

مقارنة بين مقدرات S وأصغر موجه تباين معاد الاوزان (RMVV) لتقدير معلمتي الموضع والقياس للتوزيع الطبيعي متعدد المتغيرات - بحرث تطبيقي

أ.د. حمزة اسماعيل شاهين
الجامعة المستنصرية / كلية الادارة

أحمد عباس مصلح
الباحث

P: ISSN : 1813-6729

<http://doi.org/10.31272/JAE.43.2020.126.20>

E : ISSN : 2707-1359

مقبول للنشر بتاريخ: 2018/10/10

تاريخ أستلام البحث : 2018/8/19

المستخلص

تم في هذا البحث ايجاد مقدرات حصينة وكفاءة لموجه المتوسطات (M) ومصفوفة التباين والتباين المشترك (Σ) للتوزيع الطبيعي متعدد المتغيرات عند اختراق الفرضيات المتعلقة بهذا التوزيع وعند وجود المشاهدات الشاذة (Outliers) في العينة قيد الدراسة ، من خلال استخدام طريقة (اصغر موجه تباين معاد الاوزان) ومقارنتها مع طريقة مقدرات S ، اذ تم استخدام اسلوب المحاكاة للمقارنة بين الطريقتين ومن خلال متوسط مربعات الخطأ (MSE) لموجه المتوسطات M ومصفوفة التباين والتباين المشترك Σ للوصول الى الطريقة الأفضل في التقدير ، واطهرت نتائج المحاكاة بأن طريقة (اصغر موجه تباين معاد الاوزان) هي الأفضل كونها حققت اقل متوسط مربعات الخطأ (MSE) من الطريقة الاخرى، وبأختلاف احجام العينات ونسب التلوث ، وفي الجانب التطبيقي تم استخدام كلا الطريقتين على البيانات الخاصة بمرضى الكلى وأظهرت النتائج تطابقاً مع الجانب التجريبي حيث حققت طريقة (اصغر موجه تباين معاد الاوزان) الافضليه في التقدير من خلال المؤشر الاحصائي (MSE).



• بحث مستل من رسالة ماجستير

مجلة الادارة والاقتصاد
العدد 126 / كانون الاول / 2020
الصفحات : 278-266

1. المقدمة ومشكلة وهدف البحث

اثبتت الدراسات السابقة أهمية تقدير معلمة الموقع μ (موجة المتوسطات) ومعلمة القياس Σ (مصفوفة التباين والتباين المشترك) للتوزيع الطبيعي متعدد المتغيرات (Multivariate Normal Distribution) كونها تمثل حجر الأساس في التحليل الكلاسيكي وضرورية للعديد من تقنيات تحليل البيانات متعددة المتغيرات ومنها تحليل التباين متعدد المتغيرات (Multivariate Analysis of Variance) وتحليل المركبات الرئيسية (Principal Component Analysis) والتحليل العاملي (Factor analysis) وتحليل الارتباط القويم (correlation Analysis canonical) والتحليل التمييزي والتصنيف (discriminant analysis and classification) ، والتحليل العنقودي (clustering Analysis) ، وان تقدير هذه المعلمات مهم أيضا في موضوع الاستدلال الاحصائي كونه يهتم في تقدير قيم عددية لمعلمات المجتمع والتي تكون على الاغلب غير معلومة من خلال استخدام اسلوب العينات ومن هنا تبرز أهمية التقدير في التطبيقات الاحصائية ، حيث ان المقدرات الناتجة من طريقة الامكان الاعظم (MLE) لمعلمة الموقع والقياس تكون غير حصينه عند اختراق فرضيات التوزيع الطبيعي وعند وجود المشاهدات الشاذة (Outliers) في العينة قيد الدراسة كما ان التوزيعات متعددة المتغيرات تفتقر الى طرائق التقدير المناسبة واحتمال خرق الفرضيات المتعلقة بالتوزيع الطبيعي يتضاعف بسبب ازدياد عدد المتغيرات وتداخل العلاقات بين المتغيرات لذلك سيتناول هذا البحث اسلوب التقدير الحصين لمعلمتي التوزيع الطبيعي متعدد المتغيرات من خلال استخدام مقدرات اصغر موجة تباين معاد الأوزان (RMVV) ومقارنتها مع مقدرات S-Estimators (S) بالاعتماد على المؤشر الاحصائي متوسط مربعات الخطأ (MSE) .

2. التوزيع الطبيعي متعدد المتغيرات [2],[5] Multivariate Normal Distribution

يمثل التوزيع الطبيعي متعدد المتغيرات الأساس النظري لتحليل متعدد المتغيرات وهو من التوزيعات المهمة الواسعة التطبيقات ويمتاز بالمرونة الحسابية عند استخراج النتائج ، ويعتمد عليه التحليل متعدد المتغيرات بشكل رئيسي كون الأخير يتعامل مع بيانات ذات متجهات عشوائية (Random Vectors) ، أو بيانات ذات وحدات متعددة القياسات مثل (الطول، والوزن، والعمر) للشخص الواحد أو (درجة الحرارة ، وسرعة الرياح، ومقدار الرطوبة النسبية) لنفس اليوم ، حيث إن المتجه العشوائي من الدرجة $(1 \times p)$ يتكون من المشاهدات أو القياسات لمجموعة من المتغيرات العشوائية المرتبطة في العينة العشوائية المسحوبة. على فرض ان $\underline{X} = (x_1, x_2, x_3, \dots, x_p)'$ ، $\underline{x}_j \in R \forall j$ ، فإن دالة الكثافة الاحتمالية للتوزيع الطبيعي متعدد المتغيرات ستكون بالصيغة الآتية :

$$f(\underline{X}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \right\} \dots \quad (1)$$

حيث ان $\underline{\mu}$ موجة المتوسط (معلمة الموقع) ومن الدرجة $p \times 1$.
و Σ مصفوفة التباين والتباين المشترك (معلمة القياس) من الدرجة $p \times p$ وهي مصفوفة متماثلة موجبة بالتحديد symmetric positive definite matrix .
وأیضا يمكن التعبير عن دالة الكثافة أعلاه بالرمز $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$.

3. مقدر اصغر موجة تباين معاد الأوزان [6] Reweighted Minimum Vector Variance Estimator

ان حساب موجة المتوسط (معلمة الموقع) ومصفوفة التباين المشترك (معلمة القياس) لهذا المقدر هو لزيادة الكفاءة مع الحفاظ على الحصانة حيث يتم اعطاء الوزن (0) للمشاهدات الشاذة التي تكون فيها $d_{MNV(i)}^2 > \chi_{P,0.025}^2$ واعطاء الوزن (1) للمشاهدات الغير شاذة وبالتالي يمكن تحديد مقدر (RMVV) الخام (Raw) لمعلمتي الموقع والقياس على النحو الآتي:

$$\bar{\underline{X}}_{RMVV}^{raw} = \frac{\sum_{i=1}^m w_i x_i}{m} \quad \dots (2)$$

$$S_{RMVV}^{raw} = \frac{\sum_{i=1}^m w_i (x_i - \bar{\underline{X}}_{RMVV}^{raw})(x_i - \bar{\underline{X}}_{RMVV}^{raw})'}{m} \quad \dots (3)$$

حيث ان :
 $W_i = \begin{cases} 0 & \text{if } d_{MNV(i)}^2(x_i, \bar{\underline{X}}_{MNV}) > \chi_{P,0.025}^2 \\ 1 & \text{if } d_{MNV(i)}^2 \leq \chi_{P,0.025}^2 \end{cases}$ $o . W$
m : تمثل عدد المشاهدات التي تقابل

وان d_{MNV}^2 يكون مساوياً الى :

$$d_{MNV}^2 = (\underline{x}_i - \bar{\underline{X}}_{MNV})' S_{MNV}^{-1} (\underline{x}_i - \bar{\underline{X}}_{MNV}) \quad \dots (4)$$

عادة يتم معايرة مقدرات القياس لتكون متنسقة مع التوزيع الطبيعي وبالتالي سنأخذ عامل الاتساق $C^*(m)$ والذي يكون على النحو الآتي:

$$C^*(m) = \frac{m/n}{p(X_{p+2}^2 < X_{p, m/n}^2)} \quad \dots (5)$$

وللتغلب على مشكلة عدم الكفاءة سنقوم بحساب معامل التصحيح ، $\vartheta_{m,n,p}^{*\alpha}$ ، عبر أسلوب المحاكاة لعدة أحجام (n) من العينات و p من المتغيرات ، وسنولد مجموعة بيانات $X^{(j)} \in R^{n \times p}$ من التوزيع الطبيعي القياسي ، $N_p(0,1)$ ، لكل مجموعة بيانات $X^{(j)}, j = 1, 2, \dots, K$ ، نحدد بعد ذلك مقدرات RMVV لمعلمتي الموقع والقياس كما في معادلة (2) و معادلة (3) ، ثم يليه $C^*(m)S_{RMVV}^{raw j}$ ، إذا كان المقدر غير متحيز يجب ان يكون $E[|C^*(m)S_{RMVV}^{raw}|] = I_p$ ، وبالتالي، فإننا نتوقع أن الجذر P-th من المحدد لـ $C^*(m)S_{RMVV}^{raw}$ يساوي واحد ، وان المتوسط للجذر P-th من المحدد يعرف بالصيغة الآتية :

$$K(|C^*(m)S_{RMVV}^{raw}|) = \frac{1}{K} \sum_{j=1}^K (|C^*(m)S_{RMVV}^{raw j}|)^{1/p} \quad \dots (6)$$

حيث ان $|C^*(m)S_{RMVV}^{raw}|$ تمثل محدد المصفوفة المربعة $C^*(m)S_{RMVV}^{raw}$ وسيكون اداء المحاكاة $K=1000$ لمختلف احجام العينات (n) ولابعاد p من المتغيرات ، ولمستوى معنوية $\alpha = 0.05$ ، ومعامل التصحيح لـ $C^*(m)S_{RMVV}^{raw}$ هو :

$$\vartheta_{m,n,p}^{*\alpha} = \frac{1}{K(|C^*(m)S_{RMVV}^{raw}|)} \quad \dots (7)$$

بعدها يتم تحديد مقدرات (RMVV) لمعلمتي الموقع والقياس على النحو الآتي:

$$\bar{X}_{RMVV} = \frac{\sum_{i=1}^m w_i x_i}{m} \quad \dots (8)$$

$$S_{RMVV} = \vartheta_{m,n,p}^{*\alpha} C^*(m) \frac{\sum_{i=1}^m w_i (x_i - \bar{X}_{RMVV})(x_i - \bar{X}_{RMVV})'}{m} \quad \dots (9)$$

وأخيرا مسافات مهلونوبس التربيعية الحصينة ستكون على النحو الآتي :

$$d_{i, RMVV}^2(x_i, \bar{X}_{RMVV}, S_{RMVV}) = (x_i - \bar{X}_{RMVV})' S_{RMVV}^{-1} (x_i - \bar{X}_{RMVV}) \quad \dots (10)$$

4. مقدرات (S) : [9],[8],[7],[4] (S-Estimators)

تعتبر مقدرات S (S-Estimators) من اهم طرائق التقدير الحصينة لمعلمتي الموقع والقياس في متعدد المتغيرات وسميت كذلك كون معلمتي الموقع والقياس يتم تقديرهما في وقت واحد (simultaneously) ولهذا السبب يمكن اعتبار هذه المقدرات ذات حصانة عالية في حالة اختيار دالة وزن بشكل مناسب فضلا عن خاصية ثبات التباين والاتساق .

وقد عرف كل من (Rousseeuw & Leroy) عام 1987 ، و (Lopuhaa) عام 1989 مقدرات S كحلول للموقع والقياس $(\hat{\mu}, \hat{\Sigma})$ لمسألة تصغير محدد مصفوفة التباين والتباين المشترك |S| وفقاً للقيود الآتي :

$$\frac{1}{n} \sum_{i=1}^n \rho \{d(x_i, \bar{x}, S)\} = b \quad \dots (11)$$

و تحت كل الأزواج الممكنة (\bar{x}, S) ، فإن لكل $\bar{x} \in R^p$ ، وكل $S \in SPD(p)$ ، حيث ان $SPD(p)$ جميع المصفوفات المتماثلة الموجبة التحديد symmetric positive definite ذات البعد (PXP) ، وان $d(x_i, \bar{x}, S)$ هي مسافات مهلونوبس العادية (MD_i) للملاحظات في العينة وحسب معادلة الآتية:

$$d^2 = (\underline{x} - \bar{x})' S^{-1} (\underline{x} - \bar{x}) \quad \dots (12)$$

ولتحقيق الاتساق (consistency) عند التوزيع الطبيعي فإن الثابت b يكون:

$$b = E_{F_0} \left[\rho \left(\left\| \frac{x}{c} \right\| \right) \right] = E \left(\rho \left(d/c \right) \right)$$

حيث ان :

$$F_0 = N(0, I_p)$$

و x تمثل متغير عشوائي ذات توزيع دائري (spherical distribution) .

وهناك شروط يجب ان تتوفر في الدالة ρ وهي ان تكون متماثلة حول الصفر ومستمرة ومتزايدة في الفترة $[k, 0]$ ، وثابتة في الفترة $[k, \infty)$ ، وتمتلك مشتقة مستمرة عند عدد محدود من النقاط ، وبناء على ذلك فإن $\rho(0) = 0$.

والاختيار الشائع للدالة ρ اختيرت كدالة وزن توكي (Tukey's biweight function) والتي تعرف على انها تمتلك خصائص حصينة وجيدة وحسب الصيغة الاتية [7] :

$$\rho(MD_i) = \begin{cases} 1 - \left[1 - \left(\frac{MD_i}{k}\right)^2\right]^3 & \text{if } |MD_i| \leq k_0 \\ 1 & \text{if } \left|\frac{MD_i}{k}\right| > k_0 \end{cases} \quad \dots(13)$$

حيث ان (MD_i) هي مسافات مهلونوبس للعينة Mahalanobis distance ومن اجل الحصول على نقطة انهيار عالية يتم تحديد قيمة للثابت (k_0) و هو اي قيمة ثابتة موجبة تحقق مركزية لنقطة الانهيار وتسمى بمقدر (Biweight's estimator). وان الثابت b يعتمد على القيمة k_0 .

. واول الاختيارات هي قيمة القطع للثابت (c) في حالة المتغير الاحادي (univariate) ويمثل الانحرافات المعيارية عن الوسط الحسابي والتي يكون بعدها $\rho'(MD_i)$ مساوي للصفر ، وان الثابت (c) يمكن تحويله للقيمة k على مقياس χ^2_p (chi-square) لمسافات مهلونوبس MD^2 باستخدام تحويل Wilson-Hilferty transformation وحسب الصيغة الاتية [7] :

$$k_0 = \sqrt{p \left\{ \sqrt{\left(\frac{1}{9}\right) \left(\frac{2}{p}\right) c + \left(1 - \left(\frac{1}{9}\right) \left(\frac{2}{p}\right)\right)} \right\}^3} \quad \dots (14)$$

فعلى سبيل المثال فان مقدر S (S-Estimator) يمتلك نقطة انهيار تساوي 50% عندما $(c=1.548)$ ولعدد متغيرات $(p=3)$ وهذا يعني ان قيمة $k_0=2.707$.

وبذلك فان معلمة الاتساق b في حالة كون المتغيرات تتبع التوزيع الطبيعي القياسي يتم حسابها وفق الصيغة الاتية :

$$b = \frac{p}{2} x_{p+2}^2(k_0^2) - \frac{p(p+2)}{2k_0^2} x_{p+4}^2(k_0^2) + \frac{p(p+2)(p+4)}{6k_0^4} x_{p+6}^2(k_0^2) + \frac{k_0^2}{6} (1 - x_p^2(k_0^2)) \quad \dots(15)$$

ما تقدم أعلاه فان الخوارزمية لحساب مقدرات (S) يمكن توضيحها كالآتي:

1. يتم اختيار المجموعات او العينات الجزئية (h) بحجم (P+1) عندما يكون لدينا حجم العينة (n) من المشاهدات و (p) من المتغيرات ، أي ان عدد المجاميع الجزئية مساوي الى C_{p+1}^n .
2. يتم استخراج موجه المتوسط ومصفوفة التباين والتباين المشترك لكل عينة جزئية من تلك العينات او المجموعات.
3. يتم اختيار الأفضل من تلك العينات الجزئية وفقاً الى محدد مصفوفة التباين والتباين المشترك العائدة لها ، حيث ان افضل مجموعه هي التي تمتلك اصغر محدد تباين مشترك من بين المجموعات الجزئية الأخرى.
4. يتم استخراج مسافات مهلونوبس العادية بين نقاط المشاهدات و الموقع \bar{x} بالاعتماد على مصفوفة التباين والتباين المشترك (S) التي تم اختيارها في الفقرة (3) وكالاتي:

$$d(x_i, \bar{x}, S) = \left\{ (x_i - \bar{x})' S^{-1} (x_i - \bar{x}) \right\}^{1/2} \quad \dots(16)$$

5. يتم استخراج $\rho(MD)$ حيث ρ هو دالة Biweight و المذكورة في (13)

حيث أن k_0 ثابت يحقق قيمة مرغوبة لنقطة الانهيار والمستخرج وفق معادلة (14) وبالتالي عند تطبيق معادلة (15) يتم تحديد قيمة b .

6. بعد استخراج قيمة b يتم تكرار المعادلة (14) ليتم البحث فيها عن قيمة (k_0) بطريقة التكرار لغرض تحقيق القيد المذكور في معادلة (11) .

7. بعد استخراج قيم الثوابت في أعلاه والتي تحقق الشرط او القيد في معادلة (11) نقوم باختيار مصفوفة التباين والتباين المشترك التي حققت هذا الشرط وموجه المتوسط المقابل لها والتي تمثل (مقدرات S) الحصينة للموقع والقياس.

5. المقارنة بين طريقتي التقدير: [2],[10],[11],[12]

للمقارنة بين طريقتي التقدير فإن الأسلوب المقترح هو استخدام أصغر متوسط لمربعات الخطأ Minimum Mean Square Error (MMSE) والمقدر الأفضل هو الذي يحقق اقل (MMSE) ، حيث اقترح الباحثان Forst & Saverino إستنادا الى Quadratic Frobenius norm استخدام القياس

التربيعي للمسافة بين المصفوفة المقدره والمصفوفة الحقيقية اذ انه في حالة كون المصفوفة Z مربعة متماتلة $(p \times p)$ فإن :

$$\|Z\|_F^2 = \frac{Trace(Z'Z)}{p} \quad \dots(17)$$

حيث ان قاعدة فروبينوس Frobenius norm هي:

$$\|Z\|_F = \sqrt{\sum_{i,j} |z_{ij}|^2} = \sqrt{tr(Z'Z)} \quad \dots (18)$$

وكذلك حاصل الضرب الداخلي لمصفوفتين A,B مربعتين متماتلتين $(p \times p)$ فإن

$$\langle A, B \rangle = \frac{Trace(AB')}{p} \quad \dots (19)$$

أي أن أصغر متوسط مربعات خطأ سيكون.

$$MMSE = E \left\{ \left\| \Sigma - \hat{\Sigma} \right\|_F^2 \right\} = \frac{Trace(\Sigma - \hat{\Sigma})^2}{p} \quad \dots (20)$$

إذ يكون MMSE اسلوباً جيداً للمقارنة بين مختلف مقدرات مصفوفة التباين والتباين المشترك أما بالنسبة لموجه المتوسط (معلمة الموقع) فإن الاسلوب المقترح للمقارنة هو ايضا اقل متوسط مربعات الخطأ Mean-Square Error (MSE) حيث ان متوسط مربعات الخطأ بين اي موجهين (U,V) هو ببساطة مربع المسافة الاقليدية (squared Euclidean distance) بينهما وحسب الصيغة الاتية: [12]

$$\|u - v\|^2 = (u_1 - v_1)^2 + (u_2 - v_2)^2 + \dots + (u_p - v_p)^2 \quad \dots(21)$$

وبالتالي للمقارنة بين موجه المتوسط الحقيقي (\underline{M}) وموجه المتوسط المقدر ($\hat{\underline{M}}$) يمكن استخدام الصيغة الاتية:

$$MSE = \frac{1}{p} \|\underline{M} - \hat{\underline{M}}\|^2 = \frac{1}{p} ((M_1 - \hat{M}_1)^2 + (M_2 - \hat{M}_2)^2 + \dots + (M_p - \hat{M}_p)^2) \quad \dots(22)$$

حيث ان P يمثل عدد المتغيرات (X_i)

6- الجانب التجريبي

تضمنت تجارب المحاكاة توليد بيانات تتبع التوزيع الطبيعي متعدد المتغيرات ، والتوزيع الطبيعي متعدد المتغيرات الملوث بنسب تلوث وأحجام عينات مختلفة باستخدام طريقة Box-Muller حيث تم الإعتماد على متوسط مربعات الخطأ (MSE) كمعيار للمقارنة بين الطريقتين التي تم إستخدامها في هذا البحث لتقدير موجهات المتوسطات ($\underline{\mu}$) ومصفوفات التباين والتباين المشترك (Σ) للمتغيرات التوضيحية من خلال البرنامج (R 3.5) ، وتم إيجاد التوزيعات الملوثة وذلك من خلال إستخدام النسبة (ϵ) التي تمثل نسبة التلوث المراد إستخدامها لتلويث البيانات ، كذلك النسبة ($1 - \epsilon$) التي تمثل نسبة البيانات غير الملوثة ، ويتم توليد المشاهدات غير الملوثة للمتغير (X) بنسبة ($1 - \epsilon$) ، في حين يتم توليد المشاهدات الملوثة بنسبة (ϵ) بمتوسط ($\mu + \tau$) وتباين (σ^2) ، وعليه فإنه إيجاد مشاهدات المتغير العشوائي الملوث سيكون وفق الصيغة التالية: [1]

$$X = (1 - \epsilon) N(\mu, \sigma^2) + \epsilon N(\mu + \tau, \sigma^2) , \quad \tau \neq 0 \quad (23)$$

إذ ان τ : تمثل درجة التلوث

حيث يتم تقدير المقياس الإحصائي متوسط مربعات الخطأ (MSE) لموجهات المتوسطات ومصفوفات التباين والتباين المشترك المقدره وفق طرائق التقدير المستخدمة في هذا البحث وكما يأتي :

1- يتم تقدير (MSE) لمصفوفات التباين والتباين المشترك ($\hat{\Sigma}_i$) لكل تكرار حسب الصيغة (20) ، وإن متوسط مربعات الخطأ لمقدر مصفوفة التباين والتباين المشترك ($\hat{\Sigma}$) لكل طريقة يحسب وفق الصيغة التالية :

$$MSE(\hat{\Sigma}) = \frac{1}{K} \sum_{i=1}^K \hat{\Sigma}_i \quad (24)$$

2- يتم تقدير (MSE) لموجهات المتوسطات ($\hat{\underline{\mu}}_i$) لكل تكرار حسب الصيغة (22) ، وإن متوسط مربعات

الخطأ لمقدر موجه المتوسطات ($\hat{\underline{\mu}}$) لكل طريقة يحسب وفق الصيغة التالية :

$$MSE(\hat{\underline{\mu}}) = \frac{1}{K} \sum_{i=1}^K \hat{\underline{\mu}}_i \quad (25)$$

في هذا البحث تم توليد بيانات التوزيع الطبيعي متعدد المتغيرات الملوثة ، بنسب تلوث (0.25 , 0.10 , 0.05) وبتكرار بلغ ($K = 5000$) ولأحجام العينات التالية (25 ، 50 ، 100 ، 200) تم الحصول على القيم الافتراضية لموجه المتوسطات ($\underline{\mu}$) ومصفوفة التباين والتباين المشترك (Σ) التي تم الاعتماد عليها للتوزيع الطبيعي الملوث بنسبة 5% كانت :

| μ | | Σ | X_1 | X_2 | X_3 | X_4 | X_5 |
|---------|-----|----------|-------|-------|-------|-------|-------|
| μ_1 | 12 | X_1 | 10 | | | | |
| μ_2 | 27 | X_2 | 0.99 | 25 | | | |
| μ_3 | 52 | X_3 | 0.27 | -3.81 | 50 | | |
| μ_4 | 77 | X_4 | 2.78 | -2.41 | 2.76 | 100 | |
| μ_5 | 102 | X_5 | -2.68 | 4.49 | -3.57 | 5.96 | 255 |

إن القيم الافتراضية لموجه المتوسطات (μ) ومصفوفة التباين والتباين المشترك (Σ) التي تم الحصول عليها للتوزيع الطبيعي الملوث بنسبة 10% كانت :

| μ | | Σ | X_1 | X_2 | X_3 | X_4 | X_5 |
|---------|-----|----------|-------|-------|--------|--------|--------|
| μ_1 | 14 | X_1 | 10.00 | | | | |
| μ_2 | 29 | X_2 | -0.18 | 25.00 | | | |
| μ_3 | 54 | X_3 | -1.80 | -0.33 | 50.00 | | |
| μ_4 | 79 | X_4 | -0.57 | -3.02 | -10.64 | 100.00 | |
| μ_5 | 104 | X_5 | 3.63 | -7.38 | 9.88 | -27.01 | 225.00 |

وإن القيم الافتراضية لموجه المتوسطات (μ) ومصفوفة التباين والتباين المشترك (Σ) التي تم الاعتماد عليها للتوزيع الطبيعي الملوث بنسبة 25% كانت :

| μ | | Σ | X_1 | X_2 | X_3 | X_4 | X_5 |
|---------|-----|----------|-------|-------|-------|--------|-------|
| μ_1 | 18 | X_1 | 10 | | | | |
| μ_2 | 30 | X_2 | -1.28 | 25 | | | |
| μ_3 | 60 | X_3 | 2.50 | -4.92 | 50 | | |
| μ_4 | 80 | X_4 | -2.12 | 3.70 | 6.03 | 100 | |
| μ_5 | 110 | X_5 | 1.50 | -3.31 | -7.37 | -15.74 | 225 |

1-6: عرض نتائج المحاكاة

إن عملية تحليل النتائج التي نحصل عليها من تنفيذ أسلوب المحاكاة تكون ذات أهمية كبيرة في مقارنة مقدرات طريقة (RMVV) وطريقة مقدرات (S) التي تم دراستهما في هذا البحث ، إذ تتم عملية المقارنة على أساس متوسط مربعات الخطأ (MSE) وكما موضح في الجداول التالية :

جدول (1-6)

مقدرات موجّهات المتوسطات للتوزيع الطبيعي الملوّث ($\epsilon = 5\%$) لطريقة (RMVV) و (S)

| n | \hat{M} | \hat{M}_S | \hat{M}_{RMVV} |
|-----|-------------|-------------|------------------|
| 25 | \hat{M}_1 | 13.73 | 12.90 |
| | \hat{M}_2 | 28.17 | 27.26 |
| | \hat{M}_3 | 52.29 | 52.88 |
| | \hat{M}_4 | 73.59 | 77.34 |
| | \hat{M}_5 | 103.27 | 102.67 |
| 50 | \hat{M}_1 | 12.98 | 12.78 |
| | \hat{M}_2 | 26.57 | 25.96 |
| | \hat{M}_3 | 53.96 | 52.75 |
| | \hat{M}_4 | 76.60 | 76.67 |
| | \hat{M}_5 | 102.93 | 102.33 |
| 100 | \hat{M}_1 | 12.14 | 12.04 |
| | \hat{M}_2 | 27.14 | 27.15 |

| | | | |
|-----|-------------|--------|--------|
| | \hat{M}_3 | 53.19 | 51.90 |
| | \hat{M}_4 | 79.34 | 79.03 |
| | \hat{M}_5 | 103.10 | 100.81 |
| 200 | \hat{M}_1 | 12.26 | 11.63 |
| | \hat{M}_2 | 24.60 | 25.87 |
| | \hat{M}_3 | 51.88 | 51.59 |
| | \hat{M}_4 | 78.75 | 76.55 |
| | \hat{M}_5 | 103.07 | 103.50 |

جدول (2-6)

مقدرات مصفوفات التباين والتباين المشترك للتوزيع الطبيعي الملوث ($\varepsilon=5\%$, $n=25$) لطريقة (S) و (RMVV)

| | $\hat{\Sigma}_S$ | | | | | | $\hat{\Sigma}_{RMVV}$ | | | | |
|-------|------------------|-------|-------|--------|--------|-------|-----------------------|-------|-------|--------|--------|
| | X_1 | X_2 | X_3 | X_4 | X_5 | | X_1 | X_2 | X_3 | X_4 | X_5 |
| X_1 | 10.23 | | | | | X_1 | 11.52 | | | | |
| X_2 | 1.22 | 26.13 | | | | X_2 | 0.86 | 25.65 | | | |
| X_3 | -0.45 | -4.39 | 47.98 | | | X_3 | -1.06 | -3.83 | 48.91 | | |
| X_4 | 3.43 | -2.50 | 2.64 | 102.58 | | X_4 | 0.58 | -2.35 | 2.83 | 101.71 | |
| X_5 | -3.36 | 4.43 | -4.66 | 5.32 | 223.86 | X_5 | -4.02 | 4.48 | -4.03 | 7.13 | 224.25 |

جدول (3-6)

مقدرات مصفوفات التباين والتباين المشترك للتوزيع الطبيعي الملوث ($\varepsilon=5\%$, $n=50$) لطريقة (S) و (RMVV)

| | $\hat{\Sigma}_S$ | | | | | | $\hat{\Sigma}_{RMVV}$ | | | | |
|-------|------------------|-------|-------|--------|--------|-------|-----------------------|-------|-------|-------|--------|
| | X_1 | X_2 | X_3 | X_4 | X_5 | | X_1 | X_2 | X_3 | X_4 | X_5 |
| X_1 | 11.79 | | | | | X_1 | 11.73 | | | | |
| X_2 | 0.86 | 25.14 | | | | X_2 | 0.77 | 25.02 | | | |
| X_3 | -0.89 | -3.35 | 49.61 | | | X_3 | -1.08 | -3.63 | 49.12 | | |
| X_4 | -0.27 | -4.10 | 2.29 | 101.21 | | X_4 | 0.20 | -3.39 | 3.27 | 99.98 | |
| X_5 | -4.13 | 2.77 | -5.37 | 8.60 | 227.31 | X_5 | -3.34 | 4.00 | -3.85 | 7.21 | 226.36 |

جدول (4-6)

مقدرات مصفوفات التباين والتباين المشترك للتوزيع الطبيعي الملوث ($\varepsilon=5\%$, $n=100$) لطريقة (S) و (RMVV)

| | $\hat{\Sigma}_S$ | | | | | | $\hat{\Sigma}_{RMVV}$ | | | | |
|-------|------------------|-------|-------|--------|--------|-------|-----------------------|-------|-------|--------|--------|
| | X_1 | X_2 | X_3 | X_4 | X_5 | | X_1 | X_2 | X_3 | X_4 | X_5 |
| X_1 | 11.67 | | | | | X_1 | 11.62 | | | | |
| X_2 | 0.89 | 25.24 | | | | X_2 | 0.64 | 24.88 | | | |
| X_3 | -0.73 | -3.17 | 49.86 | | | X_3 | -1.07 | -3.62 | 49.16 | | |
| X_4 | -0.69 | -4.51 | 1.77 | 102.27 | | X_4 | 0.14 | -3.42 | 3.11 | 100.67 | |
| X_5 | -3.54 | 3.04 | -5.22 | 8.49 | 226.33 | X_5 | -3.17 | 4.19 | -3.88 | 7.34 | 226.12 |

جدول (5-6)

مقدرات مصفوفات التباين والتباين المشترك للتوزيع الطبيعي الملوث ($\epsilon=5\%$, $n=200$) لطريقة (S) و (RMVV)

| | \sum_s | | | | | | \sum_{RMVV} | | | | |
|-------|----------|-------|-------|--------|--------|-------|---------------|-------|-------|--------|--------|
| | X_1 | X_2 | X_3 | X_4 | X_5 | | X_1 | X_2 | X_3 | X_4 | X_5 |
| X_1 | 12.20 | | | | | X_1 | 11.63 | | | | |
| X_2 | 0.89 | 24.63 | | | | X_2 | 0.68 | 24.78 | | | |
| X_3 | -0.67 | -3.39 | 49.79 | | | X_3 | -1.04 | -3.68 | 49.13 | | |
| X_4 | -0.61 | -3.93 | 1.99 | 101.73 | | X_4 | 0.13 | -3.33 | 3.16 | 100.60 | |
| X_5 | -4.10 | 2.04 | -5.65 | 9.36 | 225.26 | X_5 | -3.63 | 4.89 | -3.58 | 6.56 | 224.60 |

جدول (6-6)

متوسطات مربعات الخطأ لموجهات المتوسطات المقدرة لطريقة (S) و (RMVV)

| Dist. | n | RMVV | S | Best method |
|-------------------|-----|--------|--------|-------------|
| $\epsilon = 5\%$ | 25 | 1.9445 | 4.5615 | RMVV |
| | 50 | 1.8868 | 3.5691 | |
| | 100 | 1.8311 | 3.3878 | |
| | 200 | 1.2809 | 2.7065 | |
| $\epsilon = 10\%$ | 25 | 1.1038 | 1.7202 | RMVV |
| | 50 | 0.9678 | 1.4147 | |
| | 100 | 0.7235 | 1.2624 | |
| | 200 | 0.3926 | 1.1302 | |
| $\epsilon = 25\%$ | 25 | 2.4905 | 4.0827 | RMVV |
| | 50 | 1.4160 | 2.7346 | |
| | 100 | 0.9386 | 2.0463 | |
| | 200 | 0.4821 | 1.4631 | |

جدول (7-6)

متوسطات مربعات الخطأ لمصفوفات التباين والتباين المشترك المقدرة لطريقة (S) و (RMVV)

| Dist. | n | RMVV | S | Best method |
|-------------------|-----|--------|---------|-------------|
| $\epsilon = 5\%$ | 25 | 1.4795 | 2.6848 | RMVV |
| | 50 | 1.1201 | 2.0383 | |
| | 100 | 1.0065 | 1.9538 | |
| | 200 | 0.7935 | 1.6185 | |
| $\epsilon = 10\%$ | 25 | 1.0722 | 9.2199 | RMVV |
| | 50 | 0.8790 | 6.3017 | |
| | 100 | 0.5831 | 4.3547 | |
| | 200 | 0.5827 | 4.0182 | |
| $\epsilon = 25\%$ | 25 | 9.7395 | 23.3789 | RMVV |
| | 50 | 8.0142 | 18.1518 | |
| | 100 | 3.6732 | 16.4928 | |
| | 200 | 0.7994 | 12.4681 | |

من خلال النتائج المبينة في الجدول (6-6) نلاحظ الآتي:

- 1- أن قيم متوسط مربعات الخطأ MSE لموجهات المتوسطات تتناقص كلما زاد حجم العينة ولكلا الطريقتين.
- 2- من خلال المقياس (MSE) لموجهات المتوسطات يتضح لنا أن طريقة مقدرات RMVV هي افضل طريقة عند احجام العينات (n=25, n=50, n=100, n=200) ولنسب التلوث المختلفة وذلك لانها حققت اقل MSE .
- 3- أن قيم متوسط مربعات الخطأ MSE لمصفوفات التباين والتباين المشترك تتناقص كلما زاد حجم العينة ولكلا الطريقتين .
- 4- من خلال المقياس (MSE) يتضح لنا أن طريقة RMVV هي افضل طريقة عند احجام العينات (n=25, n=50, n=100, n=200) ولنسب التلوث المختلفة وذلك لانها حققت اقل MSE .

7- الجانب التطبيقي:

1-7 : وصف العينة والمتغيرات المستخدمة

جمعت البيانات الحقيقية للمرضى المصابين بأمراض الكلى من مستشفى بعقوبة التعليمي/ مركز ابن سينا للغسيل الكلوي ، وقد تم الاستعانة بمجموعة من الأطباء المتخصصين بأمراض الكلى من اجل معرفة اهم العوامل المؤثرة على المرض ،اذ تم جمع البيانات الخاصة بسنة 2017 وبعينة حجمها (100) مريض من خلال البطاقة الخاصة (الطيلة) لكل مريض، بالاعتماد على المتغيرات التوضيحية الآتية:

- 1- المتغير (X_1): يمثل الكرياتينين والمستوى الطبيعي في الدم من (5 - 1.5 mgldl) لكل 100 مل لتر دم.
- 2- المتغير (X_2): يمثل الألبومين يتراوح مستوى الألبومين من (52-36 g/l) .
- 3- المتغير (X_3): يمثل البروتين ، ويتراوح مستوى البروتين بين (86-68 g/l) .
- 4- المتغير (X_4): يمثل السكر في الدم ويقاس بوحدة (g/mnol) و النسب الطبيعية للسكر من (85-126) قبل الاكل (صيام لمدة ثمان ساعات)، ومن (165-200) بعد الغداء بساعتين.
- 5- المتغير (X_5): يمثل نسبة (اليوريا) والمستوى الطبيعي من (40 - 20 mgldl) لكل 100 مل من الدم .

جدول (1-7)

البيانات الخاصة بمرضى الكلى والعوامل المؤثرة عليه.

| المرضى | X_1 الكرياتينين | X_2 الألبومين | X_3 البروتين | X_4 السكر | X_5 اليوريا |
|--------|----------------------|--------------------|-------------------|----------------|------------------|
| 1 | 7.1 | 46.6 | 120.0 | 118.1 | 200.5 |
| 2 | 5.9 | 53.1 | 57.9 | 72.6 | 196.2 |
| 3 | 8.6 | 28.5 | 95.0 | 197.6 | 141.6 |
| 4 | 5.4 | 42.7 | 47.4 | 147.5 | 155.8 |
| 5 | 6.2 | 35.6 | 75.5 | 126.0 | 368.5 |
| 6 | 5.7 | 20.0 | 110.0 | 157.6 | 190.8 |
| 7 | 9.1 | 22.7 | 48.5 | 156.6 | 186.8 |
| 8 | 5.2 | 37.1 | 133.0 | 55.5 | 196.3 |
| 9 | 5.4 | 43.7 | 41.0 | 118.3 | 157.1 |
| 10 | 6.4 | 26.7 | 123.0 | 127.0 | 170.5 |
| 11 | 6.2 | 32.3 | 22.7 | 160.7 | 236.6 |
| 12 | 11.4 | 39.0 | 153.0 | 10.9 | 229.8 |
| 13 | 9.8 | 29.9 | 67.9 | 182.4 | 216.6 |
| 14 | 10.0 | 29.3 | 49.6 | 136.9 | 311.0 |
| 15 | 5.4 | 46.3 | 105.3 | 212.7 | 149.3 |
| 16 | 11.2 | 43.6 | 55.0 | 131.2 | 91.7 |
| 17 | 5.8 | 55.0 | 90.0 | 182.3 | 199.6 |
| 18 | 2.6 | 23.5 | 56.3 | 103.4 | 160.8 |
| 19 | 7.2 | 44.6 | 115.4 | 204.6 | 134.5 |
| 20 | 3.7 | 32.4 | 47.4 | 158.5 | 104.6 |
| 21 | 6.4 | 23.6 | 65.4 | 205.2 | 178.6 |
| 22 | 6.2 | 35.0 | 57.7 | 124.4 | 192.5 |
| 23 | 4.6 | 46.4 | 70.6 | 77.1 | 195.4 |
| 24 | 3.9 | 30.0 | 62.4 | 171.2 | 159.6 |
| 25 | 5.9 | 75.0 | 33.6 | 174.4 | 157.8 |

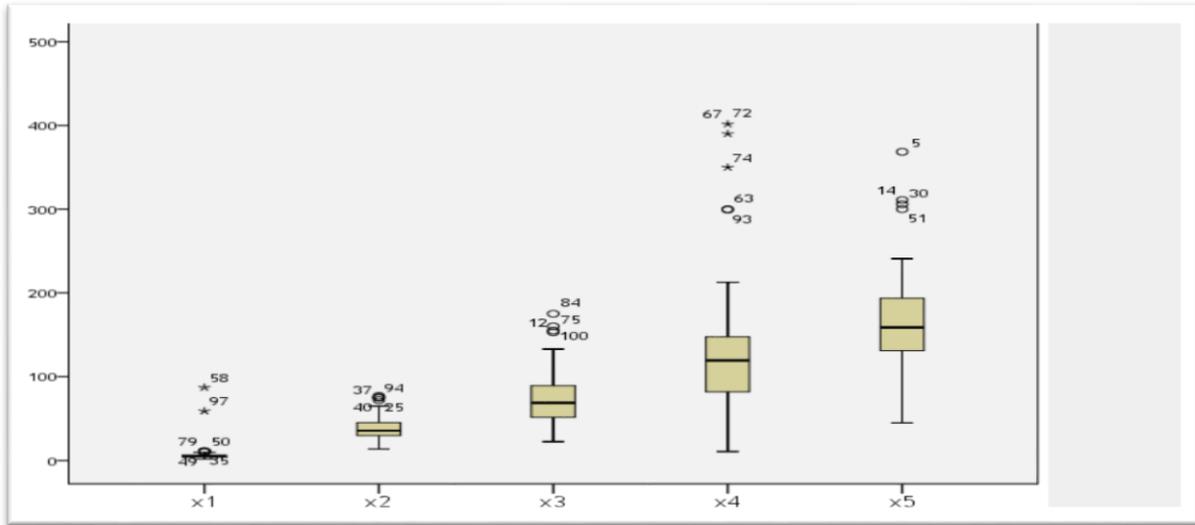
مقارنة بين مقدرات S وأسعر موجه تباين معاد الأوزان (RMVV) لتقدير معلتي الموقع والقياس للتوزيع الطبيعي متعدد المتغيرات - بعد تطبيقي

| | | | | | |
|----|------|------|-------|-------|-------|
| 26 | 4.7 | 33.3 | 82.2 | 79.0 | 45.0 |
| 27 | 3.6 | 50.3 | 72.5 | 159.2 | 172.5 |
| 28 | 5.6 | 53.1 | 110.0 | 54.5 | 174.7 |
| 29 | 5.3 | 29.9 | 82.0 | 147.2 | 241.1 |
| 30 | 4.9 | 35.8 | 93.5 | 111.9 | 305.9 |
| 31 | 5.2 | 24.2 | 63.6 | 69.5 | 153.9 |
| 32 | 5.9 | 31.9 | 76.5 | 67.1 | 188.9 |
| 33 | 2.9 | 20.9 | 58.2 | 144.3 | 86.8 |
| 34 | 6.1 | 43.2 | 105.6 | 124.5 | 160.5 |
| 35 | 11.0 | 31.0 | 57.2 | 37.3 | 157.6 |
| 36 | 6.0 | 37.6 | 86.8 | 164.6 | 137.5 |
| 37 | 5.2 | 72.0 | 60.4 | 56.8 | 122.1 |
| 38 | 3.5 | 14.8 | 38.9 | 115.3 | 192.9 |
| 39 | 4.1 | 44.6 | 51.4 | 139.6 | 144.0 |
| 40 | 6.7 | 75.0 | 80.5 | 100.7 | 111.7 |
| 41 | 5.1 | 29.0 | 102.8 | 111.1 | 136.6 |
| 42 | 3.4 | 25.3 | 59.2 | 125.6 | 162.6 |
| 43 | 3.6 | 48.6 | 87.0 | 99.5 | 125.6 |
| 44 | 2.9 | 65.0 | 44.0 | 130.8 | 132.1 |
| 45 | 4.1 | 43.9 | 75.6 | 131.3 | 174.9 |
| 46 | 4.4 | 45.5 | 116.1 | 169.6 | 200.1 |
| 47 | 3.2 | 39.7 | 62.8 | 126.9 | 138.4 |
| 48 | 4.3 | 22.3 | 62.2 | 65.7 | 125.7 |
| 49 | 10.5 | 44.5 | 45.2 | 155.3 | 146.4 |
| 50 | 10.4 | 46.8 | 93.2 | 60.7 | 201.6 |
| 51 | 4.3 | 33.9 | 79.6 | 31.9 | 300.0 |
| 52 | 2.4 | 33.0 | 46.3 | 69.4 | 170.1 |
| 53 | 2.8 | 58.0 | 108.7 | 111.5 | 221.4 |
| 54 | 1.7 | 29.8 | 80.2 | 25.6 | 177.5 |
| 55 | 4.0 | 32.5 | 51.3 | 56.9 | 144.9 |
| 56 | 2.5 | 41.9 | 43.6 | 79.2 | 63.4 |
| 57 | 3.8 | 24.8 | 54.1 | 84.7 | 233.9 |
| 58 | 87.0 | 44.1 | 86.6 | 38.1 | 162.8 |
| 59 | 8.4 | 22.5 | 78.4 | 85.0 | 123.8 |
| 60 | 4.8 | 59.9 | 72.6 | 83.4 | 152.5 |
| 61 | 5.3 | 36.5 | 97.7 | 116.5 | 139.5 |
| 62 | 4.6 | 33.4 | 60.0 | 103.0 | 128.8 |
| 63 | 7.6 | 20.0 | 52.8 | 300.0 | 170.3 |
| 64 | 4.3 | 41.4 | 53.9 | 25.3 | 132.2 |
| 65 | 5.4 | 28.3 | 46.7 | 100.5 | 85.3 |
| 66 | 3.9 | 45.8 | 93.0 | 113.5 | 211.5 |
| 67 | 4.1 | 30.0 | 69.5 | 401.6 | 90.9 |
| 68 | 5.2 | 24.9 | 43.0 | 98.0 | 155.7 |
| 69 | 5.9 | 32.8 | 44.5 | 104.5 | 171.3 |
| 70 | 5.5 | 39.1 | 56.7 | 122.5 | 198.3 |
| 71 | 5.3 | 50.0 | 33.8 | 116.1 | 181.5 |
| 72 | 10.0 | 51.2 | 42.0 | 390.0 | 80.1 |
| 73 | 6.4 | 47.4 | 48.3 | 51.4 | 152.1 |
| 74 | 5.7 | 31.4 | 62.7 | 350.0 | 114.0 |
| 75 | 6.7 | 45.2 | 155.0 | 131.2 | 70.5 |
| 76 | 5.8 | 35.2 | 81.8 | 80.1 | 148.7 |
| 77 | 5.9 | 34.8 | 69.7 | 147.7 | 180.5 |
| 78 | 6.1 | 30.5 | 82.0 | 130.5 | 134.2 |
| 79 | 10.8 | 29.1 | 62.6 | 115.3 | 207.3 |
| 80 | 8.2 | 48.6 | 88.3 | 144.4 | 161.6 |

| | | | | | |
|-----|------|------|-------|-------|-------|
| 81 | 6.8 | 36.5 | 87.1 | 136.4 | 113.5 |
| 82 | 6.7 | 77.0 | 95.0 | 125.9 | 109.8 |
| 83 | 5.9 | 39.7 | 36.3 | 64.5 | 174.5 |
| 84 | 5.8 | 35.5 | 175.0 | 103.0 | 166.2 |
| 85 | 6.3 | 37.0 | 51.8 | 106.9 | 86.9 |
| 86 | 5.8 | 36.0 | 51.0 | 139.5 | 135.3 |
| 87 | 3.9 | 36.0 | 93.2 | 128.2 | 117.8 |
| 88 | 4.6 | 25.8 | 74.0 | 121.3 | 143.9 |
| 89 | 5.6 | 30.0 | 65.8 | 41.7 | 224.7 |
| 90 | 4.3 | 36.0 | 107.0 | 69.1 | 130.1 |
| 91 | 3.6 | 13.8 | 78.2 | 151.0 | 199.9 |
| 92 | 7.8 | 32.0 | 41.4 | 94.9 | 138.7 |
| 93 | 7.8 | 35.5 | 38.3 | 299.3 | 114.7 |
| 94 | 6.3 | 75.0 | 74.1 | 54.0 | 108.3 |
| 95 | 5.0 | 55.8 | 71.2 | 194.4 | 124.9 |
| 96 | 3.6 | 47.6 | 55.8 | 120.6 | 119.6 |
| 97 | 59.0 | 33.7 | 50.0 | 85.8 | 218.1 |
| 98 | 9.5 | 33.0 | 90.0 | 129.7 | 165.5 |
| 99 | 7.7 | 25.3 | 55.5 | 165.9 | 194.1 |
| 100 | 5.4 | 40.5 | 160.0 | 91.1 | 209.8 |

3-7: اختبار توزيع البيانات

يوجد العديد من الطرائق التي بالإمكان الاعتماد عليها لفحص البيانات عند احتوائها على القيم الشاذة (Outlier Value) ومنها (الطريقة الحسابية وطريقة الرسم الاحصائي وكذلك طريقة الصندوق Box Plot التي تم الاعتماد عليها في بحثنا هذا وكما موضح في الشكل الآتي: [3]



الشكل (1-7)

يمثل القيم الشاذة في المتغيرات التوضيحية (X_1, X_2, X_3, X_4, X_5) للبيانات الخاصة بمرضى الكلى

4-7: مقدرات الطرائق المستخدمة في التقدير

سيتم في هذا الجزء تقدير معلمتي الموقع والقياس للتوزيع الطبيعي متعدد المتغيرات باستخدام مقدرات (RMVV) ومقدرات (S) المستخدمة في الجانب التجريبي، وبالاعتماد على البيانات الخاصة بمرضى الكلى. وقد تم تقدير المعلمات بأستعمال برنامج (R 3.5) والذي يعد من البرامج المتقدمة في المجالات الاحصائية والرياضية، و تم حساب متوسط مربعات الخطأ لموجهات المتوسطات ومصفوفات التباين والتباين المشترك وكما موضح في الجدول الآتي:

جدول (2-7) القيم التقديرية لمعلمتي الموقع والقياس بطريقة RMVV و S

| | $\hat{\mu}$ | \sum | X_1 | X_2 | X_3 | X_4 | X_5 |
|---------------|-------------|---------------|-------|--------|---------|---------|---------|
| $\hat{\mu}_1$ | 7.44 | \sum_{RMVV} | X_1 | 96.64 | | | |
| $\hat{\mu}_2$ | 38.28 | | X_2 | -2.75 | 167.11 | | |
| $\hat{\mu}_3$ | 73.89 | | X_3 | 1.57 | 33.55 | 846.50 | |
| $\hat{\mu}_4$ | 126.00 | | X_4 | -69.79 | -64.24 | -269.44 | 4584.78 |
| $\hat{\mu}_5$ | 162.45 | | X_5 | 40.03 | -115.81 | 157.85 | -685.28 |
| | $\hat{\mu}$ | \sum | X_1 | X_2 | X_3 | X_4 | X_5 |
| $\hat{\mu}_1$ | 8.03 | \sum_S | X_1 | 96.49 | | | |
| $\hat{\mu}_2$ | 39.48 | | X_2 | -2.81 | 167.95 | | |
| $\hat{\mu}_3$ | 74.14 | | X_3 | 1.21 | 35.61 | 846.34 | |
| $\hat{\mu}_4$ | 127.08 | | X_4 | -71.74 | -63.96 | -281.75 | 4585.26 |
| $\hat{\mu}_5$ | 163.80 | | X_5 | 37.83 | -113.21 | 151.58 | -659.09 |

جدول (3-7) متوسط مربعات الخطأ (MSE) لموجات المتوسطات ومصفوفات التباين والتباين المشترك لطريقة (RMVV) و (S)

| MSE (\hat{M}) | | MSE ($\hat{\Sigma}$) | |
|-------------------|--------|------------------------|--------|
| RMVV | 0.3021 | RMVV | 1.1621 |
| S | 1.5395 | S | 5.4884 |

من خلال النتائج المبينة في الجدولين (2-7) و (3-7) يتضح افضلية طريقة مقدرات (RMVV) لأنها حققت أقل متوسط مربعات الخطأ لكل من موجه المتوسطات ومصفوفة التباين والتباين المشترك عند وجود الشواذ في البيانات .

8- الاستنتاجات والتوصيات

1-8 : الاستنتاجات Conclusions

على ضوء ما تم التوصل إليه من نتائج في الجانبين التجريبي والتطبيقي يمكن استخلاص اهم الاستنتاجات وكالاتي:

1. اثبتت طريقة مقدرات RMVV كفاءتها في تقدير معلمتي الموقع والقياس للتوزيع الطبيعي متعدد المتغيرات في حالة تلوث البيانات اذ حققت اقل قيمة لمعيار متوسط مربعات الخطأ مقارنة مع طريقة مقدرات (S) ، كما اثبتت بأنها الافضل في التقدير في حالة حجوم العينات الصغيرة والكبيرة .
2. ان قيم متوسطات مربعات القيم (MSE) لمعلمتي الموقع والقياس تتناقص تدريجيا بزيادة حجم العينة.
3. تبين من خلال الجانب التطبيقي ان الطريقة الافضل (RMVV) اعطت نتائج تقديرية لموجه المتوسطات (\hat{M}) مقارنة للمستويات الطبيعية للعوامل المؤثرة على مرضى الكلى (X_1, X_2, X_3, X_4, X_5) والتي تمثل الكرياتين والالبومين والبروتين والسكر واليورينا في الدم على التوالي عند وجود الشواذ.

2-8 : التوصيات The Recommendations

- بناءً على ما تم التوصل إليه من استنتاجات ندرج ادناه أهم التوصيات وكالاتي:
1. اعتماد طريقة (مقدرات RMVV) الحصينة في تقدير معلمتي الموقع والقياس للتوزيع الطبيعي متعدد المتغيرات في حالة تلوث البيانات كونها اعطت اقل (MSE) لموجه المتوسطات ومصفوفة التباين والتباين المشترك من الطريقة الاخرى.
 2. ضرورة الاعتماد على النتائج الخاصة بتقديرات متوسطات مقدر RMVV بالنسبة للمتغيرات التوضيحية (X_1, X_2, X_3, X_4, X_5) الخاصة بمرضى الكلى في حالة وجود الشواذ في البيانات الطبية.
 3. توفير قاعدة بيانات حديثة من خلال ارشفة وتبويب البيانات الطبية فضلا عن استخدام انظمة حديثة من البرمجيات في المؤسسات الصحية للحصول على بيانات واقعية ودقيقة والخاصة بالمرضى ممكن الاعتماد عليها في التحليل الاحصائي واستخراج نتائج جيدة .

المصادر

1. الجشعمي, حسين علي عبد الله (2007) م, "مقارنة لبعض المقدرات الحصينة لمعالم النماذج اللاخطية", اطروحة دكتوراه في الاحصاء, كلية الادارة والاقتصاد, الجامعة المستنصرية.
2. صالح ، أحمد مهدي ، (2017)، " تطوير مقدر مقلص لتقدير مصفوفة التباين والتباين المشترك ذات الابعاد الكبيرة بأستخدام الدوال المثلية الزائدية ، المجلة العراقية للعلوم الادارية /جامعة كربلاء ، العدد (52)، المجلد (13).
3. شهاب ، ضمياء حامد. " مقارنة بعض طرائق التقدير الحصينة مع أسلوب بيز في تقدير دالة الانحدار اللوجستي مع تطبيق عملي. رسالة ماجستير . كلية الادارة والاقتصاد . جامعة بغداد . 2017 .
4. العلوي، لقاء علي محمد ، " مقارنة مقدرات التباين المشترك الحصينة في تحليل المركبات الرئيسية "، اطروحة دكتوراه . كلية الادارة والاقتصاد . جامعة بغداد . 2003
5. Olive, D. J. (2018). **Robust multivariate analysis**. Springer.
6. Ali, H., Yahaya, S. S. S., & Omar, Z. (2014, June). **The efficiency of reweighted minimum vector variance**. In *AIP Conference Proceedings* (Vol. 1602, No. 1, pp. 1151-1156). AIP.
7. Verardi, V., & McCathie, A. (2012). **The S-estimator of multivariate location and scatter in Stata**. *Stata Journal*, 12(2), 299.
8. Hossjer, O., Croux, C., & Rousseeuw, P. J. (1994). **Asymptotics of generalized S-estimators**. *Journal of Multivariate Analysis*, 51(1), 148-177.
9. Croux, C. Rousseeuw, P.J. & Hossjer, O. (1994), "Generalized S - Estimator", *JASA*, Vol. 89, No. 428, 1271 – 1281.
10. Bigot, J., Biscay, R. J., Loubes, J. M., & Muñoz-Alvarez, L. (2011). **Group lasso estimation of high-dimensional covariance matrices**. *Journal of Machine Learning Research*, 12(Nov), 3187-3225.
11. Chen, Y., Wiesel, A., Eldar, Y. C., & Hero, A. O. (2010). **Shrinkage algorithms for MMSE covariance estimation**. *IEEE Transactions on Signal Processing*, 58(10), 5016-5029.
12. https://www.cs.indiana.edu/~predrag/classes/2005springi400/lecture_notes_4_2.pdf

Comparison between S capabilities and the smallest counterweight variable wave (RMVV) to estimate the location and measurement parameters of the multivariate normal distribution: An applied Research

Ahmed Abbas Musleh, Researcher

Prof. Dr. Hamza Ismail Shaheen

Abstract

In this research, robust and efficient capabilities of the mean wave (M) and the covariance and variance matrix (Σ) of the multivariate normal distribution were found when the hypotheses related to this distribution were penetrated and when the outliers were present in the sample under study, by using the method (smaller wave). The variance of the weights equation and its comparison with the method of estimators of S, where the simulation method was used to compare the two methods and through the mean squares of error (MSE) of the averages wave and the variance matrix and the covariance Σ to reach the best method of estimation, and the simulation results showed that the method of smaller wave variation of weights is the best because it achieved the lowest mean squares of error (MSE) than the other method, and with different sample sizes and contamination ratios, and on the application side, both methods were used on the data of kidney patients and the results showed an agreement with the experimental side, where the method achieved (the smallest wave contrast re-weighted) the preference in estimation through the MSE.