n^{\hat{\lambda}})(1 - \hat{\alpha} - \hat{\beta} n^{\hat{\lambda}}) \\ n(\hat{\alpha} + \hat{\beta} \ln n)(1 - \hat{\alpha} - \hat{\beta} \ln n) \end{cases} \qquad \dots\dots\dots\dots(13)$$

5. خوارزمية التطبيق العملي لتوزيع ثنائي الحدين المشروط بحجم عينة متغير:

تم تطبيق الطريقة المقترحة في ثلاث تجارب بالاعتماد على ثلاث قيم متغيرة لـ p، إذ تم استخدام القيمة الافتراضية للمعلمة p التي ولدت على أساسها

البيانات، وحساب القيم التقديرية للمعلمة ذاتها بطريقة الترجيح الأعظم ، ومن ثم التقدير بالمنهجية المقترحة والمقارنة بين النتائج الثلاثة بالاعتماد على متوسط مربعات الخطأ (MSE) لكل طريقة وحسب الخطوات الموضحة في الخوارزمية الآتية:

1. توليد 300 مشاهدة للمتغير X تتوزع على وفق توزيع ثنائي الحدين بالمعلمة p.
2. تقدير قيمة التوقع الشرطي باستخدام القيمة الحقيقية للمعلمة p على وفق الصيغة الآتية:

$$u_n = \mathrm{E}_n(X|n) = np \quad , \quad n=1,2,...,10$$

3. تقدير قيمة المعلمة $\hat{p}$ بطريقة MLE وتعويضها في معادلة التوقع الشرطي والحصول على مقدرات جديدة لهذه المعلمة وتعويضها في معادلة التوقع الشرطي أي أن:

$$z_n = \mathrm{E}_n(X|n) = n\hat{p}$$

4. تقدير قيمة الـ MSE_1 من المعادلة الآتية:

$$MSE_1 = \frac{\sum_{n=1}^{10}(z_n - u_n)^2}{10}$$

5. تقدير معادلة الانحدار الآتية:

$$\tilde{p} = \hat{\alpha} + \hat{\beta} n^{\hat{\lambda}}$$

واختيار أفضل مقدر لـ λ الذي يعظم قيمة R^2 و دالة الترجيح الأعظم $L(\lambda)$.

6. تعويض قيم $\hat{\alpha}$ و $\hat{\beta}$ وقيم n في النموذج السابق والحصول على مقدرات $\tilde{p}$.
7. تعويض $\tilde{p}$ في معادلة التوقع الشرطي والحصول على:

$$h_i = \mathrm{E}_n(X|n) = n\tilde{p}$$

8. تقدير قيمة MSE_2 من جديد باستخدام قيم h_i وبالصيغة الآتية:

$$MSE_2 = \frac{\sum_{n=1}^{10}(h_n - u_n)^2}{10}$$

9. المقارنة بين المقدرات MSE_1 و MSE_2.

1-5 التجربـة الأولـى:

تم توليد 300 مشاهدة للمتغير العشوائي X الذي يأخذ القيم X=0,1,...,10 وبأحجام عينة متغيرة تأخذ القيم n=1,2,...,10 وعلى وفق توزيع ثنائي الحدين بالمعلمة P=0.2، والجدول (2) الآتي يمثل بيانات لتجربة واحدة من التطبيق الخاص بالتوزيع باستخدام البرنامج الإحصائي الجاهز Minitab والمبوبة في الجدول (2) على غرار الجدول (1) ووفقاً للآتي:

الجدول (2)
البيانـات الخاصـة بالتجربـة المولـدة عندما p=0.2

n	x=0	x=1	x=2	x=3	x=4	x=5	x=6	x=7	x=8	x=9	x=10	Sum
1	26	4	0	0	0	0	0	0	0	0	0	30
2	21	9	0	0	0	0	0	0	0	0	0	30
3	18	9	3	0	0	0	0	0	0	0	0	30
4	9	18	3	0	0	0	0	0	0	0	0	30
5	8	14	8	0	0	0	0	0	0	0	0	30
6	9	11	5	4	1	0	0	0	0	0	0	30
7	4	14	9	3	0	0	0	0	0	0	0	30
8	5	11	8	5	0	1	0	0	0	0	0	30
9	4	9	8	5	2	1	0	1	0	0	0	30
10	2	6	11	5	5	1	0	0	0	0	0	30
sum	106	105	55	22	8	3	0	1	0	0	0	300

أولاـ بتعويض قيمة p في معادلة التوقع الشرطي $u_n = E_n(X|n) = np$ تم الحصول على الجدول (3):

الجدول (3)
قيم التوقع الشرطي لتوزيع ثنائي الحدين المشروط بحجم عينة متغيرة عند تعويض p=0.2

n	p	$u_n = E_n(X\|n)$
1	0.2	0.2
2	0.2	0.4
3	0.2	0.6
4	0.2	0.8
5	0.2	1
6	0.2	1.2
7	0.2	1.4

8	**0.2**	**1.6**
9	**0.2**	**1.8**
10	**0.2**	**2**

ثانيا_ باستخدام طريقة الترجيح الأعظم MLE لتقدير المعلمة p وعلى وفق الصيغة الآتية:

$$\hat{p}_i = \bar{x}_j / n \qquad \mathbf{n = 1,2, \dots ,10}$$
$$\mathbf{i = 1,2, \dots ,n}$$

إذ أن $\bar{x}_j = \dfrac{\sum_j f_{ij} x_j}{\sum_j f_{ij}}$ تم الحصول على الجدول (4) الآتي:

الجدول (4)

القيم المقدرة للمعلمة p بطريقة الترجيح الأعظم وقيم التوقع الشرطي المقابلة لها عندما p=0.2

n	$\hat{p}$	$z_n = \mathrm{E}_n(\mathrm{X}\vert n) = n\hat{p}$
1	**0.1333**	**0.1333**
2	**0.15**	**0.3**
3	**0.1667**	**0.5001**
4	**0.2**	**0.8**
5	**0.2**	**1.0**
6	**0.2056**	**1.2336**
7	**0.1952**	**1.3664**
8	**0.1958**	**1.5664**
9	**0.2222**	**1.9998**
10	**0.2267**	**2.2670**

ثالثا_ بالعودة الى العلاقة بين $\hat{p}$ و n وتحت افتراض أن:

$$\hat{p}_i = \alpha + \beta n_i^{\lambda} + \varepsilon_i$$

تم استخدام طريقتي الترجيح الأعظم وطريقة Draper & Smith لتقدير معلمة التحويل، ووجد أن أفضل تقدير لـ λ هو $\lambda = 0.1$، إذ كانت أعلى قيمة لمعامل التحديد R^2=89.32 وأعلى قيمة لدالة الترجيح $L(\lambda) = 32.3864$، وعلى وفق النموذج الآتي:

$$\tilde{p}_i = -0.2128 + 0.3451 n_i^{0.1}$$

إذ أن $\tilde{p}_i$ تمثل القيم المقدرة للمتغير $\hat{p}_i$ والمعرف بتعريفه السابق نفسه.

الجدول (5)
القيم المقدرة للمعلمة p بالطريقة المقترحة وقيم التوقع الشرطي المقابلة لها عندما p=0.2

n	$\hat{p}_i$	$h_n = E_n(X\|n) = n\tilde{p}_i$
1	0.1323	0.1323
2	0.1571	0.3142
3	0.1724	0.5172
4	0.1836	0.7346
5	0.1926	0.9630
6	0.2	1.2003
7	0.2065	1.4452
8	0.2121	1.6968
9	0.2171	1.9542
10	0.2217	2.2169

رابعا_ تم تقدير متوسط مربعات الخطأ بين تقدير الترجيح الأعظم للتوقع الشرطي z_n والقيمة الحقيقية للتوقع الشرطي المستند إلى قيمة p=0.2 u_n على وفق المعادلة الآتية:

$$MSE_1 = \frac{\sum_{n=1}^{10}(z_n - u_n)^2}{10}$$

وبين منهجية التطبيق المقترحة h_n و u_n على وفق المعادلة الآتية:

$$MSE_2 = \frac{\sum_{n=1}^{10}(h_n - u_n)^2}{10}$$

تم الحصول على MSE_1=0.0139 وMSE_2=0.0107 وبالمقارنة بين هذه التقديرات يتبين أن طريقة التطبيق المقترحة أكثر كفاءة من طريقة الترجيح الأعظم.

2-5 <u>التجربـة الثانيـة</u>:

في هذه التجربة تم توليد 300 مشاهدة للمتغير X بأحجام عينة مختلفة n=1,2,...,10 وتحت افتراض أن p=0.5 من ثم تبويبها في الجدول (6) الآتي:

الجدول (6)
البيانات الخاصة بالتجربة والمولدة عندما p=0.5

n	x=0	x=1	x=2	x=3	x=4	x=5	x=6	x=7	x=8	x=9	x=10	Sum
1	19	11	0	0	0	0	0	0	0	0	0	30
2	8	17	5	0	0	0	0	0	0	0	0	30
3	5	12	11	2	0	0	0	0	0	0	0	30
4	2	7	12	8	1	0	0	0	0	0	0	30
5	0	5	5	14	5	1	0	0	0	0	0	30
6	0	4	8	10	6	2	0	0	0	0	0	30
7	1	0	5	7	9	5	3	0	0	0	0	30
8	0	0	1	10	6	6	6	1	0	0	0	30
9	1	0	2	7	7	5	6	1	1	0	0	30
10	0	0	0	4	9	8	5	4	0	0	0	30
sum	36	56	49	62	43	27	20	6	1	0	0	300

أولا_ بتعويض قيمة p في معادلة التوقع الشرطي $u_n = E_n(X|n) = np$ والنتائج مبينة في الجدول (7):

الجدول (7)
قيم التوقع الشرطي لتوزيع ثنائي الحدين المشروط بحجم عينة متغيرة عند تعويض p=0.5

n	p	$u_n = E_n(X\|n)$
1	0.5	0.5
2	0.5	1.0
3	0.5	1.5
4	0.5	2.0
5	0.5	2.5
6	0.5	3.0
7	0.5	3.5
8	0.5	4.0
9	0.5	4.5
10	0.5	5.0

ثانيا_ باستخدام طريقة الترجيح الأعظم MLE لتقدير المعلمة p تم الحصول على الجدول (8) الآتي:

الجدول (8)

القيم المقدرة للمعلمة p بطريقة الترجيح الأعظم وقيم التوقع الشرطي المقابلة لها عندما p=0.5

n	$\hat{p}$	$z_n = E_n(X\|n) = n\hat{p}_i$
1	0.36667	0.36667
2	0.450000	0.90000
3	0.44436	1.33330
4	0.491675	1.96670
5	0.526660	2.63330
6	0.466667	2.80000
7	0.523814	3.66670
8	0.537500	4.30000
9	0.477778	4.30000
10	0.486670	4.86670

ثالثا_ بتمثيل العلاقة بين $\hat{p}$ و n بنموذج انحدار خطي بسيط، وباستخدام طريقتي الترجيح الأعظم وطريقة Draper & Smith لتقدير معلمة التحويل، وجد أن أفضل تقدير لـ λ هو $\lambda = -1.1$، إذ كانت أعلى قيمة R^2=74.4% وأعلى قيمة لدالة الترجيح هي $L(\lambda) = 22.9484$ وعلى وفق النموذج الآتي:

$$\tilde{p}_i = 0.5181517 - 0.152844 n_i^{-1.1}$$

ومن ثم تم الحصول على الجدول (9) الآتي:

الجدول (9)

القيم المقدرة للمعلمة p بالمنهجية المقترحة وقيم التوقع الشرطي المقابلة لها عندما p=0.5

n	$\tilde{p}_i$	$h_n = E_n(X\|n) = n\tilde{p}_i$
1	0.365307	0.36531
2	0.446847	0.89369
3	0.472504	1.41751
4	0.484886	1.93955
5	0.492127	2.46063
6	0.496856	2.98113
7	0.500177	3.50124
8	0.502633	4.02106
9	0.504518	4.54066
10	0.506010	5.06010

رابعا_ تم تقدير متوسط مربعات الخطأ بين تقدير الترجيح الأعظم للتوقع الشرطي z_n والقيمة الحقيقية للتوقع الشرطي u_n المستندة إلى قيمة p=0.5، وكانت قيمته مساوية لـ $MSE_1=0.0290001$، وبين التقدير الناتج من تطبيق المنهجية المقترحة h_n و u_n ، وتم الحصول على النتيجة $MSE_2=0.004751$، ويتبين أن منهجية التطبيق المقترحة أكثر كفاءة من طريقة الترجيح الأعظم للتقدير.

3-5 التجربـة الثالثـة:

في هذه التجربة تم استخدام القيمة 0.7 للمعلمة p في توليد 300 مشاهدة للمتغير X على وفق توزيع ثنائي الحدين المشروط بحجم عينة متغير والجدول (10) الآتي يمثل بيانات لتجربة واحدة من التطبيق الخاص بالتوزيع باستخدام البرنامج الإحصائي الجاهز Minitab والمبوبة على وفق الجدول (1) وكما يأتي:

الجدول (10)
البيانات الخاصة بالتجربة المولدة عندما p=0.7

n	x=0	x=1	x=2	x=3	x=4	x=5	x=6	x=7	x=8	x=9	x=10	Sum
1	14	16	0	0	0	0	0	0	0	0	0	30
2	6	12	12	0	0	0	0	0	0	0	0	30
3	0	7	13	10	0	0	0	0	0	0	0	30
4	0	0	7	15	8	0	0	0	0	0	0	30
5	0	3	4	9	10	4	0	0	0	0	0	30
6	0	1	2	9	9	8	1	0	0	0	0	30
7	0	0	1	7	5	8	6	3	0	0	0	30
8	0	0	1	3	3	5	7	10	1	0	0	30
9	0	0	2	0	3	3	10	5	6	1	0	30
10	0	0	0	0	1	1	4	11	6	6	1	30
sum	20	39	42	53	39	29	28	29	13	7	1	300

أولا_ بتعويض قيمة p في معادلة التوقع الشرطي $u_n = \mathrm{E}_n(X|n) = np$ والنتائج مبينة في الجدول (11):

الجدول (11)
قيم التوقع الشرطي لتوزيع ثنائي الحدين المشروط بحجم عينة متغيرة عند تعويض p=0.7

n	p	$u_n = \mathrm{E}_n(\mathrm{X}|n)$
1	0.7	0.7
2	0.7	1.4
3	0.7	2.1
4	0.7	2.8
5	0.7	3.5
6	0.7	4.2

7	**0.7**	**4.9**
8	**0.7**	**5.6**
9	**0.7**	**6.3**
10	**0.7**	**7.0**

ثانيا_ باستخدام طريقة الترجيح الأعظم MLE لتقدير المعلمة p تم الحصول على الجدول (12) الآتي:

الجدول (12)

القيم المقدرة للمعلمة p بطريقة الترجيح الأعظم وقيم التوقع الشرطي المقابلة لها عندما p=0.7

n	$\hat{p}$	$z_n = \mathrm{E}_n(\mathrm{X}\|n) = n\hat{p}$
1	**0.5333**	**0.5333**
2	**0.6**	**1.2**
3	**0.6667**	**2.0001**
4	**0.7583**	**3.0332**
5	**0.6533**	**3.2665**
6	**0.6333**	**3.7998**
7	**0.6667**	**4.6669**
8	**0.7**	**5.6**
9	**0.6778**	**6.1002**
10	**0.74**	**7.4**

ثالثا_ بالربط بين $\hat{p}$ و n بنموذج انحدار خطي بسيط، وباستخدام طريقتي الترجيح الأعظم وطريقة Draper & Smith لتقدير معلمة التحويل، وجد أن أفضل تقدير لـ λ هو $\lambda = -1.1$ التي تقابل أعلى قيمة لـ R^2=62.08% وكذلك فانها تقابل أعلى قيمة لدالة الترجيح التي تساوي $L(\lambda) = 18.3583$ وعلى وفق النموذج الآتي: $\tilde{p}_i = 0.7116 - 0.1816 n_i^{-1.1}$

يمكن ملاحظة انه عندما يزداد حجم العينة الى $n \geq 30$ سوف تزداد المطابقة للبيانات بزيادة احتمالية النجاح حيث تؤمن اقتراب توزيع ثنائي الحدين من التوزيع الطبيعي.

الجدول (13)
القيم المقدرة للمعلمة p بالمنهجية المقترحة وقيم التوقع الشرطي المقابلة لها عندما p=0.7

n	$\tilde{p}$	$h_n = E_n(X\|n) = n\tilde{p}$
1	0.53	0.53
2	0.6269	1.2538
3	0.6574	1.9721
4	0.6721	2.6884
5	0.6807	3.4035
6	0.6863	4.1179
7	0.6903	4.8318
8	0.6932	5.5454
9	0.6954	6.2587
10	0.6972	6.9719

رابعا_ تم تقدير متوسط مربعات الخطأ بين تقدير الترجيح الأعظم للتوقع الشرطي z_n والقيمة الحقيقية للتوقع الشرطي u_n المستندة إلى قيمة p=0.7، وكانت قيمته مساوية لـ $MSE_1=0.0601$، وبين التقدير الناتج من تطبيق المنهجية المقترحة h_n و u_n ، وتم الحصول على النتيجة $MSE_2=0.0105$، ويتبين أيضا أن منهجية التطبيق المقترحة أكثر كفاءة من طريقة الترجيح الأعظم لتقدير نموذج التوقع الشرطي.

6. الاستنتاجات:

من خلال تطبيق المنهجية المقترحة توصل الباحثان الى مجموعة الاستنتاجات الاتية :

1. تطابق نتائج طريقتي الترجيح الأعظم (المستندة الى تعظيم دالة الترجيح كمعيار لتقدير معلمة القوى والتي تفترض طبيعية البيانات كما هو معروف) و Draper & Smith (المستندة على تعظيم معامل التحديد لتقدير معلمة القوى ومعلمات نموذج العلاقة بطريقة OLS) . ان هذا التطابق في تقدير نموذج التوقع الشرطي بطريقتين تختلفان في الفروض والأسس النظرية كان بمثابة التأكيد على سلامة الفروض والأسس النظرية لمنهجية التطبيق المقترحة .
2. ان تغير قيمة المعلمة p في توزيع ثنائي الحدين المشروط بحجم عينة متغير تؤثر في النموذج المقترح اذ انه كلما ازدادت قيمة المعلمة p فان مطابقة بيانات

النموذج تكون أقل أي أن قيمة R^2 تقل وكذلك فان قيم لوغاريتم دالة الترجيح الأعظم $L(\lambda)$ تقل أيضا.

3. عندما يزداد حجم العينة الى $n \geq 30$ سوف تزداد المطابقة للبيانات بزيادة احتمالية النجاح وذلك بالاستناد الى نظرية الغاية المركزية التي تؤمن اقتراب توزيع ثنائي الحدين من التوزيع الطبيعي عندما يكون حجم العينة كبيرا وقيمة p تقترب من الواحد الصحيح .

المصادر

1. الصفار، آفان سفر الياس (2008)، "استخدام تحويلات القوى في منهجية سطوح الاستجابة"، رسالة ماجستير في الإحصاء غير منشورة مقدمة إلى مجلس كلية علوم الحاسبات والرياضيات في جامعة الموصل.
2. هرمز، أمير حنا (1990)، "الإحصاء الرياضي"، دار الكتب للطباعة والنشر، جامعة الموصل.
3. اليوسف، هيثم طه، والدباغ، ظافر عاصم (1996)، "بناء نموذج لتقدير متوسط عدد الدخول الفردية في الأسرة العراقية"، مجلة وقائع المؤتمر العلمي الثامن للجمعية العراقية للعلوم الإحصائية، بغداد.
4. Draper, N.R. & Smith, H. (1981), "Applied Regression Analysis", 2nd Edition, Jhon Wiley & Sons, New York.
5. Papoulis, A. & Pillai, S.U. (2002), "Probability, Random Variables and Stochastic Processes", 4th Edition, McGraw-Hill Companies, New York
6. Poo, Juan M. Rodriguez (2003), "Computer – Aided Introduction to Econometrics",http://www.quantlet.com/mdstat/scripts/xegbo/pdf/xegbopdf.pdf
7. Soong, T.T. (2004), "Fundamentals of Probability and Statistics for engineers", John Wiley & Sons Ltd, England.
8. Teekens, R. (1985), "Mapping Personal to Household Income Distribution: A Submodel", Institute of Social Studies, The Hogue Netherlands.
9. Wooldridge, J.M. (1994), "A Simple Specification Test For the Predictive Ability of Transformation Models", The Review of Economics and Statistics, Vol. LXXVI, No.1.