

مقارنة بين طريقتي مقدرات لابلاس وهوبر الحصين لتقدير معلمات أنموذج الانحدار اللوجستي

أ.م. ايمان حسن أحمد / جامعة بغداد / كلية الإدارة والاقتصاد
الباحث / ضمياء حامد شهاب

تاريخ التقديم: 2017/4/9
تاريخ القبول: 2017/5/10

المستخلص

يعد أنموذج الانحدار اللوجستي من نماذج الانحدار المهمة، حيث يلقي اهتماماً واضحاً في معظم الدراسات التي تأخذ طابعاً أكثر تقدماً في عملية التحليل الاحصائي. أن طرائق التقدير الاعتيادية تفشل في التعامل مع البيانات التي تتضمن وجود القيم الشاذة حيث أن لها تأثير غير مرغوب على النتائج. سنستعرض في هذا البحث طرائق لتقدير معلمات انموذج الانحدار اللوجستي وهذه الطرائق هي: طريقة مقدر لابلاس (Laplace estimator) (LP-) وطريقة مقدر هوبر الحصين (Huber estimator) (H). إذ تم اجراء المقارنة بين هاتين الطريقتين من خلال أسلوب المحاكاة وبأستعمال معيار المقارنة متوسط مربعات الخطأ (MSE) بنسب مختلفة من التلوث ولحجوم عينات مختلفة للوصول الى الطريقة الأفضل في تقدير المعلمات. واتضح من خلال المقارنة أن طريقة (H) هي الأفضل في تقدير معلمات أنموذج الانحدار اللوجستي.

المصطلحات الرئيسية للبحث / الانحدار اللوجستي، البيانات الثنائية، مقدرات لابلاس (LP-)، مقدرات هوبر الحصين (H).





مقارنة بين طريقتي مقدرات لابلاس وهوبر الحصين لتقدير معلمات أنموذج الانحدار اللوجستي

Introduction

1-1 المقدمة

يعد أنموذج الانحدار اللوجستي من نماذج الانحدار اللاخطية القابلة للتحويل الى نماذج خطية، حيث يهتم بتحليل البيانات التي يكون فيها متغير الاستجابة من النوع المتقطع ويأخذ احدى القيمتين (1,0) اي ثنائي الاستجابة (Binary Response)، ويستعمل هذا الانموذج لوصف العلاقة بين متغير الاستجابة والمتغيرات التوضيحية (التفسيرية)، وتبرز اهمية استخدامه في المجالات المتعلقة بعلم الحياة والعلوم الزراعية والطبية والاجتماعية اي بشكل عام في الدراسات ذات الطابع التجريبي، نظراً الى خصوصية البيانات في هذه المجالات والتي هي غالباً ماتكون من النوع (Binary) ثنائية الاستجابة. [2]

ان طرائق التقدير التقليدية لمعلمات أنموذج الانحدار اللوجستي عند تحليل البيانات الثنائية تكون ضعيفة في معالجة المشكلات بين البيانات، ومن بين احدى تلك المشاكل وجود القيم الشاذة، الامر الذي يستوجب استخدام طرائق لمعالجة تلك المشاكل تتمثل بكونها طرائق بديلة عن الطرائق التقليدية وتكون اكثر كفاءة في التقدير تدعى (بطرائق التقدير الحصينة) (Robust estimation methods) حيث تتصف بأنها قليلة الحساسية تجاه الشواذ أي لا تتأثر كثيراً بوجود القيم الشاذة، اذ يتم الحصول من خلالها على مقدرات حصينة تمتلك كفاءة عالية. [3]

Problem of the Research

2-1 مشكلة البحث

ان مشكلة البحث تكمن في تقدير معلمات أنموذج الانحدار اللوجستي والذي يكون فيه متغير الاستجابة من النوع الثنائي (Binary) في حالة احتواء البيانات على قيم شاذة، حيث أن ظهور هذه القيم ضمن البيانات يؤثر بشكل سلبي على تحليلها، وأن استعمال طرق التقدير الاعتيادية في تقدير المعلمات تكون ضعيفة في معالجة هذه المشكلة وبالتالي تؤثر على عدم دقة النتائج في عملية التقدير، لذا يلجأ الباحثين الى استعمال طرائق اخرى لمعالجة تلك المشكلة تدعى (بطرائق التقدير الحصينة) (Robust estimation methods) حيث تتصف بأنها قليلة الحساسية تجاه الشواذ أي لا تتأثر كثيراً بوجود القيم الشاذة.

Object of research

3-1 هدف البحث

الهدف من البحث هو الحصول على افضل تقدير لمعلمات أنموذج الانحدار اللوجستي في حالة وجود القيم الشاذة في البيانات باستعمال طريقة مقدرات لابلاس (LP-) وطريقة مقدرات هوبر الحصين (H)، من خلال معيار المقارنة متوسط مربعات الخطأ (MSE) بأختلاف احجام العينات ونسب التلوث المفترضة.

2- الجانب النظري

1-2 أنموذج الانحدار اللوجستي والتحويل الخطي (دالة اللوجت)

Logistic Regression Model and Linear transformation (Logit function)

يعرف أنموذج الانحدار اللوجستي بأنه احد أنواع نماذج الانحدار والذي يصف العلاقة بين المتغير التابع (y) متغير الاستجابة وعدد من المتغيرات التوضيحية ($x_1, x_2, x_3, \dots, x_k$) بكونها علاقة غير خطية، الا انه من الممكن كتابة هذا الانموذج بصيغة خطية بدلالة ما يعرف بدالة اللوجت (logit function)، وعليه فإن أنموذج الانحدار اللوجستي هو ببساطة تحويلة لوجارتمية للانحدار الخطي. [15]

في هذا البحث سوف نأخذ المتغير التابع (y) ثنائي الاستجابة اما ان يساوي واحد (Success) بأحتمال (p_i) او يساوي صفر (Failure) بأحتمال ($1 - p_i$) لذلك يكون المتغير (y) يتوزع حسب توزيع برنولي

وبالتالي فإن دالة الكثافة الاحتمالية تكون وفق الصيغة الآتية: [11]



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير معلمات أنموذج الانحدار اللوجستي

$$P_r(Y = 1) = p_i^{y_i} (1 - p_i)^{1-y_i} \quad \dots (1)$$

$$i = 1, 2, \dots, m$$

m تمثل عدد المشاهدات (حجم العينة)
أذ أن :

y_i متغير تابع ثنائي الاستجابة (0,1)

p_i احتمال حدوث الاستجابة عندما $y_i = 1$

$1-p_i$ احتمال عدم حدوث الاستجابة عندما $y_i = 0$

لذلك فإن توقع المتغير y_i يمثل احتمال حدوث الاستجابة (p_i) وكالاتي:

$$E(y_i) = P_r(Y = 1) = p_i \quad \dots (2)$$

أما تبين المتغير y_i بالنسبة لتوزيع برنولي يكون كالاتي:

$$v(y_i) = p_i(1 - p_i) \quad \dots (3)$$

حيث أن

p_i تمثل دالة الانحدار اللوجستي أو تمثل (احتمال الاستجابة) تكتب حسب الصيغة الآتية: [17]

$$p_i = \frac{e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}}} \quad \dots (4)$$

$$1 - p_i = \frac{1}{1 + e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}}} \quad \dots (5)$$

حيث أن

$\beta_0, \beta_1, \beta_2, \dots, \beta_j$ تمثل معلمات مجهولة يراد تقديرها

x_{ij} متغيرات توضيحية

$$j = 1, 2, \dots, k$$

k تمثل عدد المتغيرات التوضيحية

$k + 1$ = تمثل عدد المعلمات

نلاحظ من المعادلة (4) ان شكل العلاقة بين المتغيرات التوضيحية (x_{ij}) واحتمال الاستجابة p_i

لا يمكن ان يكون خطياً وإنما تأخذ شكلاً منحنياً حيث يميل العديد من الاحصائين الى ازالة انحناءات دالة

الانحدار اللوجستي لان وجودها لها تأثير سلبي على خصائص مقدرات المعلمات. [8]

لتحويل العلاقة بين المتغيرات التوضيحية واحتمال حدوث الاستجابة الى علاقة بصيغة خطية يتم ذلك

من خلال ما يعرف بتحويل دالة اللوجت (logit function) والتي تتضمن بإيجاد تحويله لوغارتمية

تسمى $\frac{p_i}{1-p_i}$ نسبة الأفضلية وحسب الصيغة الآتية: [18]

$$\frac{p(y = 1)}{1 - p(y = 1)} = \frac{p_i}{1 - p_i} = \left(e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}} \right) \quad \dots (6)$$



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير معلمات أنموذج الانحدار اللوجستي

$$(p_i) = \ln \frac{p_i}{1-p_i} \quad \dots\dots(7) \text{ Logit}$$

فإن التحويل الخطي لدالة اللوجستي هو:

$$\ln \frac{p_i}{1-p_i} = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} = x'_i \beta = Z \quad \dots\dots (8)$$

حيث أن

$$\beta = \beta_0, \beta_1, \beta_2, \dots, \beta_j$$

$$x'_i = [1, x_{i1}, x_{i2}, \dots, x_{ik}]$$

$$i = 1, 2, \dots, m$$

x'_i يمثل متجه من المتغيرات التوضيحية بدرجة (1*k)

وأن $x_{i0} = 1$ وهو جزء من مصفوفة المتغيرات التوضيحية X

(Z) تمثل العلاقة الخطية الناتجة من أخذ اللوغارتم للنسبة $\frac{p_i}{1-p_i}$

وبالتالي فإن توقع Z_i يكون كالآتي:

$$E(Z_i) = x'_i \beta \quad \dots\dots (9)$$

أما تباين Z_i يكون كالآتي:

$$\text{var}(Z_i) = \frac{1}{n_i p_i (1-p_i)} = [n_i p_i (1-p_i)]^{-1} \quad \dots\dots(10)$$

أذن فإن المقدار (Z_i) الذي هو التحويل الخطي لدالة اللوجستي يتوزع توزيعاً طبيعياً تقاربياً

(Asymptotically Normal) بمتوسط ($x'_i \beta$) وتباين مقداره $[n_i p_i (1-p_i)]^{-1}$. [10]

أي أن :

$$Z_i \sim \text{Asym. N} [x'_i \beta, [n_i p_i (1-p_i)]^{-1}]$$

Methods of Estimation

2-2 طرق التقدير

لتقدير معلمات أنموذج الانحدار اللوجستي سيتم الاعتماد في هذا البحث على بعض طرق التقديرات الحصينة وهي كالآتي:

Laplace Estimates

1-2-2 مقدرات لابلاس (Lp -)

أن الأسلوب المتبع في طريقة (Lp -) للتقدير هو اعتمادها على القيم المطلقة للبواقي، حيث تستخدم هذه الطريقة لتقدير معلمات أنموذج الانحدار اللوجستي وتمتاز تقديرات هذه الطريقة بأنها لا تقتضي الحاجة إلى مقدراً معيارياً $\hat{\sigma}$ (Scale-Estimator) لأنها تمتلك درجة حرية واحدة، وأن الهدف الرئيسي في طريقة (Lp -)

(Lp) هو تصغير مقدار الخطأ العشوائي (r_i) وحسب الصيغة الآتية: [25]

$$Lp = \text{Min} \sum_{i=1}^m |r_i|^u \quad \dots\dots(11)$$

وبالاعتماد على معادلة البواقي

$$r_i = y_i - x'_i \beta \quad \dots\dots(12)$$



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير معلمات أنموذج الانحدار اللوجستي

حيث أن y_i تمثل المشاهدة (i) للمتغير المعتمد (الاستجابة)

$$x_i' = [x_{i1}, x_{i2}, \dots, x_{ik}]$$

$$i = 1, 2, \dots, m$$

x_{i0} يمثل متجه صفي من المتغيرات التوضيحية حيث أن $x_{i0} = 1$ وهو جزء من مصفوفة المتغيرات التوضيحية X

β يمثل متجه المعلمات المراد تقديرها

وأن

$$1 \leq u \leq 2$$

حيث أن

u تمثل القيم المثلى والتي تعطي تقديرات حصينة.

تشكل طريقة (Lp -) عائلة من التقديرات عند كل قيمة من قيم u .

في حالة $u = 1$ فإن المقدرات تسمى (L.A.R) (Least Absolute Residual) ويمثل الخطوة الأولى باتجاه المقدر الأكثر حصانة.

أما في حالة $u = 2$ فإن المقدرات التي تحقق من المعادلة (11) تكون متكافئة لمقدرات المربعات الصغرى. للحصول على مقدرات طريقة (Lp -) والتي تصغر المقدار في المعادلة (11) يتم استخدام طريقة المربعات

الصغرى الموزونة (WLS) (Weighted Least Squares) وحسب الصيغة الآتية: [23]

$$\hat{\beta} = (X'WX)^{-1}X'WZ \quad \dots (13)$$

حيث أن:

W تمثل مصفوفة قطرية عناصر القطر الرئيسي فيها معطاة بالصيغة الآتية:

$$W_i = |r_i|^{u-2} \quad \dots (14)$$

وبالاعتماد على المعادلة رقم (12)

β يمثل متجه المعالم الابتدائية ويقدر من طريقة المربعات الصغرى الاعتيادية (OLS).

وللحصول على المقدرات الحصينة من هذه الطريقة فقد تم استخدام القيمتين ($u_1=1.5, u_2=1.8$) التي تمثل

القيم المثلى لـ (u) والتي تعطي تقديرات حصينة. [22]

Huber Estimates

2-2-2 مقدر Huber الحصين (H)

يعتمد هذا المقدر في عملية حسابه على تقديرات الأوساط الحسابية والتباينات الموزونة، وأن فكرة هذه الطريقة تستند على إعطاء المشاهدات الشاذة أوزان (موازنة المشاهدات) للتقليل من تأثير الشواذ وبالاعتماد على مسافات مهلونوبس (Mahalanobis Distances) حيث يتم الحصول على مقدر طريقة (H-Estimator) باستخدام طريقة المربعات الصغرى الموزونة أو إحدى الطرائق العددية (Numerical

Methods) [19] كما في الصيغة (13) ويمكن توضيح هذه الطريقة وكالاتي:

1- تحسب مسافة مهلونوبس $D(x_i)$ لكل مشاهدة وحسب الصيغة الآتية: [15]

$$D(x_i) = \left[(x_i - T(x))' S^{-1}(x) (x_i - T(x)) \right]^{1/2} \quad \dots (15)$$



مقارنة بين طريقتي مقدرات لابلاس وهوبر الحصين لتقدير معلومات أنموذج الانحدار اللوجستي

$$i = 1, 2, \dots, m$$

كما يمكن أن نحصل على t_i والتي تمثل دالة وفق الصيغة الآتية:

$$t_i = D(x_i) = \left[(x_i - T(x))' S^{-1}(x) (x_i - T(x)) \right]^{1/2} \quad \dots (16)$$

2- حساب كل من $T(x)$ و $S(x)$ وبالصيغ الموضحة وكالاتي: [20]

$$T(x) = \frac{1}{m} \sum_{i=1}^m x_i \quad \dots (17)$$

$$S(x) = \frac{1}{m} \sum_{i=1}^m [(x_i - T(x))(x_i - T(x))'] \quad \dots (18)$$

$T(x)$ تمثل التقديرات الاعتيادية لموجه الوسط الحسابي بدرجة $k*1$

$S(x)$ تمثل التقديرات الاعتيادية لمصفوفة التباين المشتركة بدرجة $k*k$

أن التقديرات في المعادلتين (17) (18) تستخدم في حساب الصيغة في المعادلة (15).

3- إعادة احتساب كل من المتوسطات $T(x)$ ومصفوفة التباينات $S(x)$ حيث تبدل $(T(x), S(x))$ بـ

$(T^*(x), S^*(x))$ بأستخدام اوزان مقدرة وكالاتي: [24]

$$T^*(x) = \frac{\sum_{i=1}^m w_i x_i}{\sum_{i=1}^m w_i} \quad \dots (19)$$

$$S^*(x) = \frac{\sum_{i=1}^m w_i^2 [(x_i - T^*(x))(x_i - T^*(x))']}{\sum_{i=1}^m w_i^2} \quad \dots (20)$$

حيث أن

$T^*(x)$ تمثل متجه موقع حصين بدرجة $(k*1)$

$S^*(x)$ تمثل مصفوفة قياس حصين بدرجة $(k*k)$

w_i تمثل دالة الوزن

لدالة الوزن w_i عدة صيغ حيث تم استخدام احدى هذه الصيغ وهي صيغة دالة هوبر (Huber) [16] حيث

يتميز المقدر الناتج من هذه الصيغة بكونه مقدر كفوء وقليل الحساسية وكما في الصيغة الآتية: [19] [4]

$$w_i = \begin{cases} 2 t_i & \text{if } t_i > 2 \\ 1 & \text{if } t_i \leq 2 \end{cases} \quad \dots (21)$$

وتمثل الخطوات اعلاه التكرار الاول للطريقة، حيث أن استخدام هذا الأسلوب يمكننا من الحصول على مقدرات كفوءة وحصينة، ثم يعاد احتساب الخطوة (3) وبصورة تكرارية معتمدين على نتائج التكرار السابق ويتم التوقف عن العملية التكرارية عندما يكون الفرق بين نتائج عمليتين تكرارين متعاقبين) عندما يصبح الفرق بين التقديرات المتعاقبة (اللاحقة والسابقة)) في تقدير المعلومات قليل ووفق مستوى معين من الدقة.



3- الجانب التجريبي

1-3 المقدمة Introduction

أن استخدام أسلوب المحاكاة هو لتطبيق ماجاء في الجانب النظري من اجل محاكاة اكبر قدر من الحالات التي تواجهنا في الواقع العملي للوصول الى نتائج اكثر دقة وعمومية، إذ أن استخدامه يوفر على الباحثين الكثير من الوقت والجهد والمال كما أن استخدام هذا الاسلوب يؤدي الى تطوير أنموذج النظام من خلال ملاحظة التغييرات التي تطرأ على صياغة المشكلة عند تنفيذها عملياً [4]

ويتمثل أسلوب المحاكاة بكونه وسيلة رياضية لحل الكثير من المعادلات والعلاقات الرياضية في الجوانب النظرية والتي تحتاج الى جهد نظري لاشتقاقها فضلاً عن صعوبة ايجاد حلول لبعض التكاملات والمعادلات التفاضلية المعقدة، غالباً ما يستخدم أسلوب المحاكاة في حالة فشل جميع الطرائق المستخدمة

لايجاد حل لمشكلة ما او في حالة صعوبة الحصول على البيانات اللازمة لدراسة تلك الظاهرة [5] ومما ذكر اعلاه يمكن التوصل الى ان اسلوب المحاكاة هو عملية تمثيل سلوك الظاهرة الحقيقية قيد الدراسة بشكل يكون اقرب للواقع الحقيقي الذي يمثل الانموذج المدروس، وكلما كانت نتائج المحاكاة دقيقة دل ذلك على ان اسلوب المحاكاة يكون اكثر قرباً للواقع المدروس. [2]

2-3 مراحل تطبيق تجارب المحاكاة Stages of the application of simulation experiments

لقد تضمنت تجارب المحاكاة لهذه الدراسة كتابة عدد من البرامج بلغة الـ (Matlab) في توليد البيانات لغرض المقارنة بين الطرائق بأختلاف احجام العينات ونسب التلوث، مع الاخذ بنظر الاعتبار أن الانموذج الذي تم الاعتماد عليه في هذا البحث يكون وفق الصيغة الآتية:

$$Z = Ln \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} \quad \dots (22)$$

1- تعيين القيم الافتراضية للمعلمات، وتعد هذه الخطوة مهمة واساسية تعتمد عليها الخطوات اللاحقة، إذ تم فيها اختيار قيم المعلمات والنموذج المفترض كما مبين في أدناه:

β_0	β_1	β_2	β_3	β_4	β_5
0.6551	-0.0031	0.0021	-0.0902	-0.0006	0.0008

2- ولتنفيذ تجارب المحاكاة جرى استخدام العوامل وكالاتي:

- حجوم العينات، حيث تم استخدام أربع احجام مختلفة للعينات المفترضة وهي (n=15,25,50,100).

- اعتماد نسب مختلفة لتلوث البيانات وهي (20%, 10% = τ).

- يتم تكرار تجربة المحاكاة (1000) مرة.

3- تقدير معلمات أنموذج الانحدار اللوجستي وفق طرائق التقدير التي تم عرضها في الجانب النظري من البحث وهي كالاتي:

1- طريقة مقدرات لابلاس (LP).

2- طريقة مقدرات هوبر الحصين (H).

3- المقارنة بين طرائق التقدير بالاعتماد على المقياس الاحصائي متوسط مربعات الخطأ (MSE) للانموذج لمعرفة افضلية الانموذج تمثيلاً للبيانات وحسب الصيغة الآتية:

$$MSE = \frac{\sum_{i=1}^m (Y_i - \hat{Y}_i)^2}{m - k} \quad \dots (23)$$



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير
معلمات أنموذج الانحدار اللوجستي

3-3 تحليل نتائج المحاكاة

جدول (1-3) تقدير المعلمات بجميع الطرائق، وحجوم العينات عندما تكون نسبة التلوث ($\tau=0\%$)

N	Param. Methods	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
15	Lp(1)	0.5896	-0.0047	0.0030	-0.1330	-0.0009	0.0012
	Lp(2)	0.5896	-0.0048	0.0031	-0.1343	-0.0011	0.0013
	H	0.5615	-0.0038	0.0024	-0.1078	-0.0009	0.0010
25	Lp(1)	0.5679	-0.0045	0.0030	-0.1297	-0.00012	0.0011
	Lp(2)	0.5653	-0.0046	0.0030	-0.1309	-0.0011	0.0012
	H	0.5491	-0.0034	0.0023	-0.1012	-0.0009	0.0010
50	Lp(1)	0.5358	-0.0043	0.0029	-0.1232	-0.0010	0.0011
	Lp(2)	0.5265	-0.0045	0.0030	-0.1235	-0.0010	0.0010
	H	0.5352	-0.0033	0.0023	-0.0948	-0.0008	0.0009
100	Lp(1)	0.5354	-0.0042	0.0027	-0.1221	-0.0009	0.0010
	Lp(2)	0.5239	-0.0043	0.0028	-0.1219	-0.0008	0.0009
	H	0.5382	-0.0033	0.0022	-0.0953	-0.0009	0.0008

أولاً : في حالة عدم وجود تلوث

من خلال النتائج المبينة في الجدول (1-3) نلاحظ ما يلي:

- 1- أظهرت الطرائق أفضليتها من حيث اقتراب القيم التقديرية من القيم الافتراضية.
- 2- لأحجام العينات المفترضة جميعها، أظهرت تقديرات (H) أفضليتها في اقترابها من القيم الافتراضية مقارنة بطريقة (LP-)، أي أن طريقة (H) هي الأفضل.



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير
معلمات أنموذج الانحدار اللوجستي

جدول (2-3) تقدير المعلمات بجميع الطرائق، وحجوم العينات عندما تكون نسبة التلوث $\tau(=10\%)$

N	Param. Methods	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
15	Lp(1)	0.4293	-0.0035	0.0033	-0.1033	-0.0008	0.0011
	Lp(2)	0.4312	-0.0037	0.0030	-0.1046	-0.0007	0.0012
	H	0.4279	-0.0036	0.0031	-0.1017	-0.0007	0.0010
25	Lp(1)	0.3979	-0.0035	0.0027	-0.0878	-0.0007	0.0010
	Lp(2)	0.3973	-0.0033	0.0025	-0.0865	-0.0006	0.0011
	H	0.4061	-0.0032	0.0029	-0.0976	-0.0007	0.0009
50	Lp(1)	0.3910	-0.0031	0.0025	-0.0896	-0.0007	0.0010
	Lp(2)	0.3723	-0.0027	0.0022	-0.0758	-0.0005	0.0009
	H	0.3687	-0.0026	0.0021	-0.0736	-0.0005	0.0008
100	Lp(1)	0.3517	-0.0024	0.0018	-0.0651	-0.0004	0.0007
	Lp(2)	0.3449	-0.0022	0.0016	-0.0611	-0.0004	0.0006
	H	0.3833	-0.0029	0.0027	-0.0861	-0.0006	0.0007

ثانياً: في حالة وجود تلوث $\tau(=10\%)$

من خلال النتائج المبينة في الجدول (2-3) نلاحظ ما يلي:

- 1- أظهرت الطرائق أفضليتها من حيث اقتراب القيم التقديرية من القيم الافتراضية.
- 2- لأحجام العينات المفترضة جميعها، أظهرت تقديرات (H) أفضليتها في اقترابها من القيم الافتراضية مقارنة بطريقة (LP-)، أي أن طريقة (H) هي الأفضل عند هذه النسبة من تلوث .



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير
معلمات أنموذج الانحدار اللوجستي

جدول (3-3) تقدير المعلمات بجميع الطرائق ، وحجوم العينات عندما تكون نسبة التلوث ($\tau = 20\%$)

N	Param. Methods	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
15	Lp(1)	0.3890	-0.0031	0.0027	-0.0862	-0.0008	0.0009
	Lp(2)	0.3831	-0.0030	0.0026	-0.0860	-0.0007	0.0010
	H	0.4205	-0.0029	0.0024	-0.0929	-0.0006	0.0009
25	Lp(1)	0.3660	-0.0024	0.0020	-0.0728	-0.0007	0.0007
	Lp(2)	0.3567	-0.0023	0.0023	-0.0720	-0.0006	0.0009
	H	0.4068	-0.0026	0.0021	-0.0786	-0.0005	0.0008
50	Lp(1)	0.3394	-0.0020	0.0016	-0.0585	-0.0005	0.0006
	Lp(2)	0.3253	-0.0019	0.0015	-0.0557	-0.0004	0.0005
	H	0.3909	-0.0024	0.0020	-0.0721	-0.0005	0.0007
100	Lp(1)	0.3312	-0.0018	0.0013	-0.0522	-0.0003	0.0005
	Lp(2)	0.3157	-0.0017	0.0012	-0.0486	-0.0003	0.0004
	H	0.3904	-0.0023	0.0019	-0.0694	-0.0005	0.0007

ثالثاً: في حالة وجود تلوث ($\tau = 20\%$)

من خلال النتائج المبينة في الجدول (3-3) نلاحظ ما يلي:

- 1- أظهرت الطرائق أفضليتها من حيث اقتراب القيم التقديرية من القيم الافتراضية.
- 2- لأحجام العينات المفترضة جميعها، أظهرت تقديرات (H) أفضليتها في اقترابها من القيم الافتراضية مقارنة بطريقة (LP-)، أي أن طريقة (H) هي الأفضل عند هذه النسبة من تلوث .

جدول (4-3) الخاص بقيم (MSE) للأنموذج بجميع الطرائق وحجوم العينات وعند نسب معينة من التلوث

نسب التلوث	n	MSE Lp(1)	MSE Lp(2)	MSE H	Best
0%	15	0.6874	0.6826	0.6927	Lp(2)
	25	0.5834	0.5858	0.5812	H
	50	0.5149	0.5220	0.5179	Lp(1)
	100	0.4926	0.4923	0.4896	H
10%	15	0.6885	0.6995	0.6856	H
	25	0.5773	0.5836	0.5851	Lp(1)
	50	0.5290	0.5217	0.5181	H
	100	0.5028	0.5037	0.4915	H
20%	15	0.6940	0.6998	0.6960	Lp(1)
	25	0.5844	0.5825	0.5760	H
	50	0.5240	0.5299	0.5169	H
	100	0.5031	0.5028	0.4956	H



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير معلمات أنموذج الانحدار اللوجستي

من الجدول (3-4) نلاحظ مايلي:

أولاً: في حالة عدم وجود تلوث ($\tau=0\%$)

أظهرت النتائج ومن خلال المقياس (MSE) للانموذج بأن طريقة (Lp(2)) هي افضل طريقة عند حجم العينة (n=15) أي في حالة حجوم العينات الصغيرة وذلك لانها حققت اصغر (MSE) للانموذج، بينما كانت طريقة (Lp(1)) هي الافضل عند حجم العينة (n=50) أي في حالة حجوم العينات المتوسطة، كما أظهرت طريقة (H) كفاءتها عند احجام العينات (n=25, n=100).

ثانياً: في حالة وجود تلوث $\tau(=10\%)$

أظهرت النتائج ومن خلال المقياس (MSE) للانموذج بأن طريقة (H) هي افضل طريقة عند حجم العينة (n=15, n=50, n=100) وذلك لانها حققت اصغر (MSE) للانموذج، بينما كانت طريقة (Lp(1)) هي الافضل عند حجم العينة (n=25) أي في حالة حجوم العينات الصغيرة.

ثالثاً: في حالة وجود تلوث $\tau(=20\%)$

أظهرت النتائج ومن خلال المقياس (MSE) للانموذج بأن طريقة (Lp(1)) هي افضل طريقة عند حجم العينة (n=15) أي في حالة حجوم العينات الصغيرة، كما أظهرت طريقة (H) كفاءتها في حالة حجوم العينات الصغيرة والمتوسطة والكبيرة وذلك لانها حققت اصغر (MSE) للانموذج.

4- الاستنتاجات والتوصيات

1-4 الاستنتاجات

بعد تنفيذ تجارب المحاكاة وما تم عرضه من نتائج وتحليل في الجانب التجريبي استنتج الباحث مايلي:

- 1- اثبتت طريقة (H) كفاءتها في تقدير معلمات أنموذج الانحدار اللوجستي في حالة تلوث البيانات كما اثبتت بأنها الافضل كلما زادت نسبة التلوث.
- 2- اثبتت طريقة (H) بانها الافضل في التقدير في حالة حجوم العينات الكبيرة ونسب التلوث العالية.
- 3- اشارت نتائج المحاكاة بأن طريقة (Lp(2)) تكون افضل بالنسبة لحجوم العينات الصغيرة في حالة عدم وجود تلوث في البيانات.
- 4- اثبتت طريقة (Lp(1)) كفاءتها في تقدير المعلمات في حالة وجود نسبة من التلوث ولحجوم العينات الصغيرة.
- 5- تتناقص قيمة MSE عند زيادة حجوم العينات ونسب التلوث.

2-4 التوصيات

على ضوء الاستنتاجات التي تم التوصل اليها من خلال البحث يمكن ان ندرج التوصيات وكالاتي:

- 1- استعمال طريقة (H) في تقدير معلمات أنموذج الانحدار اللوجستي وبأختلاف احجام العينات لما تبيده من كفاءة ومرونة في التطبيق.
- 2- اعتماد طريقة (H) في تقدير معلمات أنموذج الانحدار اللوجستي لانها اعطت اقل (MSE) للانموذج في حالة حجوم العينات الكبيرة ونسب التلوث العالية.
- 3- استعمال طريقة (Lp(1)) في تقدير معلمات أنموذج الانحدار اللوجستي في حالة حجوم العينات الصغيرة وعند نسبة معينة من التلوث.
- 4- اعتماد طريقة (Lp(2)) في حالة عدم وجود تلوث في البيانات بالنسبة لحجوم العينات الصغيرة.



مقارنة بين طريقتي مقدرات لابلاس وهوير الحصين لتقدير معلمات أنموذج الانحدار اللوجستي

5-المصادر

- 1-الجشعبي حسين علي عبد الله (2007) م. "مقارنة لبعض المقدرات الحصينة لمعالم النماذج اللاخطية" اطروحة دكتوراه في الاحصاء كلية الادارة والاقتصاد الجامعة المستنصرية.
- 2-العاني اميمة دهام شعبان (1999) م. "دراسة لصيغ التحيز النظري لتقديرات معلمات نموذج اللوجستك" رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد جامعة بغداد.
- 3-العزاوي احمد نياي (2005) م. "المقارنة بين بعض طرائق تقدير انموذج الانحدار اللوجستي والطرائق الحصينة للتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة" رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد جامعة بغداد.
- 4-النداوي سري صباح كيتب (2008) م. "مقارنة بعض المقدرات الحصينة في الدوال التمييزية مع تطبيق عملي" رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد جامعة بغداد.
- 5-حسين شرين علي (2009) م. "مقدرات الامكان الاعظم الموزونة الحصينة ومقارنتها مع طرائق اخرى لانموذج اللوجستك مع تطبيق عملي" رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد جامعة بغداد.
- 6-يلدا نازك هرمز (1995) م. "مقارنة طرائق التقدير الحصينة لبعض تصاميم التجارب" رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد جامعة بغداد.

7-Ben-Gal,I.(2005),"Outlier Detection", Department of Industrial Engineering, Tel-Aviv University, Kluwer Academic Publishers.

8-Berkson,J.(1944),"Application of the Logistic Function to Bio- Assay", JASA Vol. 39,No. 227,pp. 357-465.

9-Box,G.E.P.(1953),"Non-Normality and Tests on Variances", Biomet. 40, pp. 318-335.

10-Chatterjee,S.,Hadi,A.S.(2012),"Regression Analysis By Example",Fifth Edition John Wiley & Sons, Inc ,Hoboken, New Jersey.

11_Cook,D.,Dixon,P.,Duckworth,W.M.,Kaiser,M.S.,Koehler,K.,Meeker,W .Qand Stephenson ,W.R. (2001),"Binary Response and Logistic Regression Analysis", University NSF/ILI Project Beyond Traditional Statistical Methods, grant.

12-Hadi,A.S.and Chatterjee ,S.(1988),"Sensitivity Analysis in Linear Regression", Joun Wiley and Sons ,New York.

13-Hansen,L.P.and Sargent,T.J.(2010),"The Changing History of Robustness", The American Statistician Vol.64,No. 4,pp. 277-281.

14-Hosmer,D.W.,Lemeshow,S.and Sturdivant ,R.X.(2013),"Applied Logistic Regression" University of Massachusetts ,Third Edition.

15-Hubert,M.andDebruyne ,M.(2010)."Minimum Covariance Determinant", Wiley Interdiscipl Rev :Comput Stat, Vol,2,pp.36-43.

16-Huber,P.J.(1981),"Robust Statistics", John Wiley & Sons, New York.

17-Kleinbaum,D.G.and Klein ,M.(2002),"Logistic Regression", book ,Springer – Verlag New York ,Second Edition.



- 18-Magnac,T.(2005),"Logit Models of individual Choices",Universite de Toulouse,Prepared for the New Palgrave,First Version.
- 19-Maronna,R.A.Martin,R.Dand Yohai,V.J.(2006)."Robust Statistics", Theory and Method, Jone Wiley & Sons, Ltd,The Atrium, Southern Gate, Chichester,West Sussex, England.
- 20-Muller,ch.H and Neykov,N.(2003)."Breakdown Points of Trimmed Likelihood Estimators and Related Estimators in Generalized Linear Models",J.Statistic.Planning Inference 116,pp. 503-519.
- 21-Pearl,R.and Reed,L.J.(1920),"On the Rate of Growth of the Population of the United States Since 1790 and its Mathematical Representation",National Academy of Sciences,Vol. 6,No. 6,pp. 275-288.
- 22-Ramsy,J.O.(1977)."A Comparative Study of Several Robust Estimates of Slope,Intercept and Scale in Linear Regression", JASA Vol. 72,No. 359,pp. 608-615.
- 23-Rousseeuw,P.J.(1984)."Least Median of Squares Regression",Journal of the American Statistical Association,Vol. 79,No. 388,pp. 871-880.
- 24-Rousseeuw ,P.J.and Huber,M.(2011)."Robust Statistics for Outlier detection" John Wiley &Sons,Ins,Vol, 1,pp.73-79.
- 25-Rousseeuw,P.J.and Leroy,A.(1987)."Robust regression and Outlier Detection",John Wiley and Sons .
- 26-Wilcox,R.R.(2005)."Introduction to Robust Estimation and Hypothesis Testing",Second Edition ,Academic Press,New York.



A Comparison between Methods of Laplace Estimators and the Robust Huber for Estimate parameters logistic regression model

Abstract

The logistic regression model regarded as the important regression Models ,where of the most interesting subjects in recent studies due to taking character more advanced in the process of statistical analysis .

The ordinary estimating methods is failed in dealing with data that consist of the presence of outlier values and hence on the absence of such that have undesirable effect on the result.

We will review in this research to estimate parameters of logistic regression model these methods are Laplace Estimators (LP-) and Huber estimator (H).

Was conducted to compare between two methods through the simulation and using comparison criteria mean square error (MSE) for proportion different of contamination and sample sizes for determinant to reach the best method to estimate the parameter.

It was found that method (H) is better in estimate parameters of logistic regression model.

Keywords: Logistic Regression, Binary data, the Laplace estimators(LP-), the Robust Huber Estimators(H).