KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

توظيف بعض الطرائق الجزائية لاختيار اهم التاثيرات والتفاعلات الرئيسية لتجربة عاملية باستخدام المحاكاة

أ.م.د. بحر كاظم محمد
الباحثة ناهدة لطيف خضير
جامعة القادسية / كلية الادارة والاقتصاد

DOI: https://doi.org/10.36322/jksc.178(B).21551

المستخلص

يهتم موضوع "تصميم التجارب العاملية" بدراسة تأثير العوامل والتفاعل بين مستويات الظاهرة الأساسية. كما يهدف تصميم التجارب العاملية إلى دراسة تفاعل الوحدات التجريبية مع العوامل المعنية. ان إجراء اختيارالمتغيرات في دراسة تصميم التجارب العاملية يعد موضعع اهتمام الباحثين العاملين في مجال تصميم التجارب العاملية مع وجود عدد كبير من المستويات. دراسة نموذج الانحدارفي موضوع تصميم التجارب مع عدد كبير من العوامل وتفاعلاتها أنتج نموذجًا overfitted، لذا فإن إزالة العوامل التي ليس لها تاثيرعلى متغير الاستجابة من تجارب العاملية تعد أمرًا مهمًا للغاية للحصول على نموذج تقديري أكثر دقة. ركزهذا البحث على طرق التنظيم لاختيار المتغيرات لإيجاد الحل الصفري (sparse solution) للمعلمات المعنية, مثل طرائق (lasso، lenth، الشبكة المرنة، و group lasso). تم إجراء دراسة محاكاة، وتم إجراء مثالين للمحاكاة لفحص قدرة طرائق اختيار المتغيرات أعلاه في التجارب العاملية، وأظهرت النتائج أن طريقة group lasso تليها الشبكة مرنة هي أفضل الطرق من حيث اختيار المتغيرات لأنها تعطي الحل الصفري مع أصغر قيم للمعايير EER وMSE كمقاييس لجودة التقدير.

الكلمات المفتاحية: تجربة عاملية , الحل الصفري , تاثيرات رئيسية , محاكاة.

Employing Some Penalized Methods for Selecting the Main Effects and Interaction in Factorial Design via Simulation

Assoc. Prof. Dr. Bahr Kazim Mohammed
Researcher Nahida Latif Khudair
University of Al-Qadisiyah / College of Business and Economics

Abstract

The well-known scientific research "factorial experiments design" studied the effect of the factors, the levels, and the interaction of levels of the underlying

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

phenomenon. Also, factorial experiments design aims to study the reaction of the experimental units to the interested factors. Variable section procedure in the study of the factorial experiments design it is of interest to researchers working in the field of designing factorial experiments. Design regression model with large number of levels and its interactions produced overfitted model, so removing the relevant levels of the designing factorial experiments are very important to yield the more accurate estimated model. This paper focused on variable selection regularization methods to find the sparse solution of the interested parameters. Such as (lasso, length, Elastic net, and group lasso). Simulation study have conducted, two simulation examples have conducted to examine the ability of the above variable selection method in factorial experiments, and the results show that the group lasso and followed by elastic net are the best methods in term of variable selection method because it give the sparse solution with smallest EER and MSE values as measures of estimate quality.

Keywords: Factorial design, sparse solution, main effects, Simulation.

١ـ المقدمة (Introduction)

التصميم التجريبي هو فرع من فروع الإحصاء يهتم بتطبيق الأساليب الإحصائية في التجارب العلمية. وهو من أهم طرق البحث العلمي الحديث ويأخذ دورا "رئيسيا" في مختلف مجالات العلوم التطبيقية. ومن أهم فروعها التجارب العاملية. يعد موضوع التجارب العاملية من المواضيع المهمة لتصميم التجارب التي تهدف إلى معرفة تأثير العوامل ومستوياتها الجزئية وكذلك تفاعل تلك المستويات على الظاهرة المدروسة وكذلك معرفة مدى استجابة العوامل الوحدات التجريبية لتلك العوامل المعنية. استخدمت التجارب العاملية لدراسة تأثير أكثر من عامل على الخاصية محل الدراسة، بغرض الحصول على معلومات عن كل عامل في التجربة من ناحية، وتحديد التفاعل في هذه العوامل من ناحية أخرى. تستخدم التجارب العاملية مع أي من التصاميم البسيطة المعروفة تصميم عشوائي كامل (CRD)، تصميم القطاعات العشوائية الكاملة (CRBD)، تصميم المربع اللاتيني (LSD) وغيرها من التصاميم. تعتبر

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

التجارب العاملية أكثر كفاءة من التجارب البسيطة، والتي تتطلب أموالاً. والجهد والوقت وتكرار التجربة لكل عامل في التجربة وكذلك شرط استقلالية العوامل.

يهتم التصميم التجريبي بإجراء التجارب الزراعية، والصناعية، والطبية، وغيرها. يهدف أسلوب التصميم التجريبي إلى إيجاد أفضل توافق ممكن بين مستويات العامل والوحدات التجريبية المراد اختبارها على المعاملات. ويجب الوصول إلى الاستجابة الأفضل وتحسينها مستقبلاً، وكذلك معرفة التباينات وإعادتها إلى مصادرها الحقيقية وتحديد أسبابها من خلال تحليل جدول التباين، حيث تعتبر هذه من الأهداف الرئيسية لإقامة التجربة. وفقا لتصميم معين. في هذه الدراسة تم اقتراح طريقة جديدة من بعض طرق اختيار العوامل وهي (الشبكة المرنة، اللاسو، مجموعة لا سو، lenth) بنموذج تجربة عامليه 2^4. وتم استخدام طرق المحاكاة لقياس دقة عمل هذه الطرق بالإضافة إلى البيانات الحقيقية. كما تم استخدام بعض الاختبارات لقياس EER وMSE لمقارنة هذه الطرق. هناك العديد من المشكلات التي تواجه الباحثين عند دراسة التجارب العاملية، بما في ذلك اختيار العوامل المؤثرة وكذلك تفاعلات العوامل باستخدام الطرق الكلاسيكية. ولحل هذه المشكلة تم استخدام طريقة اختيار العوامل ومنها الشبكة المرنة. تهدف هذه الدراسة إلى دراسة واختيار العوامل الرئيسية وكذلك التفاعلات بين العوامل باستخدام طرق الاختزال الحديثة بما في ذلك طريقة الشبكة المرنة، مع تطبيق هذه الدراسة على بيانات حقيقية.

في عام ١٩٨٩، قدم lenth طريقة جديدة تعتبر طريقة حل صفري (sparse) لتضمين التأثيرات الرئيسية والتفاعل لنموذج تصميم التجارب العاملية. تفترض هذه الطريقة المقترحة وجود مجموعة خطية من متجهات التركيبات الخطية المعلمات. تعمل هذه الطريقة بشكل جيد خاصة في ظل وجود عدد كبير من التأثيرات النشطة في الطريقة العاملية. وفي عام ١٩٢٦، أجرى العالم فيشر تجربة عاملية وأطلق عليها الاسم الذي تُعرف به حاليًا (Fisher، ١٩٢٦). وفي عام ١٩٣٧ كتب العالم (Yates) عن تحليل التجارب العاملية من النوع (٢) و(٣) وطور أساليب التحليل الإحصائي الخاصة به. ووجد أنها تصبح أكثر صعوبة عند زيادة عدد العوامل في التجارب . في (١٩٩٦) قدم Tibshirani طريقة اللاسو كطريقة اختيار متغيرات عندما تكون المشكلة قيد الدراسة تحتوي على عدد كبير من المتغيرات المستقلة وحجم عينة صغير. تعتبر Lasso طريقة جزائية تقلل من مجموع الأخطاء المربعة تحت قيد الجزاء L1−norm وهي دالة جزائية. أيضًا، تقوم طريقة اللاسو باختيار المتغير المستقل غير المهم ووجع

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

قيمته تمامًا على الصفر. لذلك، لاسو هي طريقة حل صفري. في عام ٢٠٠٥، قدم Zou وHastie ما يسمى بالشبكة المرنة، والتي تعتبر طريقة تنظيم تجمع بين لاسو و Ridge. إنها أيضًا طريقة ناجحة لاختيار المتغيرات، والشبكة المرنة تعمل بشكل "جيد" مع تأثير تجمع المتغيرات المستقلة والمترابطة، على عكس اللاسو. في عام ٢٠٢٢ أجرى الباحث (Saleem) دراسة هدفت إلى معالجة مشكلة البيانات عالية الأبعاد في مصفوفة تصميم التجربة العاملية التطبيقية باستخدام التصميم العشوائي الكامل بعد تحويل النموذج الرياضي لتصميم التجارب ومن ثم إلى نموذج انحدار من خلال تمثيل مستويات العوامل والتفاعلات بينها. حيث ان مستويات هذه العوامل تمثل المتغيرات تفسيرية ويتم اختيار العوامل المهمة باستخدام الأساليب الجزائية. في هذا البحث تم استخدام طريقة lenth وبعض طرق اختيار المتغيرات، مثل اللاسو ومجموعة لاسو والشبكة المرنة لنمذجة التجارب العاملية، وقد تم عرض نتائج عمل هذه الطرق من خلال بعض تجارب المحاكاة.

٢: نموذج الانحدار للتجربة العاملية الكاملة

Regression Model of Full Factorial Experiment

يعتبر تصميم التجارب العاملية اداة فعالة لدراسة عاملين او أكثر من عاملين. ان أفضل وصف لدراسة تأثير العامل هو عى انه التغيير في الاستجابة الناتج عن التغيير في مستوى العامل؛ وهذا ما يسمى التأثير الرئيسي. في العديد من التجارب العاملية، يكون التمييز في وحدة الاستجابة بين مستويات عامل واحد مختلفًا عن المستويات الكاملة للعوامل الأخرى. وهذا ما يسمى تاثير التفاعل بين العوامل. وبالتالي، يشار إلى التأثيرات الرئيسية للعوامل ونتيجة التفاعل باسم التاثيرات العاملية (Wu and Hamada, 2009). الميزة الأكثر أهمية في التصميم العاملي الكامل هي أنه يمكنه تقدير جميع التأثيرات الرئيسية بالإضافة إلى التفاعلات ذات الرتب العالية. تحليل الانحدار هو أداة إحصائية معروفة يمكن استخدامها لتمثيل التأثيرات الرئيسية للعوامل وتأثيراتها التفاعلية لتصميم عاملي ذو مستويين. للتوضيح، لنفترض أن لدينا تجربة عاملية تتكون من أربعة عوامل بمستويات تساوي اثنين لكل عامل, $٤^{٢}$. تمثل الأحرف الكبيرة A وB وC وD العوامل، ولدينا $١٦=٤^{٢}$ معالجة. نموذج الانحدار الذي يمثل التأثيرات والتفاعلات الرئيسية للتجربة العاملية الكاملة هو:

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_4 x_4 + \beta_{12} x_1 x_2 + y_i$$
$$\cdots + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 + \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 +$$
$$\cdots + \beta_{234} x_2 x_3 x_4 + \beta_{1234} x_1 x_2 x_3 x_4 + \varepsilon \dots \dots . (1)$$

حيث y_i؛ i=1,2,…,n هو متغير الاستجابة، وβ هي معاملات غير معروفة، X1,X2,X3,X4 تمثل العوامل A وB وC وD، وε هو حد الخطأ العشوائي. يتم ترميز المتغيرات المشتركة X1،X2،X3،X4، على أنها ١ و−١، وهي تمثل عوامل المستوى العالي والمستوى المنخفض. أيضًا، يتم تعريف التفاعل بين X1 وX2 على أنه X1X2، وهكذا بالنسبة للعوامل الأخرى. طريقة المربعات الصغرى المستخدمة لإيجاد تقديرات المعلمات لنموذج الانحدار. يمكن تعريف مقدر المربعات الصغرى العادية (OLS) على النحو التالي:

$$\hat{\beta}_{ols} = (x^T x)^{-1} x^T y$$

من حيث تدوين المصفوفة يمكن تعريف نموذج الانحدار على النحو التالي:

$$,Y = X\beta + \epsilon$$

$$, \quad X = \begin{bmatrix} 1 & X_{11} X_{12} \dots X_{1k} \\ 1 & X_{21} X_{22} \dots X_{1k} \\ . & . \quad . \dots \quad . \\ . & . \quad \dots \quad . \\ 1 & X_{n1} X_{n1} \dots X_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

X عبارة عن مصفوفة يمكن إعادة كتابتها في سياق متغيرات التصميم بترتيب قياسي في التصميم العاملي 2^K وان (k=4)،

$$X = [1\ C_1 C_2 C_3 C_4 C_{12} C_{13} C_{14} C_{23} C_{24} C_{34} C_{123} C_{124} C_{134} C_{234} C_{1234}]$$

حيث يكون كل C_i عبارة عن متجه عمودي (تركيبة خطية) لتأثير معين. بسبب الطبيعة المتعامدة للتصميم العاملي، يمكننا كتابة الحقيقة التالية:

$$X^T X = m\ 2^k I_{2^k}$$

$$(X^T X)^{-1} = \frac{1}{2^k} I_{2^k}$$

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

$$\begin{bmatrix} 1^T Y \\ C_{11}^T Y \\ C_{12}^T Y \\ . \\ . \\ . \\ C_{nk}^T Y \end{bmatrix}, \qquad \hat{\beta} = (X^T X)^{-1} X^T Y = \frac{1}{2^k}$$

الاستجابة المتوقعة هي $\hat{Y}$=Xβ̂=، و $\hat{y}_i = C_i^T \hat{\beta}$ هي الاستجابة المتوقعة لأي تجربة (Montgemry، ٢٠٠١).

٣: إجراء اختيار المتغيرات في الانحدار Variables Selection procedure in regression

في معظم الأبحاث، يقوم الباحثون عادة بالتحقيق في ارتباط متغير الاستجابة بمجموعة المتغيرات المتنبئة، كما أن تشخيص المتغيرات الأكثر تفسيرية والمتغيرات الغير مهمه يعني اختيار النموذج الأفضل لتمثيل البيانات. تعتبر مسألة اختيار المتغيرات التفسيرية من المواضيع المحورية في عملية بناء النموذج حيث يقوم علم الإحصاء بدراسة أهم المتغيرات التفسيرية المتضمنة في الدراسة واختيار أهم المتغيرات من إجمالي المتغيرات التفسيرية وبأقل خطأ في التقدير أو الخطأ في التنبؤ، بالإضافة إلى ايجاد النموذج ذو القدرة التفسيرية العالية، مما ينعكس على تقليل تكلفة العمليات الحسابية، بالإضافة إلى زيادة دقة تنبؤات [Tibshirani,1996] . ان طرائق اختيار المتغيرات التفسيرية الكلاسيكية، مثل الحذف الأمامي والخلفي، والاختيار التدريجي للأمام، والاختيار الخلفي هي مكلفة حسابيا وتتطلب وقتا طويلا عندما يكون عدد المتغيرات التوضيحية في النموذج كبيرا. ان صعوبة حسابها تزداد أضعافاً مضاعفة مع عدد ازدياد المتغيرات المفسرة. علاوة على ذلك [Zou and Hastie,2005] قد يؤدي تغيير بسيط في البيانات إلى اختيار نموذج مختلف اخر. بخلاف ذلك، إذا كان عدد المتغيرات المستقلة أكبر من حجم العينة، فإن ذلك يعتبر عائقًا أمام الطرق التقليدية لاختيار النموذج الافضل.

٣-١ طريقة لاسو Lasso Method

تم تطوير هذه الطريقة في الأصل للتعامل مع اجراء اختيار المتغيرات، ومع ذلك، تعمل اللاسو إلى جعل العديد من تقديرات المعلمات قريبة من الصفر، مما يؤدي إلى نموذج إحصائي أكثر تفسيرًا. لقد طور Tibshirani طريقة اللاسو في عام ١٩٩٦، ويتم تعريف مقدر اللاسو بالشكل التالي،

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

$$\widehat{\beta}_{\text{lasso}} = (y - X\beta)'(y - X\beta) + \lambda \sum_{k=1}^{p} |\beta_k| \dots\dots (2)$$

حيث ≤٠ هي معلمة الانكماش. تحدد معلمة الانكماش ما إذا كانت β متفرقة أم لا. ومن المعروف أن تقديرات مقدر لاسو قد لا تكون فريدة من نوعها.

٢-٣ Group Lasso

تعد Group Lasso امتدادًا طبيعيًا لـ Lasso وتختار المتغيرات بطريقة مجمعة. حيث تعاني lasso من ان تقديرها غير كفوء وغير المتسق. ولمعالجة هذه المشكلة، تم اقتراح طريقة المجموعة Lasso. الطريقة الجديدة قادرة على تحديد النموذج الحقيقي ، ويكون المقدر الناتج كفوء ويتميز بخصائص Oracle. تمت دراسة هذه الطريقة (Group Lasso) عام ٢٠٠٦ من قبل الباحثين (Yuan و Lin) والتي تستخدم مع المتغيرات التنبؤية التي لها بنية هيكلية مجمعة. وليس من الضروري أن تكون المتغيرات التنبؤية مرتبطة ببعضها البعض بالضرورة، ولكن يجب أن يكون هيكلها عبارة عن هيكل مجمع. وبسبب عيوب Lasso، كان اللجوء إلى Group Lasso مناسبًا للنماذج ذات الاستجابات متعددة الطبقات. وهذا يعني أن جميع المعاملات المرتبطة بمتنبئ مشترك تشكل مجموعة وتضطر قيمها إلى التخفيض إلى الصفر في وقت واحد في آلية اختيار المتغير. ان مقدر مجموعة Lasso هو كما يلي:

$$\hat{\beta} = argmin\beta \, \|y - x\beta\|_2^2 + \lambda \sum_{g=1}^{G} \|\beta_g\| s_g \dots\dots (3)$$

٣-٣ طريقة الشبكة المرنة Elastic Net

تم اقتراح الشبكة المرنة (EN) من قبل الباحثين (Zou and Hastie, 2005) لاختيار مجموعات من المتغيرات المترابطة، حيث إن انحدار الشبكة المرنة (EN) يعزز القدرة التنبؤية لـ Lasso من خلال ربط طريقة Lasso مع التأثير التجميعي لنموذج ridge. علاوة على ذلك، ينتج Lasso تنبؤًا ضعيفًا عندما يكون p > n ومع الارتباط الزوجي بين المتغيرات المشتركة، كما لوحظ تجريبيًا أن أداء تنبؤات Lasso يتم التغلب عليه من خلال انحدار ridge (Zou and Hastie، 2005). اقترحوا طريقة الشبكة المرنة (EN) وكشفوا أن الشبكة المرنة (EN) تتفوق على اللاسو عندما تكون المتغيرات المشتركة مرتبطة بشكل كبير. ان مقدر طريقة (EN) معرف بالاتي:

$$\hat{\beta} = argmin_{\beta} \|y - x\beta\|_2^2 + \lambda_1 \|\beta_k\|_1 + \lambda_2 \|\beta_k\|_2^2 \dots\dots (4)$$

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

٣-٤ طريقة Lenth

قدم Lenth في عام (١٩٨٩) مقدرًا قويًا لخطأ التباين القياسيτ ،والذي أطلق عليه اسم تقدير الخطأ القياسي الزائف أوPSE (Pseudo Standard Error) (Hamada and Balakrishnan,1998):

$$PSE = 1.5\ median\{|ci| < 2.5\ S_0\}|ci|$$

$$S_0 = 1.5\ median|ci|.$$

أي أن PSE عبارة عن وسيط مشذب يحاول إزالة التباينات المقابلة للتأثيرات النشطة. ثم يتم تحديد التأثيرات النشطة باستخدام الهامش الخطأ PSE $ME = t_{0.975;df}$ مع درجات الحرية$df = k/3$ أو المتزامنة هامش الخطأ$SME = t_{\gamma;d}$ PSE ، اذ ان $\gamma = (1 + 0.951/k)/2$ لاحظ أن PSE هو طبيعي مقارب ويتوافق مع τ . ان درجات الحرية $k/3$ تأتي من تقريب PSE بواسطة توزيع$\chi 2$. باستخدامPSE . تعتمد طريقة lenth على ايجاد حل صفري للتأثيرات, أنه من المتوقع أن يكون عدد التأثيرات غير الصفرية أو النشطة بين التباينات في تجربة عاملية أو تجربة عاملية غير متكررة صغيرًا. بافتراض وجود عدد قليل من التأثيرات النشطة، تظهر التأثيرات النشطة كقيم متطرفة على مخطط الاحتمال. وبطبيعة الحال، فإن ما يشكل بالضبط "الحالة المتطرفة" هو أمر ذاتي إلى حد ما. (من المهم أن نلاحظ هنا أن تفسير التأثيرات التي يجب اعتبارها "حقيقية" يعتمد على فهم الجوانب الموضوعية للمشكلة، أي، كما هو الحال مع جميع التحليلات، يجب دمج المعرفة والتوقعات بالموضوع في أي تحليل للبيانات.

٤- المحاكاة Simulation

في هذا المبحث قمنا باختبار أداء إجراء اختيار المتغيرات المذكور في الفصل السابق وبالطراق (لاسو، group lasso، الشبكة المرنة، lenth) مع تصميم تجربة عاملية 2^4. تم إجراء مقارنة الأداء بين الطريقة السابقة من خلال تطبيق سيناريوهات محاكاة مختلفة من خلال لغة البرمجة R. تم تقييم طرق اختيار المتغيرت بناءً على معيار EER ومعيار متوسط مربع الخطأ (MSE). يتم تعريف صيغة MSE على النحو التالي؛

$$MSE = (\hat{\beta} - \beta)^T V (\hat{\beta} - \beta)$$

تم توليد (1000) عينة من مجموعة العوامل A وB وC وD مع تفاعلاتها. يتم تنفيذ الخوارزمية من خلال الخطوات التالية:

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

الخطوة ١: توليد متغير الاستجابة من التوزيع الطبيعي، $Y \sim N(\mu, \sigma^2)$.

الخطوة ٢: توليد مشاهدات عشوائية لحد الخطأ من التوزيع الطبيعي القياسي، $(\epsilon \sim N(0, \sigma^2)$, حيث $\sigma^2 \in (0.1, 0.3)$

الخطوة ٣: توليد مصفوفة تصميم مكونة من خمسة عوامل ومستويين (المستوى العالي [+١] والمستوى المنخفض [-١]).

الخطوة ٤: توليد عدد مختلف من العينات في كل تكرار باستخدام (n = 50,n=100,n=200).

الخطوة ٥: حساب MSE وEER وr^2 لجميع طرق الدراسة.

٤-١ تجربة المحاكاة الاولى

في تجربة المحاكاة العاملية هذه، قمنا توليد ١٥ متغيرًا (العوامل الرئيسية وتفاعلاتها) بأحجام عينات مختلفة تحت (σ^2=0.1)، وباتباع خطوات خوارزمية مونت كارلو، كانت النتائج كما يلي.

الجدول (1): المحاكاة الأولى المعلمات المقدرة لطرق التجربة العاملية عندما

$$\sigma^2 = 0.1, \quad n = 50$$

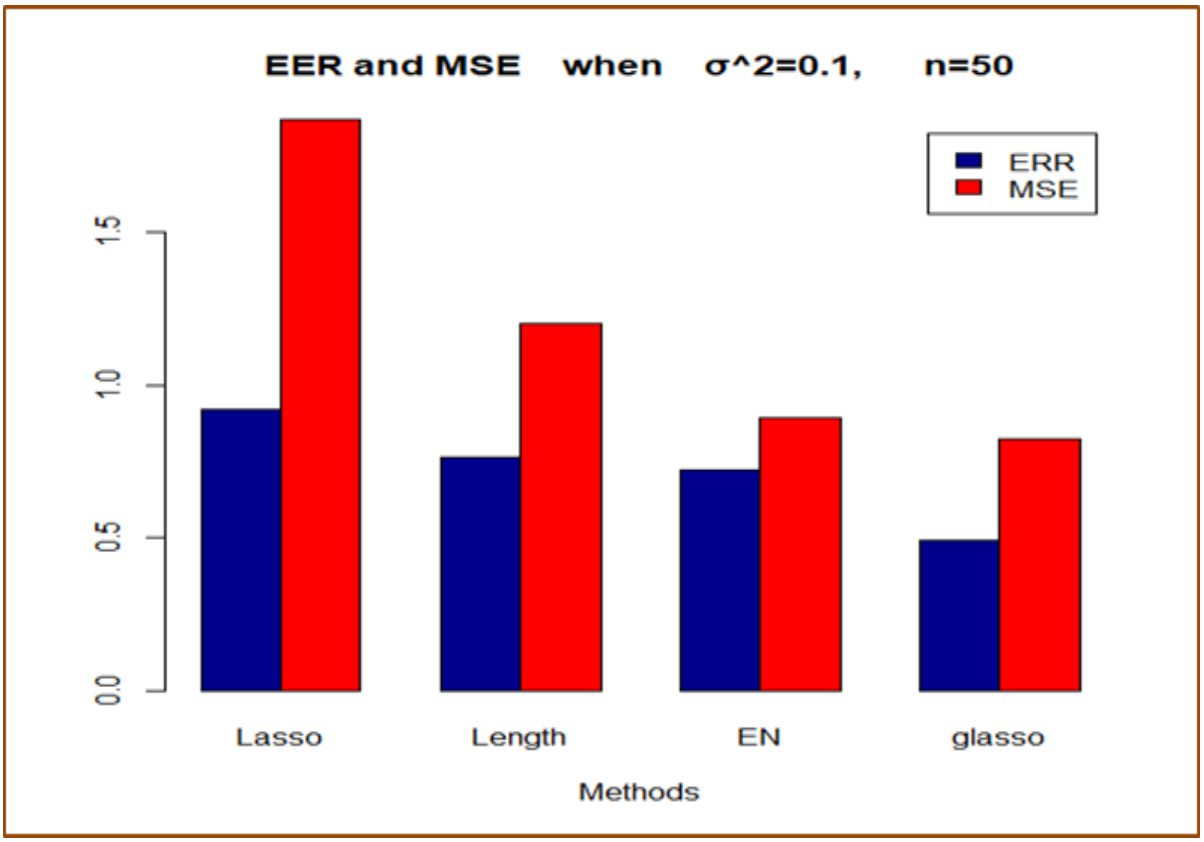
Coefficients	Variable selection methods			
	Lasso	Length	EN	G lasso
Intercept	1.280	0.643	1.185	-0.209
x_1	0.928	1.871	1.431	-0.032
x_2	2.262	-0.723	-0.686	0.806
x_3	-0.030	0.322	0.381	-1.032
x_4	2.042	0.450	1.979	0.899
x_1x_2	1.129	-0.051	1.981	-1.152
x_1x_3	-1.937	1.703	-0.358	2.636
x_1x_4	1.095	1.887	1.535	0.264
x_2x_3	2.025	0.707	3.048	2.875
x_2x_4	1.281	1.426	-0.699	0.861
x_3x_4	1.218	0.338	0.200	1.719
$x_1x_2x_3$	3.219	-0.165	-0.516	2.613
$x_1x_2x_4$	0.174	2.156	1.102	-0.320
$x_1x_3x_4$	2.485	1.933	0.919	2.049
$x_2x_3x_4$	1.978	1.135	-0.071	-0.633
$x_1x_2x_3x_4$	1.280	0.643	1.185	-0.209
EER	0.921	0.762	0.74	0.49
MSE	1.948	1.238	0.972	0.943

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

يتضح من الجدول (١) أعلاه أنه عندما تكون القيمة توليد 0.1 = مع 50 = n، نلاحظ أن أفضل طريقة لتقدير هي طريقة glasso، حيث تكون قيمة (MSE = 0.943)، مع قيمة (EER) = 0.069). وبالمقارنة مع الطرق (Lasso، Length، EN)، تأتي طريقة EN بعد القيمة الأقل وهي (MSE = 0.972)، بقيمة (EER = 0.972). ويبين الشكل (١) قيم المعايير (MSE, EER,) التي تعكس كفاءة الطرق. كما يوضح الشكل (٢) أن القيم تتمركز بالقرب من متوسطات العوامل.

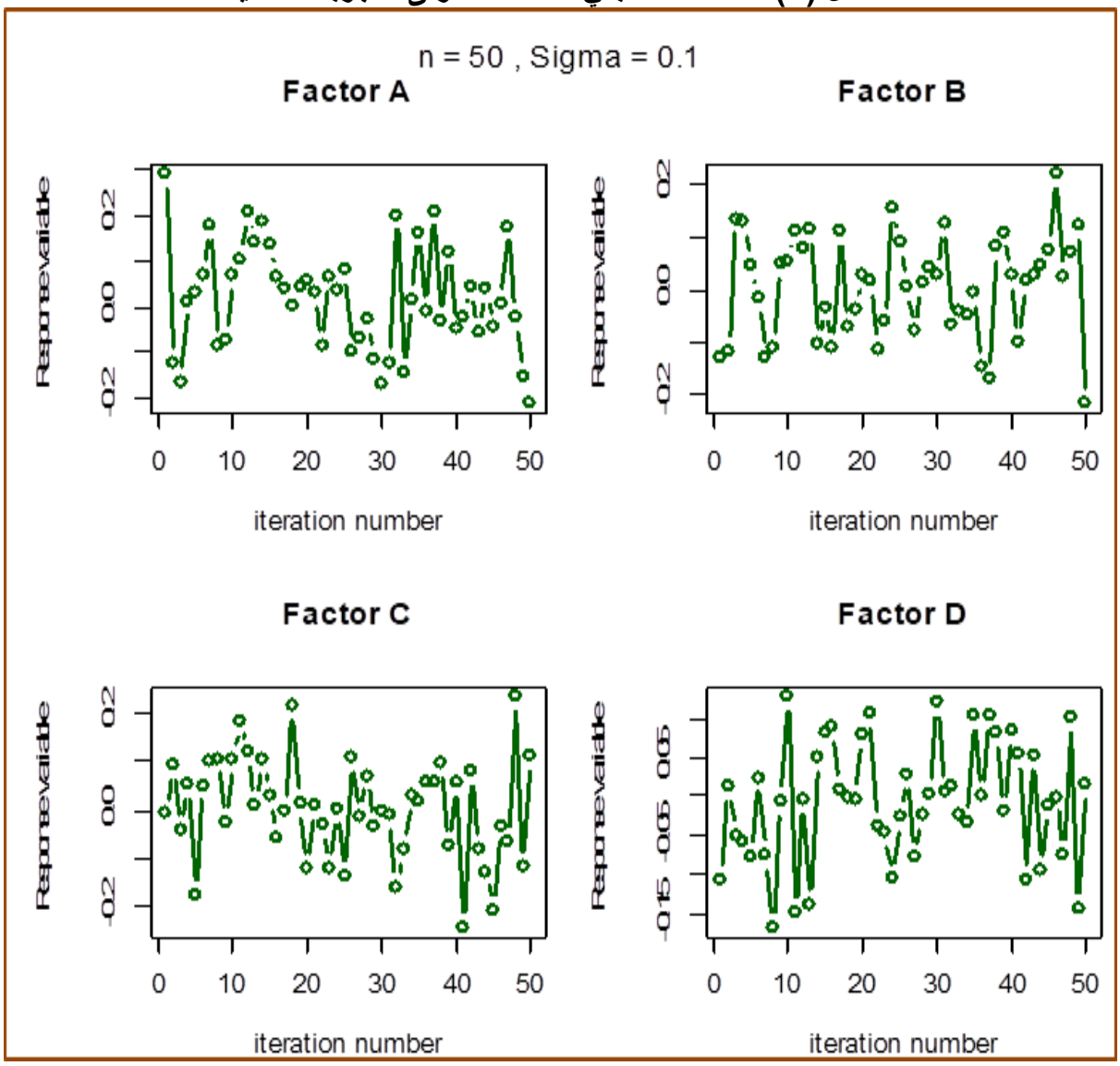
الشكل (١): محاكاة قيم MSE وEER عندما تكون σ^2=0.1



This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

الشكل (٢): المخطط التتبعي للمحاكاة الأولى للتجربة العاملية



This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

٤-٢ المحاكاة الثانية

في تجربة المحاكاة العاملية هذه، قمنا بتوليد ١٥ متغيرًا (العوامل الرئيسية وتفاعلاتها) بأحجام عينات مختلفة تحت (σ^2=0.3)، وباتباع خطوات خوارزمية مونت كارلو، كانت النتائج كما يلي:

$$\sigma^2 = 0.3, \quad n = 50$$

الجدول (٢): محاكاة اثنين من المعلمات المقدرة لطرق التجربة العاملية عندما

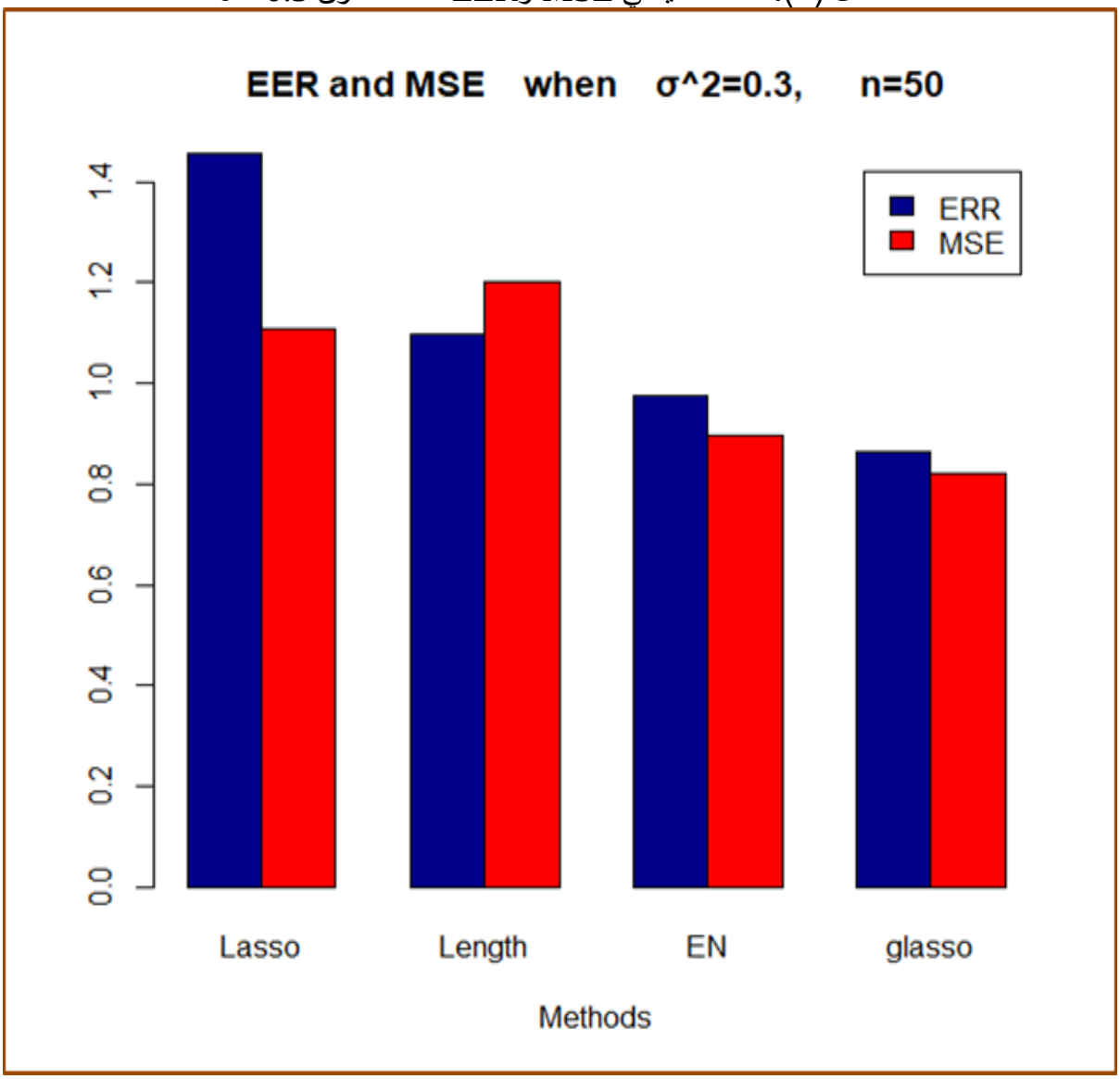
Coefficients	Variable selection methods			
	Lasso	Length	EN	Glasso
Intercept				
x_1	0.583	1.625	1.669	1.681
x_2	0.433	0.715	1.947	1.535
x_3	0.954	2.848	1.776	2.313
x_4	-0.791	0.918	-0.469	1.062
x_1x_2	2.992	0.620	1.197	3.227
x_1x_3	2.178	-0.855	2.438	2.109
x_1x_4	1.550	1.140	1.388	1.132
x_2x_3	1.024	0.759	0.849	1.636
x_2x_4	0.246	1.481	0.165	0.325
x_3x_4	-0.640	1.577	0.226	1.322
$x_1x_2x_3$	0.036	-0.752	1.312	-0.102
$x_1x_2x_4$	2.667	2.012	-0.314	0.973
$x_1x_3x_4$	-0.151	0.093	-1.315	3.139
$x_2x_3x_4$	1.701	1.787	0.899	0.792
$x_1x_2x_3x_4$	-0.338	1.426	0.324	3.182
EER	1.456	1.096	0.975	0.863
MSE	2.453	1.945	1.754	1.054

من الجدول (٢) أعلاه يتضح أنه عندما تكون قيمة σ^2= 0.3 بحجم عينة n = 50 نلاحظ أن أفضل طريقة مقترحة هي g lasso بقيمة (MSE = 1.054)) بينما بلغت قيمة (EER = 1.054). وبالمقارنة مع الطرق (Lasso، Length، EN)، وصلت طريقة EN إلى قيمة (MSE = 1.754)، في حين أن قيمة (EER = 1.754). حيث أن الشكل (٣) أدناه يوضح قيم معايير (MSE، EER) كفاءة الطريقة المقدمة مقارنة بالطرق الأخرى، ويوضح الشكل (٤) أن القيم تتمركز بالقرب من متوسطات العوامل.

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

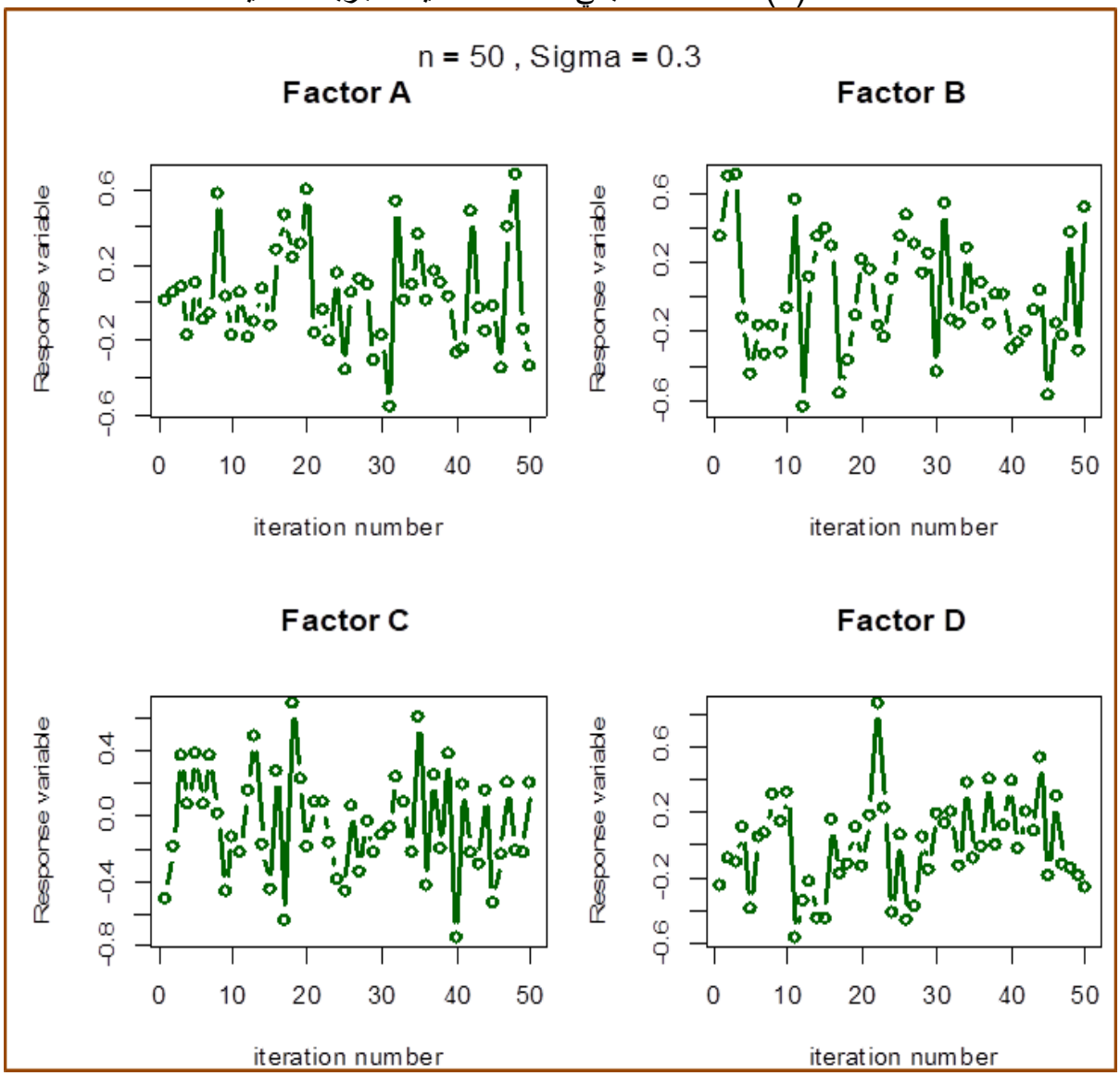
الشكل (٤): محاكاة قيمتي MSE وEER عندما تكون $\sigma^2=0.3$



This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

الشكل (٥): المخطط التتبعي للمحاكاة الثانية للتجربة العاملية



This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

٥-الاستنتاجات

استناداً إلى الجزء النظري وجزء المحاكاة ، يعتبر تصميم التجارب العاملية أداة هامة في تحليل البيانات في مجالات مختلفة مثل البيانات الزراعية والطبية والاقتصادية. ومن منظور اختيار المتغيرات نستنتج من دراسة المحاكاة أن طريقة (group lasso) تفوقت على الطرائق الأخرى وتتبعها طريقة الشبكة المرنة، من حيث اختيار أهم العوامل والتفاعل وسط هذه العوامل بالاعتماد على قيم EER ومعيار MSE. نستنتج من تجارب المحاكاة أن طريقة مجموعة لاسو تليها طريقة الشبكة المرنة أعطت أصغر قيم لكل من EER وMSE وبالتالي اثبات تفوق هذه الطرق من حيث دقة التقدير واختيار المتغيرات التي لها تاثير معنوي على عوامل التجربة العاملية.

المصادر References

1- LENTH, R.V. 1989. Quick and Easy Analysis of Unreplicated Factorials in Technometrics, Vol. 31 No.4. November 1989. pp 469-473.
2- Fisher, R.A. (1935). The Design of Experiments. Edinburgh and London: Oliver and Boyd.
3- Yates,F.(1937) ,"Design analysis of factorial Experiment", Imprial Burean of soil scienes Harpenden Engeland , Vol. 35 , pp.77.
4- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288.
5- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67, 301-320.
6- Saleem , Wisam Wadullah ,)2022(, "Modeling of High-Dimensional Factors in the Design of Factorial Experiments Using Penal Methods with Practical Application", A Dissertation Submitted to the: Council of College of Administration and Economics /University of Baghdad in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics.
7- Wu, C. F. J. and Hamada, M. (2009) Experiments: Planning, Analysis and

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

Parameter Design Optimization, 2nd edition. Wiley, New York.
8- Montgomery , D.C.,"Design-and-analysis-of-experiments" ,2001.
9- Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society Series B, 68, 49{67}.

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.