

استخدام البرمجة الخطية في تقدير الانحدار الربيعي مع تطبيق

م. د مزاحم محمد يحيى الهاشمي

جامعة الموصل/كلية علوم الحاسوب والرياضيات

Linear programming methods for estimating quantile regression with the application

Muzahem Mohammed Yahya AL-Hashimi

University of Mosul

College of Computer Science & Mathematics

muzahim63@gmail.com

المستخلص:

يعرض البحث طريقة الانحدار الربيعي في تقدير الربيعات الشرطية، وكذلك يعرض البحث طريقتين في الكشف عن المشاهدات ذات القوة الرافعة والمشاهدات الشاردة. لتحديد الخوارزمية الافضل في ايجاد تقديرات معلمات الانحدار الربيعي لبيانات ثلاثيميا الاطفال في الموصل، فقد تمت المقارنة بين كُل من الخوارزمية المبسطة (Simplex algorithm)، الخوارزمية الممهدة (Smoothing algorithm) وخوارزمية النقطة الداخلية (Interior point algorithm). تم استخدام طريقة اعادة المعاينة بووتستراب الهامشية لسلسلة ماركوف (Markov Chain Marginal Bootstrap) (MCMB) لحساب حدود الثقة.

الكلمات الافتتاحية: الانحدار الربيعي، الخوارزمية الممهدة، الخوارزمية المبسطة، خوارزمية النقطة الداخلية، بووتستراب الهامشية لسلسلة ماركوف، ثلاثيميا الأطفال.

Abstract:

This paper deals with a quantile regression for estimating conditional quantiles, also it deals with two methods for detecting leverage and outlier observations. To determine the best algorithm for estimating the parameters of the quantile regression for the thalassemia data in Mosul city. Three algorithms compared for regression estimation, which are simplex algorithm, smoothing algorithm, and interior point algorithm. Markov chain marginal bootstrap used to compute the confidence intervals.

Keywords:

Quantile regression, Smoothing algorithm, Simplex algorithm, Interior point algorithm, Markov chain marginal bootstrap, Thalassemia

1. المقدمة والهدف من البحث:

يحظى مرض الثلاثيميا بعناية واهتمام الباحثين، لما يمثله هذا المرض من اهمية محورية في مجال الصحة العامة، متمثلة في أن لها اثار مستقبلية سريرية ونفسية جسيمة يمكن أن تؤدي الى عاهة في العظام، قصر القوام، تأخر النمو، تأخر سن البلوغ، عجز في القلب، وغيرها من الاثار. الامر الذي يتطلب البحث عن سبل في التخفيف من معاناة المرضى من خلال الكشف عن العوامل التي لها الأثر الايجابي في حياة مرضى ثلاثيميا الاطفال.

تظافرت جهود الباحثين في تناول العديد من العوامل المختلفة المحيطة بمرض ثلاثيميا الاطفال للكشف عن مدى تأثيرها أو تأثرها بتلك العوامل. من بين تلك العوامل ما يتعلق بالعمر من العظم والمتغيرات المؤثرة فيه. وتعد دراسة

العمر من العظم، من الدراسات الجوهرية والأساسية في الكشف عن المساهمات الايجابية للمتغيرات التوضيحية لكي يطلع عليها اصحاب الاختصاص من اجل وضع الحلول وللتخفيف من وطاءة المرض للمصابين بتلاسيميا الاطفال. غالباً ما تتضمن البيانات الاحصائية التي تتناول البحث في الامراض التي تصيب الانسان قيم شاردة، وهي تمثل مشكلة احصائية إذا ما اختير ادراجها في تحليل البيانات أوفي حالة اتخاذ القرار باستبعادها من التحليل. في هذا البحث تم التعامل مع القيم الشاردة على أنها قيم جوهرية يُعتمد عليها في التحليل. وتم استخدام كُلاً من طريقتي مسافة مهلائبس (Mahalanobis distance) والمسافة الحصينة (Robust distance) في الكشف عن القيم الشاردة. من المعروف أن طرائق الانحدار الحصين (Robust regression)، المتمثلة بتقديرات M هوبر (Huber M estimation)، تقدير المربعات المشذبة الصغرى (Least trimmed squares)، تقدير S (S estimation)، تقدير MM (MM estimation). هذه الطرائق لها القدرة في الكشف عن القيم الشاردة ومن ثم توفير نتائج مستقرة في حالة وجود القيم الشاردة. الا أن هذه الطرائق لا توفر امكانية دراسة علاقة تأثيرات المتغيرات التوضيحية على الرُبعيات الشرطية لمتغير الاستجابة.

تعد طريقة الانحدار الرُبعي من الطرائق الحصينة، إذ لا توجد اي فروض توزيعية حول حد الخطأ في الأنموذج، فضلا عن كونها حصينة تجاه المشاهدات الشاردة، كما أنها تستخدم مع البيانات غير المتجانسة، وأنها توفر امكانية دراسة العلاقة بين متغير توضيحي واحد أو أكثر وبين الرُبعيات الشرطية لمتغير الاستجابة، مثل الوسيط أو المئينات (the percentile)، كالمئين 90 مثلاً. عليه فقد تمت مقارنة ثلاثة خوارزميات من خوارزميات البرمجة الخطية لطريقة الانحدار الرُبعي وهي الخوارزمية المبسطة (Simplex Algorithm)، الخوارزمية الممهدة (Smoothing Algorithm) وخوارزمية النقطة الداخلية (Interior point algorithm) من اجل تحديد الخوارزمية الافضل في تقدير البيانات ذات الحجم الصغير نسبياً في ظل وجود بيانات شاردة وذات قوة رافعة والمتمثلة في بيانات مرضى تلاسيميا الاطفال في الموصل.

2. الجانب النظري:

المشاهدة المتطرفة : هي مشاهدة أو مجموعة من المشاهدات في مجموعة من البيانات، لا تبدو متنسقة مع بقية البيانات. وعدم الاتساق هذا يعود الى قيمة المشاهدة، اي أنها اما أن تكون اعلى أو اقل بكثير من المشاهدات الاخرى. الا أنما يبدو على أنها مشاهدة غير متنسقة لباحث ما، يمكن أن لا يبدو كذلك الى اخر من ذوي الخبرة بالظاهرة قيد الدراسة (اي أن المشاهدات الشاردة هي جزء من الظاهرة قيد الدراسة، وهو ما يسمى بالتوزيع ثقيل الذيل (Kleiber and Zeileis. 2008: 98)، إذ لا يُصنف المشاهدة غير المتسقة على أنها قيمة متطرفة (ان لم تكن المشاهدة ناتجة عن خطأ في القياس).

يمكن تصنيف المشاهدة المتطرفة الى نوعين بحسب المتغير الذي تظهر فيه، وكالاتي (Weisberg, S. 2005: 194):

1.2.1 المشاهدة ذات القوة الرافعة (Leverage point) وهي القيمة المتطرفة في المتغيرات التوضيحية والتي

تعمل كعتلة أو رافعة (Lever) وتغير من الميل لأنموذج خط الانحدار. هذه المشاهدة تقع بعيدة عن بقية البيانات وبالتحديد في نهاية أحد طرفي خط الانحدار وتؤدي إلى ميلانه باتجاهها.

2.2.1 المشاهدة الشاردة (Outlier) وهي القيمة المنطرفة في متغير الاستجابة والتي لها باقي (Residual)

عالي، بمعنى آخر، أنها المشاهدة التي قيمة متغير الاستجابة فيها غير عادية. لغرض الكشف عن المشاهدات ذات القوة الرافعة للبيانات المستخدمة في هذا البحث، فقد تم استخدام كل من طريقتي مسافة مهلائبس (Mahalanobis) والمسافة الحصينة (Robust distance). تُعرّف مسافة مهلائبس (Mahalanobis) بأنها المسافة بين النقطة x والوسط الحسابي للعينة، مع الأخذ بنظر الاعتبار كل المعلومات المتعلقة بمصفوفة التباين-التباين المشترك للعينة، وتأخذ الصيغة الآتية (Wilcox, 1997: 204):

$$d = [(x - \bar{X})' S^{-1} (x - \bar{X})]^{1/2} \dots (1)$$

إذ أن $\bar{X} = (\bar{X}_1, \dots, \bar{X}_p)'$ ، وان $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ و $S = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'}{n-1}$ هما الوسط الحسابي (الموقع متعدد المتغيرات التجريبي (empirical multivariate location) ومعكوس مصفوفة التباين-التباين المشترك (القياس (scale)).

وتعرف المسافة الحصينة على أنها (Rousseeuw and Zomeren, 1990):

$$RD(x) = [(x - \hat{\mu}_R)' \hat{\Sigma}_R^{-1} (x - \hat{\mu}_R)]^{1/2} \dots (2)$$

إذ أن $\hat{\mu}_R(X)$ و $\hat{\Sigma}_R(X)$ هما موقع متعدد المتغيرات الحصين وتقديرات القياس (scale estimates)، وصيغتهما الرياضية كالآتي.

$$\hat{\mu}_R(X) = \left(\sum_{i=1}^n w_i x_i \right) / \sum_{i=1}^n w_i \dots (3)$$

وان

$$\hat{\Sigma}_R(X) = \left(\sum_{i=1}^n w_i (x_i - \hat{\mu}_R)(x_i - \hat{\mu}_R)' \right) / \sum_{i=1}^n w_i - 1 \dots (4)$$

لنكن قيمة القطع (cutoff value)، كالآتي:

$$C(p) = \sqrt{\chi_{(p, 1-\alpha)}^2} \dots (5)$$

إذ يتم تحديد المشاهدة ذات القوة الرافعة، كالآتي:

$$\left. \begin{array}{l} RD(x) \leq C(p) \text{ إذا كان } 0 \\ \text{ماعد ذلك } 1 \end{array} \right\} = \text{المشاهدة ذات القوة الرافعة}$$

وتستخدم البواقي r_i (إذ أن $i = 1, \dots, n$) التي تستند الى تقديرات الأنحدار الربيعي للكشف عن المشاهدات الشاردة العمودية (vertical outliers)، اي المشاهدات الشاردة في متغير الاستجابة، إذ أن:

$$\left. \begin{array}{l} \text{المشاهدة الشاردة في متغير الاستجابة} \\ 0 \\ \text{إذا كان } |r_i| \leq k\sigma \\ 1 \\ \text{ماعد ذلك} \end{array} \right\}$$

يتم احتساب σ على أنها الوسيط للبقايا المطلقة، أي أن

$$\sigma = \text{median} \{ |r_i| / \beta_0 \} \dots (6)$$

إذ أن $(i = 1, \dots, n)$ ، $\beta_0 = \Phi^{-1}(0.75)$ هي الحد الثابت المعدل، إذ تم تعديله لكي يتسق مع التوزيع الطبيعي.

2.1 المفاهيم الأساسية Fundamental Concepts

يعد الانحدار الرُّبِيعي علم منهجي مهم (Jiansheng et.al, 2010). إذ من خلال تقدير الدوال الرُّبِيعية الشرطية المختلفة، فإن الانحدار الرُّبِيعي يُكْمَل انحدار المربعات الصغرى التقليدية على الوسط الشرطي، ويوفر استراتيجية منهجية للبحث في حجم تأثير التباين المشترك على توزيع متغير الاستجابة. كما انه يوفر انموذجا احصائيا أكثر تكاملا من الانحدار التقليدي، مما يجعل منه ذا تطبيقات واسعة الانتشار (Keming et.al, 2003). من المعروف أن مقدر طريقة المربعات الصغرى الاعتيادية (Ordinary least-squares estimator) يجعل من مجموع مربعات البواقى (sum squares residuals) اقل ما يمكن وهو ما يطلق عليه بالانحدار الخطي، والطريقة البديلة هي مقدر الانحرافات المطلقة الصغرى (least absolute deviations estimator) والذي يجعل من مجموع القيم المطلقة للبقاوي (Sum of absolute values of residuals) اقل ما يمكن، وهو ما يطلق عليه بالانحدار الرُّبِيعي.

ان الانحدار الخطي يُنمذج العلاقة بين متغير توضيحي واحد أو أكثر (X) وبين الوسط الحسابي لمتغير الاستجابة (Y) مشروطا بـ X . اما الانحدار الرُّبِيعي (Quantile regression)، فهو يُنمذج العلاقة بين متغير توضيحي واحد أو أكثر وبين متغير الاستجابة، مشروطا بالرُّبِيعات مثل الوسيط (أو المئينات مثل المئين 90 (90th percentile)) لمتغير الاستجابة. بمعنى اخر، أنه يعمل على ايجاد مقدار التغير في العلاقة بين متغير الاستجابة والمتغيرات التوضيحية بالاعتماد على قيم متغير الاستجابة. عليه، يختلف الانحدار الرُّبِيعي عن الانحدار الخطي التقليدي في أن الأخير يعتمد فقط على الوسط الحسابي الشرطي، بينما الانحدار الرُّبِيعي يسمح بتقدير الاعتماد الدالي بين المتغيرات لكل اجزاء التوزيع الشرطي لمتغير الاستجابة. أي أن الانحدار الرُّبِيعي يوسع إطار التقدير من سلوك الجزء المركزي من نقاط البيانات الى كل اجزاء التوزيع الشرطي.

يستخدم الانحدار الرُّبِيعي عندما تكون نسبة التغير في الرُّبِيعي الشرطي تعتمد على الرُّبِيعات. كما يستخدم الانحدار الرُّبِيعي مع البيانات غير المتجانسة (heterogeneous)، بمعنى أن طرفي (tails) التوزيع والموقع المركزي للتوزيعات الشرطية تتفاوت بشكل مختلف مع المتغيرات التوضيحية (كمثال على ذلك، العلاقة بين الاعمار والأوزان للفئات العمرية من 2 سنة الى أكثر من 80 سنة، إذ أن الأوزان تزداد بوتيرة متسارعة مع تقدم العمر من 2-20 سنة، ومن ثم تصبح مسنقرة نسبيا في منتصف العمر، ثم ما تلبث أن تتناقص ما بعد سن الـ 60 سنة). على ضوء هذا المثال، فإن الموقع المركزي للتوزيعات الشرطية هو الوزن في منتصف العمر، وأن طرفي التوزيع هي الأوزان للفئة العمرية من 2-20 سنة فضلا عن أوزان الفئة العمرية من 60-80 سنة. عليه، إذا ما تم استخدام الانحدار الرُّبِيعي لنمذجة العلاقة

بين الاعدار (متغير توضيحي) والأوزان (متغير الاستجابة)، فان معاملات الانحدار ستختلف من ربيعي شرطي لأخر، وخصوصا للفئة العمرية التي هي اقل من 20 سنة.

تظهر البيانات غير المتجانسة في العديد من حقول المعرفة بما في ذلك الطب الحيوي (Biomedicine)، والقياس الاقتصادي (Econometrics)، العلوم المالية (finance) وتحليل البقاء (Survival analysis)، وفي الدراسات البيئية (environmental). وعليه، فان الانحدار الربيعي أصبح ذا فائدة كبيرة في التطبيقات العملية لهذه المجالات (Chen and Wei, 2005).

يعد الانحدار الوسيط (median regression) حالة خاصة من الانحدار الربيعي، إذ يقدم صورة كاملة عن العلاقة بين متغير الاستجابة والمتغيرات التوضيحية، كما يوفر القدرة على بيان الميزات الهامة للبيانات والتي يمكن أن تغيب في حالة نمذجة المعدل على التوزيع الشرطي للوسط الحسابي.

يعد الانحدار الربيعي مرنا (flexible) تجاه خطأ التوزيع (error distribution)، إذ لا يوجد اي فروض توزيعية حول حد الخطأ في الأنموذج (اي ممكن أن يكون توزيع الخطأ أي توزيع غير التوزيع الطبيعي)، على خلاف طريقة المربعات الصغرى الاعتيادية التي تستند الى فرضية أن حد الخطأ يتبع التوزيع الطبيعي. كما أنها حصينة (robust) تجاه النقاط الشاردة في البيانات، فضلا عن أنها لا تتضمن دالة رابطة (link function) والتي تربط التباين والوسط الحسابي لمتغير الاستجابة كما هو الحال بالنسبة للنماذج الخطية المعممة (Generalized linear models)، فضلا عن توصيف التوزيع الذي يقع ضمن توزيعات العائلة الأسية.

من الجدير بالملاحظة أنه لا يمكن تنفيذ الانحدار الربيعي وذلك بتجزئة التوزيع غير المشروط لتوزيع متغير الاستجابة ومن ثم الحصول على توفيق للمربعات الصغرى للمجموعات الجزئية. إذ تسمى هذه العملية بقطع (truncation) متغير الاستجابة، وهي تؤدي الى نتائج مضللة عندما تتضمن البيانات مشاهدات متطرفة، في حين أن الانحدار الربيعي يستخدم جميع البيانات لتوفيق الربيعات، ومن ضمنها الربيعات المتطرفة.

2.2 مفهوم الانحدار الربيعي Concept of Quantile regression

ان اول من قدم مفهوم الانحدار الربيعي هو العالم Koenker في العام 1978 (Koenker and Bassett) (1,978). وعزفه (Møller, 2006: 9) على انه مسألة خطية امثليه (linear optimization problem). والآتي مفهوم الانحدار الربيعي.

ليكن Y متغير الاستجابة (response variable)، و X يمثل مصفوفة تصميم (Koenker, 2005) (design matrix) لمشاهدات المتغيرات التوضيحية. ولتكن $F(y|X = x) = P(Y \leq y|X = x)$ تمثل دالة التوزيع التراكمية الشرطية (conditional cumulative distribution function) لـ Y علما أن $X = x$ ، فانه يمكن تعريف الربيعي الشرطي q لمتغير الاستجابة Y على أنه الدالة المعكوسة لدالة التوزيع التراكمية، وكالاتي:

$$F^{-1}(q) = \inf \{y: F(y|X) \geq q\} \dots (7)$$

إذ أن $0 < q < 1$. وإذا كانت $q = 1/2$ ، فان الوسيط هو $F^{-1}(1/2)$.

إذا كان لدينا عينة عشوائية مكونة من n من المشاهدات $\{y_1, \dots, y_n\}$ ، فان من المعلوم أن وسيط العينة يجعل من مجموع الانحرافات المطلقة في نهايتها الصغرى، اي أن

$$\text{median} = \arg \min_{\varepsilon \in R} \sum_{i=1}^n |y_i - \varepsilon| \dots (8)$$

وبنفس الطريقة، فان الربيعي q العام $(\varepsilon(q))$ ، والذي هو مناظر لـ $F(q)$ ، يمكن صياغته كنهاية صغرى وكالاتي:

$$\varepsilon(q) = \arg \min_{\varepsilon \in R} \sum_{i=1}^n \rho_q(y_i - \varepsilon) \dots (9)$$

إذ أن ρ_q هي دالة الخسارة (loss function) (Koenker, 2005)، وتأخذ الآتي:

$$\rho_q(e_i) = \begin{cases} qe_i & , e_i \geq 0 \\ (q-1)e_i & , e_i < 0 \end{cases} \dots (10)$$

يستفاد من دالة الخسارة في تعيين الوزن لـ q فيما يتعلق بالبقايا الموجبة $(y_i - \varepsilon)$ والوزن $(1 - q)$ للبقايا السالبة، علماً أن $\rho_q(z) = z(q - I(z < 0))$ ، إذ أن $0 < q < 1$ ، وأن $I(\cdot)$ دالة مؤشر (Indicator function). باستخدام دالة الخسارة، فإن الدالة الربيعية الشرطية الخطية (linear conditional Quantile function) توسع رُبعي العينة $q(\varepsilon(q))$ إلى الانحدار بنفس الطريقة التي توسع بها دالة الوسط الحسابي الشرطية الخطية الوسط الحسابي للعينة.

من المعروف أن طريقة المربعات الصغرى الاعتيادية تعمل على إيجاد تقدير لدالة الوسط الحسابي الشرطية الخطية $E(Y|X = x) = x'\beta$ وذلك بإيجاد حل للمعادلة الآتية:

$$\hat{\beta} = \arg \min_{\beta \in R^k} \sum_{i=1}^n (y_i - x_i'\beta)^2 \dots (11)$$

إن تقدير المعلمات $\hat{\beta}$ يجعل من مجموع مربعات البقايا في نهايته الصغرى بنفس الطريقة التي يجعل الوسط الحسابي للعينة $\hat{\mu}$ من مجموع مربعات البقايا في نهايته الصغرى.

$$\hat{\mu} = \arg \min_{\mu \in R} \sum_{i=1}^n (y_i - \mu)^2 \dots (12)$$

بنفس الطريقة، فإن الانحدار الربيعي يقدر الدالة الربيعية الشرطية الخطية، $F(q / X = x) = x'\beta_q$ ، وذلك بإيجاد حل للمعادلة الآتية:

$$\hat{\beta}_q = \arg \min_{\beta \in R^k} \sum_{i=1}^n \rho_q(y_i - x_i'\beta) = \arg \min_{\beta} \sum_{i=1}^n \rho_q(e_i) \dots (13)$$

لأي رُبعي $q \in (0,1)$. وبناء على الصيغة (10)، وبما أن

$$\rho_q(y_i - x_i'\beta) = \rho_q(e_i) \dots (14)$$

إذن يمكن كتابة الصيغة (14) بالاستناد إلى الصيغة (10)، كالآتي:

$$= \rho_q(u_i - v_i) \dots (15)$$

$$= qu_i + (q-1)(-v_i) \dots (16)$$

$$= qu_i + (1-q)(v_i) \dots (17)$$

عليه فإنه يمكن كتابة المعادلة (13)، كالآتي:

$$\sum_{i=1}^n \rho_q(y_i - x_i'\beta) = \sum_{i=1}^n [qu_i + (1-q)(v_i)] \dots (18)$$

وفقاً للقيود الآتية:

$$u_i \geq 0, \quad v_i \geq 0, \quad y_i - x_i'\beta = u_i - v_i$$

عليه فإن أفضل تقدير لـ β هو

$$\hat{\beta}(q) = \arg \min_{\beta} \sum_{i=1}^n [qu_i + (1 - q)(v_i)] \dots (19)$$

ان المقدار $\hat{\beta}(q)$ يطلق عليه برئبي الانحدار (regression quantile) q وهو افضل تقدير لـ β . وعندما $q = 0.5$ ، والتي تجعل من مجموع البواقي المطلقة اقل ما يمكن، فإنها تقابل الانحدار الوسيط وتعرف ايضا بانحدار L_1 . ان مجموعة ربيعات الانحدار $\{\hat{\beta}(q): q \in (0,1)\}$ يطلق عليها بعمليات الرئبيعي (quantile process).

تعد المعادلة (19) مسألة برمجة خطية (Linear programming)، ولغرض ايجاد الحل لها يتم استخدام احد الخوارزميات الاتية: الطريقة المبسطة (Simplex method)، خوارزمية النقطة الداخلية (Interior point algorithm) والخوارزمية الممهدة (Smoothing algorithm).

2.3 التقدير Estimation

1.2.4 الطريقة المبسطة (Simplex method) (Chen and Wei, 2005):

لتكن $\mu = [y - A'\beta]_+$ ، $v = [A'\beta - y]_+$ ، $\phi = [\beta]_+$ ، $\varphi = [-\beta]_+$ ، حيث أن $[z]_+$ هي الجزء الموجب من z .

ليكن $D_{LAR}(\beta) = \sum_{i=1}^n |y_i - x'_i \beta|$ ، بالنسبة لمسألة L_1 ، فان الطريقة المبسطة تعمل على ايجاد $\min_{\beta} D_{LAR}(\beta)$ باعادة صياغتها كمسألة صغرى مقيدة (constrained minimization problem) وكالاتي:

$$\min_{\beta} \{e'\mu + e'V|y = A'\beta + \mu - V, \{\mu, V\} \in R_+^n\} \dots (20)$$

إذ أن، e تمثل موجه عمودي $(n \times 1)$ جميع قيمه الواحد.

لتكن $B = [A' - A'I - I]$ ، $\theta = (\phi' \varphi' \mu' V')'$ ، $d = (0' 0' e' e')'$ ، إذ أن $0' = (0 \ 0 \dots 0)_p$. ان اعادة الصياغة كمسألة برمجة خطية هو كالاتي:

$$(p) \min_{\theta} d' \theta$$

$$\text{subject to } B\theta = y$$

$$\theta \geq 0$$

هذه المسألة لها صيغة مقابلة (dual) وهي كالاتي:

$$(D) \max_{d} y' z$$

$$\text{subject to } B'z \leq d$$

والتي يمكن تبسيطها كالاتي:

$$\max_z y' z : \text{subject to } Az = 0, z \in [-1,1]^n \dots (21)$$

وبوضع $\eta = \frac{1}{2}z + \frac{1}{2}e$ ، $b = \frac{1}{2}Ae$ ، فان المسألة ستكون كالاتي:

$$\max_{\eta} y' \eta : \text{subject to } A\eta = b, \eta \in [0,1]^n \dots (22)$$

فيما يتعلق بالانحدار الرئبيعي، فانه لجعل المسألة اقل ما يمكن هو $\min_{\beta} \sum_{i=1}^n \rho_q(y_i - x'_i \beta)$ ، وبنفس مجموعة الخطوات اعلاه، فإنها تؤدي الى الصيغة المقابلة (dual formulation)، وكالاتي:

$$\max_z y' z : \text{subject to } Az = (1 - \tau)Ae, z \in [0,1]^n \dots (23)$$

2.2.4 خوارزمية النقطة الداخلية (Interior Point Algorithm):

تستخدم خوارزمية النقطة الداخلية لإيجاد حل لمسألة برمجة خطية كبيرة الى كبيرة جدا. وأشار (Chen and Wei 2005) بان أداء هذه الخوارزمية في اسوأ احوالها هو افضل من الخوارزمية المبسطة. هناك العديد من خوارزميات النقطة الداخلية، الا ان الخوارزمية الاكثر شيوعا والمستخدمة في الانحدار الربيعي هي خوارزمية التنبؤ المقابلة المصححة الأولية primal-dual predictor-corrector، وهي كالاتي (Chen SAS Institute, 2008: 5380; and Wei, 2005):

لتكن $b = (1 - q)Ae$ ، $c = -y$ ، ولتكن u هي الحد الأعلى العام (general upper bound). وان المسألة الخطية المراد حلها هي الاتي:

$$\begin{aligned} & \max\{c'z\} \\ & \text{subject to } Az = b \\ & 0 \leq z \leq u \end{aligned}$$

نفرض لدينا n من المتغيرات، وليكن المؤشر i يشير الى رقم المتغير، و k يشير الى رقم التكرار. لتكن v مكتملة اولية (primal slack)، عليه فان $z + v = u$ يقترن بالمتغيرات المقابلة w مع هذه القيود. تعمل خوارزمية النقطة الداخلية على حل منظومة المعادلات لتحقيق شروط الامثلية لكاروش-كوهن-توكر (Karush-Kuhn-Tucker (KKT)):

$$\begin{aligned} Az &= b \\ z + v &= u \\ A't + s - w &= c \\ ZSe &= 0 \\ VWe &= 0 \\ z, s, v, w &\geq 0 \end{aligned}$$

إذ أن

$$\begin{aligned} W &= \text{diag}(w) \\ & \text{(إذا كان } i = j \text{ فان } W_{i,j} = w_i \text{، وبخلافه فان } W_{i,j} = 0) \\ S &= \text{diag}(s), Z = \text{diag}(z), V = \text{diag}(v) \end{aligned}$$

تعد هذه هي شروط جدوى (feasibility)، مع إضافة شروط التكاملية (conditions complementarity) $ZSe = 0$ و $VWe = 0$ ، التي يجب أن تظهر في الحل الأمثل. تعمل الشروط التكاملية على أن تتساوى الأهداف المثلى للنماذج الأولية مع النماذج المقابلة، $C'z_{opt} = b't_{opt} - u'w_{opt}$ ، وكالاتي:

$$\begin{aligned} 0 &= v'_{opt}w_{opt} = (u - z_{opt})'w_{opt} = u'w_{opt} - z'_{opt}w_{opt} \dots (24) \\ 0 &= z'_{opt}s_{opt} = s'_{opt}z_{opt} = (c - A't_{opt} + w_{opt})'z_{opt} = \\ & c'z_{opt} - t'_{opt}(Az_{opt}) + w'_{opt}z_{opt} = c'z_{opt} - b't_{opt} + u'w_{opt} \dots (25) \end{aligned}$$

عليه

$$0 = c'z_{opt} - b't_{opt} + u'w_{opt} \dots (26)$$

ان الفجوة الثنوية ($duality\ gap$)، $c'z - b't + u'w$ تستخدم لقياس تقارب الخوارزمية.

قبل الوصول الى الحل الامثل، فانه يمكن أن يكون هناك اخلاخل بشروط KKT للحل (z, t, s, v, w) ، وكالاتي:

- قد يكون هناك اخلاخل بقيود الحد الأولي (Primal bound constraints)، $\delta_0 = u - z - v \neq 0$.
- قد يكون هناك اخلاخل بالقيود الأولية (Primal constraints)، اي ان، $\delta_c = b - Az \neq 0$.
- قد يكون هناك اخلاخل بالقيود المقابلة (Dual constraints)، اي ان، $\delta_d = c - A't - s + w \neq 0$.
- قد لا تتحقق شروط التكاملية (Complementarity conditions)، اي ان، $z't \neq 0$ و $v'w \neq 0$.

للعثور على الاتجاه $(\Delta z^k, \Delta t^k, \Delta s^k, \Delta v^k, \Delta w^k)$ ، يتم استخدام طريقة نيوتن (Newton method) للانتقال من

الحل الحالي $(z^k, t^k, s^k, v^k, w^k)$ نحو حل افضل:

$$(z^{k+1}, t^{k+1}, s^{k+1}, v^{k+1}, w^{k+1}) = (z^k, t^k, s^k, v^k, w^k) + k(\Delta z^k, \Delta t^k, \Delta s^k, \Delta v^k, \Delta w^k) \dots (27)$$

إذ أن k تمثل طول الخطوة.

تستخدم طريقة نيوتن لإيجاد الاتجاه $(\Delta z_{aff}^k, \Delta t_{aff}^k, \Delta s_{aff}^k, \Delta v_{aff}^k, \Delta w_{aff}^k)$ الذي بموجبه يتم التحرك. تعرف

هذه الخطوة باسم الخطوة أفيني (affine step).

في التكرار k ، فان منظومة الخطوة أفيني (affine step system) التي يتطلب ايجاد الحل لها هي:

$$\Delta z_{aff} + \Delta v_{aff} = \delta_b$$

$$A\Delta z_{aff} = \delta_c$$

$$A'\Delta t_{aff} + \Delta s_{aff} - \Delta w_{aff} = \delta_d$$

$$S\Delta z_{aff} + Z\Delta s_{aff} = -ZSe$$

$$V\Delta w_{aff} + W\Delta z_{aff} = -VWe$$

عليه، فإن الحسابات المعنية في ايجاد حل الخطوة أفيني هي

$$\theta = SZ^{-1} + WV^{-1}$$

$$\rho = \theta^{-1}(\delta_d + (S - W)e - V^{-1}W\delta_b)$$

$$\Delta t_{aff} = (A\theta^{-1}A')^{-1}(\delta_c + A\rho)$$

$$\Delta z_{aff} = \theta^{-1}A'\Delta t_{aff} - \rho$$

$$\Delta v_{aff} = \delta_b - \Delta z_{aff}$$

$$\Delta w_{aff} = -We - V^{-1}W\Delta z_{aff}$$

$$\Delta s_{aff} = -Se - Z^{-1}\Delta z_{aff}$$

$$(z_{aff}, t_{aff}, s_{aff}, v_{aff}, w_{aff})$$

$$= (z, t, s, v, w) + k(\Delta z_{aff}, \Delta t_{aff}, \Delta s_{aff}, \Delta v_{aff}, \Delta w_{aff}) \dots (28)$$

تمثل k طول الخطوة كما تمت الاشارة اليه سابقا.

يقاس نجاح الخطوة أفيني عن طريق حساب التكاملية لـ $z's$ و $v'w$ عند $(z_{aff}^k, t_{aff}^k, s_{aff}^k, v_{aff}^k, w_{aff}^k)$ ومقارنتها مع التكاملية عند نقطة البداية $(z^k, t^k, s^k, v^k, w^k)$. فإذا كانت الخطوة أفيني ناجحة في الحد من التكاملية بمقدار كبير، فإن الحاجة إلى التمرکز (centering) ليست كبيرة، ويتم تعيين قيمة قريبة من الصفر لـ σ في المنظومة الخطية الثانية، والتي تستخدم لتحديد موجه مركزي (centering vector). ولكن، إذا كانت خطوة أفيني غير ناجحة، فإن المركزي يعتبر مفيدا، ويتم تعيين قيمة قريبة من 1.0 إلى σ . بعبارة أخرى، يتم تبديل قيمة σ اعتمادا على التقدم المحرز نحو الأمثلية.

يتم ايجاد الحل للمنظومة الخطية الاتية وذلك لتحديد الموجه المركزي $(\Delta z_c, \Delta t_c, \Delta s_c, \Delta v_c, \Delta w_c)$ من $(z_{aff}, t_{aff}, s_{aff}, v_{aff}, w_{aff})$.

$$\Delta z_c + \Delta v_c = 0$$

$$A\Delta z_c = 0$$

$$A'\Delta t_c + \Delta s_c - \Delta w_c = 0$$

$$S\Delta z_c + Z\Delta s_c = -Z_{aff}S_{aff}e + \sigma\mu e$$

$$V\Delta w_c + W\Delta v_c = -V_{aff}W_{aff}e + \sigma\mu e$$

إذ أن $\xi_{start} = z's + v'w$ هي التكاملية عند بداية التكرار.

وإن $\xi_{aff} = z'_{aff}s_{aff} + v'_{aff}w_{aff}$ هي افين التكاملية.

وإن $\mu = \xi_{aff}/2n$ ، هي معدل التكاملية.

وإن $\sigma = (\xi_{aff}/\xi_{start})^3$.

عليه، فإن الحسابات التي تتضمن ايجاد الحل لخطوة المركزي هي:

$$\rho = \theta^{-1}(\sigma\mu(Z^{-1} - V^{-1})e - Z^{-1}Z_{aff}S_{aff}e + V^{-1}V_{aff}W_{aff}e)$$

$$\Delta t_c = (A\theta^{-1}A')^{-1}A\rho$$

$$\Delta z_c = \theta^{-1}A'\Delta t_c - \rho$$

$$\Delta v_c = -\Delta z_c$$

$$\Delta w_c = \sigma\mu V^{-1}e - V^{-1}V_{aff}W_{aff}e - V^{-1}W_{aff}\Delta v_c$$

$$\Delta s_c = \sigma\mu Z^{-1}e - Z^{-1}Z_{aff}S_{aff}e - Z^{-1}S_{aff}\Delta z_c$$

وبالتالي فإن

$$(\Delta z, \Delta t, \Delta s, \Delta v, \Delta w)$$

$$= (\Delta z_{aff}, \Delta t_{aff}, \Delta s_{aff}, \Delta v_{aff}, \Delta w_{aff})$$

$$+ (\Delta z_c, \Delta t_c, \Delta s_c, \Delta v_c, \Delta w_c)(z^{k+1}, t^{k+1}, s^{k+1}, v^{k+1}, w^{k+1}) =$$

$$(z^k, t^k, s^k, v^k, w^k) + k(\Delta z, \Delta t, \Delta s, \Delta v, \Delta w) \dots (29)$$

إذ أن k هي طول الخطوة.

3.2.4 الخوارزمية الممهدة (Smoothing Algorithm):

استخدمت هذه الخوارزمية لإيجاد تقديرات الانحدار الربيعي من قبل كل من (Clark and Osborne, 1986) و (Madsen and Nielsen, 1993) لجعل مجموع البواقي المطلقة لـ $D_{LAR}(\beta)$ اقل مايمكن، فان الخوارزمية الممهدة تعمل على تقريب الدالة غير القابلة للاشتقاق (nondifferentiable) D_{LAR} بواسطة الدالة الممهدة الآتية (المزيد من التفاصيل، يمكن الرجوع الى (Chen and Wei, 2005):

$$D_{\gamma}(\beta) = \sum_{i=1}^n H_{\gamma}(r_i(\beta)) \dots (30)$$

إذ أن

$$H_{\gamma}(t) = \begin{cases} t^2/(2\gamma) & \text{if } |t| \leq \gamma \\ |t| - \gamma/2 & \text{if } |t| > \gamma \end{cases} \dots (31)$$

اذ ان $r_i(\beta) = y_i - x_i'\beta$ ، وأن العتبة (threshold) γ هي عدد حقيقي موجب (positive real number). ان الدالة D_{γ} هي دالة مستمرة وقابلة للاشتقاق (differentiable) وأن المصغر B_{γ} (minimizer) لـ D_{γ} هو قريب من المصغر $\bar{B}_{LAR}(\beta)$ لـ $\bar{B}_{LAR}(\beta)$ عندما γ تقترب من الصفر.

ان الفائدة من الخوارزمية الممهدة هي في امكانية ايجاد حل لـ $L_1(\bar{B}_{LAR})$ عندما تكون قيمة ($\gamma > 0$) صغيرة. بمعنى اخر، انه ليس من الضروري أن تقترب γ من الصفر لغرض ايجاد المصغر لـ $D_{LAR}(\beta)$. تنتهي الخوارزمية عندما لا يكون هناك اي تغيير في وضع البواقي $r_i(\beta)$ بموجب سلسلة من قيم γ التي يتم توليدها في الخوارزمية (Madsen and Nielsen, 1993).

ان الخوارزمية الممهدة تتوسع بشكل طبيعي من الانحدار L_1 الى الانحدار الربيعي العام (Chen, 2007). ان الدالة

$$D_{\rho_{\tau}}(\beta) = \sum_{i=1}^n \rho_{\tau}(y_i - x_i'\beta) \dots (32)$$

يمكن أن تكون متقاربة بواسطة الدالة الممهدة

$$D_{\gamma,\tau}(\beta) = \sum_{i=1}^n H_{\gamma,\tau}(r_i(\beta)) \dots (33)$$

إذ أن

$$H_{\gamma,\tau}(t) = \begin{cases} t(\tau - 1) - \frac{1}{2}(\tau - 1)^2\gamma & \text{if } t \leq (\tau - 1)\gamma \\ \frac{t^2}{2\gamma} & \text{if } (\tau - 1)\gamma \leq t \leq \tau\gamma \\ t\tau - \frac{1}{2}\tau^2\gamma & \text{if } t \geq \tau\gamma \end{cases} \dots (34)$$

ان الدالة $H_{\gamma,\tau}$ يتم تحديدها بواسطة $r_i(\beta) \leq (\tau - 1)\gamma$ ، $r_i(\beta) \geq \tau\gamma$ ، أو $(\tau - 1)\gamma \leq r_i(\beta) \leq \tau\gamma$. هذه المتباينات تقسم R^p الى مناطق فرعية بواسطة مستويات متوازية $r_i(\beta) = (\tau - 1)\gamma$ و $r_i(\beta) = \tau\gamma$. يشار الى مجموعة المستويات بواسطة $B_{\gamma,\tau}$:

$$B_{\gamma,\tau} = \{\beta \in R^p | \exists i: r_i(\beta) = (\tau - 1)\gamma \text{ or } r_i(\beta) = \tau\gamma\} \dots (35)$$

يتم تعريف موجه الاشارة (sign vector) $s_\gamma(\beta) = (s_1(\beta), \dots, s_n(\beta))'$ على أنه

$$s_i = s_i(\beta) = \begin{cases} -1 & \text{if } r_i(\beta) \leq (\tau - 1)\gamma \\ 0 & \text{if } (\tau - 1)\gamma \leq r_i(\beta) \leq \tau\gamma \\ 1 & \text{if } r_i(\beta) \geq \tau\gamma \end{cases}$$

ويختزل الى $w_i = w_i(\beta) = 1 - s_i^2(\beta)$

عليه، فان

$$H_{\gamma,\tau}(r_i(\beta)) = \frac{1}{2\gamma} w_i r_i^2(\beta) + s_i \left[\frac{1}{2} r_i(\beta) + \frac{1}{4} (1 - 2\tau)\gamma + s_i(r_i(\beta)) \left(\tau - \frac{1}{2} \right) \right] - \frac{1}{4} (1 - 2\tau + 2\tau^2)\gamma$$

يؤدي الى

$$D_{\gamma,\tau}(\beta) = \frac{1}{2\gamma} r' W_{\gamma,\tau} r + v'(s)r + c(s) \dots (36)$$

إذ أن $W_{\gamma,\tau}$ هي مصفوفة قطرية $n \times n$ بالعناصر القطرية $w_i(\beta)$ ،

$$v'(s) = (s_1((2\tau - 1)s_1) + 1)/2, \dots, s_n((2\tau - 1)s_n + 1)/2,$$

$$c(s) = \sum \left[\frac{1}{4} (1 - 2\tau)\gamma s_i - \frac{1}{4} s_i^2 (1 - 2\tau + 2\tau^2)\gamma \right] \dots (37)$$

و

$$r(\beta) = (r_1(\beta), \dots, r_n(\beta))' \dots (38)$$

التدرج (gradient) $D_{\gamma,\tau}$ تعطى بواسطة الاتي:

$$D_{\gamma,\tau}^{(1)}(\beta) = -A[W_{\gamma,\tau}(\beta)r(\beta) + v(s)] \dots (39)$$

ول $\beta \in \mathbb{R}^P \setminus \beta_{\gamma,\tau}$ ، فان هيسين (Hessian) معطى بواسطة الاتي:

$$D_{\gamma,\tau}^{(2)}(\beta) = \frac{1}{\gamma} A W_{\gamma,\tau}(\beta) A' \dots (40)$$

ان التدرج هو دالة مستمرة في \mathbb{R}^P ، بينما الهيسين هو ثابت القطع الحكيم (piecewise).

الموجه s يشير على أنه موجه الاشارة الممكنة (γ) *feasible sign vector* فيما إذا كان هناك $\beta \in \mathbb{R}^P \setminus \beta_{\gamma,\tau}$ وان $s_\gamma(\beta) = s$ ، وإذا كانت s هي الممكنة (*feasible*)، فان Q_s تعرف على أنها دالة تربيعية (*quadratic function*) $Q_\gamma(\alpha)$ والتي اشتقت من $D_{\gamma,\tau}(\beta)$ وذلك باستبدال s حيثما كان s_γ . عليه، لاي β حيث ان

$s_\gamma = s$ فان

$$Q_\gamma(\alpha) = \frac{1}{2} (\alpha - \beta)' D_{\gamma,\tau}^{(2)}(\beta) (\alpha - \beta) + D_{\gamma,\tau}^{(1)}(\beta) (\alpha - \beta) + D_{\gamma,\tau}(\beta) \dots (41)$$

في المجال

$$C_s = \{\alpha | s_\gamma(\alpha) = s\}$$

$$D_{\gamma,\tau}(\alpha) = Q_s(\alpha)$$

لكل $\gamma > 0$ و $\theta \in \mathbb{R}^P$ ، اذ يمكن أن يكون هناك واحد أو عدة درجات تربيعية مقابلة Q_s . وإذا كانت $\theta \in \beta_{\gamma,\tau}$ فان Q_s متميزة بواسطة θ و γ ، ولكن عندما $\theta \in \beta_{\gamma,\tau}$ فان الدرجة التربيعية تكون غير وحيدة (unique). لذلك، فان

المرجعية (reference)

$$(\gamma, \theta, s)$$

تحدد الدرجة التربيعية. لتكن (γ, θ, s) مرجعية ممكنة (feasible reference) إذا كانت s هي موجه اشارة ممكنة (γ) عندما $\theta \in C_s$ ، ولتكن (γ, θ, s) مرجعية حل (solution reference) إذا كانت ممكنة وأن θ تجعل من $D_{\gamma, \tau}$ اقل ما يمكن. ان الخوارزمية الممهدة تجعل من D_0 اقل ما يمكن بالاستناد الى ان $D_{\gamma, \tau}$ اقل ما يمكن لمجموعة متناقصة من γ . لأي قيمة جديدة لـ γ ، يتم استخدام المعلومات من الحل السابق. وعندما تكون γ صغيرة بما فيه الكفاية، فإنه يمكن العثور على حل بواسطة خوارزمية نيوتن-رافسون المعدلة (modified Newton-Raphson algorithm) وكالاتي:

أوجد الحل الأولي المرجعي $(\gamma, \beta_\gamma, s)$

كرر

انقاص γ

أوجد الحل الأولي المرجعي $(\gamma, \beta_\gamma, s)$

حتى تكون $\gamma = 0$

وان β_0 هي الحل.

عمليا يتم ايجاد الحل الأولي المرجعي (β_γ) بواسطة المربعات الصغرى الاعتيادية.

2.5 حساب حدود الثقة باستخدام طريقة اعادة المعاينة بووتستراب الهامشية لسلسلة ماركوف MCMB (Markov chain marginal bootstrap)

تعمل MCMB على ايجاد حلول لـ p من المعادلات الأحادية البعد (one-dimensional) بدلا من ايجاد الحل لمعادلة ذات البعد p لكل تكرار من تكرارات بووتستراب.

ليكن لدينا انموذج الانحدار الربيعي الاتي:

$$\hat{\beta}_q = x_i' \beta(q) \dots (42)$$

اذ ان $\beta(q) \in R^k$.

لغرض ايجاد حدود الثقة باستخدام MCMB نتبع الخوارزمية الاتية:

1. لتكن $\hat{\beta}$ تقدير لـ β .
2. احسب $e_i = y_i - x_i' \hat{\beta}(q)$. عرّف $z_i = x_i \psi_q(e_i) - \bar{z}$ اذا ان $\bar{z} = \sum_{i=1}^n x_i \psi_q(e_i) / n$ ، وان $\psi_q(e) = q - I(e < 0)$.

3. الخطوة 1: لتكن $\beta^0 = \hat{\beta}(q)$ وان $k = 1$.
4. الخطوة k: اسحب عينة بوونستراب من z_1, \dots, z_n للحصول على $z_1^{j,k}, \dots, z_n^{j,k}$ لكل عدد صحيح $j = 1, \dots, p$. اوجد الحل لـ $\beta_j^{(k)}$ من الصيغة الآتية:

$$\sum_{i=1}^n x_i \psi_q \left[y_i - \sum_{l < j} x_{i,l} \beta_l^{(k)} - \sum_{l > j} x_{i,l} \beta_l^{(k-1)} - x_{i,j} \beta_j^{(k)} \right] = \sum_{i=1}^n z_i^{j,k} \dots (43)$$

5. كرر الخطوة (3) الى K من التكرارات للحصول على سلسلة (MCMB) $\beta^{(k)}$ ، اذ ان $k = 1, \dots, K$. ومن ثم ايجاد تقدير تباين العينة $\hat{\beta}^{(k)}(q)$ ، اذ ان $k = 1, \dots, K$.

3. الجانب العملي:

3.1 مشكلة البحث:

تتضمن مشكلة البحث جانبين اثنين، أحدهما مشكلة صحية عامة يراد تسليط الضوء على بعض جوانبها، متمثلة بمرض ثلاسيميا الاطفال في الموصل. والثاني مشكلة احصائية تتمثل في عدم تجانس البيانات فضلا عن وجود قيم شاردة وذات قوة رافعة عدت على انها جوهرية ويعتمد عليها في التحليل.

3.2 محددات البحث:

في البحوث الاحصائية التي تتناول الامراض الحياتية ومنظومة العلاقات التي تؤثر وتتأثر بها، فانه لغرض الخروج باستنتاجات يعول عليها لابد وأن تتوفر قاعدة بيانات دقيقة فضلا عن اسلوب احصائي يناسب وطبيعة البيانات ومتسقا مع الهدف من البحث.

3.3 طبيعة المرض:

مرض الثلاسيميا (Thalassemia) هو مرض دم وراثي، ويعد من المشاكل الصحية العامة والخطيرة في المجتمع. ينتشر هذا المرض في منطقة البحر الابيض المتوسط، الشرق الأوسط وشبه القارة الهندية، فضلا عن جنوب شرق آسيا.

الثلاسيميا هو عيب جيني ينتج عن اختزال نسبة في تكوين أحد سلاسل بروتين الكلوبين الذي يكون الهيموغلوبين وهذا يسبب فقر الدم وعدة أمراض ذات العلاقة مثل اضطرابات عظمية مثل العاهات (تشوهات العظام)، ونخر في العظام التي تؤدي في نهاية المطاف الى الموت (Mohammad, et.al., 2010).

3.4 متغيرات البحث:

تم تطبيق الانحدار الربيعي على بيانات حقيقية (النعيمي، 2005) جمعت من مستشفى ابن الاثير التعليمي في الموصل (150) مريضا من المرضى مصابين بالثلاسيميا. يتضمن البحث 14 متغيرا توضيحيا وكما موضح ادناه (لمزيد من التفاصيل عن متغيرات البحث، أنظر النعيمي، 2005: 51).

Age of the Bone (Month)

y : العمر من العظم مقاساً (بالشهر)

Real Age (Month)

X₁ : العمر الحقيقي مقاساً (بالشهر)

X_2 : عمر المريض عند المرض مقاسا (بالشهر)
Age of patient at the onset of Disease (Month)

Enlargement of liven (c.m)

X_3 : تضخم الكبد مقاسا (بالسنتمتر)

Hemoglobin

X_4 : هيموكلوبين الدم

Packed cell volume

X_5 : مكداس الدم (خلايا الدم المضغوط)

Reticulocyte

X_6 : الخلايا الشبكية

Norm oblast

X_7 : ارومة حمراء

Fetal Hemoglobin

X_8 : الهيموكلوبين الجنيني

Numbers of Blood units

X_9 : عدد وحدات الدم

X_{10} : بداية نقل الدم حسب العمر مقاسا (بالشهر)
Onset of Blood Transform according to age (Month)

Sex: الجنس وهو متغير ثنائي يتضمن مستويان (0 الاناث و 1 يمثل الذكور).

Mang.: الشكل المنغولي وهو متغير ثنائي يتضمن مستويان (0 الشكل غير المنغولي و 1 يمثل الشكل المنغولي)

Mongoloid feature

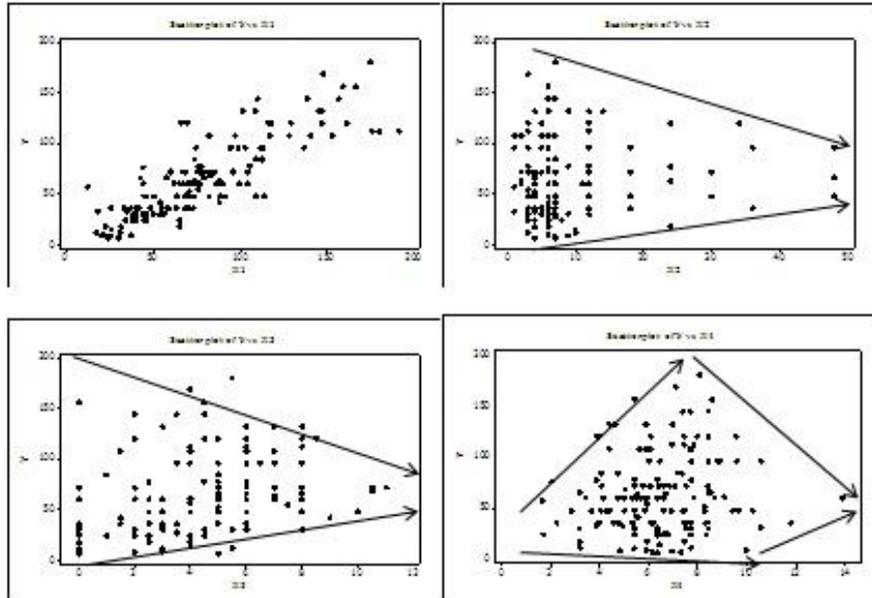
Height (c.m)

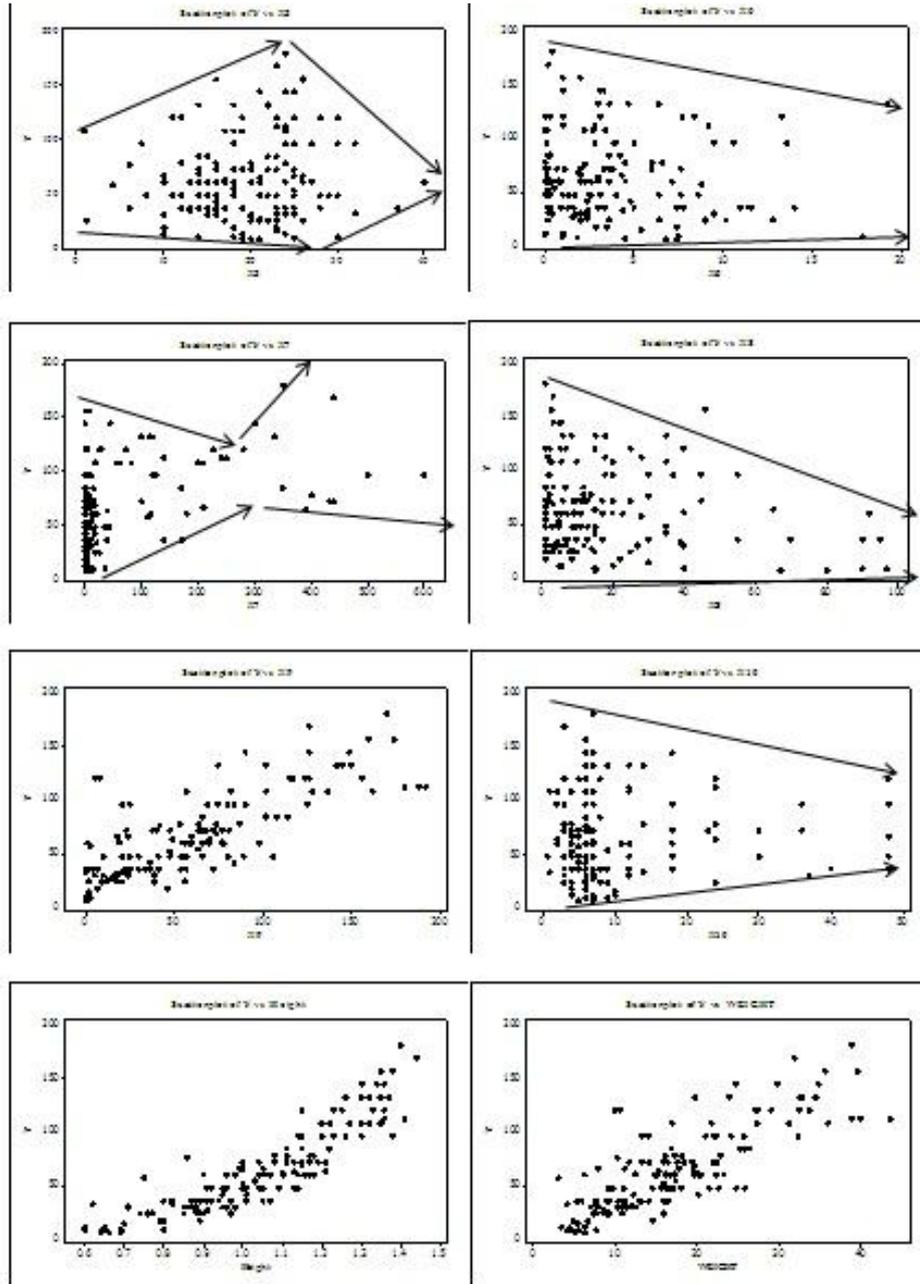
Height: الطول مقاسا (سنتمتر)

Weight (k.g)

Weight: الوزن مقاسا (كغم)

من خلال الشكل البياني (1) الذي يبين شكل انتشار متغيرات البحث، يلاحظ ان هناك انعطافات هامة في مسار هذه المتغيرات وخصوصا بالنسبة للمتغيرات ($X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_{10}$). اذ يلاحظ ان انتشار البيانات في الذبول يختلف مما هو عليه في موقع تركز البيانات وحسب ما مؤشر باتجاه السهم، مما يدل على عدم تجانس البيانات.





الشكل (1): يبين شكل انتشار متغير الاستجابة مقابل كل متغير من المتغيرات التوضيحية للبحث
المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي Minitab 16

3.5 نتائج البحث:

يلاحظ من الجدول (1) أن الوسط الحسابي والانحراف المعياري للمتغيرات (X_2, X_7, X_8, X_{10}) هو أكبر من المقاييس (الوسيط والانحراف المطلق الوسيط (MAD) (median absolute deviation)، مما قد يشير الى وجود قيم شاردة.

جدول (1) الوسط الحسابي، الوسيط، الانحرافات المعيارية والانحراف المطلق الوسيط والرُبيعات

Variable	Q1	Q3	Mean	Standard Deviation	Median	MAD
X_1	45.0000	99.0000	76.6499	39.215	70.7615	38.5477
X_2	4.0000	10.0000	9.0000	8.950	6.0000	2.9652
X_3	3.0000	6.0000	4.4433	2.469	4.2500	2.5946
X_4	5.3667	7.7333	6.4324	1.979	6.3333	1.9768
X_5	16.0000	24.0000	19.4547	6.367	20.0000	5.9304
X_6	0.7000	4.8000	3.5740	3.773	2.4500	3.0393
X_7	1.0000	44.0000	63.8200	121.100	8.5000	12.6021
X_8	4.2000	23.0000	17.8417	20.991	9.9500	10.4523
X_9	20.0000	80.0000	57.8527	47.177	54.0000	46.7020
X_{10}	5.0000	12.0000	10.2404	10.159	7.0000	4.4478
Sex	0.0000	1.0000	0.5733	0.4962	1.0000	0.0000
Height	0.8900	1.1800	1.0433	0.2036	1.0500	0.2224
Mang.	1.0000	1.0000	0.8333	0.3739	1.0000	0.0000
Weight	9.3167	22.0000	16.8749	9.0888	16.0084	9.3281
Y	36.0000	84.0000	63.7267	38.4634	60.0000	35.5825

يتبين من الشكل (1) عدم تجانس البيانات، ويبدو من الجدول (1) ان البيانات قد تحتوي على قيم شاردة. وبما ان الهدف من البحث هو دراسة تأثيرات المتغيرات التوضيحية على الرُبيعات الشرطية لمتغير الاستجابة. فقد تم استخدام الانحدار الرُبيعي في تقدير معلمات نموذج الانحدار، باعتباره من الطرائق الحصينة تجاه المشاهدات الشاردة فضلا عن انه يستخدم في حالة عدم تجانس التباين كما انه يحقق الهدف من البحث.

كما لاحظنا في الجانب النظري، فانه لغرض تقدير معلمات نموذج الانحدار، فان هناك ثلاث خوارزميات رئيسة للتقدير وهي الخوارزمية المبسطة، الخوارزمية الممهدة وخوارزمية النقطة الداخلية. وبما ان الهدف هو في الوصول الى أفضل تقدير لمعلمات الانحدار، فقد تم استخدام الخوارزميات الثلاث عند الرُبيعات من 0.1 الى 0.9. الجداول (2) الى (4) يبين عينة لنتائج تقدير معلمات الانحدار الرُبيعي بوساطة الخوارزميات الثلاث عند الرُبيع 0.5 والتي اظهرت انها أفضل تقدير وفقا لمعيار الخطأ المعياري لتقدير المعلمات فضلا عن المدرج الإحصائي للبوافي المعيارية، المنحنى الطبيعي والمنحنى اللبي.

يلاحظ من الجداول (2) الى (4) ان تقدير المعلمات التي تم الحصول عليها باستخدام الخوارزميات الثلاث عند الرُبيع 0.5 متطابقة وهذا يعني أن مشكلة الأمثلية للبيانات لها حل مثالي وحيد. لغرض تحديد الخوارزمية الافضل في تقدير المعلمات، فقد لوحظ أن عدد المعلمات باستخدام خوارزمية النقطة الداخلية والتي لها خطأ معياري اقل، هذا العدد هو اكبر مما هو عليه بالنسبة للخوارزميتين الأخرتين مما انعكس على قيم t المحسوبة وعدد المعلمات المعنوية فضلا عن مستوى المعنوية للعديد من المعلمات. عليه سيتم الاعتماد على تقدير المعلمات باستخدام خوارزمية النقطة الداخلية في تحليل البيانات.

جدول (2) يبين نتائج تقدير المعلمات باستخدام الانحدار الرُبيعي عند الرُبيع 0.5 بوساطة الخوارزمية المبسطة

Parameter	Level	DF	Estimate	Standard Error	95% Confidence Limits	t Value	Pr > t/
Intercept		1	-45.4774	16.8010	-78.7047 -12.250	-2.7100	0.0077
X1		1	1.9529	0.5720	0.8216 3.084	3.4100	0.0008

X2		1	-0.2991	0.2765	-0.8461	0.248	-1.0800	0.2813
X3		1	-0.1223	0.5932	-1.2954	1.051	-0.2100	0.8370
X4		1	-3.6493	5.9684	-15.4530	8.155	-0.6100	0.5419
X5		1	0.4831	1.1409	-1.7733	2.739	0.4200	0.6727
X6		1	0.2755	0.3474	-0.4115	0.963	0.7900	0.4291
X7		1	0.0425	0.0169	0.0090	0.076	2.5100	0.0134
X8		1	-0.1326	0.0781	-0.2870	0.022	-1.7000	0.0919
X9		1	0.2358	0.8570	-1.4591	1.931	0.2800	0.7836
X10		1	0.2074	0.2913	-0.3688	0.783	0.7100	0.4779
Sex	0	1	8.5866	2.6040	3.4366	13.737	3.3000	0.0012
Sex	1	0	0.0000	0.0000	0.0000	0.000	*	*
Mang.	0	1	7.3632	3.7630	-0.0789	14.805	1.9600	0.0524
Mang.	1	0	0.0000	0.0000	0.0000	0.000	*	*
Height		1	78.2375	26.8963	25.0448	131.430	2.9100	0.0042
Weight		1	-7.4929	5.7009	-18.7676	3.782	-1.3100	0.1910

جدول (3) يبين نتائج تقدير المعلمات باستخدام الانحدار الربيعي عند الربع 0.5 بواسطة الخوارزمية النقطية الداخلية

Parameter	Level	DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > t
Intercept		1	-	14.8408	-	-16.127	-3.0600	0.0026
			45.4774		74.8280			
X1		1	1.9529	0.4866	0.9907	2.915	4.0100	<.0001
X2		1	-0.2991	0.3068	-0.9059	0.308	-0.9800	0.3313
X3		1	-0.1223	0.5197	-1.1501	0.905	-0.2400	0.8143
X4		1	-3.6493	5.8569	-	7.934	-0.6200	0.5343
X5		1	0.4831	1.0942	-1.6809	2.647	0.4400	0.6596
X6		1	0.2755	0.2988	-0.3154	0.866	0.9200	0.3581
X7		1	0.0425	0.0175	0.0078	0.077	2.4200	0.0167
X8		1	-0.1326	0.0519	-0.2351	-0.030	-2.5600	0.0117
X9		1	0.2358	0.8248	-1.3954	1.867	0.2900	0.7754
X10		1	0.2074	0.2978	-0.3816	0.796	0.7000	0.4875
Sex	0	1	8.5866	2.1722	4.2906	12.883	3.9500	0.0001
Sex	1	0	0.0000	0.0000	0.0000	0.000	*	*
Mang.	0	1	7.3632	2.7414	1.9415	12.785	2.6900	0.0081
Mang.	1	0	0.0000	0.0000	0.0000	0.000	*	*
Height		1	78.2375	25.0665	28.6636	127.811	3.1200	0.0022
Weight		1	-7.4929	5.4896	-	3.364	-1.3600	0.1746
					18.3497			

جدول (4) يبين نتائج تقدير المعلمات باستخدام الانحدار الربيعي عند الربع 0.5 بواسطة الخوارزمية الممهدة

Parameter	Level	DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > t
Intercept		1	-	15.9553	-	-13.923	-2.8500	0.0051
			45.4774		77.0322			
X1		1	1.9529	0.5874	0.7912	3.115	3.3200	0.0011
X2		1	-0.2991	0.2826	-0.8580	0.260	-1.0600	0.2917
X3		1	-0.1223	0.5210	-1.1527	0.908	-0.2300	0.8148
X4		1	-3.6493	5.9383	-	8.095	-0.6100	0.5399
					15.3935			

X5		1	0.4831	1.1080	-1.7083	2.674	0.4400	0.6636
X6		1	0.2755	0.3213	-0.3599	0.911	0.8600	0.3927
X7		1	0.0425	0.0199	0.0031	0.082	2.1300	0.0348
X8		1	-0.1326	0.0790	-0.2888	0.024	-1.6800	0.0956
X9		1	0.2358	0.8067	-1.3596	1.831	0.2900	0.7705
X10		1	0.2074	0.2630	-0.3127	0.728	0.7900	0.4318
Sex	0	1	8.5866	2.3611	3.9171	13.256	3.6400	0.0004
Sex	1	0	0.0000	0.0000	0.0000	0.000	*	*
Mang.	0	1	7.3632	3.9673	-0.4830	15.209	1.8600	0.0656
Mang.	1	0	0.0000	0.0000	0.0000	0.000	*	*
Height		1	78.2375	25.7629	27.2864	129.189	3.0400	0.0029
Weight		1	-7.4929	5.4457	-	3.277	-1.3800	0.1711
					18.2627			

بعد ان تم تحديد الخوارزمية الافضل من بين الخوارزميات الثلاث، الجداول من 5 الى 8 تبين تقدير معاملات متغيرات البحث، الانحراف المعياري، قيم t ، $|t| > Pr$ عند مستويات مختلفة من الرُبيعات باستخدام خوارزمية النقطة الداخلية، وذلك لبيان كيفية تعيين قيمة الرُبيع.

جدول (5) يبين تقدير معاملات متغيرات البحث عند مستويات مختلفة من الرُبيعات باستخدام خوارزمية النقطة الداخلية

	Quantile =0.1	Quantile =0.2	Quantile =0.3	Quantile =0.4	Quantile =0.5	Quantile =0.6	Quantile =0.7	Quantile =0.8	Quantile 9=0.
Intercept	-61.3791	-56.9850	-62.5649	-52.9311	-45.4774	-38.1981	-31.7627	-42.1870	-27.0881
X1	0.6732	0.9318	1.3057	1.5747	1.9529	2.1457	2.3868	2.4079	3.2231
X2	0.0850	0.1684	0.2561	0.1385	-0.2991	-0.4165	-0.5732	-0.5155	-0.2936
X3	0.6316	0.6272	-0.2504	0.0225	-0.1223	0.2128	0.1806	0.5413	0.5720
X4	-0.1783	-10.6984	-6.7921	-8.8797	-3.6493	-4.2009	-3.5260	-0.9419	-0.1546
X5	-0.1568	1.8436	1.1821	1.7007	0.4831	0.3962	0.2552	-0.0683	-0.3705
X6	0.8976	0.2096	0.2290	-0.0312	0.2755	0.2552	0.3025	0.2970	0.4205
X7	0.0349	0.0328	0.0161	0.0217	0.0425	0.0376	0.0319	0.0207	0.0072
X8	-0.1283	-0.1018	-0.0744	-0.0779	-0.1326	-0.0222	0.0031	-0.0263	-0.0421
X9	-0.0768	-0.9672	-0.5474	-0.4697	0.2358	0.1566	0.2642	0.7701	1.0194
X10	-0.0312	-0.1737	-0.3572	-0.3118	0.2074	0.1424	0.3251	0.3164	0.1455
SEX 0	8.9891	7.4556	9.3339	8.7113	8.5866	9.5943	6.8816	1.0000	1.0000
SEX 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MANG. 0	1.4963	7.6632	4.7896	5.7229	7.3632	5.9585	4.9664	1.0000	1.0000
MANG. 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Height	74.8791	69.9097	82.8374	74.1782	78.2375	66.4876	59.4800	80.4736	67.3231
Weight	-1.3856	2.6512	-1.1025	-2.1796	-7.4929	-7.3210	-8.6436	-11.6875	-15.9849

جدول (6) يبين تقدير الانحراف المعياري لمعاملات متغيرات البحث عند مستويات مختلفة من الرُبيعات باستخدام خوارزمية النقطة الداخلية

	Quantile =0.1	Quantile =0.2	Quantile =0.3	Quantile =0.4	Quantile =0.5	Quantile =0.6	Quantile =0.7	Quantile =0.8	Quantile 9=0.
Intercept	21.6627	17.6408	17.2953	15.5323	14.8408	15.2278	14.4482	15.0460	28.2254
X1	0.5535	0.4848	0.5224	0.5457	0.4866	0.5351	0.4922	0.4759	0.8799
X2	0.3690	0.3436	0.3333	0.3482	0.3068	0.2587	0.2440	0.2369	0.4107
X3	0.9146	0.6172	0.5492	0.5180	0.5197	0.5997	0.5612	0.7286	1.0426
X4	8.4969	7.2726	6.5060	6.2817	5.8569	5.4882	4.3304	5.0660	8.8005
X5	1.9427	1.5566	1.3054	1.2324	1.0942	1.0198	0.8274	0.9144	1.8464
X6	0.7392	0.5421	0.4174	0.3725	0.2988	0.3779	0.3184	0.4178	0.6413
X7	0.0199	0.0187	0.0172	0.0191	0.0175	0.0205	0.0191	0.0178	0.0333
X8	0.1036	0.0941	0.0755	0.0798	0.0519	0.0708	0.0555	0.0692	0.1344
X9	0.8253	0.7655	0.8098	0.8825	0.8248	0.7573	0.5483	0.6674	1.0416

<i>X10</i>	0.2915	0.2982	0.2966	0.3042	0.2978	0.2337	0.2307	0.2248	0.3579
<i>SEX 0</i>	3.7309	2.8333	2.6101	2.5509	2.1722	2.4398	2.2586	7.2044	4.6103
<i>SEX 1</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
<i>MANG. 0</i>	6.0263	4.5897	4.1588	3.6215	2.7414	3.3846	3.4523	7.1739	7.0606
<i>MANG. 1</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
<i>Height</i>	32.2105	27.3827	26.5013	24.0450	25.0665	22.5069	21.9268	23.4292	40.9299
<i>Weight</i>	5.4266	4.8372	5.3912	5.9854	5.4896	4.8630	3.5736	4.3258	7.9950

جدول (7) يبين تقدير قيم لمعاملات متغيرات البحث عند مستويات مختلفة من الرُّبوعات باستخدام خوارزمية النقطة

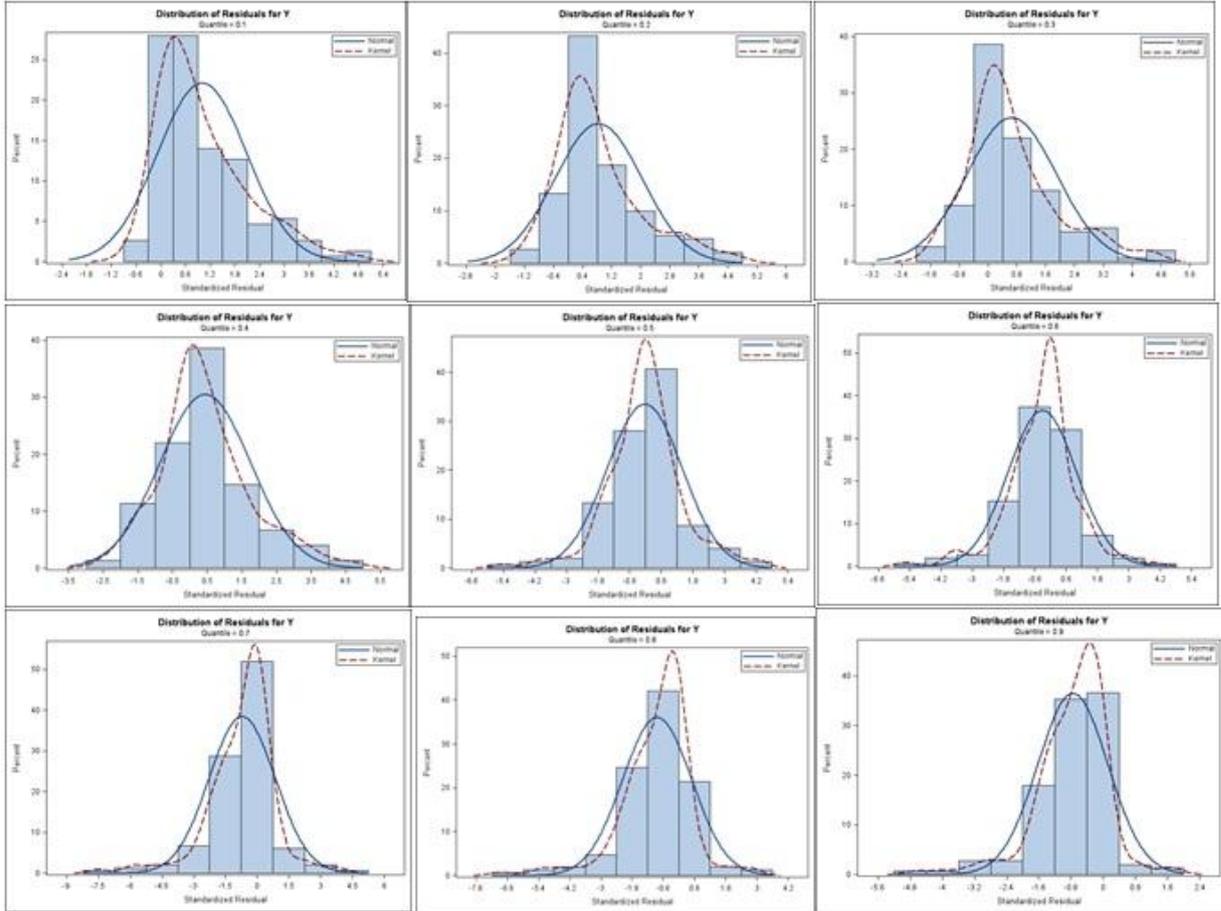
الداخلية

	Quantile =0.1	Quantile =0.2	Quantile =0.3	Quantile =0.4	Quantile =0.5	Quantile =0.6	Quantile =0.7	Quantile =0.8	Quantile 9=0.
Intercept	-2.8300	-3.23	-3.62	-3.4100	-3.0600	-2.51	-2.20	-2.80	-0.96
<i>X1</i>	1.2200	1.92	2.50	2.8900	4.0100	4.01	4.85	5.06	3.66
<i>X2</i>	0.2300	0.49	0.77	0.4000	-0.9800	-1.61	-2.35	-2.18	-0.71
<i>X3</i>	0.6900	1.02	-0.46	0.0400	-0.2400	0.35	0.32	0.74	0.55
<i>X4</i>	-0.0200	-1.47	-1.04	-1.4100	-0.6200	-0.77	-0.81	-0.19	-0.02
<i>X5</i>	-0.0800	1.18	0.91	1.3800	0.4400	0.39	0.31	-0.07	-0.20
<i>X6</i>	1.2100	0.39	0.55	-0.0800	0.9200	0.68	0.95	0.71	0.66
<i>X7</i>	1.7500	1.75	0.94	1.1400	2.4200	1.84	1.67	1.17	0.22
<i>X8</i>	-1.2400	-1.08	-0.99	-0.9800	-2.5600	-0.31	0.06	-0.38	-0.31
<i>X9</i>	-0.0900	-1.26	-0.68	-0.5300	0.2900	0.21	0.48	1.15	0.98
<i>X10</i>	-0.1100	-0.58	-1.20	-1.0200	0.7000	0.61	1.41	1.41	0.41
<i>SEX 0</i>	2.4100	2.63	3.58	13.7563	3.9500	3.93	3.05	2.91	0.91
<i>SEX 1</i>	*	*	*	0.0000	*	*	*	*	*
<i>MANG. 0</i>	0.2500	1.67	1.15	1.5800	2.6900	1.76	1.44	1.22	0.61
<i>MANG. 1</i>	*	*	*	*	*	*	*	*	*
<i>Height</i>	2.3200	2.55	3.13	3.0800	3.1200	2.95	2.71	3.43	1.64
<i>Weight</i>	-0.2600	0.55	-0.20	-0.3600	-1.3600	-1.51	-2.42	-2.70	-2.00

جدول (8) يبين تقدير قيم $Pr > |t|$ لمعاملات متغيرات البحث عند مستويات مختلفة من الرُّبوعات

	Quantile =0.1	Quantile =0.2	Quantile =0.3	Quantile =0.4	Quantile =0.5	Quantile =0.6	Quantile =0.7	Quantile =0.8	Quantile 9=0.
Intercept	0.0053	0.0016	0.0004	0.0009	0.0026	0.01330	0.02960	0.00580	0.3389
<i>X1</i>	0.2260	0.0567	0.0136	0.0046	<.0001	0.0001	0.0001	0.0001	0.0004
<i>X2</i>	0.8183	0.6249	0.4435	0.6915	0.3313	0.10980	0.02020	0.03130	0.4759
<i>X3</i>	0.4910	0.3113	0.6491	0.9654	0.8143	0.72330	0.74810	0.45880	0.5842
<i>X4</i>	0.9833	0.1436	0.2984	0.1598	0.5343	0.44530	0.41690	0.85280	0.9860
<i>X5</i>	0.9358	0.2383	0.3668	0.1699	0.6596	0.69830	0.75820	0.94060	0.8413
<i>X6</i>	0.2268	0.6997	0.5842	0.9334	0.3581	0.50070	0.34370	0.47850	0.5131
<i>X7</i>	0.0820	0.0825	0.3506	0.2583	0.0167	0.06850	0.09780	0.24510	0.8294
<i>X8</i>	0.2176	0.2813	0.3262	0.3307	0.0117	0.75410	0.95510	0.70470	0.7544
<i>X9</i>	0.9260	0.2086	0.5002	0.5955	0.7754	0.83650	0.63060	0.25060	0.3295
<i>X10</i>	0.9150	0.5612	0.2306	0.3072	0.4875	0.54340	0.16110	0.16150	0.6850
<i>SEX 0</i>	0.0173	0.0095	0.0005	3.4100	0.0001	0.00010	0.00280	0.00430	0.3652
<i>SEX 1</i>	*	*	*	*	*	*	*	*	*
<i>MANG. 0</i>	0.8043	0.0973	0.2515	0.1164	0.0081	0.08060	0.15260	0.22440	0.5433
<i>MANG. 1</i>	*	*	*	*	*	*	*	*	*
<i>Height</i>	0.0216	0.0118	0.0022	0.0025	0.0022	0.00370	0.00750	0.00080	0.1023
<i>Weight</i>	0.7988	0.5845	0.8383	0.7163	0.1746	0.13450	0.01690	0.00780	0.0476

يلاحظ من الجداول (5) الى (8)، انه لا يمكن تعيين قيمة الرُبيع الذي يؤدي الى الوصول الى أفضل التقديرات بالاعتماد على الخطأ معياري، اذ يلاحظ عدم ارجحية اي رُبيع على اخر. عليه، فقد تم اللجوء الى المدرج الإحصائي للبقاقي المعيارية فضلا عن المنحنى الطبيعي والمنحنى اللبي، وكما موضح ذلك في الشكل (2) والذي يشير بأن الانحدار الرُبيعي قد وفق البيانات بصورة جيدة عندما $q = 0.5$. عليه، سيتم تفسير المعلمات وفقا لخوارزمية النقطة الداخلية عندما $q = 0.5$.



الشكل (2) يوضح المدرج الإحصائي للبقاقي المعيارية فضلا عن المنحنى الطبيعي والمنحنى اللبي

المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي SAS 9.2

من الجدول (3) يلاحظ أن معلمات المتغيرات ($X_1, X_7, X_8, Sex, Mang., length$) معنوية، بينما بقية المتغيرات غير معنوية. ان معاملات الانحدار تشير الى تأثير المتغير التوضيحي على العمر من العظم بوحدة واحدة وذلك بفرض أن المتغيرات الاخرى ثابتة. وعليه، فان زيادة المتغيرات بمقدار ($X_1, X_5, X_6, X_7, X_9, X_{10}, sex, Mang., Height$) بوحدة واحدة يؤدي الى زيادة العمر من العظم بالترتيب. ويلاحظ أن في مقدمة المتغيرات المؤثرة على العمر من العظم هو الطول، إذ أن معامل الانحدار لهذا المتغير هو 78.2375 وقد اظهر معنوية عالية عند مستوى معنوية 0.0022. وفي المرتبة الثانية والثالثة جاء المتغيران ($Sex, Mang$) على الترتيب، وكل منهما يتضمن مستويان وهما (0, 1)، عليه فان النموذج يبين معلمة الذكور نسبة للإناث، ومعلمة الشكل غير المنغولي نسبة للشكل المنغولي (لمتغير الشكل المنغولي). عليه فان معلمة الجنس ذا

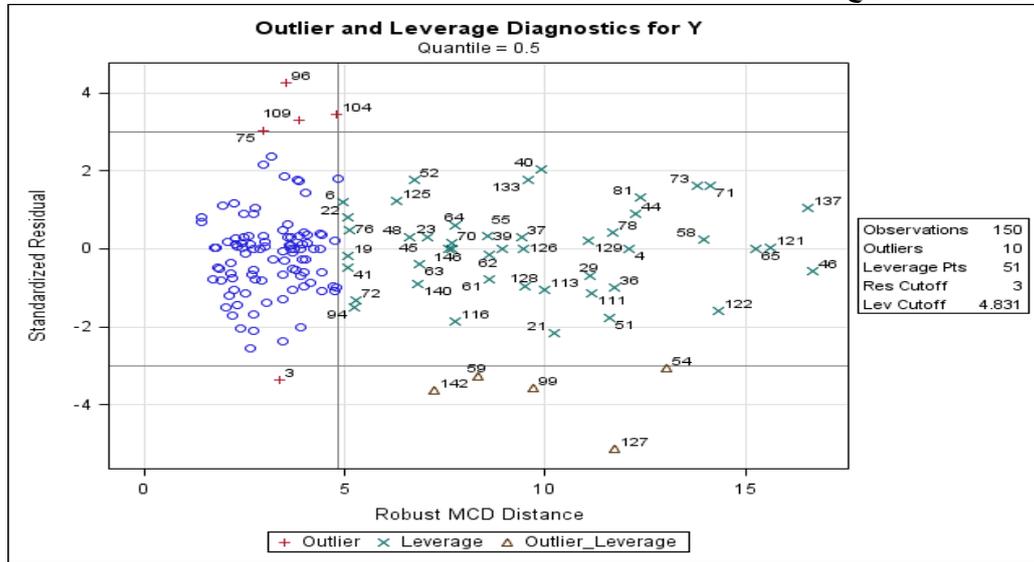
المستوى 0 (الاناث) معنوية عند مستوى المعنوية 0.0001 وأنها أكبر بمقدار 8.5866 من الذكور (التي تعد المجموعة المرجعية). كذلك، فإن معلمة الشكل المنغولي ذا المستوى غير المنغولي معنوية أيضا عند مستوى المعنوية 0.0081 وأنها أكبر بمقدار 7.3632 من الشكل المنغولي (التي تعد المجموعة المرجعية). احتل المتغير X_1 المرتبة الرابعة من حيث الاهمية في التأثير على y ، وبمعامل انحدار قدره 1.9529 واطهر معنوية عالية ايضا عند مستوى المعنوية <0.0001 . وجاء المتغير X_7 في المرتبة الخامسة في التأثير على المتغير y ، وبمعامل انحدار قدره 0.0425 وهو معنوي بمستوى معنوية قدره 0.0167. اما بالنسبة للمتغيرات (X_5, X_6, X_9, X_{10}) فكانت غير معنوية في التأثير على المتغير y .

من الجدول (3) يلاحظ أن المتغيرات ($X_2, X_3, X_4, X_8, Weight$) لها اشارات سالبة، وهذا يعني أن زيادة وحدة واحدة في هذه المتغيرات يؤدي الى نقصان العمر من العظم بمقدار (0.2991, 0.1223, 3.6493, 0.1326, 7.4929) على الترتيب. ان في مقدمة هذه المتغيرات هو المتغير X_8 بمعامل انحدار قدره 0.1326 وقد اظهر معنوية عالية بمستوى معنوية قدره 0.0117، اما بالنسبة للمتغيرات ($X_2, X_3, X_4, Weight$) فكانت غير معنوية في التأثير على المتغير y .

يبين الشكل (3) البواقي المعيارية مقابل المسافة الحصينة، محدد التباين المشترك الادنى (minimum covariance determinant) (MCD)، والذي من خلاله يتم الكشف عن القيم الشاردة في متغير الاستجابة (العمودية) والقيم ذات القوة الرافعة في المتغيرات التوضيحية (الافقية).

يبين الشكل (4) المسافة الحصينة MCD مقابل مسافة مهلائبس والذي من خلاله يتم الكشف عن القيم الشاردة القيم ذات القوة الرافعة في المتغيرات التوضيحية (الافقية).

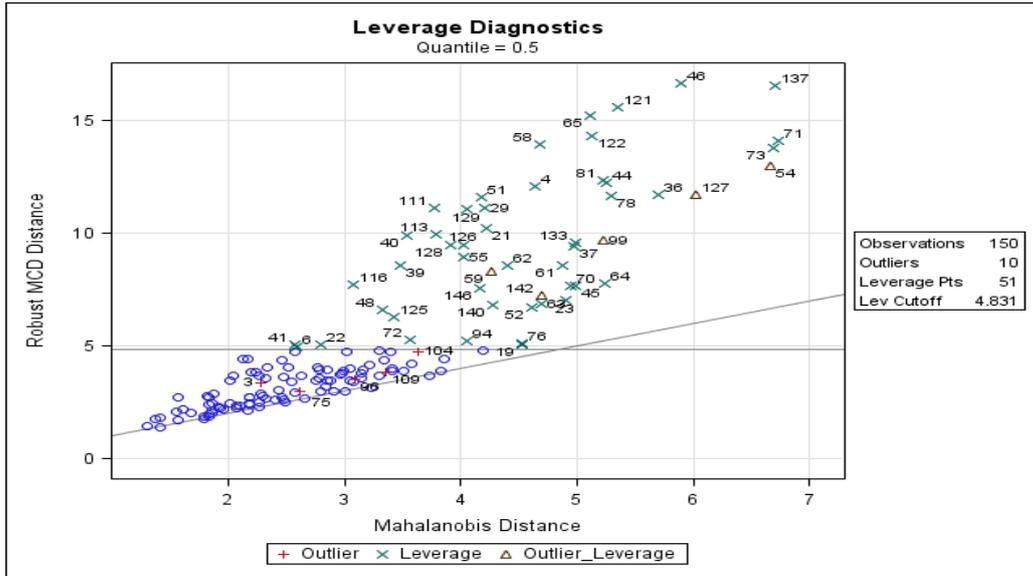
يلاحظ من الشكلين اعلاه أنه تم الكشف عن 51 مشاهدة ذات قوة رفع عالية، و10 مشاهدات شاردة و5 مشاهدات تتصف بكونها ذات قوة رفع عالية فضلا عن كونها شاردة.



الشكل (3) يبين البواقي المعيارية مقابل المسافة الحصينة، محدد التباين المشترك الادنى (minimum

(MCD) (covariance determinant)

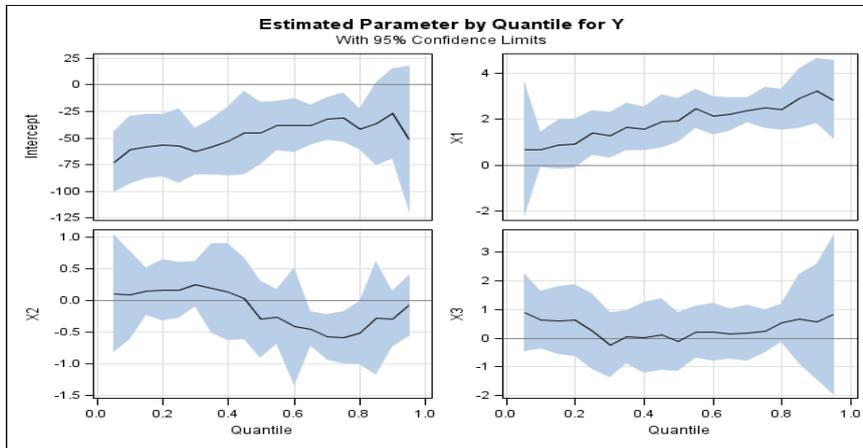
المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي SAS 9.2



الشكل (4) يبين المسافة الحصينة MCD مقابل مسافة مهلائبس

المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي SAS 9.2

تم استخدام طريقة اعادة المعاينة بووتستراب الهامشية لسلسلة ماركوف (Markov chain marginal bootstrap) (MCMB) لحساب حدود الثقة للعمليات الربيعية. يبين الشكل (5) عمليات الربيعي مع نطاق حدود الثقة 95% للحد الثابت مع المتغيرات (X_1, X_2, X_3) .

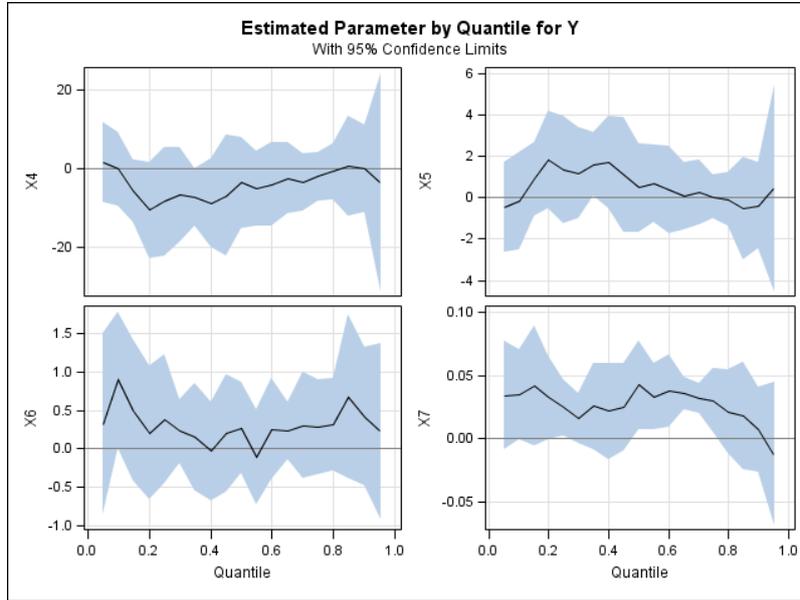


الشكل (5) يبين عمليات الربيعي مع نطاق حدود الثقة 95% للحد الثابت مع المتغيرات (X_1, X_2, X_3)

المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي SAS 9.2

يظهر من رسم المتغير X_1 (الشكل (5)) بأن ميل معلمة هذا المتغير تتصاعد كلما ازدادت الرُبيعات، وهذا يعني ان العمر الحقيقي للمريض (مقاسا بالشهر) له أثر كبير على عمر العظم عند اي رُبيع من رُبيعات توزيع عمر العظم، ويكون اكبر ما يكون عند الرُبيع الذي في الجهة اليمنى (عند التقدم بالعمر) وأقلها عند الرُبيع الذي في الجهة اليسرى (عند الولادة). بالنسبة للمتغير X_2 ، يلاحظ أن إشارة هذا المتغير تتغير من الموجب الى السالب، ويكون تغير الإشارة عند الرُبيعي (0.45937). ويتضح من الرسم ان تأثير عمر المريض عند المرض مقاساً (بالشهر) له أثر سلبي كبير على عمر العظم وخصوصا في الجهة اليمنى من توزيع عمر العظم، اي ما بين الرُبيع 0.66 والرُبيع 0.805. وذا أثر

قليل ما عدا ذلك. فيما يتعلق بالمتغير X_3 (تضخم الكبد مقاسا (بالسنتمتر))، يلاحظ أن له تأثيرا منتظما وضعيفا وبمقدار 0.12 على امتداد توزيع عمر العظم.

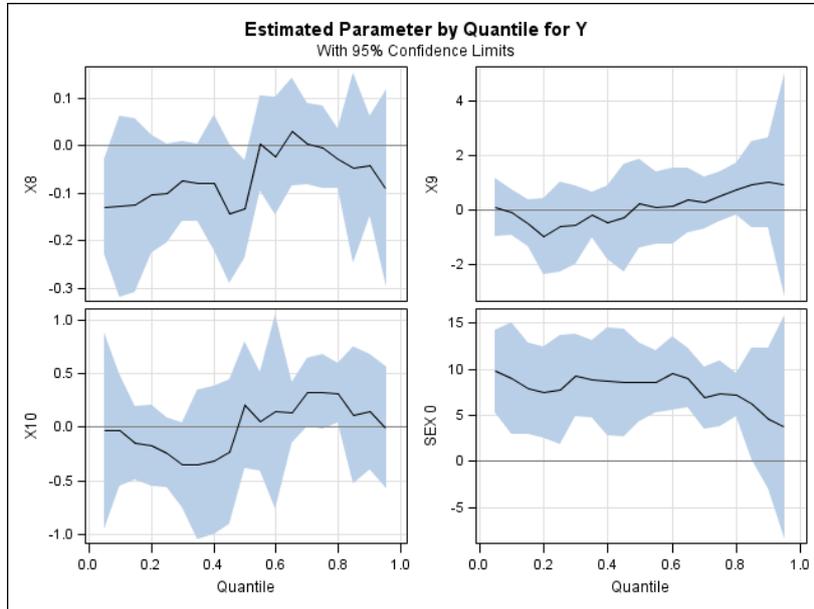


الشكل (6) يبين عمليات الرُّبَيعي مع نطاق حدود الثقة 95% للمتغيرات (X_4, X_5, X_6, X_7)
المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي SAS 9.2

يبين الشكل (6) عمليات الرُّبَيعي مع نطاق حدود الثقة 95% للمتغيرات (X_4, X_5, X_6, X_7). فيما يتعلق بالمتغير X_4 (هيموكلوبين الدم)، يلاحظ أن اشارة هذا المتغير تتغير من الموجب الى السالب عند الرُّبَيع (0.105)، ومن ثم تتحول الاشارة الى الموجب عند الرُّبَيع (0.82). والملاحظ ان أثر هذا المتغير ضعيف على عمر العظم عند اي رُّبَيع من الرُّبَيعات. بالنسبة للمتغير X_5 (مكداس الدم (خلايا الدم المضغوط))، فيلاحظ أن اشارة هذا المتغير تتغير من السالب الى الموجب عند الرُّبَيع (0.115)، ومن الموجب الى السالب عند الرُّبَيع (0.76)، ومن ثم الى الموجب عند الرُّبَيع (0.92). علما ان أثر هذا المتغير ضعيف عند اي رُّبَيع من رُّبَيعات توزيع عمر العظم. فيما يتعلق بالمتغير X_6 (الخلايا الشبكية)، يلاحظ أن اشارة هذا المتغير تتغير من الموجب الى السالب عند الرُّبَيع (0.37)، ومن السالب الى الموجب عند الرُّبَيع (0.415)، ثم ما تلبث ان تتغير الى السالب عند الرُّبَيع (0.535) ومن ثم الى الموجب عند الرُّبَيع (0.565). كما يلاحظ ان أثر هذا المتغير ضعيف على عمر العظم عند اي رُّبَيع من الرُّبَيعات. بالنسبة للمتغير X_7 فان له أثر ايجابي كبير، وخصوصا عند الرُّبَيعات من (0.135) الى (0.17) والرُّبَيعات من (0.485) الى (0.53) فضلا عن الرُّبَيعات من (0.605) الى (0.64). وذا اثر ضعيف فيما عدا ذلك، علما ان اشارة المتغير تتغير من الموجب الى السالب عند الرُّبَيع (0.935).

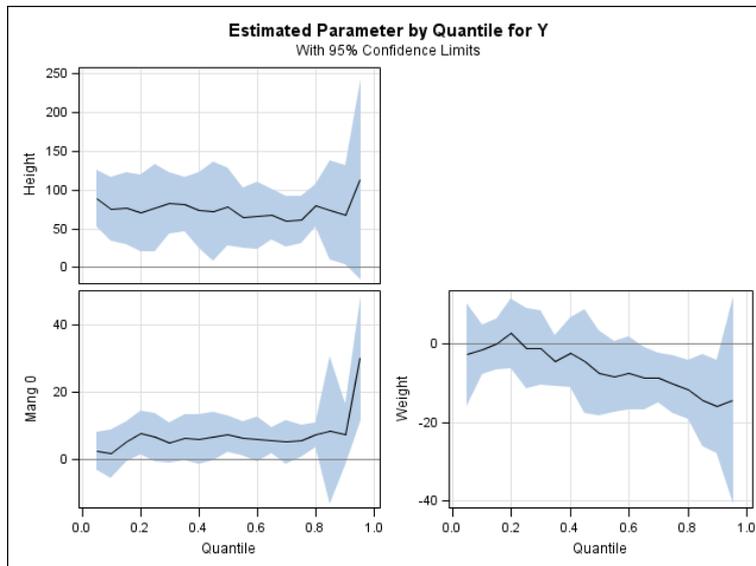
الشكل (7) يبين عمليات الرُّبَيع مع نطاق حدود الثقة 95% للمتغيرات ($X_8, X_9, X_{10}, sex 0$). المتغير X_8 (الهيموكلوبين الجيني) له أثر ايجابي كبير على عمر العظم وخصوصا عند الرُّبَيع (0.46). كما يلاحظ من الرسم الارتفاع الحاد عند الرُّبَيع (0.53)، فضلا عن تغير الاشارة من السالب الى الموجب عند الرُّبَيع (0.54) ومن ثم تغير الاشارة الى السالب عند الرُّبَيع (0.565). بالنسبة للمتغير X_9 (عدد وحدات الدم)، فيلاحظ تغير الاشارة من الموجب الى السالب عند الرُّبَيع (0.056)، ومن ثم الى الموجب عند الرُّبَيع (0.49). كما يلاحظ ان أثر هذا المتغير ضعيف على عمر العظم عند اي رُّبَيع من الرُّبَيعات. يلاحظ من الشكل (7) ان المتغير X_{10} (بداية نقل الدم حسب العمر مقاسا

(بالشهر)) له أثر ضعيف على عمر العظم عند اي ربيع من الربيعات، كما يلاحظ تغير الاشارة من السالب الى الموجب عند الربيع (0.48). (Sex) الجنس (الاناث) لها تأثير كبير على عمر العظم مما هو عليه عند الذكور ويتجلى هذا في الربيع الذي في الجهة اليسرى مما هو عليه في الربيع الذي في الجهة اليمنى وخصوصا في الربيعات من (0.1) الى (0.855).



الشكل (7) يبين عمليات الربيعي مع نطاق حدود الثقة 95% للمتغيرات $(X_8, X_9, X_{10}, sex 0)$

المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي SAS 9.2



الشكل (8) يبين عمليات الربيعي مع نطاق حدود الثقة 95% للمتغيرات $(Height, Mang.0, Weight)$

المصدر: من عمل الباحث بالاستناد الى مخرجات البرنامج الإحصائي SAS 9.2

الشكل (8) يبين عمليات الربيع مع نطاق حدود الثقة 95% للمتغيرات $(Height, Mang.0, Weight)$. ان المتغير $(Height)$ له أثر كبير في عمر العظم عند الربيعات من (0.1) الى (0.85)، وذا اثر قليل ماعدا ذلك. الشكل

غير المنغولي ($Mang. 0$) له أثر كبير على عمر العظم مما هو عليه بالنسبة للمرضى ذات الشكل المنغولي ويتجلى هذا عند الرُبيعات من (0.16) الى (0.23)، ومن (0.27) الى (0.29)، ومن (0.35) الى (0.36)، ومن (0.445) الى (0.665) واخيرا من (0.715) الى (0.815). بالنسبة للوزن ($Weight$)، لها أثر سلبي ضعيف في عمر العظم وخصوصا في الرُبيعات التي في الجهة اليسرى، وذا أثر كبير عند الرُبيعات التي في الجهة اليمنى وخصوصا ما بعد الرُبيع (0.62).

الاستنتاجات والتوصيات:

في ضوء ما ورد في هذا البحث يمكن اجمال الاستنتاجات الآتية:

1. ان تقدير المعلمات باستخدام خوارزمية النقطة الداخلية اعطت كفاءة اعلى مقارنة بالخوارزمية الممهدة والخوارزمية المبسطة، وهو خلاف ما توصل اليه الباحث ((Chen, C. (2004)) بان كلا من خوارزمية النقطة الداخلية والممهدة تعطي نتائج افضل مقارنة بالخوارزمية المبسطة عندما يكون حجم البيانات كبيرا والعكس صحيح. ونرجح ان يكون للملاحظات المتطرفة (الشاردة، ذات قوة الرفع العالية، فضلا عن الملاحظات التي تتمتع بالخاصيتين المشار اليهما معا) تأثير على كفاءة الخوارزميات. عليه يوصي الباحث بإجراء مقارنة بين الخوارزميات الثلاث وفق حجوم عينات مختلفة في ظل وجود مشاهدات متطرفة.
2. الكشف عن الملاحظات المتطرفة، فضلا عن حالة عدم تجانس البيانات لهما الاثر الكبير في تحديد الطريقة الانسب في الوصول الى أفضل التقديرات.
3. ساهم كل من المدرج الإحصائي للبواقي المعيارية فضلا عن المنحنى الطبيعي والمنحنى اللبي في تعيين قيمة الرُبيع الذي يؤدي الى اختيار النموذج الافضل.
4. يمكن تحديد الخوارزمية الافضل في حالة وجود حل مثالي وحيد للمسألة قيد الدراسة من خلال مقارنة عدد المعلمات التي لها خطأ معياري اقل، الذي سينعكس على قيم t المحسوبة ومن ثم على عدد المعلمات المعنوية فضلا عن مستوى المعنوية. يمكن اتخاذ ذلك معيارا في اختيار الخوارزمية الافضل، فضلا عن المدرج الإحصائي للبواقي المعيارية، المنحنى الطبيعي والمنحنى اللبي.
5. باستخدام طريقة اعادة المعاينة بوونستراب الهامشية لسلسلة ماركوف يمكن معرفة نسبة التغير في الرُبيعات الشرطية.

المصادر:

1. النعيمي، أسوان محمد طيب (2005): "اختيار المتغيرات في انحدار الحرف"، رسالة ماجستير غير منشورة، كلية علوم الحاسبات والرياضيات، جامعة الموصل.
2. Chen, C. (2004), "An Adaptive Algorithm for Quantile Regression," Theory and Applications of Recent Robust Methods, ed. M. Hubert, G. Pison, A. Struyf, and S. Van Aelst, Series: Statistics for Industry and Technology, Birkhauser, Basel, 39–48.
3. **Chen, C. and Wei, Y. (2005), "Computational Issues for Quantile Regression," The Indian Journal of Statistics, Special Issue on Quantile Regression and Related Methods, Volume 67, Part 2, pp 399-417.**
4. **Chen, C. (2007), "A Finite Smoothing Algorithm for Quantile Regression," Journal of Computational and Graphical Statistics, 16, 136–164.**
5. **Clark D. I. and Osborne, M.R. (1986), "Finite algorithms for Huber's M-estimator," SIAM J. Sci. Statist. Comput., 6, 72-85.**
6. Jan K. Møller (2006), "Modeling of Uncertainty in Wind Energy Forecast," Master Thesis, Informatics and Mathematical Modeling, Technical University of Denmark. URL: <http://www.imm.dtu.dk/pubdb/p.php?4428>.
7. Jiansheng, W. and Liangyong, H., Xiaoming P., (2010), "A Novel Bayesian Additive Regression Trees Ensemble Model Based on Linear Regression and Nonlinear Regression for Torrential Rain Forecasting," Third International Joint Conference on Computational Science and Optimization. IEEE, Pages 466-470.
8. Keming, Y., Zudi L. and J. Stander, (2003), "Quantile regression: applications and current research areas," The Statistician, Vol. 52, No. 3, pp: 331–350.
9. Kleiber, C. and Zeileis, A. (2008), "Applied Econometrics with R". Springer.
10. Koenker, R. and Bassett G. (1978) "Regression Quantile Econometrica," Vol. 46, No 1, Jan (33-50).
11. Koenker, R. (2005), "Quantile Regression", Cambridge University Press, Cambridge
12. Rousseeuw PJ and van Zomeren BC (1990), "Unmasking multivariate outliers and leverage points," Journal of the American Statistical Association; 85(411): 633-9.
13. Madsen, K. and Nielsen, H. B. (1993), "A Finite Smoothing Algorithm for Linear Estimation," SIAM Journal on Optimization, 3, 223–235.
14. **Mohammad A., et.al. (2010), "Nuclear magnetic resonance-based screening of thalassemia and quantification of some hematological parameters using chemometric methods," Talanta, Vol. 81, Issues 4-5, 15 June 2010, Pages 1229-1236.**
15. SAS Institute. SAS/STAT 9.2 User's Guide: The QUANTILREG Procedure (book Excerpt). Edited by Publishing SAS Publishin g. SAS Institute, 2008.
16. Weisberg, S. (2005), "Applied Linear Regression". 3d ed. New York: John Wiley and Sons.
17. Wilcox, R. R. (1997), "Introduction to robust estimation and hypothesis testing". San Diego, CA: Academic Press.