AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

المقارنة بين مقدري Lasso و MCP في نموذج الانحدار التقسيمي الجزائي باستعمال المحاكاة مع تطبيق عملي .

Jiddah Wally Zaher[1] **Ali Hameed Yousif [2]**

Wasit University Iraq

giddahz1402@uowasit.edu.iq ahameed@uowasit.edu.iq

المستخلص: - **Abstract**

يُعدّ الانحدار التقسيمي الجزائي طريقة مثالية لاختيار المتغيّرات في ظلّ وجود عدد كبير من المتغيّرات التوضيحية ومن ضمن هذه المتغيّرات تكون عديمة الفائدة أي ليس لها تأثير على متغيّر الاستجابة فوجودها يسبّب في تعقيد النموذج ويصعب تفسيره . في هذا البحث نركز على اختيار المتغيّرات بالاعتماد على المقدرات الجزائية (lasso) و (MCP) ، والمقارنة بينهما بالاعتماد على معيار متوسط مربعات الخطأ MSE من إذّ التقدير وبالاعتماد على معياري معدل الإيجابية الزائف FNR ومعدل السلبية الزائف FPR لبيان كفاءة المقدّر من إذّ اختيار المتغيّرات . وبعد تطبيق تجارب المحاكاة والبيانات الحقيقية تبيّن أنّ مقدر MCP هو الأفضل من ناحية التقدير واختيار المتغيّرات .

الكلمات المفتاحية : الانحدار التقسيمي ، الانحدار التقسيمي الجزائي ، اختبار المتغيّرات ، مقدر Lasso ، مقدر MCP .

Abstract

Quintile regression is an excellent method for selecting variables in the presence of a large number of explanatory variables, many of which are useless, meaning they have no effect on the response variable and make the model complicated and difficult to interpret. In this research, we focus on selecting variables based on penalized estimators (lasso) and (MCP) and comparing them based on the MSE mean square error criterion in terms of estimation, and based on the FNR and FPR criteria to demonstrate the efficiency of the estimator in terms of

selecting variables. After applying simulation experiments and real data, it was found that the MCP estimator is the best in terms of estimation and selection of variables.

1-المقدمة Introduction

إنّ إحدى طرائق التحليل الإحصائية التي تستعمل في دراسة وتحليل العلاقة بين المتغيّرات المدروسة ولكافة العلوم هو الانحدار الخطي .والذي يُعدّ وسيلة إحصائية لدراسة العلاقة بين متغيّر توضيحي واحد او أكثر و يصف العلاقة بين المتغيّرات على هيئة معادلة. إنّ للانحدار استعمالات عديدة منها : وصف البيانات، وتقدير المعلمات والتَّنبؤ والسيطرة وغيرها.

وفي أغلب الآحيان لا يمكن استعمال الانحدار الخطي الاعتيادي في العديد من الدراسات العلمية والظواهر المختلفة لأنّ شروط الانحدار غير قابلة للتطبيق ، و البديل المناسب لذلك هو الانحدار التقسيمي (Quantile Regression) ، والذي تَمّ اقتراحة بواسطة (koenker and Basset) في عام 1978 [12] ويُعدّ البديل المناسب للانحدار الخطي ،وقد زاد الاهتمام به في السنوات الأخيرة . وبفرض لدينا عينة عشوائية $(Y_1, X_1), \dots, (Y_n, X_n)$ ، $(0 < \theta < 1)$ فإن نموذج الانحدار الخطي يمكن صياغته بالشكل الآتي [7]

$$Y_i = \acute{X}_i \beta + \epsilon_i \qquad i = 1,2, \dots, n \qquad (1)$$

إذّ إنّ Y_i يمثل متغيّر الاستجابة أمّا X_i تمثّل المتغيّرات التوضيحية (المتغيّرات المستقلة) .أمّا β تعني معلمات النموذج $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ أبعادها. p أمّا ϵ_i. فيمثّل الخطا العشوائي الذي يتوزّع طبيعيا بمتوسط صفر وتباين σ^2 $\epsilon_i \sim N(0, \sigma^2)$

يمكن تقدير معلمات انموذج الانحدار التقسيمي $\in \mathcal{R}^P$ $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ باستعمال الصيغة الآتية :-

$$\hat{\beta}_{(\theta)} = \arg min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\theta \left(y_i - \acute{X}_i \beta \right) \qquad (2)$$

إذّ إنّ $\rho_\theta(t)$ تدعى بدالة الضبط والتي يعبر عنها بالصيغة الآتية :-[12]

$$\rho_\theta(t) = \begin{cases} \theta t & if \quad t \geq 0 \\ -(1-\theta)t & if \quad t < 0 \end{cases} \qquad (3)$$

إذّ إنّ كُلّ $(0 < \theta < 1)$. [19]

يعدّ اختيار المتغيّرات مهم جدا في ظلّ وجود عدد كبير من المتغيّرات التوضيحية (explanatory variables)، اذ اننا لا نعلم أياً من المتغيّرات التوضيحية من الممكن أنْ تؤثّر على متغيّر الاستجابة، لأن هنالك العديد من المتغيّرات التوضيحية تكون عديمة التأثير أو لها تأثير قليل على متغيّر الاستجابة و الهدف الرئيس هو الحصول على نموذج المقلص (parsimonious model)أيْ نموذج يحتوي على عدد قليل من المتغيّرات التوضيحية ,والذي يسمّى بالنموذج المبعثر (sparse) إذّ يصعُب تحديد المتغيّرات المعنوية التي تؤثر على متغيّر الاستجابة (response variable) ويحتمل أنْ تؤدّي إلى إسقاط بعض المتغيّرات التوضيحية الهامة ويجب معالجة هذه المشاكل بشكل صحيح بطريقة مثالية . ويُعدّ الانحدار التقسيمي الجزائي أداة مفيدة لمعالجة هذه المشاكل [1] . في عام 2009 قام الباحثان (Wu and Liu) بدراسة الانحدار التقسيمي الجزائي بالاعتماد على مقدر جزاء (SCAD) و (ADAPTIVE LASSO) وباستعمال خوارزمية (DCA) كما أكّدتْ نتائج البيانات الحقيقية والمحاكاة بأنْ هذه المقدرات المستعملة تتمتع بخصائص اوراكل [2] ،في عام 2012 قام الباحثان (Wang and LI) بدراسة الانحدار التقسيمي الجزائي في حالة البيانات عالية الأبعاد، وغالبا ما تُعرّض هذه البيانات إلى مشكلة عدم التجانس ، بسبب الزيادة في عدد المتغيّرات التوضيحية . ولاختيار المتغيّرات المعنوية تَمّ الاعتماد على دالة جزاء (SCAD) [10].

إنّ مشكلة البحث في الانحدار الخطي لا يمكن تقدير التوزيع الشرطي لمتغيّر الاستجابة عند نقاط مختلفة ، لذلك نستعمل الانحدار التقسيمي كما لا يمكن استعمال الانحدار التقسيمي في حالة كون عدد المتغيّرات التوضيحية كبيره لمعالجة هذه المشكلة نستعمل الانحدار التقسيمي الجزائي .

يهدف هذا البحث إلى المقارنة بين مقدري Lasso و MCP لنموذج الانحدار التقسيمي والحصول على أفضل مقدر من بين المقدّرات الأخرى ومن ثَمّ الحصول على أفضل تقدير باستعمال المحاكاة ، وذلك بالاعتماد على معيار متوسط مربعات الخطأ ، معدّل الايجابية الزائف ومعدّل السلبية الزائف.

تَمّ ترتيب باقي تفاصيل البحث على النحو الآتي في القسم الثاني يَتَمّ فيه التطرّق إلى الانحدار التقسيمي الجزائي يَتَمّ فيه توضيح مقدري Lasso و MCP . أمّا في القسم الثالث يَتَمّ فيه التطرّق إلى تجارب المحاكاة . وفي القسم الرابع يَتَمّ فيه التطرّق إلى الجانب التطبيقي تطبيق البيانات الحقيقية .وفي القسم الخامس يَتَمّ التطرّق إلى أهْم الاستنتاجات والتوصيات .

2-الانحدار التقسيمي الجزائي **Penalized quantile regression**

عند الزيادة في عدد المتغيّرات التوضيحية على حساب حجم العينة تكون عملية تقدير معلمات النموذج صعبة التنفيذ . كذلك يصعب اختيار المتغيّرات المعنوية التي تُعدّ أحد المسائل المهمّة في النمذجة الإحصائية .وفي هذه الحالة لا يمكن تطبيق الانحدار التقسيمي ، وعليه فإنّ البديل المناسب هو استعمال طريقة انحدار ملائمة للتعامل مع البيانات ذات الابعاد العالية. تسمّى طريقة الانحدار التقسيمي الجزائي [11]

يَتَمّ إضافة دالة الجزاء إلى دالة الخسارة التقسيمية فنحصل على دالة الهدف وكما في الصيغة الآتية [5]

$$Q_T(\beta) = \sum_{i=1}^{n} \rho_\theta\big(y_i - \acute{X}_i\beta\big) + \lambda \sum_{j=1}^{p} p_\lambda(|\beta_j|) \qquad (4)$$

إذّ إنّ$p_\lambda(.)$ تمثل دالة الجزاء .

λ :- تعبر عن معلمة الجزاء والتي تكون قيمتها $\lambda \geq 0$. وتسمى بمعلمة الضبط.

وتوجد هنالك العديد من الدوال الجزائية التي يَتَمّ إضافتها إلى دالة الخسارة التقسيمية لتقدير معلمات نموذج الانحدار التقسيمي وكالآتي :-.

1-2 مقدر Least absolute shrinkage and selection operator (Lasso)

تُعدّ هذه الطريقة أحد دوال الجزاء التي تستعمل لتقدير معلمات انموذج الانحدار التقسيمي. وتعدّ كمختصر لـ (Least Absolute Shrinkage and Selection Operator) . لقد تَمّ اقتراحها من قبل (Tibshirani) في عام 1996 التي تعمل على التقدير واختيار المتغير(Variable Selection). إنّ المبدأ الأساس لهذه الطريقة هي تصغير مجموع مربعات الأخطاء وفقا للقيد الذي يمثّل المجموع المطلق للمعلمات [3] .

$\sum_{j=1}^{p}|\beta_j|$

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

إذّ إنّ دالة الجزاء (lasso) تجعل بعض المعلمات مساوية بالضبط إلى الصفر عندما تكون معلمة الجزاء كبيرة وتقليص المعلمات الاخرى وفقا لمقدار معين. يَتَمّ الاعتماد على معلمة الجزاء λ للتحكّم بكمية التقليص (Shrinkage) .

اقترح الباحثان (Li and Zhu) دالة جزاء (lasso) في الانحدار التقسيمي (QR) في عام (2008) وذلك بإضافة دالة الجزاء (Lasso) إلى دالة الخسارة (loss function)التقسيمية للحصل على دالة الهدف التقسيمية.

من خلالها نحصل على مقدرات (lasso) $\hat{\beta}_{LASSO} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_P)$ في لانحدار التقسيمي (QR) وفق للصيغة الاتية.**[5]**

$$\hat{\beta}_{(LASSO,QR)} = arg\ min_\beta \left\{ \sum_{i=1}^{n} \rho_\theta(y_i - \acute{X}_i\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \qquad (5)$$

إذّ إنّ :

λ :تمثّل معلمة الجزاء (Penalty parameter) وتسمّى أيضا بمعلمة الضبط (tuning parameter) تعمل على السيطرة والمفاضلة بين دالة الخسارة التقسيمية وحدود الجزاء أيْ تتحكم بحجم المعاملات .

$\lambda p(.)$: تمثّل دالة الجزاء (penalty function).

لحساب مقدر Lasso نستعمل خوارزمية (SNCD) [17]

(4-4-2)مقدر Minimax convex Penalty (MCP)

وهي دالة جزاء مقعرة ، اقترحت من قبل الباحث (Zhang) في عام 2010 . الذي يعبر عنها بالمصطلح (Minimax convex Penalty) ، ويرمز لها بالمختصر (MCP) . هذه الطريقة تتميّز بالسرعة في تقدير المَعْلمات (Parameters) وتكون مستمرة وغير متحيزه ودقيقة ، فضلاً عن ذلك تقوم بتبسيط الانموذج وتحسين دقته في وقت واحد . إذْ إنّها تفوق طريقة (Lasso) من إذّ الدقّة وسرعة التقدير والاختزال .[15]

يمكن كتابة طريقة جزاء (MCP) كما في الصيغة الNتية:

$$p_{\lambda.\gamma}(|\beta|) = \begin{cases} \lambda|\beta| - \frac{\beta^2}{2\gamma} & |\beta| \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 & |\beta| > \gamma\lambda \end{cases} \qquad (6)$$

إذّ إنّ $\gamma > 1$

أمّا دالة الهدف للانحدار الجزائي التقسيمي لمقدر (MCP) يمكن صياغته بالصيغة الآتية . **[16]**

$$\hat{\beta}_{MCP,QR} = argmin_{\beta}\{\sum_{i=1}^{n} \rho_{\theta}(y_i - X_i^T \beta) + \lambda \sum_{j=1}^{p} p_{\lambda\gamma}(\beta)\} \qquad (7)$$

نستعمل خوارزمية (LLA) لحساب مقدر MCP.[18]

3-تجربة المحاكاة

من أجل المقارنة بين المقدرات الجزائية (Lasso , MCP) يَتّمّ تنفيذ تجارب المحاكاة بطريقة Monte Carlo بالاعتماد على برنامج R .

1-3 توليد المتغيّرات التوضيحية:

نعتمد على التوزيع الطبيعي متعدّد المتغيّرات بمتوسط (mean=0) ومصفوفة التباين والتباين المشترك (Σ) لتوليد المتغيّرات التوضيحية P كما في الصيغة الآتية :-

$$X \sim MN(0, \Sigma)$$

إذّ إنّ :-

$\Sigma_{ij} = \rho^{|i-j|} \quad , \qquad \rho = 0.5$

2-3 توليد الأخطاء العشوائية :

أمّا عملية توليد الأخطاء العشوائية يَتّمّ توليدها وفق التوزيع الطبيعي بمتوسّط (mean=0) وتباين (σ^2)

$\varepsilon_i \sim N(0, \sigma^2) \qquad , \quad i = 1,2, \dots, n$

3-3 حساب المتغيّر المعتمد

أمّا المتغيّر المعتمد يمكن حسابة عن طريق ضرب مصفوفة المتغيّرات التوضيحية X التي تَمّ توليدها في الفقرة (1-3) بموجة المعلمات الافتراضية مضاف لهما حد الخطأ العشوائي ε_i الذي تَمّ توضيح طريقة توليده في الفقرة (2-3) .

إنّ تجارب المحاكاة يكون وفقا للأسلوب الآتي :

σ : يأخذ القيمتين (0.5,1) .

نعتمد على النموذج الأول : $\beta = (3,1.5,0,2,0, \dots ,0)$ ، (P=10, n=25)

النموذج الثاني : β=(1,1,1,-2,-2,-2,0,…,0) ، (P=25, n=100)

أمّا بالنسبة للقيم الافتراضية (Quantile) تكون $\theta = (0.3,0.5,0.7)$

وتَتّمّ المقارنة بين المقدرات بالاعتماد على معيار متوسّط مربع الأخطاء MSE الذي من خلاله يمكن المفاضلة بين المقدرات من ناحية التقدير . أمّا من ناحية اختيار المتغيّرات يَتّمّ الاعتماد على معيار معدل الإيجابية الزائف FNR ومعيار معدّل السلبية الزائف FPR .

الجدول (1)

يمثّل نتائج المحاكاة للنموذج الأول عندما (P=10, n=25)

θ	σ	Estimators	MSE	FPR	FNR
0.3	0.5	LASSO	**0.0326**	**0.131**	**0**
		MCP	**0.0228**	**0.005**	**0**
		LASSO	**0.5695**	**0.302**	**0.106**

	1	MCP	**0.4368**	**0.057**	**0.24**
0.5	0.5	LASSO	**0.0633**	**0.148**	**0**
		MCP	**0.0204**	**0.005**	**0**
	1	LASSO	**0.2154**	**0.24**	**0**
		MCP	**0.1284**	**0.028**	**0.013**
0.7	0.5	LASSO	**0.0403**	**0.006**	**0**
		MCP	**0.0401**	**0**	**0**
	1	LASSO	**0.0931**	**0.194**	**0**
		MCP	**0.0399**	**0.057**	**0**

الجدول (2)

يمثّل نتائج المحاكاة للنموذج الأول عندما (P=25, n=100)

θ	σ	Estimators	MSE	FPR	FNR
0.3	0.5	LASSO	**0.0066**	**0.167**	**0**
		MCP	**0.0015**	**0.14**	**0**
	1	LASSO	**0.0508**	**0.16**	**0**
		MCP	**0.0367**	**0.136**	**0.086**
0.5	0.5	LASSO	**0.0123**	**0.174**	**0**
		MCP	**0.0022**	**0.136**	**0**

		LASSO	**0.0911**	**0.1454**	**0.08**
	1	MCP	**0.1049**	**0.133**	**0.229**
0.7	0.5	LASSO	**0.0849**	**0.152**	**0**
		MCP	**0.0671**	**0.136**	**0**
	1	LASSO	**0.0614**	**0.165**	**0.02**
		MCP	**0.0308**	**0.136**	**0.066**

من خلال نتائج المحاكاة الواردة في الجداول (1) و(2) المذكورين انفاً في مختلف حجوم العينات وعدد المتغيّرات التوضيحية وعند القيم الافتراضية كافة quantile ان مقدر MCP هو الأفضل من ناحية التقدير لأنّه يعطي أقلّ قيم لمتوسط مربعات الأخطاء MSE . كذلك هو الأفضل من ناحية اختيار المتغيّرات لأن يعطي أقلّ قيم لمعيار معدّل الإيجابية الزائف FNR ومعدّل السلبية الزائف FPR .

4-الجانب التطبيقي

خُصّصتْ هذه الفقرة إلى الجانب التطبيقي الذي يتمثّل بالبيانات الطبية جمعت من مختبر تحليل الدم للمرضى المصابين بأمراض الدم الوراثي (الثلاسيميا) من مستشفى الكوت للنسائية والأطفال في محافظة واسط وكان حجم العينة (30) مريضا . من خلال جمع البيانات والدراسة المستفيضة لطبلات المرضى و بعد المراجعة وسؤال العديد من أطباء الاختصاص الكفوئين في هذا المجال تَمّ حصر تفاصيل العينة التي تتألّف من متغيّر معتمد Y يتمثّل بعدد خلايا الدم البيضاء (WBC) ومجموعة من المتغيّرات التوضيحية X

التي تتمثّل (Gender يمثّل الجنس ، Age يمثّل العمر، Blood Type نوع فصيلة الدم . RBC (Red Blood Cells) عدد خلايا الدم الحمراء . HGB (Hemoglobin) الهيموغلوبين . (Hematocrit) HCT تركيز كريات الدم الحمراء في الدم . MCV (Mean Corpuscular Volume) متوسّط حجم كريات الدم الحمراء . (Mean Corpuscular Concentration) MCHC (Mean Corpuscular Hemoglobin . هيموغلوبين الكرية الوسطى Hemoglobin) متوسط تركيز الهيموغلوبين. PLT (Platelets) عدد الصفائح الدموية . Lym % (Lymphocytes) الخلايا الليمفاوية . LYM نسبة كريات الدم البيضاء الحية . NEUT% (Neutrophils) خلايا مختلفة . (MXD% (Mixed Cells) عدد الليمفاويات (Lymphocyte Absolute) . MXD (Mixed Absolute count) يعني عدد الخلايا المختلفة . NEUT) Neutrophil Absolute Count) عدد الخلايا الحية . RDW-SD (Red Cell Distribution Width) قياس توزيع خلايا الدم الحمراء. RDW-CV مدى تباين توزيع خلايا الدم الحمراء. PDW (Platelet Distributions

(Widthقياس توزيع الصفائح الدموية . MPV (Mean Platelet Volume)حجم الصفائح الدموية . P-LCR (Platelet Larger cell Ratio)نسبة الخلايا الكبيرة للصفائح الدموية. PCT (Procalcitonin)البروكالسيتونين.)

وبعد تطبيق البيانات الحقيقية على المقدرات الجزائية بالاعتماد على برنامج R حصلنا على النتائج الآتية كما موضحة في الجدول الآتي :-

الجدول (3)

يمثّل نتائج الجانب التطبيقي لمعاملات المتغيّرات التوضيحية

Variables	$\boldsymbol{\theta = (0.3)}$		$\boldsymbol{\theta = (0.5)}$		$\boldsymbol{\theta = (0.7)}$	
	Lasso Covariates	**MCP Covariates**	**Lasso Covariates**	**MCP Covariates**	**Lasso Covariates**	**MCP Covariates**
Gender	0	0	0	0	0	0
Age	-0.0037	-0.015	0	0	-0.003	0
Blood Type	0	0	0	0	0	-0.015
RBC	-0.047	-0.033	0	0	0	0
HGB	0.009	0.039	0	0	0.007	0
HCT	-0.011	0	0	0	0	0
MCV	0	0	0	0	0	0
MCH	0	0	0	0	0	0

MCHC	0	0	0	0	0	0
PLT	0.063	0	0	0	0.022	0.035
Lym %	0	0	0	0	0	0
MXD %	0	0.009	0	0	0	0
NEUT %	-0.015	0	0	0	0	0
LYM	0.194	0.225	0.174	0.216	0.198	0.197
MXD	0.056	0.062	0.005	0	0.049	0.035
NEUT	0.217	0.213	0.049	0.039	0.150	0.172
RDW-SD	0	0	0	0	-0.0006	0
RDW-CV	0	0	0	0	0	0
PDW	-0.003	0	0	0	0	0
MPV	-0.005	0	0	0	0	0
P-LCR	0	0	0	0	0	0
PCT	0.029	0.028	0.014	0.024	0.0309	0.019
عدد المعلمات الصفرية	**10**	**14**	**18**	**18**	**14**	**16**

من خلال نتائج الجدول أعلاه نلاحظ عند المستويات التقسيمية $(\theta = 0.3 \ , \ \theta = 0.7)$ ~Hkمقدر MCB هو الأفضل من ناحية اختيار المتغيّرات ، لأنّ عدد معاملاتها الصفرية أكثر من المعاملات عند مقدر Lasso .وهذا يعيني ان مقدر MCB قد اختار 8 متغيرات معنوية أمّا مقدر Lasso قد اختار 12 متغيّر معنوي عند المستوى التقسيمي $\theta = 0.3$. أمّا عند المستوى التقسيمي $\theta = 0.7$ إنّ مقدر MCB قد اختار 6 متغيرات معنوية أمّا مقدر Lasso قد اختار 8 متغيرات معنوية. أمّا عند المستوى التقسيمي $\theta = 0.5$ فإنّ كلا المقدرين قد اختارا 4 متغيرات معنوية .

الجدول (4)

يبيّن متوسّط مربعات الأخطاء المقدرات الجزائية كافة

Quantile	$\theta = 0.3$		$\theta = 0.5$		$\theta = 0.7$	
Methods	Lasso	MCP	Lasso	MCP	Lasso	MCP
MSE	**0.1917**	**0.2091**	**0.1798**	**0.1819**	**0.1686**	**0.2149**

يَتّضِحَ من خلال نتائج الجدول رقم 4 أنّ مقدر Lasso هو الأفضل من ناحية التقدير، لأنّه اعطى أقلّ قيمة لمتوسط مربعات الأخطاء عند المستويات التقسيمية كافة .

5-الاستنتاجات والتوصيات

نستنتج على ضوء ما ورد في الجانب التجريبي والجانب التطبيقي ما يأتي .

1- أفضلية مقدر MCP في اغلب تجارب المحاكاة من ناحية التقدير لأنّه يعطي أقل قيمة لمتوسّط مربعات الخطأ MSE ، كذلك أفضل مقدّر من ناحية اختيار المتغيّرات، لأن يعطي اقل قيمة لمعيار FPR.

2-أفضلية مقدر MCP في الجانب التطبيقي من ناحية اختيار المتغيّرات .إذّ إنه اختار أقلْ عدد ممكن من المتغيّرات التوضيحية.

3- افضلية مقدر Lasso في الجانب التطبيقي من ناحية التقدير لأنّه يعطي أقل قيمة لمتوسّط مربعات الخطأ MSE .

4-نوصي الأشخاص الذين يعانون من نقص في عدد كريات الدم البيضاء بضرورة إجراء الفحوصات الطبية الآتية (LYM,MXD,NEUT,PCT).لما لها تأثير على عدد كريات الدم البيضاء.

6-المصادر :-

[1] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96, 1348-1360.

[2] Wu, Y., & Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 801-817.

[3] Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B 58, 267-288.

[4] An, L. T. H. and Tao, P. D. (1997). Solving a class of linearly constrained indefinite quadratic problems by d.c. algorithms. J. Global Optim. 11, 253-285.

[5] Li, Y., & Zhu, J. (2008). L 1-norm quantile regression. Journal of Computational and Graphical Statistics, 17(1), 163-185.

[6] Buhlmann , Peter. Geer, Sara (2011) . Statistics for High Dimensional Data Methods, Theory and Applications. Springer

[7] Amin, M., Song, L., Thorlie, M. A., & Wang, X. (2015). SCAD-penalized quantile regression for high-dimensional data analysis and variable selection. Statistica Neerlandica, 69(3), 212-235.

[8] Peng, B., & Wang, L. (2015). An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression. Journal of Computational and Graphical Statistics, 24(3), 676-694

[9] Kim, D., & Jung, Y. (2019). A numerical study on group quantile regression models. Communications for Statistical Applications and Methods, 26(4), 359-370.

[10] Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. Journal of the American Statistical Association, 107(497), 214-222.

[11] Yousif, A. H., & Housain, W. J. (2021, March). Atan Regularized in Quantile Regression for High Dimensional Data. In Journal of Physics: Conference Series (Vol. 1818, No. 1, p. 012098). IOP Publishing.

[12] Koenker, R., and Bassett, G. (1978), .Regression Quantiles. Econometrica, 46, 33–50.

[13] Zou, H. (2006). The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc., 101, 1418-1429.

[14] E. R. Lee, H. Noh, and B. U. Park,(, 2014) .Model selection via Bayesian information criterion for quantile regression models. J. Am. Stat. Assoc., vol. 109, no. 505, pp. 216–229

[15] Zhang CH (2010). Nearly unbiased variable selection under minimax concave penalty, The Annals of Statistics, 38, 894–942

[16] Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. Journal of the American Statistical Association, 107(497), 214-222

[17] C. Yi and J. Huang, .Semismooth newton coordinate descent algorithm for elastic-net penalized huber loss regression and quantile regression.J. Comput. Graph. Stat., vol. 26, no. 3, pp. 547–557, 2017.

[18] Peng, B., & Wang, L. (2015). An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression. *Journal of Computational and Graphical Statistics*, *24*(3), 676-694.

[19] Zaher, J., & Yousif, A. H. (2022). Shrinkage Estimator of SCAD and Adaptive Lasso penalties in Quantile Regression Model. *Mathematical Statistician and Engineering Applications*, *71*(4), 5945-5953.