استعمال طريقتي المركبات الرئيسية والمربعات الصغرى الجزئية لتقدير معلمات أنموذج الانحدار اللوجستي ثنائي الاستجابة في حالة وجود مشكلة التعدد الخطي

أ.د.محمود مهدي البياتي/كلية الادارة والاقتصاد/جامعة بغداد
الباحث/ هديل حميد شاكر/كلية الادارة والاقتصاد/جامعة بغداد

تاريخ التقديم:2018/7/8
تاريخ القبول:2018/7/25

المستخلص

يعد أنموذج الانحدار اللوجستي من النماذج اللاخطية الذي يهدف الى الحصول على مقدرات تمتلك كفاءة عالية ، كما انه يعطي الباحث فكره عن مقدار تأثير المتغير التوضيحي على متغير الاستجابة الثنائية.
أن العدد الكبير لمتغيرات توضيحية تستعمل عادة لتوضيح الاستجابة ادى الى ظهور مشكلة التعدد الخطي بين المتغيرات التوضيحية التي تجعل تقدير معلمات النموذج ليست دقيقة جدا.
يتم عرض في هذا البحث طريقتين لتقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي (Multicollinearity) وهما : طريقة المركبات الرئيسية للانحدار اللوجستي (PCLR)، وطريقة انحدار المربعات الصغرى الجزئية(PLSR).
اذ تم اجراء المقارنة بين هاتين الطريقتين من خلال اسلوب المحاكاة وبأستعمال معيار المقارنة متوسط مربعات الخطأ(MSE) للوصول الى الطريقة الأفضل في تقدير المعلمات في حالة وجود مشكلة التعدد الخطي، وقد بينت نتائج المحاكاة أن طريقة ((PCLR(3pc's) هي الافضل في تقدير معلمات أنموذج الانحدار اللوجستي ثنائي الاستجابة في حالة وجود مشكلة التعدد الخطي.

المصطلحات الرئيسية للبحث/ الانحدار اللوجستي، البيانات الثنائية، المركبات الرئيسية، المربعات الصغرى الجزئية، مشكلة التعدد الخطي

*بحث مستل من رسالة ماجستير

مجلة العلوم الاقتصادية والإدارية
العدد 106 المجلد 24
الصفحات 338ـ 355

1- المقدمة: Introduction

يعرف الانحدار بشكل عام بأنه أحد الأساليب الاحصائية المهمة التي تستخدم بشكل واسع جدا ومنذ فترات طويلة لتحديد وتوضيح التأثيرات بين المتغيرات التوضيحية (التفسيرية) والمتغير التابع (متغير الاستجابة) ويستخدم ايضا للتنبؤ عن قيمة المتغير التابع بدلالة المتغيرات التوضيحية بعد أيجاد معادلة الانحدار الخطية [1].

يعد أنموذج الانحدار اللوجستي الاسلوب الاحصائي المستخدم لتوفيق العلاقة بين المتغير التابع ثنائي القيمة وعدة متغيرات توضيحية أيا كان نوعها، ويسمى الأنموذج في هذه الحالة أنموذج الانحدار اللوجستي الثنائي، وتعد نماذج الانحدار اللوجستي حالة خاصة من حالات نماذج الانحدار العامة (Generalized linear models)، اذ تستخدم عندما نرغب في التنبؤ بوجود صفة معينة أو خاصية أو ظاهرة معينة بالاعتماد على قيم متغير أو مجموعة من المتغيرات التوضيحية الاخرى التي لها علاقة بالمتغير التابع تماما كما هو الحال في نماذج الانحدار العامة، ويستخدم أنموذج الانحدار اللوجستي في كثير من مجالات الحياة، مثلا في مجال الطب والجيولوجيا، والبايولوجيا والزراعة.

تواجه الباحث مشكلات عديدة أغلبها عدم توفر افتراضات التحليل عند استخدام طريقة المربعات الصغرى الاعتيادية ومنها مشكلة التعدد الخطي التي تؤثر على نتائج التقديرات وتظهر هذه المشكلة نتيجة وجود ارتباط بين اثنين أو أكثر من هذه المتغيرات، أو بين جميع المتغيرات التي تؤدي الى أعطاء تقديرات ضعيفة لا يمكن أن يعول عليها، في هذا البحث تم أستخدام طريقتي المركبات الرئيسية والمربعات الصغرى الجزئية لمعالجة هذه المشكلة وللحصول على مقدرات دقيقة في الأنموذج اللوجستي.

2- الأستعراض المرجعي

1- في عام (1990م) قدم الباحث (Marx)[14] مقدرات معلمات المركبات الرئيسية ذات الخطوة الواحدة والتكرارية في حالة وجود مشكلة التعدد الخطي كبديل الى مقدرات الامكان الاعظم التكرارية في حالة أنموذج الانحدار اللوجستي وتبين أن مقدرات المركبات الرئيسية أفضل من مقدرات الامكان الاعظم التكرارية في معالجة مشكلة التعدد الخطي.

2- في عام (2005م) قام (Anne and Korbinian)[10] بأستخدام طريقة المربعات الصغرى الجزئية وهي أسلوب احصائي مهم في تحليل البيانات للأبعاد العليا لبيانات الالاف من الجينات وكذلك لتقليص الابعاد حيث أن المرحلة الأولى، أستخدم التصنيف بالطريقة الكلاسيكية(الانحدار اللوجستي) بأستخدام المركبات الرئيسية للمربعات الصغرى الجزئية.

3- في عام (2013م) درس الباحثان (Muhammed and Tuti)[15] طريقة الانحدار اللوجستي الترتيبية الصغرى (SOLR) وطريقة المركبات الرئيسية حيث تم أستخدام هذه الدراسة على بيانات حقيقية المتمثلة في حالات الفقر في جزيرة جافا حيث تم التوصل الى أن أهم العوامل المؤثرة على حالات الفقر هي (عدد العمال، الكهرباء)، حيث كانت هذه العوامل هي الأكثر العوامل السلبية.

4- في عام (2014م) درس الباحث (Zhenjiell)[18] أنموذج الانحدار اللوجستي وطريقة المركبات الرئيسية حيث تم اختيار المتغيرات من البيانات المالية لعام 2012 وتم سحب عينة تحتوي على (56) شركة حيث توصل الباحث الى أن طريقة المركبات الرئيسية تعطي نتائج أكثر دقة.

3- مشكلة البحث Problem of Research

في حالة كون متغير الاستجابة من النوع الثنائي أي ثنائي الاستجابة (0,1) مع متغيرات توضيحية عديدة 7 فأكثر فأنه لايمكن اعتماد النموذج اللوجستي بصيغته العادية وذلك بسبب ظهور مشاكل في المتغيرات التوضيحية ، ومن هذه المشاكل هي مشكلة التعدد الخطي (Multicollinearity) لذلك تم اللجوء الى استعمال طرائق تلائم واقع هذه البيانات.

4- **هدف البحث** Object of research

يهدف هذا البحث الى معالجة مشكلة التعدد الخطي التي من الممكن أن تظهر بين المتغيرات التوضيحية من خلال أستخدام طريقة المركبات الرئيسية للانحدار اللوجستي **(Principal Components Logistic Regression: PCLR)** وطريقة انحدار المربعات الصغرى الجزئية **(Partial least square regression: PLSR)** في الأنموذج اللوجستي للحصول على مقدرات دقيقه وبأستخدام اسلوب المحاكاة تم المقارنه بين طريقتي التقدير من خلال معيار متوسط مربعات الخطأ**(MSE)**.

5- **الجانب النظري**:

1-5 **أنموذج الانحدار اللوجستي**: Logistic regression model

يبنى أنموذج الانحدار اللوجستي على افتراض أساسي هو أن المتغير التابع **(Y)** ثنائي الاستجابة يأخذ أحدى القيمتين **(0,1)** أما النجاح **(Success)** بأحتمال (π_i) أو الفشل **(Failure)** بأحتمال $(1-\pi_i)$ لذلك يكون المتغير (y_i) يتوزع حسب توزيع برنولي **Ber** (π_i) [13] .
أي أن

$$y_i \sim Ber(\pi_i)$$
i=1,2,… ,n

ومن ثم فأن دالة الكثافة الاحتمالية تكون وفق الصيغة الاتية :

$$\pi_i^{y_i}(1-\pi_i)^{1-y_i} \qquad \text{.....(1)}$$

y_i**=0,1**

أذ أن :
y_i متغير تابع ثنائي الاستجابة **(0,1)**
π_i أحتمال حدوث الاستجابة عندما $y_{i=1}$
$1-\pi_i$ أحتمال عدم حدوث الاستجابة عندما $y_{i=0}$
لذلك فأن توقع المتغير y_i يمثل أحتمال حدوث الاستجابة (π_i) كالاتي:

$$E(y_i) = pr(y=1) = \pi_i$$

أما تباين المتغير y_i بالنسبة لتوزيع برنولي كالاتي:

$$V(y_i) = \pi_i(1-\pi_i)$$

ليكن $X_1, X_2, \dots, X_p$ مجموعة من المتغيرات التوضيحية ولتكن **n** تمثل عدد المشاهدات لهذه المتغيرات التي تكون المصفوفة **X** [9] .

$$X = (X_{ij})_{n\times p} \qquad \text{.....(2)}$$

أذ أن:
i=1,2,…,n ، **(n)** تمثل حجم العينة.
j=1,2,…,p ، **(p)** تمثل عدد المتغيرات التوضيحية.
فأذا كان $y_i = [y_1, y_2, \dots, y_n]$ عينة عشوائية من المتغير ثنائي الاستجابة وأن $y_i \in \{0,1\}$ وبالتالي فأن أنموذج الانحدار اللوجستي يكتب بالصيغة الاتية:

$$y_i = \pi_i + \varepsilon_i \qquad \text{.....(3)}$$

اذ أن π_i تمثل دالة الانحدار اللوجستي (احتمال الاستجابة)

$$\pi i = \frac{\exp\{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}\}}{1+\exp\{\beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}\}} \qquad \text{.....(4)}$$

أو يمكن كتابته بالصيغة الاتية:

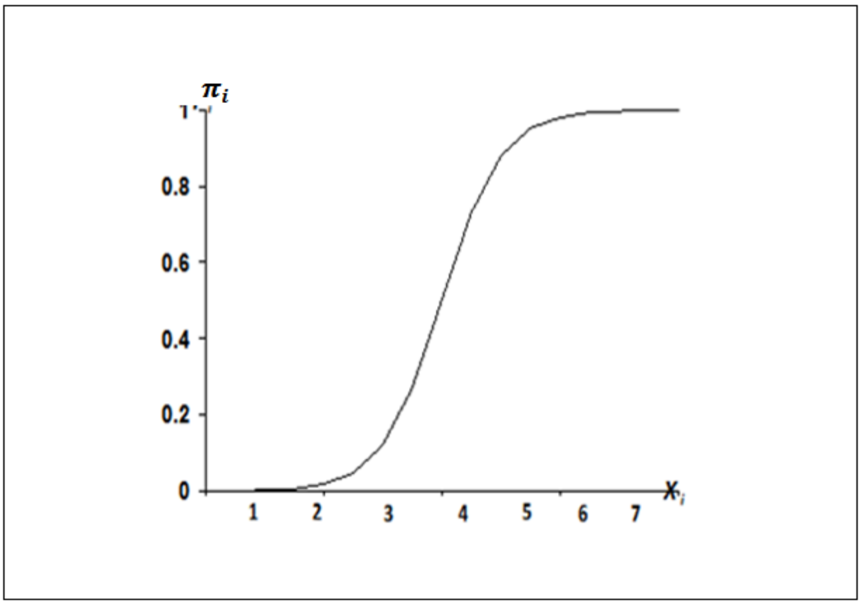
$$\pi i = \frac{\exp\{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}\}}{1+\exp\{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}\}} \qquad \ldots..(5)$$

$$(1-\pi i) = \frac{1}{1+\exp\{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}\}} \qquad \ldots..(6)$$

حيث $\beta_0, \beta_1, \ldots, \beta_p$ هي معلمات النموذج، وأن ε_i يمثل الخطأ العشوائي بمتوسط صفر و تباين $\pi_i(1-\pi_i)$

يتضح من المعادلة (5) أن شكل العلاقة بين المتغيرات التوضيحية (X_{ij}) واحتمال الاستجابة π_i لايمكن أن يكون خطيا وهي تأخذ شكلا منحنيا والمخطط في شكل رقم (1) يوضح ذلك [4]

شكل رقم(1): العلاقة بين احتمال الاستجابة π_i والمتغير التوضيحي X_i



لقد اقترح (Berkson) عام 1944م بأنه يمكن تحويل دالة الانحدار اللوجستي الى دالة خطية وحسب الصيغة الاتية [11]

$$\frac{\pi_i}{(1-\pi_i)} = \exp\{\beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}\} \qquad \ldots.(7)$$

وبأخذ اللوغاريتم الطبيعي لكلا الطرفين نحصل على:

$$Z_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} \qquad \ldots(8)$$

أذ أن

(Z_i) تمثل العلاقة الخطية الناتجة من اخذ اللوغاريتم الطبيعي ل $\left(\frac{\pi_i}{1-\pi_i}\right)$ والذي يتبع التوزيع الطبيعي بمتوسط $X_i\beta$) وتباين $([n_i \pi_i(1-\pi_i)]^{-1}$ اي أن [11]:

$$Z_i \sim N\left((X_i\beta), [n_i \pi_i (1-\pi_i)]^{-1}\right) \qquad \ldots..(9)$$

2-5 طريقة المركبات الرئيسية للانحدار اللوجستي:

Principal components Logistic Regression Method

أن تحليل المركبات الرئيسية هو من اساليب متعدد المتغيرات ، والذي عرفه الباحث (Karl Pearson) سنة (1901) التي تتضمن هذه الطريقة للتخلص من مشكلة التعدد الخطي التام بين المتغيرات التوضيحية ، والذي تم تحسينه من قبل (Harold Hoteling) عام (1933)[9] ، تقوم طريقة المركبات الرئيسية على تحويل المتغيرات التوضيحية الاصلية المرتبطة دون حذف أي منها الى متغيرات جديدة متعامدة تسمى بالمركبات الرئيسية (Principal Components) ، وكل مركبة رئيسية عبارة عن تركيب خطي في المتغيرات المستقلة الاصلية[17] ، يتم ترتيب المركبات الرئيسية وفقا لحجم التباين الذي تستطيع كل مركبة تفسيره بواسطة المتغيرات التي تتضمنها ، فالمركبة الاولى هي المركبة ذات التباين الأكبر ويليه المركبة الثانية وهكذا.

نفترض انه لدينا (P) من المتغيرات المستمرة ، وكل متغير يحتوي على (n) من المشاهدات ، معطاة في المصفوفة $X = (X_{ij})_{n*p}$ ، أن متجهات العمود لمثل هكذا مصفوفة مشار اليها من خلال $(X_1 , X_2 , \dots , X_p)$، أن مصفوفة التباينات والتباينات المشتركة ($S = (S_{jk})_{P\times P}$) والتي عناصرها معرفة من خلال: $S_{jk} = \frac{1}{n-1}\ (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$ يمكن أن تعطى بالصيغة[9]:

$$S = \frac{1}{n-1}\ \acute{X}X \qquad \dots (10)$$

ويمكن كتابة المصفوفة S بالصيغة:

$$S = V\ \Delta\ \acute{V}$$

أذ أن

$\Delta = diag(\theta_1 , \dots , \theta_k)$: تمثل مصفوفة الجذور المميزة للمصفوفة S

V: تمثل مصفوفة متعامدة من المتجهات المميزة المناظرة للجذور المميزة للمصفوفة S

نفترض أن (Z) مصفوفة أعمدتها تمثل المركبات الرئيسية (Principal Components) وتعطى بالصيغة:

$$Z = X\ V \qquad \dots (11)$$

اذ ان $(V_1 , V_2 , \dots , V_p)$ هي المتجهات المميزة للمصفوفة S المناظرة للجذور المميزة $(\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p)$ على الترتيب ، ويمكن التعبير عن مصفوفة المشاهدات (X) بالصيغة:

$$X = ZV' \qquad \dots\dots(12)$$

ويمكن كتابة المصفوفة X على النحو الاتي:

$$X_j = \sum_{k=1}^{s} Z_k\ V_{jk} \quad , \quad j = 1,2, \dots , p \qquad \dots.(13)$$

التي تمثل نسبة عالية من التباين الكلي والذي تعطى بالعلاقة الاتية:

$$\left[\frac{\sum_{j=1}^{s} \theta_j}{\sum_{j=1}^{P} \theta_j} \times 100\right] \quad , \quad S \leq P \qquad \dots.(14)$$

كما في الخطوة السابقة لتحديد أنموذج انحدار المركبات الرئيسية اللوجستي (PCLR) ، سنقوم بصياغة الأنموذج اللوجستي لجميع المركبات الرئيسية المناظرة لمصفوفة مشاهدات المتغيرات التوضيحية المستقلة X ، فأننا نفترض بدون فقدان للعمومية ، وأن احتمالية النجاح للنموذج اللوجستي المعطى في المعادلة (4) يمكن التعبير عنه لجميع المركبات الرئيسية [9]

$$\pi_i = \frac{exp\left[\beta_0 + \sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik}\ V_{jk}\ \beta_j\right]}{1 + exp\left[\beta_0 + \sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik}\ V_{jk}\ \beta_j\right]}$$

$$= \frac{exp\left[\beta_0 + \sum_{k=1}^{p} z_{ik}\gamma_k\right]}{1 + exp\left[\beta_0 + \sum_{k=1}^{p} z_{ik}\gamma_k\right]} \qquad \dots. (15)$$

حيث

$Z = X\,V$ هي عناصر مصفوفة المركبات الرئيسية , (i=1,…,n ; k=1,…,p)Z_{ik}
و K=1, … , p , $y_k = \sum_{j=1}^{p} V_{jk}\beta_j$. والنموذج اللوجستي يمكن التعبير عنه بصيغة المصفوفات
من حيث التحويلات لنموذج المركبات الرئيسية اللوجستي وكمايلي:

$$L = X\,\beta = Z\,V\grave{B} \;= Zy \qquad ....(16)$$

L : تمثل المصفوفات

$$Z = (1|Z)\,, \begin{pmatrix} 1 & 0' \\ 0 & V \end{pmatrix}\,, 0 = (0\,,\dots,0)'\;,\; 1 = (1\,,\dots,1)'$$

لذلك يمكن الحصول على المعلمات في الأنموذج اللوجستي من حيث تلك التي نحصل عليها من الأنموذج الذي يحتوي على جميع المركبات الرئيسية على النحو الاتي:

$$\beta = \hat{V}\,\grave{y} \qquad .....(17)$$

ونتيجة لخاصية الثبات لتقديرات الامكان الاعظم (MLE):

$$\hat{\beta} = V\,\hat{y} \qquad .....(18)$$

والمعادلة التخمينية هي:

$$\hat{y} = \hat{\pi} \qquad ....(19)$$

ولكي يتم تحسين تقدير المعلمات الاصلية في حالة التعدد الخطي ، سنقدم أنموذج انحدار المركبات الرئيسية اللوجستي (PCLR) الذي يتم الحصول عليه عن طريق اتخاذ متغيرات الانموذج اللوجستي من مجموعة مخفضه من المركبات الرئيسية من المتغيرات الاصلية.

سنقوم بتجزئة المصفوفات Z و V كالاتي:

$$Z = \left[\begin{array}{cccc|ccc} 1 & z_{11} & \cdots & z_{1s} & z_{1(s+1)} & \cdots & z_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & z_{n1} & \cdots & z_{ns} & z_{n(s+1)} & \cdots & z_{np} \end{array}\right] = \left(Z_{(s)} \mid Z_{(r)}\right)$$

$$V = \left[\begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & v_{11} & \cdots & v_{1s} & v_1(s+1) & \cdots & v_{1p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & v_{1p} & \cdots & v_{ps} & v_p(s+1) & \cdots & v_{pp} \end{array}\right] = \left(V_{(s)} \middle| V_{(r)}\right)$$

حيث $(r = p - s)$

فأن

$$Z_{(s)} = X\, V_{(s)}$$

$$Z_{(r)} = X\, V_{(r)} \qquad ..... (20)$$

اذ أن:

Z : مصفوفة ذات رتبة $(n \times p+1)$ وهي مصفوفة أعمدتها تمثل المركبات الرئيسية.

V : مصفوفة ذات رتبة $(p+1)(p+1)$ وهي مصفوفة عناصرها تمثل معاملات المركبات الرئيسية.

لذا فأن المعلمات الاصلية يمكن التعبير عنها كالاتي:

$$\beta = V\, y = V_{(s)}\, y_{(s)} + V_{(r)}\, y_{(r)} \qquad ..... (21)$$

$$y = \left(y_0\ y_1 \cdots\ y_s | y_{s+1} \cdots\ y_p\right)' = \left(y'_{(s)} | y'_{(r)}\right)'$$

مع الاخذ في الحسبان انه نموذج اللوجستي من ناحية جميع المركبات الرئيسية والذي يعطى بالمعادلة (16) ، ويمكن أن يتحلل ويكتب بالصيغة الاتية [9]:

$$L = Zy = Z_{(s)}\, y_{(s)} + Z_{(r)}\, y_{(r)} \qquad .....(22)$$

أن أنموذج انحدار المركبات الرئيسية اللوجستي (PCLR) من ناحية (S) من المركبات الرئيسية (PCLR(S)) يتم الحصول عليه عن طريق حذف (r) من المركبات الرئيسية من المعادلة (22) لذا سيكون لدينا :

$$yi = \pi_{i(s)} + \varepsilon_{i(s)}$$

$$\pi_{i(s)} = \frac{exp\left[y_0 + \Sigma_{j=1}^{s} z_{ij}\, y_j\right]}{1 + exp\left[y_0 + \Sigma_{j=1}^{s} z_{ij}\, y_j\right]} \quad , \quad i = 1, \dots, n \qquad ..... (23)$$

هذا الأنموذج ممكن أن يصاغ بشكل مكافئ بصيغة المصفوفات من خلال تحويلات متجه اللوجستي كالاتي:

$$L_s = Z_{(s)}\, y_{(s)} = X\, V_{(s)}\, y_{(s)} = X\beta_{(s)} \qquad ....(24)$$

لذا لقد حصلنا على اعادة بناء المعلمات الاصلية والتي تعطى بالصيغة:

$$\beta_{(s)} = V_{(s)}\, y_{(s)} \qquad .... (25)$$

اي أن معلمات أنموذج انحدار المركبات الرئيسية اللوجستي (PCLR) تعتمد على متغيرات أو (S) من المركبات الرئيسية. أن تقديرات الامكان الاعظم (MLE) لنموذج (PCLR) سوف يتم تقدير المعلمات الاصلية لها:

$$\hat{\beta}_{(s)} = V_{(s)}\, \hat{y}_{(s)} \qquad ..... (26)$$

والتي من شأنها تحسين التقدير $\hat{\beta}$ والتي تم الحصول عليه من المتغيرات الاصلية في حالة التعدد الخطي اي ان:

$$\hat{y}_{(s)} = \left[\hat{y}_{0(s)}, \hat{y}_{1(s)}, \dots, \hat{y}_{s(s)}\right]' \neq \left[\hat{y}_0, \hat{y}_1, \dots, \hat{y}_s\right]' \qquad ..... (27)$$

ونتيجة لذلك فأن الاحتمالات $\hat{\pi}_{i(s)}$ المقدرة بنموذج انحدار المركبات الرئيسية اللوجستي (PCLR(s)) تختلف عن تلك التي يتم الحصول عليها عن طريق أقتطاع الاحتمالات المقدرة بطريقة الامكان الاعظم للنموذج الذي استخدم جميع المركبات الرئيسية اي انه:

$$\hat{\pi}_{i(s)} = \frac{exp\left[\hat{y}_{0(s)} + \Sigma_{j=1}^{s} z_{ij}\, \hat{y}_{j(s)}\right]}{1 + exp\left[\hat{y}_{0(s)} + \Sigma_{j=1}^{s} z_{ij}\, \hat{y}_{j(s)}\right]} \neq \frac{exp\left[\hat{y}_0 + \Sigma_{j=1}^{s} z_{ij}\, \hat{y}_j\right]}{1 + exp\left[\hat{y}_0 + \Sigma_{j=1}^{s} z_{ij}\, \hat{y}_j\right]} \qquad ....(28)$$

وهذا يعني زيادة كبيرة في الجهد الحسابي لان نموذج انحدار المركبات الرئيسية اللوجستي (PCLR) يجب ان يتم تعديله في كل مره ندخل أو نزيل مركبا رئيسيا في الأنموذج وفي حالة كوننا استحدمنا اول S من المركبات الرئيسية (اكثرها تفسيرية) لصياغ نموذج انحدار المركبات الرئيسية اللوجستي (PCLR) ، مع ذلك في انحدار المركبات الرئيسية (PCR) ليست بالضرورة أن تكون المركبات الرئيسية ذات التباينات الاكبر هي افضل المتغيرات التخمينية (Predictors) لان مركبات رئيسية طفيفه بتباينات صغيرة يمكن أن ترتبط ارتباطا كبيرا مع متغير الاستجابة بحيث يجب عدها من المتغيرات التوضيحية اي الانموذج الامثل . أي أن هذه المركبات الرئيسية يمكن ادراجها في الأنموذج وفقا لقدراتها التنبوئية.

ولكي يتم الحصول على اعادة بناء مثالية للمعالم الاصلية مع عدد قليل من المركبات الرئيسية ، سوف يتم الاعتماد على طريقة اختيار المركبات الرئيسية في نموذج انحدار المركبات الرئيسية اللوجستي (PCLR) ولا تأخذ في الحسبان متغير الاستجابة ونضم المركبات الرئيسية بالاعتماد على تفسيرها للتباين[9].

3-5 **طريقة انحدار المربعات الصغرى الجزئية**: Partial Least Square Regression (PLSR)

أن طريقة انحدار المربعات الصغرى الجزئية (Partial least square) والتي يرمز لها بالرمز PLS هي تقنية عممت خصائصها من خلال تحليل المركبات الرئيسية (Principal Components Analysis) والتي يرمز لها بالرمز PCA والانحدار المتعدد (Multiple Regression) هذا اذا كان لدينا متغير الاستجابة واحد ، كذلك اذا كان لدينا عدة متغيرات استجابة فأن طريقة انحدار المربعات الصغرى الجزئية مفيدة عندما نحتاج الى التنبؤ لمجموعة من متغيرات الاستجابة عن طريق مجموعة كبيرة من المتغيرات التوضيحية، تستخدم طريقة انحدار المربعات الصغرى الجزئية للتخلص من مشكلة التعدد الخطي التام بين المتغيرات التوضيحية[8] ، و أول من استخدم هذه التقنية هو العالم المتخصص بعلم الاقتصاد Herman Wold عام (1966) ثم أصبحت هذه الطريقة تستخدم في مجال الكيمياء عندما أستخدمها الكيميائيون في تحليل تركيب عينة كيميائية ، كذلك تم استخدام PLS في مجال الكهرباء والحاسبات ولكنها في النهاية وضعت في اطار احصائي من قبل Friedman عام (1993)[3].

أن طريقة المربعات الصغرى الجزئية تعتمد على خطوتين اساسيتين الاولى هي ايجاد المتغيرات الكامنه (latent variable) بين X و$\underline{Y}$ من خلال تعظيم مصفوفة التباين والتباين المشترك والخطوة الثانية هي انحدار $\underline{Y}$ على المركبات t ، توجد عدة خوارزميات فيما يتعلق بالمربعات الصغرى الجزئية ومن الخوارزميات التي تم الاعتماد عليها هي (NIPALS (PLS1,PLS2 ، حيث أن PLS1 تستعمل في حالة متغير الاستجابة واحد اما PLS2 فتستعمل في حالة وجود عدة متغيرات استجابة وهي من اولى الخوارزميات ثم تليها خوارزمية Kernel حيث الاثنان تعطيان نفس النتائج لكن الفرق هو في كيفية حساب المركبات ففي الاول يتم بصورة تكرارية اما الخوارزمية الاخرى يتم عن طريق ايجاد المتجهات الذاتية .

نفرض لدينا المصفوفة $Xn.p$ والمتجه $\underline{Y}$n.1 فطريقة المربعات الصغرى الجزئية تعتمد على النموذج الثنائي بين X و $\underline{Y}$ وكالاتي[12] [16] :

$$X = T\acute{P} + E \qquad \text{....(29)}$$

$$\underline{Y} = Uq' + f \qquad \text{....(30)}$$

حيث

T : مصفوفة درجات- مصفوفة x-score ذات رتبة nxr حيث ان T هي مصفوفة متماثلة symmetric

U : مصفوفة درجات -Y-score ذات رتبة nxr

P' : مصفوفة تحميلات- x-loading ذات رتبة pxr

q' : متجه تحميلات- Y-loading ببعد 1xr

E : مصفوفة البواقي- x-residual ذات رتبة nxp

$\underline{F}$: متجه البواقي- Y-residual ببعد nx1

المصفوفة P' والمتجه q' له r من الأعمدة وهو محدد بما يأتي:

$$(r < \min\,(n, p))$$

حيث أن:
P: عدد المتغيرات
n: عدد المشاهدات
r: عدد المركبات
والعلاقة الداخلية التي تربط بين scores تعطى كالاتي:

$$U = TD + H \quad .....(31)$$

حيث D مصفوفة قطرية ذات رتبة rxr
H مصفوفة البواقي ذات رتبة nxr
الفكرة الاساسيه في المربعات الصغرى الجزئية هو في كيفية ايجاد المتجه w من مجال X والمتجه c من المجال Y بحيث ان

$$\text{Max } COV(Xw, Yc)$$

$$with\|t\|=\|Xw\|=1 \quad \text{and} \quad \|u\| = \|Yc\| = 1 \quad ....(32)$$

حيث ان COV هو تقدير التباين المشترك
وان t,u هي اعمدة في المصفوفتين T,U ويتم تنفيذ التكرارات بطريقة متسلسلة وهذا يعني ان المتجهات scores يتم احتسابها الواحد بعد الاخر حتى يتم استخراج كافة المتجهات الى r تحت القيد عدم الارتباط بين المتجهات وتوجد طرائق عدة لحل المعادلة (32) منها خوارزمية Kernel وخوارزمية SIMPLS وخوارزمية NIPALS وغيرها من الخوارزميات وفي هذا البحث تم الاعتماد على خوارزمية التكرار غير الخطي للمربعات الصغرى الجزئية Non-linear Iterative partial least squares NIPALS(PLS1)

<u>1-3-5 خوارزمية التكرار غير الخطي للمربعات الصغرى الجزئية</u>
Non-linear Iterative partial least squares NIPALS(PLS1)

فيما يأتي الخطوات الاساسية لخوارزمية NIPALS(PLS1) لحساب اول مركبة [8][6]
1- في الخطوة الاولى يتم تهيئة U_1 عن طريق $\underline{Y}$ بحيث

$$U_1 = \underline{Y} \quad ......(33)$$

حيث ان U_1 متجه ببعد nx1
2- حساب X-weight

$$\underline{W1} = X'U_1/(U'_1U_1) \quad .....(34)$$

حيث ان $\underline{W_1}$ متجه ببعد px1
3- حيث ان $\underline{W_1}$ يكون normalize بالشكل $\underline{W1} = \underline{W1}/\|W1\|$
4- اسقاط البيانات X على x-weight لحساب x-scores وكالاتي

$$t_1 = XW1 \quad .....(35)$$

حيث ان t1 متجه ببعد nx1
5-حساب y-weight

$$C_1 = \underline{\acute{Y}}t_1/(t'_1t_1) \quad .....(36)$$

حيث c1 متجه ببعد 1x1
6- حيث ان C_1 تكون normalize بالشكل $C_1 = C_1/\|C1\|$
7- اسقاطات بيانات Y على Y-weight لحساب y-scores

$$U^*_1 = Y\, C_1 \quad .....(37)$$

حيث U^*_1 متجه ببعد nx1
8- نحدد U^*_1 بحيث تحقق مايلي

$$\Delta U = (U\Delta)'(U\Delta) \quad ....(38)$$

$U\Delta = U^*_1 - U_1$

-9 اذا كانت $\Delta U < \varepsilon$ وجدنا اول مركبة ونتوقف حيث ε قيمه صغيره عدا ذلك نذهب الى الخطوة الاولى ونستعمل U^*_1 بدل U_1 ونستمر بالخطوات.

-10 ايجاد X-loading وثم ايجاد الاوزان

$$P_1 = X' t_1 / (t'_1 t_1) \quad .....(39)$$

حيث ان P_1 متجه ببعد px1

$$P = P_1 / ||P_1|| \quad ......(40)$$

$$t_1 = t ||P_1|| \quad ......(41)$$

$$W = W ||P_1|| \quad ....(42)$$

-11 ايجاد التداخل الخطي للمعالم بواسطة انحدار OLS

$$d_1 = U'_1 t_1 / (t'_1 t_1) \quad .....(43)$$

حيث ان d1 متجه ببعد1x1

-12 عمل تفريغ deflate الى بيانات X

$$X_1 = X - t_1 P'_1 \quad .....(44)$$

-13 عمل تفريغ deflate الى بيانات Y

$$Y_1 = Y - d_1\, t_1\, C'_1 \quad .....(45)$$

ونستمر بالخطوات من (1-13) عدة مرات وبأستعمال البيانات المفرغه الى X وY حتى نحصل على كل المركبات المحددة ونستطيع ان نجد معاملات الانحدار بواسطة العلاقة الاتيه

-14 ايجاد معاملات الانحدار

$$\beta = w(p'w)^{-1}\, c' \quad .....(46)$$

حيث ان W هي مصفوفة برتبة pxr

P مصفوفة برتبة pxr

C مصفوفة برتبة rxr

-6 **الجانب التجريبي**:

1-6 **المقدمة** Introduction

تعرف المحاكاة بصورة عامة بأنها عبارة عن الحلول للمشكلات الرياضية من خلال بناء نموذج مشابه للمشكلة الأصلية ، يتم استخدام المحاكاة كأسلوب للتحليل عند عدم التمكن من استخدام أساليب التحليل الأخرى كما وانه يوفر على الباحثين الكثير من الجهد والكلفة والوقت.[5]

وتوجد طرائق مختلفة للمحاكاة ومن هذه الطرائق هي الطريقة التناظرية (Analog Method)، والطريقة المختلطة (Mixed Method)، وطريقة مونت كارلو (Monte Carlo Method) وتعد طريقة مونت كارلو من أهم هذه الطرائق وأكثرها شيوعا، حيث تستخدم في حل المسائل التي تتخللها عمليات عشوائية حيث يصعب عمل تجارب طبيعية وكذلك يصعب وضع صيغة رياضية معينة لها، لذا يستخدم اسلوب المحاكاة بواسطة العينة حيث يتم اخذ عينة عشوائية من المجتمع لتمثيل الظاهرة بدلا من المجتمع الحقيقي[7].

2-6 **مراحل تطبيق تجربة المحاكاة**

<u>Stages of the application of simulation experiment</u>

لقد تضمنت تجارب المحاكاة كتابة عدد من البرامج بلغة(MATLAB 2017)، اذ أن الأنموذج الذي تم الاعتماد عليه يكون وفق الصيغة(3) الواردة في الجانب النظري من هذا البحث.
اذ يتم وصف مراحل تجربة المحاكاة من خلال الخطوات الاتية:

الخطوة الاولى

تعيين القيم الافتراضية للمعلمات وهذه المرحلة من اهم المراحل التي يعتمد عليها، اذ أختيرت قيم المعلمات والنموذج المفترض كما مبين في ادناه:

جدول رقم (1): القيم الافتراضية للمعلمات والنموذج المفترض

β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
-1.11	0.74	1.15	0.92	2.32	-0.69	0.31	-4.01

الخطوة الثانية

توليد المتغيرات التوضيحية من خلال استعمال اسلوب مونت-كارلو(Mont-Carlo) في المحاكاة حيث يتم توليد سبعة متغيرات توضيحية وفق التوزيع الطبيعي المعياري وحدوث مشكلة التعدد الخطي.

الخطوة الثالثة

توليد قيم متغير الخطأ العشوائي في أنموذج الانحدار اللوجستي تبعا لتوزيع برنولي.

الخطوة الرابعة

حساب المتغير التابع(y) ثنائي الاستجابة الذي يتوزع توزيع برنولي، وفق طريقة توليد التحويل المعكوس(invers transformation) بالاعتماد على دالة الانحدار اللوجستي(π_i) وحد الخطأ العشوائي.

الخطوة الخامسة

من أهم العوامل الأخرى التي يتم اختيارها والمؤثرة هي كالاتي:

1- اختيار اربعة أحجام للعينات المفترضة وهي(200,100,50,25)

2- اختيار القيم الافتراضية لمعاملات الارتباط وهي(0.99,0.90,0.80)

الخطوة السادسة

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

1- طريقة المركبات الرئيسية للانحدار اللوجستي (PCLR).

2- طريقة انحدار المربعات الصغرى الجزئية (PLSR).

الخطوة السابعة

أما في هذه المرحلة، تتم المقارنة بين طرائق التقدير المدروسة بالاعتماد على المقياس الاحصائي متوسط مربعات الخطأ(MSE) للأنموذج.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{\pi}_i - \pi_i)^2 \qquad ......(47)$$

واخيرا سوف يتم تكرار تجربة المحاكاة (1000) مره.

3-6 نتائج تجربة المحاكاة

سيتم عرض نتائج تجربة المحاكاة وتحليلها حسب القيم الافتراضية للمعلمات والنموذج المفترض من خلال طريقتي تقدير معلمات أنموذج الانحدار اللوجستي ومتوسط مربعات الخطأ لكل طريقة في وجود مشكلة التعدد الخطي حيث تم التأكد من وجود هذه المشكلة من خلال اختبار معامل تضخم التباين (Variance inflation factor)حيث أنه اذا كانت قيمة معامل التضخم (VIF>4) دل ذلك على وجود تعدد خطي بين المتغيرات[2].

جدول رقم (2) يبين تقديرات المعلمات و قيم متوسط مربعات الخطأ لأنموذج الانحدار اللوجستي في حالة $n=25$ وبارتباطات مختلفة

coefficients		Methods		
		PCLR All PC'S	PCLR 3 PC'S	PLS-LR
$\rho = 0.80$	$\hat{\beta}_0$	-0.1498	0.0231	0.5298
	$\hat{\beta}_1$	0.2157	-0.0168	-0.1864
	$\hat{\beta}_2$	-0.6337	-0.9614	-3.6010
	$\hat{\beta}_3$	0.5415	-1.3376	1.0792
	$\hat{\beta}_4$	0.0036	-0.0318	0.3937
	$\hat{\beta}_5$	0.1180	0.2877	4.1230
	$\hat{\beta}_6$	-1.0731	0.7965	-0.4159
	$\hat{\beta}_7$	0.8699	0.0858	-0.6745
MSE		0.0002	0.2267	0.2243
$\rho = 0.90$	$\hat{\beta}_0$	0.5025	0.6185	1.6361
	$\hat{\beta}_1$	-2.7041	-0.5721	-0.6737
	$\hat{\beta}_2$	0.4853	0.1463	-2.3224
	$\hat{\beta}_3$	0.9408	0.0996	0.1259
	$\hat{\beta}_4$	-0.2281	-0.5504	-0.8778
	$\hat{\beta}_5$	-1.0504	0.2232	1.2712
	$\hat{\beta}_6$	0.1841	-0.2945	1.6119
	$\hat{\beta}_7$	4.0555	-0.2799	1.5070
MSE		0.0016	0.0517	1.0790
$\rho = 0.99$	$\hat{\beta}_0$	0.6233	0.6020	1.6329
	$\hat{\beta}_1$	8.5495	-2.5195	-2.3308
	$\hat{\beta}_2$	-2.6632	5.3243	-0.5445
	$\hat{\beta}_3$	8.9589	-2.3377	-0.3447
	$\hat{\beta}_4$	-1.7458	-0.6398	1.2087
	$\hat{\beta}_5$	4.5908	-0.0966	2.2933
	$\hat{\beta}_6$	2.0426	-0.1623	0.6480
	$\hat{\beta}_7$	-17.3583	-1.0585	-0.2977
MSE		0.0407	0.0751	1.6954

جدول رقم (3) يبين تقديرات المعلمات و قيم متوسط مربعات الخطأ لأنموذج الانحدار اللوجستي في حالة $n = 50$ وبارتباطات مختلفة

coefficients		PCLR All PC'S	PCLR 3 PC'S	PLS-LR
$\rho = 0.80$	$\hat{\beta}_0$	-0.0336	-0.1601	1.2961
	$\hat{\beta}_1$	-3.1145	1.1760	1.3432
	$\hat{\beta}_2$	1.3477	-0.7631	-0.9160
	$\hat{\beta}_3$	1.2889	0.5587	-1.2313
	$\hat{\beta}_4$	-0.0550	-0.2834	-1.4452
	$\hat{\beta}_5$	0.8439	0.4204	3.3509
	$\hat{\beta}_6$	-0.8068	0.1866	1.0446
	$\hat{\beta}_7$	-0.4410	-1.2370	0.0824
MSE		0.2207	0.2374	0.7891
$\rho = 0.90$	$\hat{\beta}_0$	0.3059	0.4935	1.5279
	$\hat{\beta}_1$	0.7360	0.0584	0.5120
	$\hat{\beta}_2$	-1.2710	-0.8340	-0.9895
	$\hat{\beta}_3$	0.4342	-0.6704	1.5037
	$\hat{\beta}_4$	1.2264	0.0273	0.0844
	$\hat{\beta}_5$	-0.5473	0.5433	1.9922
	$\hat{\beta}_6$	1.8526	0.2620	-1.8660
	$\hat{\beta}_7$	-0.8988	0.0033	0.4797
MSE		0.1266	0.1435	1.9055
$\rho = 0.99$	$\hat{\beta}_0$	0.7435	0.2206	1.7149
	$\hat{\beta}_1$	-11.7146	-0.6409	-0.2526
	$\hat{\beta}_2$	8.8317	0.1520	-2.0224
	$\hat{\beta}_3$	-2.3180	0.8260	2.0901
	$\hat{\beta}_4$	10.5606	0.1944	0.9373
	$\hat{\beta}_5$	3.4069	-1.5000	-0.0630
	$\hat{\beta}_6$	-6.5536	0.1569	-0.7323
	$\hat{\beta}_7$	-2.6321	-0.4608	0.3760
MSE		0.1197	0.0160	2.9814

جدول رقم (4) يبين تقديرات المعلمات و قيم متوسط مربعات الخطأ لأنموذج الانحدار اللوجستي في حالة $n = 100$ و بارتباطات مختلفة

Coefficients		Methods		
		PCLR All PC'S	PCLR 3 PC'S	PLS-LR
$\rho = 0.80$	$\hat{\beta}_0$	-0.4198	0.3643	1.1624
	$\hat{\beta}_1$	-0.1125	-0.0673	2.9594
	$\hat{\beta}_2$	0.2387	-0.0487	-2.4117
	$\hat{\beta}_3$	0.6603	-0.1027	-2.6019
	$\hat{\beta}_4$	0.2846	-0.4316	-0.3998
	$\hat{\beta}_5$	-0.7143	-0.0661	1.1412
	$\hat{\beta}_6$	0.8846	-0.1965	1.0934
	$\hat{\beta}_7$	-0.0620	-0.3897	-0.2157
MSE		1.6025	0.0364	1.3342
$\rho = 0.90$	$\hat{\beta}_0$	0.4031	0.3979	1.4873
	$\hat{\beta}_1$	-0.9688	0.1621	-0.3000
	$\hat{\beta}_2$	0.5485	-0.2029	-0.9526
	$\hat{\beta}_3$	1.0647	-0.2410	-0.0955
	$\hat{\beta}_4$	0.3807	-0.2167	-0.7994
	$\hat{\beta}_5$	-1.1333	-0.2915	1.9513
	$\hat{\beta}_6$	1.5300	-0.0599	-0.0493
	$\hat{\beta}_7$	-0.3908	-0.1156	1.8067
MSE		0.0170	0.0017	3.9235
$\rho = 0.99$	$\hat{\beta}_0$	0.4378	0.3691	1.7331
	$\hat{\beta}_1$	-1.0828	0.2031	0.0684
	$\hat{\beta}_2$	-0.7384	0.9490	-1.0823
	$\hat{\beta}_3$	1.6442	-0.2469	-0.7895
	$\hat{\beta}_4$	2.8495	-0.9889	-0.5130
	$\hat{\beta}_5$	2.3071	-0.1720	2.0271
	$\hat{\beta}_6$	-2.6945	-0.6907	-0.2279
	$\hat{\beta}_7$	-3.1790	-0.0652	1.3162
MSE		0.0236	0.0008	6.0445

جدول رقم (5) يبين تقديرات المعلمات و قيم متوسط مربعات الخطأ لأنموذج الانحدار اللوجستي في حالة $n = 200$ وبارتباطات مختلفة

Coefficients		Methods		
		PCLR All PC'S	PCLR 3 PC'S	PLS-LR
$\rho = 0.80$	$\hat{\beta}_0$	-0.1490	0.4672	1.0635
	$\hat{\beta}_1$	0.2401	-0.0322	-0.5595
	$\hat{\beta}_2$	-0.3608	0.0353	-2.4166
	$\hat{\beta}_3$	0.4647	-0.4349	-0.2173
	$\hat{\beta}_4$	-0.7113	0.2077	0.9166
	$\hat{\beta}_5$	-0.3545	0.0100	1.2216
	$\hat{\beta}_6$	0.5392	-0.5141	-0.6322
	$\hat{\beta}_7$	-1.2270	0.0402	-0.8248
MSE		2.2398	0.5282	2.2575
$\rho = 0.90$	$\hat{\beta}_0$	-0.4535	0.2696	1.7102
	$\hat{\beta}_1$	-1.0766	0.9587	-0.2430
	$\hat{\beta}_2$	1.8349	-1.4398	-0.4857
	$\hat{\beta}_3$	-1.0601	-0.2122	2.1358
	$\hat{\beta}_4$	-0.0994	-0.5258	-0.2084
	$\hat{\beta}_5$	-0.4181	0.8445	1.3770
	$\hat{\beta}_6$	0.2595	-0.0262	-0.9076
	$\hat{\beta}_7$	1.2210	-0.7748	-0.3782
MSE		4.5064	0.0018	11.6709
$\rho = 0.99$	$\hat{\beta}_0$	-0.4983	0.2394	1.6590
	$\hat{\beta}_1$	1.9868	-0.2197	0.7591
	$\hat{\beta}_2$	-0.9657	-0.2948	-1.2731
	$\hat{\beta}_3$	-2.9580	-0.7046	-0.7084
	$\hat{\beta}_4$	-0.3734	0.4451	-0.6410
	$\hat{\beta}_5$	-1.2423	0.8650	0.7049
	$\hat{\beta}_6$	2.9272	-0.7303	1.9199
	$\hat{\beta}_7$	1.2511	-0.6595	0.3066
MSE		5.6328	0.0786	13.7107

من خلال النتائج المبينة في الجدول (2) و (3) و (4) و (5) تم ملاحظة الاتي:

1- عند حجم العينة (25) ولجميع قيم معاملات الارتباط(0.99 ,0.90 ,0.80) نلاحظ بأن طريقة PCLR(All pc's) هي افضل طريقة في معالجة مشكلة التعدد الخطي لأنموذج الانحدار اللوجستي وذلك لانها تمتلك اقل (MSE).

2- عند حجم العينة (50) وعندما يكون معامل الارتباط (0.90 ,0.80) نلاحظ بأن طريقة PCLR(All pc's) تمتلك اقل (MSE) بينما عندما يكون معامل الارتباط (0.99) نلاحظ بأن طريقة PCLR(3pc's) تمتلك اقل (MSE).

3- عند حجم العينة (100) ولجميع قيم معاملات الارتباط (0.99 ,0.90 ,0.80) نلاحظ بأن طريقة PCLR(3pc's) هي افضل طريقة في معالجة مشكلة التعدد الخطي لأنموذج الانحدار اللوجستي وذلك لانها تمتلك اقل (MSE).

4- عند حجم العينة (200) ولجميع قيم معاملات الارتباط (0.99 ,0.90 ,0.80) نلاحظ بأن طريقة PCLR(3pc's) هي افضل طريقة في معالجة مشكلة التعدد الخطي لأنموذج الانحدار اللوجستي وذلك لانها تمتلك اقل (MSE).

7- الاستنتاجات والتوصيات

7-1 الاستنتاجات

1- أثبتت طريقة PCLR(All pc's) كفاءتها في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي في حالة حجوم العينات الصغيرة ولجميع قيم معاملات الارتباط (0.99 , 0.90 , 0.80).

2- أثبتت طريقة PCLR(All pc's) كفاءتها في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي في حالة حجوم العينات المتوسطة ولقيم معاملات الارتباط (0.90 , 0.80).

3- أثبتت طريقة PCLR(3pc's) كفاءتها في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي في حالة حجوم العينات المتوسطة ولقيمة معامل الارتباط (0.99).

4- أثبتت طريقة PCLR(3pc's) كفاءتها في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي في حالة حجوم العينات الكبيرة ولجميع قيم معاملات الارتباط (0.99 , 0.90 , 0.80).

5- ظهر أن طريقة PLSR اقل كفاءة في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي من حيث قيم (MSE) للأنموذج بالمقارنة مع طريقة PCLR(All pc's) و PCLR(3pc's) .

7-2 التوصيات

1- استعمال طريقة PCLR(All pc's) في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي في حالة حجوم العينات المتوسطة والصغيرة.

2- استعمال طريقة PCLR(3pc's) في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي في حالة حجوم العينات الكبيرة والمتوسطة.

8- المصادر

1- البياتي ، محمود مهدي حسن . (2012) ، " تطبيق عملي لتحليل البيانات الاحصائية " ، الجزيرة للطبع والنشر/ جامعة بغداد.

2- حسن ، زينة إبراهيم . (2017) ، " استعمال الأنحدار المعكوس الشرائحي مع اساليب اخرى في اختزال الأبعاد " ، أطروحة دكتوراه في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد.

3- حسين ، الهام عبد الكريم . (2012) م ، " مقارنة بين استعمال أنموذج المربعات الصغرى الجزئي PLSR وانحدار المكونات الرئيسية PCR في العوامل المؤثرة على تمدد الاسمنت " ، مجلة التربية والعلم ، المجلد 25 ، العدد 2 .

4- حسين ، شيرين علي . (2009) ، "مقدرات الامكان الاعظم الموزونة الحصينة ومقارنتها مع طرائق اخرى لأنموذج اللوجستك مع تطبيق عملي" ، رسالة ماجستير في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد .
5- خطار ، جبران عبد الأمير ومحمد ، كوركيس شهيد واسماعيل ، محمد سالم . (2016)، " تقدير معلمات الانحدار الخطي المتعدد بأستخدام الطرائق الحصينة (دراسة مقارنة)"، مجلة القادسية لعلوم الحاسوب والرياضيات ، المجلد 8 ، العدد 1.
6- صالح ، رباب عبد الرضا . (2016) ، " مقارنة بين طرائق المربعات الصغرى الجزئية والمركبات الرئيسية بأستعمال المحاكاة " ، مجلة العلوم الاقتصادية والادارية ، المجلد 22 ، العدد 87 .
7- كاظم ، اموري هادي ومسلم ، باسم شليبه.(2002)م ، " القياس الاقتصادي المتقدم النظرية والتطبيق " ، مطبعة الطيف ، بغداد.

8- Abdi , Hervi , (2010). " Partial Least Squares Regression and Projection on Latent Structure Regression (PLS Regression) " , John Wiley & Sons, 1-10.

9- Aguilera , A.M., and Escabias , M and M.J. Valderrama , (2006). " Using principal components for estimating Logistic regression with high dimensional Multicollinearity data ", Computational statistics Data Analysis 50 , 1905-1924.

10- Anne Laure Boulesteix , Korbinian Strimmer , (2005). "Partial Least Squares: a Versatile Tool for the analysis of high-dimensional genomic data " , Briefings in Bioinformatics , oxford University , P.32-44.

11- Berkson, J.(1944). " Application of the Logistic Function to Bioassay " . JASA Vol . 39,PP.357-365.

12- Chong , I., Jun T , C., 2005. " Performance of Some Variable Selection Methods When Multicollinearity Present " , Chemometrics and Intelligent Laboratory Systems 78 , 103-112.

13- Cook , D., Dixon , P., Duckworth , W.M., Kaiser , M.S., Koehler , K., Meeker , W.Q and Stephenson , W.R., (2001). " Binary Response and Logistic Regression Analysis", University NSF/ILI Project Beyond Traditional statistical Methods , grant.

14- Marx , B.D., (1990) . " A macro For Principal Component Logistic Regression" , Louisiana State University , pp 1438-1443.

15- Muhammed , Nur , Aidi. M.S. and Tuti , Purwaningsih , S. Stat (2013) " Modeling Spatial Ordinal Logistic Regression and The Principal Component to Predict Poverty Status of Districts in Java Island " , International Journal of Statistics and Applications , 3(1) : 1-8.

16- Roon , P., Zakizadeh , J., Chartier, S. (2014) . " Partial Least Squares tutorial for analyzing neuroimaging data ".

17- Wirtschaftsmathematik , (2009). " Using Penalized Logistic Regression Models for predicting the Effects of Advertising Material " , Vienna University of Technology.

18- Zhenjie , Li. (2014) , " Research on Distinguish The Accounting Information Distortion Based on The Principle Components Logistic Regression Model " , International Journal of Security and Its Applications Vol.8 , No.4 , pp. 37-50.

The use of the Principal components and Partial least squares methods to estimate the parameters of the logistic regression model in the case of linear multiplication problem

Abstract

The logistic regression model is one of the nonlinear models that aims at obtaining highly efficient capabilities, It also the researcher an idea of the effect of the explanatory variable on the binary response variable.

The large number of explanatory variables usually used to illustrate the response led to the problem of linear multiplicity between the explanatory variables that make estimating the parameters of the model not very accurate.

In this paper, examined methods for estimating the parameters of the logistic regression model in the case of the problem of linear multiplicity These methods are: Principal components of logistic regression method and Partial least square regression method.

The results of the simulation showed that the method (PCLR(3pc's)) is best for estimating the parameters of the binary logistic regression model response in the case of a problem of linear multiplicity.

Keywords **Logistic regression, binary data, Principal components, Partial least square, multicollinearity.**