مجلة

كلية التـــراث الجامعة

مجلة علمية محكمة

متعددة التخصصات نصف سنوية

العدد السابع والثلاثون

15 حزيران 2023

ISSN 2074-5621

رئيس هيئة التحرير

أ.د. جعفر جابر جواد

مدير التحرير

أ. م. د. حيدر محمود سلمان

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم (ب 3059/4) والمؤرخ في (2014/ 4/7)

دراسة مقارنة لبعض طرائق الانحدار الحصين لتقدير متوسط المجتمع في المعاينة العشوائية الطبقية

مصطفى حبيب مهدي **أ.د. سجى محمد حسين**
قسم الاحصاء، كلية الادارة والاقتصاد جامعة بغداد

المستخلص
في هذا البحث ، تم عرض مقدرات الانحدار المنفصل لتقدير متوسط المجتمع في المعاينة العشوائية الطبقية من خلال معيار المقارنة MSE ، تمت مقارنة مقدرات متوسط المجتمع للانحدار المنفصل باستعمال الطرائق الحصينة مع مقدرات الانحدار المنفصل التقليدية . بالإضافة الى ذلك ، تم مقارنة هذه التقديرات مع المقدرات الكلاسيكية باستعمال معيار الكفاءة (RE). يعتبر مقدر الطريقة (OLS) فعالا لأنه يأخذ في الاعتبار قيم التباين المشترك للأخطاء ،ولكنه أيضا حساس جدا للقيم الشاذة . وجدنا أن المقدرات الحصينة أدت إلى تحسين جودة تقديرات الانحدار المنفصل لتقدير الوسط الحسابي للمجتمع للمعاينة العشوائية الطبقية بشكل كبير من خلال تقليل تأثير القيم الشاذة باستعمال مقدرات طرائق الانحدار الحصينة (Huber M , LMS , LAD) ومصفوفات التباين والتباين المشترك الحصينة(MVE , MCD) ، بالإضافة إلى ذلك أعطت النتائج ان طريقة (LAD و LMS) كانت الأفضل في التعامل مع القيم الشاذة الموجودة في مجموعة البيانات، حيث تمتلك قيم أقل (MSE) وأكبر (RE) .

نوع البحث : ورقة بحثية
المصطلحات الرئيسة للبحث : مقدر الانحدار المنفصل ($\bar{y}_{lrs}$) ، القيم الشاذة ، طرائق الانحدار الحصينة ، الكفاءة ، المعاينة العشوائية الطبقية .

1. المقدمة (Introduction)

في أدبيات أساليب المعاينة ، يمكن أن يؤدي استعمال المعلومات المساعدة إلى زيادة دقة المقدر عندما يرتبط متغير الدراسة **y** بشكل إيجابي مع المتغير المساعد **x** ، معلومات المجتمع للمتغيرات المساعدة ، تستعمل بشكل شائع لزيادة كفاءة التقدير لمتوسط المجتمع .
ومع ذلك ، فمن المعروف أن عند وجود القيم الشاذة في البيانات ، تتأثر المقدرات التقليدية بهذه القيم الشاذة ، وتقل كفاءتها.
عام (1996) أستعمل الباحث (Mehta)[13] طريقة جديدة لإيجاد الحدود المثلى للطبقات في المعاينة العشوائية الطبقية ، تهدف الى إيجاد كل من مقدرات النسبة والانحدار والضرب ، مستخدما أسلوب توزيع العينة المتناسب مع مجموع مجتمع .
عام (2001) اقترح الباحث (المعلم)[1] طريقة $Cum\ f_x^{\frac{3}{4}}$ لإيجاد الحدود الطبقية المثلى مستخدما كل من حصص التوزيع المتساوي والامثل ، وتبين له أن الطريقة المقترحة لتقدير تباين متوسط المعاينة الطبقية باستخدام متغيرين إضافيين للطبقية أعطت نتائج افضل منها في حالة ان متغيري الطبقية مستقلين ويخضعان للتوزيع الطبيعي .
عام 2016م اقترح الباحثون (نور الأمين وآخرون) [14] مُقَدِّرات النسبة المُقترحة لمتوسط المجتمع باستعمال الانحدار الحصين في ظل أسلوب المعاينة المزدوجة .
عام 2018م اقترح الباحثان (Zaman and Bulut) [16] مقدرات جديدة من نوع النسبة للعينات العشوائية البسيطة للتقليل من تأثير القيم الشاذة في مجموعة البيانات باستعمال طرائق الانحدار الخطي الحصينة .
عام 2020م طبق الباحثون (Zhang et al) [4] مقدر M الحصين ومقدر MM لتقدير متوسط عينة طبقية من بيانات جودة المياه . وتوصلت الى أن كلا المقدرين قدموا تقديرات أكثر دقة للمتوسط مقارنة بالمقدرات التقليدية ، مثل مقدّر النسبة ومقدّر الانحدار ، لا سيما عندما تكون العينة ذات درجة عالية من عدم تجانس التباين (heteroscedasticity) .
الهدف هو لتقدير معلمة الانحدار الخطي لتقدير متوسط المجتمع بطريقة الانحدار المنفصل باستعمال بعض الطرائق الحصينة مثل(Huber M , LMS , LAD) وتقدير مصفوفة التباين والتباين المشترك الحصينة (MVE , MCD) ، في المعاينة العشوائية الطبقية وذلك للتقليل من تأثير القيم الشاذة وزيادة كفاءة هذه الطرائق ومن ثم اجراء مقارنة لمقدرات هذه الطرائق والطريقة التقليدية (OLS) للوصول الى افضل مقدر من خلال معيار متوسط مربعات الخطأ (MSE) .

2 . تصميم المعاينة العشوائية الطبقية Stratified Random Sampling Design (STRS)

عندما تكون وحدات المجتمع غير متجانسة بشكل كلي ، عندها يمكن وضع بعض القيود على SRS للزيادة في دقة التقدير ، وذلك بالتقليل من تأثير عدم التجانس . وأبسط القيود هو تجزئة المجتمع الى مجموعات جزئية تسمى طبقات Strata إذ

تكون كل طبقة Stratum متجانسة في داخلها وتختلف عن الطبقات الأخرى . وعند تحديد عدد الطبقات تسحب عينة من كل طبقة ويتم السحب بصورةٍ مستقلة وبشكل عشوائي في الطبقات المختلفة . [2] [20]

3 . مقدرات الانحدار الخطي لتقدير متوسط المجتمع في المعاينة العشوائية الطبقية

Linear Regression Estimators to Estimate the population Mean in Stratified Random Sampling

ان التقدير لمتوسط المجتمع في المعاينة العشوائية الطبقية ، بالاستناد على الانحدار الخطي لـ Y_i على X_i ، يكون من خلال مقدر الانحدار المنفصل ($\bar{y}_{lrs}$) : - [21]

3.1 مقدر الانحدار المنفصل Separate regression Estimator ($\bar{y}_{lrs}$)

في هذا المقدر ، تم استعمال مقدر الانحدار للعينات العشوائية البسيطة على مستوى الطبقات وهذا يعني أن لكل طبقة على حدة ، يتم تقدير متجه مع معاملات الانحدار المجتمع b_h ، يتم بعد ذلك تجميع تقديرات الانحدار الخاصة بكل طبقة عن طريق حساب المتوسط المرجح ، باستعمال الأحجام النسبية للطبقات كأوزان :[12]

وهذا النوع من التقدير يستخدم في حالة وجود ادلة كافية للافتراض بأن القيم الحقيقية (B_h) في كل الطبقات متجانسة فيما بينها .

نحسب تقدير الانحدار لمتوسط كل طبقة ، أي :

$$\bar{y}_{lrh} = \bar{y}_h + b_h (\bar{X}_h - \bar{x}_h) \qquad (1)$$

حيث أن :

معلمة مقدر الانحدار المنفصل b_h ، يتم تقديرها باستعمال طريقة المربعات الصغرى (OLS) .

$$b_h = \frac{S_{xyh}}{S_{xh}^2}$$

s_{xh}^2 هو تباين عينة المتغير المساعد في الطبقة h .

$$\boldsymbol{s_{xh}^2 = \frac{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}{n_h}}$$

s_{xyh} هو التباين المشترك بين المتغير المساعد ومتغير الدراسة في الطبقة h .

$$\boldsymbol{s_{xyh} = \frac{\sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{n_h - 1}}$$

متوسطات العينة الطبقية للمتغيرين (**x** و **y**) .

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

$$\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$$

متوسط المجتمع الطبقي للمتغير X .

$$\bar{X}_h = \frac{\sum_{i=1}^{N_h} X_{hi}}{N_h}$$

وزن الطبقة للطبقة h .

$$W_h = \frac{N_h}{N}$$

وعندئذ يكون متوسط الانحدار الطبقي المنفصل :

$$\hat{\bar{Y}}_{lrs} = \sum_h W_h \bar{y}_{lrh} \qquad (2)$$

$\hat{\bar{Y}}_{lrs}$ هو تقدير غير متحيز لمتوسط المجتمع $\bar{Y}$.

ويكون هذا التقدير مناسبا عندما يبدو لنا ان معامل الانحدار للمجتمع B_h يتغير من طبقة الى طبقة .

وبما أن المعاينة مستقلة في الطبقات المختلفة فإن تباين هذا المقدر :

$$V (\hat{\bar{Y}}_{lrs}) = \sum_h \frac{W_h^2 (1 - f_h)}{n_h} (S_{yh}^2 - 2 b_h S_{yxh} + b_h^2 S_{xh}^2) \qquad (3)$$

ويكون أصغر ما يمكن عندما يكون مساوياً لمعامل الانحدار للمجتمع ، وبذلك فان اصغر تباين على الشكل التالي :

$$V\left(\hat{\bar{Y}}_{lrs}\right) = \sum_{h} \frac{W_h^2\,(1-f_h)}{n_h}\left(S_{yh}^2 - \frac{S_{yxh}^2}{S_{xh}^2}\right) \qquad (4)$$

حيث ان :

f_h كسر المعاينة ضمن الطبقة h .

$$f_h = \frac{n_h}{N_h}$$

S_{yh}^2 هو تباين المجتمع لمتغير الدراسة في الطبقة h .

$$\boldsymbol{S_{yh}^2 = \frac{\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2}{N_h - 1}}$$

S_{xh}^2 هو تباين المجتمع للمتغير المساعد في الطبقة h .

$$\boldsymbol{S_{xh}^2 = \frac{\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)^2}{N_h - 1}}$$

S_{yxh} هو التباين المشترك بين المتغير المساعد ومتغير الدراسة في الطبقة h .

$$\boldsymbol{S_{xyh} = \frac{\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)(y_{hi} - \bar{Y}_h)}{N_h - 1}}$$

الدرجة الأولى من التقريب إلى الـ $\mathrm{MSE}\,(\hat{\bar{Y}}_{lrs})$ كالاتي : [5]

$$\mathrm{MSE}\left(\hat{\bar{Y}}_{lrs}\right) \cong \sum_{h=1}^{l} W_h^2\,\lambda_h\left[S_{yh}^2 + \beta_h^2\,S_{xh}^2 - 2\beta_h S_{yxh}\right] \qquad (5)$$

4 . القيم الشاذة في البيانات outliers in data

تعد مشكلة القيم الشاذة *outliers* واحدة من اهم واقدم المشكلات في علم الاحصاء , ولعلها احدى المشكلات الرئيسية لطرائق تقدير الانحدار , حيث يؤثر ظهور القيم الشاذة في مجموعة البيانات على نتيجة التحليل الاحصائي للبيانات , مما يؤثر على اتخاذ القرار المناسب , المشاهدات التي تكون منفصلة عن الجزء الاكبر من البيانات تسمى (القيم الشاذة outliers), ان الوعي بمشكلة القيم الشاذة في البيانات نشأ في وقت مبكر جداً ارتبط بأسلوب (OLS) .[17]

وفي عام 1986 عرفها كل من (Hampel Ronchetti , Rousseuw and Stahel) [19]على انها المشاهدة التي تحيد عن النمط الذي تسلكه غالبية البيانات .

وفي عام **1990** عرفها كل من (**Staudte and Sheather**) **[18]**على انها المشاهدة التي تكون ابعد ما يكون عن تجمع الجزء الأكبر من البيانات وبمعنى اخر يمكن تعريفها على انها تلك المشاهدة التي تبدوا غير منطقية اذا ما قورنت مع بقية بيانات المجموعة أي انها تلك المشاهدة التي تأتي من مجتمع يختلف عن بقية المشاهدات وهذا الاختلاف نتيجة عدة عوامل منها :

1. خطأ التسجيل .
2. خطأ القياس .
3. خطأ المعاينة .

5 . الطرائق الحصينة لتقدير معلمة الانحدار المنفصل

للحصول على تقدير لمتوسط المجتمع من خلال مقدرات الانحدار الطبقي المنفصل باستعمال معلمة انموذج الانحدار الحصين الاتي :

$$\hat{\bar{\mathbf{Y}}}_{\mathbf{lrs(zb)}} = \sum_{\mathbf{h=1}}^{\mathbf{l}} \mathbf{W_h}\left[\bar{\mathbf{y}}_{\mathbf{h(j)}} + \mathbf{b_{hrob(zb)}}\left(\bar{\mathbf{X}}_{\mathbf{h(j)}} - \bar{\mathbf{x}}_{\mathbf{h(j)}}\right)\right] \qquad (6)$$

حيث ان : $\mathbf{b_{hrob(zb)}}$ هو معامل الانحدار الذي يتم الحصول عليه من طرائق الانحدار الحصينة :
(Huber M , LMS , LAD)

والـ (**zb**) تعني اختصاراً لهذه الطرائق المذكورة اعلاه ، وايضاً الـــ ($\bar{\mathbf{X}}_{\mathbf{h(j)}}\bar{\mathbf{y}}_{\mathbf{h(j)}}$, $\bar{\mathbf{X}}_{\mathbf{h(j)}}$, $\bar{\mathbf{x}}_{\mathbf{h(j)}}$) يتم حسابها باستعمال مقدرات (MCD و MVE) .

$$\mathbf{MSE}\left(\hat{\bar{\mathbf{Y}}}_{\mathbf{lrspr(zbj)}}\right) \cong \sum_{\mathbf{h=1}}^{\mathbf{l}} \mathbf{W_h^2}\,\lambda_{\mathbf{h}}\left[\mathbf{S^2_{yh(j)}} + \boldsymbol{\beta}^{\mathbf{2}}_{\mathbf{hrob(zb)}}\,\mathbf{S^2_{xh(j)}} - \mathbf{2}\boldsymbol{\beta}_{\mathbf{hrob(zb)}}\,\mathbf{S_{yx(j)}}\right] \qquad (7)$$

5.1 طريقة M

تعد طريقة M واحدة من الطرائق الحصينة الاكثر شيوعاً والمستعملة في تقدير معلمات نموذج الانحدار الخطي ، والتي اقترحت من قبل هوبر(Huber 1973) حيث تعمل على اختزال تأثير الشواذ حيث انه بدلاً من تقليل مجموع مربعات الخطأ في دالة الهدف يتم تصغير دالة الهدف [11] .

بالنسبة لدالة الهدف ρ سوف نستعمل دالة Huber كالاتي :

$$\rho(e) = \begin{cases} \dfrac{e^2}{2} & , \quad |e| \le k \\ k|e| - \dfrac{k^2}{2} & , \quad |e| > k \end{cases} \qquad (8)$$

5.2 طريقة انحدار الانحرافات المطلقة الصغرى Least Absolute Deviations Regression (LAD)

يعتبر انحدار الانحرافات المطلقة الأقل (**LAD**) أحد البدائل الرئيسية لطريقة المربعات الصغرى العادية (**OLS**) عندما يسعى المرء إلى تقدير معاملات الانحدار . [3]

الهدف من انحدار **LAD** هو توفير مقدر قوي يتم تصغيره إلى الحد الأدنى من مجموع الاخطاء المطلقة . [5]

$$\min \sum_{i=1}^{n} |r_i| \qquad (9)$$

5.3 طريقة انحدار مربعات الوسيط الصغرى (LMS)

تتكون طريقة انحدار المربعات الصغرى العادية (OLS) من تقليل مجموع مربعات الخطأ ، لذلك قام العديد من الباحثين باقتراح طرق أكثر حصانة من هذا المقّدر عن طريق استبداله بالمقّدرLMS وهو مجموع متوسط مربعات البواقي ، تم اقتراحه من قبل Rousseeuw وتم تحسينه بواسطة (Rousseeuw and Leroy , 1987) . يمكن لهذا المقّدر ان يقاوم تأثير ما يقارب من 50% من تلوث البيانات . (Rousseeuw , 1984) . [8] [10]

بمعنى آخر ، الدالة التي سيتم تصغيرها في طريقة LMS هي

$$\text{Min median}\,(\varepsilon_i^2) \qquad (10)$$

6 . طرائق تقدير مصفوفة التباين والتباين المشترك الحصينة

6.1 مقدر اصغر قطع بيضوي Minimum Volume Ellipsoid Estimator (MVE)

من المعروف جيداً ان العمليات الاحصائية التقليدية تكون حساسة جداً تجاه الشواذ في البيانات، لذلك هناك طرائق بديلة طورت للتعامل مع الاخطاء في النموذج والتلويث في البيانات .

المشكلة الرئيسية في الاحصاءات الحصينة هي في مقدرات الموقع والتشتت في توزيعات متعددة المتغيرات .

ومن هذه المقدرات مقـــدر اصغـــر قطـــع بيضوي Minimum Volume Ellipsoid Estimator (MVE) غالباً ما يستخدم لتقدير الموقع والتشتت في متعدد المتغيرات. مقدر (MVE) لمصفوفة التباين المشترك يعرف على انه اصغر شكل بيضوي يحوي نصف المشاهدات، بينما تقدير الموقع في (MVE) فهو نقطة الوسط لذلك الشكل البيضوي .

مقدرات (MVE) بالإمكان حسابها بتصغير مقياس معين في الابعاد العالية . في التطبيق بالإمكان استخدام خوارزمية تعتمد على تقليل دالة الهدف Object Function في سلسلة من محاولات التقدير .

وهذا المقدر MVE يمتلك نقطة انهيار تساوي $\frac{(\lfloor \frac{n}{2} \rfloor - P + 1)}{n}$ وعندما لا تكون $n \to \infty$ فأن نقطة انهيار مقدر MVE تساوي 50 % . (Rousseeuw 1985) . [6]

6.2 مقدر اصغر محددة تباين مشترك Minimum Covariance Determinant estimator (MCD)

الحد الأدنى لمقدّر التغاير (MCD) هو طريقة إحصائية حصينة تُستخدم لتقدير معلمات التوزيع الطبيعي متعدد المتغيرات. تم تقديمه بواسطة (Peter J. Rousseeuw, 1984) كتحسين على تقديرات التباين والتباين المشترك الحصينة الحالية .

تم تصميم مقدر MCD لتحديد المجموعة الفرعية من المشاهدات الأكثر تمثيلا لمصفوفة التباين والتباين المشترك الحقيقي للتوزيع الأساسي . تعمل الطريقة عن طريق حساب محدد مصفوفة التباين والتباين المشترك لجميع المجموعات الفرعية المحتملة للحجم p ، حيث p هو عدد المتغيرات في البيانات. يتم اختيار المجموعة الفرعية ذات المحدد الأصغر كتقدير لمصفوفة التباين والتباين المشترك . [9]

يتميز مقدر MCD بالعديد من المزايا مقارنة بمقدرات التباين والتباين المشترك الحصينة الأخرى . إنها عالية الكفاءة ، مما يعني أنها تتطلب حجم عينة صغير نسبيًا لتحقيق نتائج دقيقة. كما أنه مقاوم للغاية للقيم الشاذة ، مما يعني أنه يمكنه تقدير مصفوفة التباين المشترك بدقة حتى لو كانت نسبة كبيرة من البيانات ملوثة بقيم شاذة .

يستخدم مقدر MCD بشكل شائع في التمويل ، حيث يتم استعماله لتقدير مصفوفة التباين والتباين المشترك لعائدات الأصول . يتم استعماله أيضًا في مجالات أخرى ، مثل الهندسة وعلم الأحياء وعلم النفس ، حيث يتم العثور على البيانات متعددة المتغيرات بشكل شائع . [15] [7]

7 . معيار المقارنة بين الطرائق الحصينة Criterions of Comparison Between Robust Methods

لغرض التوصل للمقدر الاكفأ وجدت عدة معايير (مقاييس) للمقارنة بين طرائق التقدير وان افضلها من يمتلك اقل خطأ ممكن ، لذلك هنالك العديد من مقاييس المقارنة المستعملة بين المقدرات الحصينة واحد من هذه المعايير هو :

متوسط مربعات **(MSE)** . [5] pp 5

$$MSE = \frac{1}{500}\sum_{i=1}^{500}(\hat{\hat{Y}}_i - \bar{Y})^2 \qquad (11)$$

8. الجانب التجريبي

في ها الجانب استعملنا برنامج اللغة R (الإصدار (4.1.3)) في إجراء التحليل الإحصائي لتحقيق أهداف الدراسة .للمقارنة بين المقدرات الحصينة واختبار الافضل منها بالاعتماد على معيار متوسطات مربعات الخطأ (MSE) .

يتم وصف تجارب المحاكاة على النحو الآتي :

1. تم تحديد القيم الافتراضية لمعلمة الانحدار المنفصل من البيانات الحقيقية عبر تقديرها باستعمال طريقة (OLS) .
2. اختيار خمسة احجام مختلفة للعينات المفترضة (200 ,75, 35) بحجم مجتمع (700) .
3. توليد المتغيرات التوضيحية X_{ij} لخمس طبقات (متغيرات) توضيحية واعتماد نسب مختلفة لتلوث (5% , 10% , 30%) .
4. توليد الأخطاء العشوائية وفقاً للتوزيع الطبيعي للمشاهدات الاعتيادية $\varepsilon_i \sim N(0,1)$ والمشاهدات الشاذة : $\varepsilon_i \sim N(18,10)$

9. مناقشة نتائج المحاكاة

ومن خلال النتائج المبينة في الجداول (1) و(2) و(3) نلاحظ ما يلي :

عند حساب (MSE) لمتوسط العينة الطبقية بالانحدار المنفصل وعند توظيف مقدر المعلمة β بطريقة OLS ، وطرائق الانحدار الحصينة (Huber M , LMS, LAD) بتطبيق الصيغة (7) يكون الاتي :

1. باستعمال مصفوفات التباين والتباين المشترك الحصينة (MCD) عند تطبيق الصيغة (7) نحصل على ان طريقة (LAD و LMS) افضل كفاءة من طرائق الانحدار الحصينة الأخرى من حيث امتلاكهم إلى أقل (MSE) عند مقارنتهم مع باقي الطرائق ولجميع احجام العينات ونسب القيم الشاذة (5% , 10% , 30%)
2. باستعمال مصفوفات التباين والتباين المشترك الحصينة (MVE) عند تطبيق الصيغة (7) نحصل على ان طريقة (LAD و LMS) افضل كفاءة من طرائق الانحدار الحصينة الأخرى من حيث امتلاكهم إلى أقل (MSE) عند مقارنتهم مع باقي الطرائق ولجميع احجام العينات ونسب القيم الشاذة (5% , 10% , 30%)
3. في جميع الجداول اعلاه ، نلاحظ ان قيم الكفاءة النسبية لمقدرات طرائق الانحدار الحصينة (LAD ، LMS، Huber M) ومصفوفات التباين المشترك الحصينة(MCD وMVE) ، فيما يتعلق بمقدر الانحدار المنفصل لكل نسب القيم الشاذة وأحجام العينات المفترضة ، تم حساب هذه القيم وفقاً للصيغة (11) .

9. نتائج المحاكاة في حالة تساوي احجام الطبقات :

$$(N_1 = N_2 = N_3 = N_4 = N_5 = 140)$$

الجدول (1)

قيم (MSE , RE) لتقدير متوسط المجتمع بالانحدار المنفصل ($\widehat{\bar{Y}}_{lrs}$) عند تقدير معلمة الانحدار b_h بالطرائق (MCD , MVE) حسب حجوم العينات (n=35,75, 200) و نسبة القيم الشاذة (5 %)

n	Methods	The outliers rate is 5%			
		MCD Estimators		MVE Estimators	
		MSE	RE	MSE	RE
35	OLS	0.030286	19.09	0.030681	18.84
	LAD	0.024936	23.18	0.025209	22.93
	LMS	0.025749	22.45	0.02606	22.18
	Huber M	0.234306	2.47	0.239773	2.41
	Huber MM	0.247965	2.33	0.254407	2.27
75	OLS	0.011603	21.94	0.011727	21.70
	LAD	0.010899	23.35	0.011023	23.09
	LMS	0.011243	22.64	0.01138	22.37
	Huber M	0.087453	2.91	0.087696	2.90
	Huber MM	0.090973	2.80	0.091229	2.79
200	OLS	0.003675	20.68	0.003704	20.52
	LAD	0.003298	23.05	0.003323	22.87
	LMS	0.003405	22.32	0.003432	22.15
	Huber M	0.022116	3.44	0.022211	3.42
	Huber MM	0.022196	3.42	0.022304	3.41

الجدول (2)

قيم (MSE , RE) لتقدير متوسط المجتمع بالانحدار المنفصل ($\widehat{\bar{Y}}_{lrs}$) عند تقدير معلمة الانحدار b_h بالطرائق (MCD , MVE) حسب حجوم العينات (n=35, 75,200) و نسب القيم الشاذة (10 %)

n	Methods	The outliers rate is 10 %			
		MCD Estimators		MVE Estimators	
		MSE	RE	MSE	RE
35	OLS	0.030463	35.81	0.030784	35.43
	LAD	0.025154	43.36	0.025411	42.93
	LMS	0.025956	42.02	0.026232	41.58
	Huber M	0.283099	3.85	0.288233	3.78
	Huber MM	0.288896	3.78	0.294482	3.70
75	OLS	0.016536	28.46	0.016689	28.20
	LAD	0.011074	42.50	0.011213	41.97
	LMS	0.011383	41.35	0.011532	40.81
	Huber M	0.161809	2.91	0.163393	2.88
	Huber MM	0.173371	2.71	0.175146	2.69
200	OLS	0.00405	35.32	0.004102	34.88
	LAD	0.003323	43.05	0.003371	42.44
	LMS	0.003422	41.81	0.003471	41.22
	Huber M	0.049171	2.91	0.049549	2.89
	Huber MM	0.053351	2.68	0.053756	2.66

الجدول (3)

قيم (MSE , RE) لتقدير متوسط المجتمع بالانحدار المنفصل ($\widehat{\bar{Y}}_{lrs}$) عند تقدير معلمة الانحدار b_h بالطرائق (MCD , MVE) حسب حجوم العينات (n = 35 , 75 , 200) و نسب القيم الشاذة (30 %)

n	Methods	The outliers rate is 30 %			
		MCD Estimators		MVE Estimators	
		MSE	RE	MSE	RE
35	OLS	0.037148	72.51	0.037358	72.10
	LAD	0.026671	100.99	0.026794	100.53
	LMS	0.027113	99.34	0.027236	98.90
	Huber M	0.393154	6.85	0.393735	6.84
	Huber MM	0.374704	7.19	0.375218	7.18
75	OLS	0.015582	75.26	0.015644	74.96
	LAD	0.011675	100.44	0.011709	100.15
	LMS	0.011894	98.59	0.011929	98.30
	Huber M	0.245018	4.79	0.244701	4.79
	Huber MM	0.248411	4.72	0.248089	4.73

200	OLS	0.004729	74.60	0.004753	74.22
	LAD	0.003498	100.85	0.003526	100.05
	LMS	0.003558	99.15	0.003588	98.32
	Huber M	0.057363	6.15	0.05759	6.13
	Huber MM	0.055586	6.35	0.055816	6.32

. الاستنتاجات (Conclusions)

1. ان تقديرات متوسط المجتمع عند استعمال تقديرات الانحدار الحصينة ومصفوفات التباين المشترك الحصينة في المعاينة العشوائية الطبقية هي تقديرات أكثر كفاءة من التقديرات التقليدية عندما تكون هناك قيم شاذة في مجموعات البيانات .
2. أن المقدرات الحصينة هي أكثر كفاءة من مقدرات الانحدار المنفصلة التقليدية لجميع نسب القيم الشاذة وأحجام العينات .

المصادر :

1. المعلم ، سعيد عوض سعيد (2001) " الأمثلية في استراتيجية متوسط المعاينة الطبقية دكتوراه فلسفة في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد .
2. الناصر ، حمزة عبدالمجيد، صفاء يونس الصفاوي ، (2001) العينات نظري وتطبيقي ، جامعة بغداد ، جامعة الموصل ، وزارة التعليم العالي .

References

3. Nadia, H., and A. A. Mohammad. 2013. "Model of robust regression with parametric and nonparametric methods" . Mathematical Theory and Modeling .
4. Zhang, Y., Chen, X., Li, L., & Zhao, L. (2020). "A review of the active compounds, pharmacological effects", and potential therapeutic target of Lycium barbarum fruit (Goji). Evidence-Based Complementary and Alternative Medicine, 2020, 1-16. https://doi.org/10.1155/2020/6762374
5. Zaman, T., and H. Bulut. 2019. " Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling" , Communications in Statistics - Theory and Methods .
6. Rousseeuw, P.J. (1985) " Multivariate estimation with high breakdown point " . Math Stat Appl B, pp. 283–297 .
7. Bulut, H., and Y. Oner. 2017." The evaluation of socio-economic development of development agency regions in Turkey using classical and robust principal component analyses". Journal of Applied Statistics 44 (16):2936–48.
8. Rousseeuw, P.J., Leroy ,A. M.(1987) " Robust Regression and Outliers Detection " John Wiley , New York .
9. Rousseeuw, P. J., Driessen, K. V. (1999) "A fast algorithm for the minimum covariance determinant estimator". Technometrics, vol 41(3) , pp. 212-223
10. Rousseeuw, P. J. (1984). " Least median of squares regression". Journal of the American Statistical Association, 79, 871–880 .
11. Chen C., (2002), " Robust regression and outlier detection with the Robustreg procedure " In Proceedings of the Twenty Seventh Annual SAS Users Group International Conference; SAS Institute: Cary, NC.
12. Dick J. Brus . " Spatial Sampling with R " . CRC Press is an imprint of Taylor & Francis Group, LLC, Boca Raton London New-York , 2022 .
13. Mehata , S.K.,Kish0re,L.and Singh,R.(1996) "On optimum stratification for allocation proportional to strata totals ", Journal of Indian Statistical association , 34 , 9 – 19 .

14. Noor-Ul-Amin, M., M. Q. Shahbaz, and C. Kadılar. 2016. " Ratio estimators for population mean using robust regression in double sampling ". Gazi University Journal of Science 29 (4):793–8
15. Zaman, T., and H. Bulut. 2019. " Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling " , Communications in Statistics - Theory and Methods,
16. Zaman, T., and H. Bulut. 2018. " Modified ratio estimators using robust regression methods ". Communications in Statistics - Theory and Methods.
17. Paul, R. K. (2001). " Some Methods of Detection of Outliers in Linear Regression Model " . Iasri, Library Avenue, New Delhi-110012.63 .
18. Robert G. Staudte, and Simon J. Sheather. (1990). "Robust Estimation and Testing" A Wiley-Interscience Publication John Wiley & Sons, Inc . New York. [71]- Whitaker , R.T.(2001) . " Maximum Likelihood surface estimation " 1-9 .
19. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) " Robust Statistics: The Approach Based on Influence Functions " John Wiley & Sons, New York .
20. Ahmed, Rikan A., and Saja Mohammad Hussein. "Generalized modified ratio-cum-product kind exponentially estimator of the populations mean in stratified ranked set sample." *International Journal of Nonlinear Analysis and Applications* 13.1 (2022): 1137-1149.
21. Ahmed, Rikan A., and Saja Mohammad Hussein. "Some ratio estimators of finite population variance using auxiliary information in ranked set sampling." *International Journal of Nonlinear Analysis and Applications* 13.1 (2022): 3537-3549.