

# **مقارنة بعض طرائق التعويض الأحادي للبيانات المفقودة لدالة الكثافة الاحتمالية للتوزيع الطبيعي ثانوي المتغيرات**

أ. م. د. تهاني مهدي عباس      م. د. محمد جاسم محمد      م. د. قتيبة نبيل نايف  
جامعة بغداد- كلية الادارة والاقتصاد- قسم الاحصاء

## **المستخلص**

في هذا البحث تم اقتراح طريقة جديدة للتعويض عن القيم المفقودة في حالة دالة التوزيع الطبيعي ثانوي المتغيرات ومقارنة كفاءة هذا الطريقة مع طريقي التعويض الأحادي، بما طريقة تعويض المتوسط الشرطي وطريقة تعويض المتوسط غير الشرطي وباستخدام المحاكاة التي اظهرت ان الطريقة المقترنة اكثر كفاءة من الطريقتين.

### **Abstract:**

In this paper we suggest new method to estimate the missing data in bivariate normal distribution and compare it with Single Imputation method (Unconditional mean and Conditional mean) by using simulation.

## **١. المقدمة**

من المشاكل التي تواجه الباحثين أو القائمين على اتخاذ القرار هو وجود مشكلة المشاهدات المفقودة ضمن العينة قيد الدراسة وخاصة في حالة التوزيع الطبيعي ثانوي المتغيرات، لذلك كان لابد من التفكير في حل لمعالجة مثل هذه المشاكل وتعدت الطرائق المستخدمة في تحليل هكذا نوع من المشاكل وحسب آلية فقدان ونمط فقدان ومن الطرائق الشائعة الاستعمال هي طريقة التعويض الأحادي والمتمثلة بطريقتي التعويض بالمتوسط الشرطي وطريقة التعويض بالمتوسط غير الشرطي، علماً ان هناك طرائق كثيرة للتعويض لكن هذه الطريقتين اثبتتا كفاءة عند تقدير دالة الكثافة الاحتمالية للتوزيع، ولكن نقل كفاءة هذه الطرائق بالتقدير كلما زادت نسبة فقدان وخاصة عندما تتجاوز نسبة فقدان للبيانات  $20\%$  ، لذلك في هذا البحث سوف نقدم طريقة مقترنة جديدة في تعويض البيانات المفقودة وذات كفاءة عالية وخاصة عندما تتجاوز نسبة فقدان  $20\%$  من البيانات، وتم استخدام المحاكاة لمقارنة هذه الطريقة مع طريقي التعويض الأحادي وباعتماد معيار متوسط مربعات الخطأ.

## **٢. الجانب النظري**

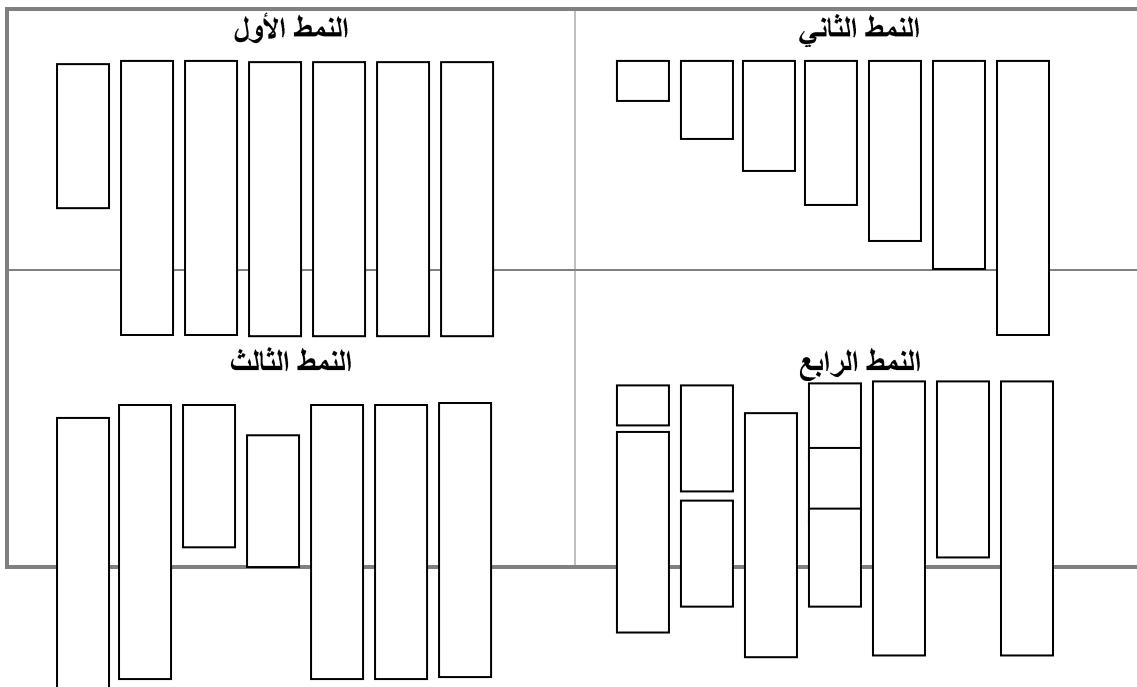
### **١. آليات وأنماط البيانات المفقودة**

ان مشكلة المشاهدات المفقودة تنتج من أحدى آليات فقدان الآتية<sup>[5]</sup>:

١. فقدان البيانات تماماً بشكل عشوائي (MCAR)
٢. فقدان البيانات بشكل عشوائي (MAR)
٣. فقدان لبيانات بشكل غير عشوائي (Not MAR)

أن لأي آلية من الآليات المبينة أثراً مسبباً في الظهور وعلى ضؤها تظهر مشكلة البيانات المفقودة على شكل أربعة أنماط مختلفة هي<sup>[5]</sup>:

١. النمط الأول، فقدان في أحد المتغيرات Pattern of univariate Missing Data
٢. النمط الثاني، النمط المرتب او المتداخل Monotone or Nested Missing Data
٣. النمط الثالث، عدم تطابق المعالم Missing Data with Unidentified Parameters
٤. النمط العام General Pattern



شكل (١)  
أنماط البيانات المفقودة

**٢. طرائق تقدير القيم المفقودة للتوزيع الطبيعي ثانوي المتغيرات :**  
إذا كان لدينا متوجه المتغيرات  $\underline{x}$  ذو البعد  $p$  يتبع التوزيع الطبيعي ثانوي المتغيرات<sup>[3]</sup>:

$$\underline{x}' = [x_1, x_2, \dots, x_p]$$

وأن

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

$$\underline{\mu}' = [\mu_1, \mu_2]$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

حيث ان

$\bar{x}_j$  يمثل متاجه المتوسطات

$\Sigma$  مصفوفة التباينات والتباينات المشتركة Variance – Covariance Matrix، وهي مصفوفة متتماثلة Symmetric.

وعلى فرض أن متاجه المتغيرات  $x_j$  يعني من وجد مشاهدات مفقودة وتحت شرط آلية فقدان MCAR ، ولتعويض قيم بدل عن القيم المفقودة سوف نستعمل طرائق التعويض الأحادي الآتية:

### 2.2.1 التعويض بالمتاجه الشرطي [6]: Unconditional Mean (UCM)

يتم التعويض عن كل قيم مفقودة لأي متاجه بمتوسط القيم للبيانات المتاحة (غير المفقودة) للمتغير الذي يعني من فقدان في مشاهداته :

$$\tilde{x}_j = \frac{\sum x_{\text{obs.}}}{n_j} \dots (2)$$

$n_j$  : تمثل عدد القيم المشاهدة فعلاً للمتغير  $x_j$ .

وكما هو واضح عند تعويض  $\tilde{x}_j$  محل القيمة المفقودة  $x_{miss}$  لمتغير  $x_j$  فإن المتوسط العام  $\bar{x}_j$  لهذا المتاجه يكون مساوياً إلى  $\tilde{x}_j$  ، وبما أن آلية فقدان MCAR فإن التباين  $\tilde{\sigma}_{jj}$  للمتغير  $\tilde{x}_j$  هو (تقدير متسلق للتباين الحقيقي) [5].

أن هذا الأسلوب يمكن استعماله في التعويض ولجميع أنماط فقدان تحت شرط آلية فقدان MCAR.

### 2.2.2 التعويض بالمتاجه الشرطي [6]: Conditional Mean (CM)

تعتمد هذا الطريقة على استخدام أنموذج الانحدار الخطى، ولفهم آلية هذه الطريقة نفرض لدينا متاجه المتغيرات  $[x_1, x_2] = x$  وأن المتاجه  $x_1$  متغير تام المشاهدات والمتاجه  $x_2$  يعني من وجود فقدان في بعض مشاهداته، وعليه فإن خطوات طريقة CM ستكون كما يلى:

1. يتم فرض المتاجه  $x_2$  الذي يعني من فقدان في بعض مشاهداته متغير استجابة (معتمد) response variable و  $x_1$  متغير توضيحي (مستقل) ، وبحذف كل قيمة للمتغير  $x_1$  التي تقابل قيمة مفقودة للمتغير  $x_2$  ، يصبح لدينا معادلة الانحدار الخطى التالية:

$$x_{i'2} = \alpha + \beta x_{i'1} + \varepsilon_{i'} \dots (3)$$

$$i' = 1, 2, \dots, n_c ; n_c < n$$

حيث أن  $n_c$  تمثل عدد قيم البيانات المتاحة

2. تستعمل طريقة المرربعات الصغرى OLS للحصول على تقدير لمعلمات معادلة الانحدار  $\alpha, \beta$  وكما يلى:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (X'_1 X_1)^{-1} X'_1 \underline{X}_2 \dots (4)$$

3. يتم حساب المتوسط الشرطي للمتغير  $x_2$  على المتاجه  $x_1$  وكما في المعادلة الآتية وأن هذا المتوسط الشرطي ما هو إلا القيمة التقديرية  $\hat{x}_{ij(miss.)}$  للقيمة المفقودة  $x_{ij(miss.)}$  وكما يلى:

$$E(x_2 / x_1) = \tilde{x}_2 + \tilde{\beta}(x_{i1} - \tilde{x}_1) \dots (5)$$



حيث أن:

$$\left. \begin{array}{l} \tilde{x}_1 = \frac{\sum_{i'=1}^{n_c} x_{i'1}}{n_c} \\ \tilde{x}_2 = \frac{\sum_{i'=1}^{n_c} x_{i'2}}{n_c} \end{array} \right\} \dots \quad (6)$$

٤. يتم التعويض بقيمة المتوسط الشرطي عن كل قيمة مفقودة أي

$$x_{i2(\text{miss.})} = E(x_2 / x_1)$$

أن هذا الأسلوب في التعويض لا يمكن استعماله لجميع انماط فقدان تحت شرط آلية فقدان MCAR وانما يمكن استخدامه فقد بالنمط الأول والثاني.

### 2.2.3 طريقة مقترحة: Proposed Method

تستند فكرة الطريقة على توليد قيم تحل محل القيم المفقودة وتتبع هذه القيم المولدة التوزيع الطبيعي ثانوي المتغيرات ذو معالم محدد يتم الحصول عليها من بيانات العينة المتاحة، وتم اقتراح أسلوبان للحصول على المعالم وحسب آلية فقدان وهذا الأسلوبان هما:

#### ١.٣.١ الأسلوب الأول

يستعمل هذا الأسلوب في حالة النمط الأول والثاني لفقدان وذلك لإمكانية الحصول على عينة جزئية من بيانات العينة الأصلية لا تتضمن بيانات مفقودة وتتلخص خطوات هذا الأسلوب بما يأتي:

١. يتم حذف صف المشاهدات الذي يحوي قيمة مفقودة ومن كل المتغيرات لحصل على عينة جزئية تامة من العينة الكلية.

٢. استعمال العينة الجزئية التي تم الحصول عليها في الخطوة (١) للحصول على تقديرات لمعلمات التوزيع الطبيعي ثانوي المتغيرات ( $\hat{\theta}, \hat{\tau}$ ) باستخدام طريقة الإمكان الأعظم.

٣. استعمال تقديرات لمعلمات التوزيع الطبيعي ثانوي المتغيرات ( $\hat{\theta}, \hat{\tau}$ ) التي تم الحصول عليها في الخطوة (٢) كمعلومات أولية في توليد مشاهدات تتبع التوزيع الطبيعي ثانوي المتغيرات وبحجم عينة مساوياً إلى حجم العينة الأصلية.

٤. يتم مقارنة البيانات الأصلية مع البيانات المولدة باستخدام الخطوة (٣) ويتم أحلال القيمة المولدة بدلاً عن كل قيمة مفقودة تقابلها.

ان هذا الأسلوب ينجح فقط في حالة النمط الأول والثاني ولا ينجح للنمطين الثالث والرابع وذلك لتعذر الحصول على عينة جزئية من العينة الكلية. وعليه في حالة النمط الثالث والرابع يتم الاعتماد على حدود الثقة Confidence Intervals لمعامل كل متغير وكما يلي:

## الطبيعي ثانوي المتغيرات

## ١.٣.٣.٢ الأسلوب الثاني

يستعمل هذا الأسلوب في حالة النمط الثالث والرابع لفقدان وذلك لتعذر الحصول على عينة جزئية من بيانات العينة الأصلية لا تتضمن بيانات مفقودة وتتلخص خطوات هذا الأسلوب بما يأتي:

١. يتم الحصول على عينة جزئية لكل متغير.
٢. استعمال العينة الجزئية للحصول على تقديرات لمعلمات التوزيع الطبيعي وكل متغير باستخدام طريقة الإمكان الأعظم.
٣. يتم الحصول على معلمات التوزيع الطبيعي ثانوي المتغيرات كما يأتي:  
ن. يتم الاعتماد على حدود الثقة **Confidence Intervals** لمتوسط كل متغير للحصول على قيمة متوجه المتوسطات.

$$\bar{x}_i \pm t_{\left(\frac{\alpha}{2}, n_i-1\right)} \cdot \frac{s_i}{\sqrt{n}} \dots (7)$$

حيث أن  $p, \dots, i=1, 2, \dots$  وأن  $\bar{x}_i$  و  $s_i^2$  تمثلان الوسط الحسابي والتباين المحسوبان

من المشاهدات المتوفرة لكل متغير

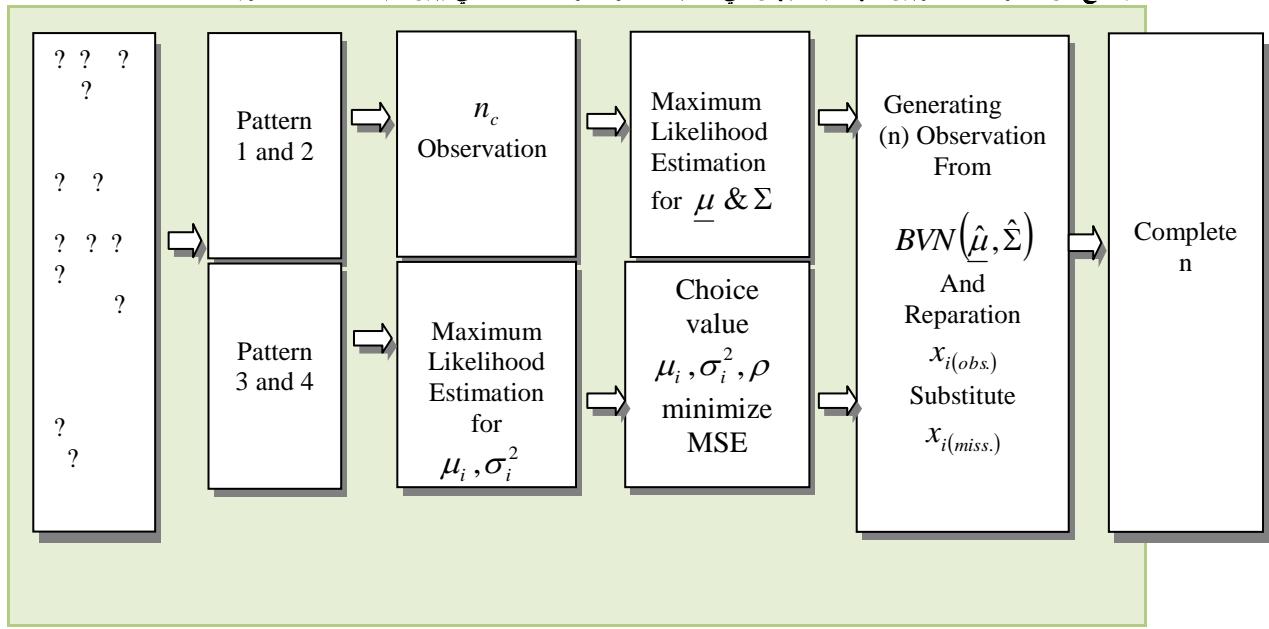
- ii. يتم الاعتماد على حدود الثقة **Confidence Intervals** لتباين كل متغير للحصول على التباين.

$$\frac{(n-1)s_i^2}{\chi_{\frac{\alpha}{2}}^2} \leq \sigma_i^2 \leq \frac{(n-1)s_i^2}{\chi_{(1-\frac{\alpha}{2})}^2} \dots (8)$$

- iii. يتم اخذ قيم لارتباط بين المتغيرات ضمن الفترة  $1 \leq \rho \leq 1$  – ويتم اعتماد القيمة التي تعطي اقل متوسط مربعات خطأ ومن خلال استعمال المحاكاة

٤. استعمال المعلمات المقدرة والتي تم الحصول عليها في الخطوة السابقة في توليد مشاهدات تتبع التوزيع الطبيعي ثانوي المتغيرات.

٥. يتم التعويض عن كل قيمة مفقودة بقيمة يتم توليدها كما في الخطوة السابقة.  
يتضح من خطوات الأسلوبين أنهما يتشابهان في اغلب الخطوات والمخطط الآتي يبين آلية عمل هذه الطريقة:



مخطط (١)  
يبين آلية عمل الطريقة المقترنة

### ٣. الجانب التجاري

لغرض معرفة كفاءة الطريقة المقترحة (التي تم الرمز لها بالرمز Prop.) في التعويض عن القيم المفقودة بالنسبة لطريقي التعويض بالمتوسط غير الشرطي (UCM) والمتوسط الشرطي (CM) وبيان تأثير هذه الطرائق بنسب الفقدان ونمط الفقدان وأختلاف حجوم العينة وتغير قيم التباينات والارتباطات تم اللجوء الى اسلوب المحاكاة وكما يلى:

❖ توليد توزيع طبيعي ثانوي المتغيرات باستخدام دالة التوليد (mvnrnd) المتوفرة في البرنامج الجاهز Matlab ، وتم استخدام تباينات مختلفة للمتغيرات المستخدمة وهي على التوالي :

$$X \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$X \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sqrt{2} \\ \rho\sqrt{2} & 2 \end{pmatrix}\right)$$

$$X \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & \rho\sqrt{15} \\ \rho\sqrt{15} & 3 \end{pmatrix}\right)$$

وقد تم استعمال قيم مختلفة للارتباطات هي<sup>(\*)</sup>  $\rho = 0.3, 0.5, 0.8$  كذلك تم استخدام أحجام مختلفة للعينات  $n = 25, 50, 100$

❖ تم توليد مصفوفة الفقدان تحت شرطالية فقدان من نوع MCAR<sup>[5]</sup> وللنظام الاول والثاني وبنسبة فقدان (10% , 20% , 30% , 40%).

❖ ولغرض المقارنة تم استعمال معيار متوسط مربعات الخطأ MSE . والجدوى الاتية توضح نتائج المحاكاة وفق المعطيات المذكورة اعلاه ، كذلك تم عرض بعض الاشكال لعدد من الحالات المستخدمة.

<sup>(\*)</sup> تم اخذ القيم الموجبة للارتباطات وأهمال القيم السالبة وذلك لعدم تأثير قيمة متوسط مربعات الخطأ باشرارة الارتباط لمزيد من التفاصيل راجع المصدر [ ١ ].



الطبيعي ثانوي المتغيرات

جدول (١)

يبين قيمة متوسط مربعات الخطأ لتقدير دالة الكثافة للتوزيع الطبيعي ثانوي المتغيرات ( $MSE(\hat{f}(\underline{x}))$ )  
وبحسب النمط الأول للقيم المفقودة وقيم التباينات وحجوم العينات المستخدمة ونسبة الفقدان  
(العدد الناتج مضروب في ١٠٠٠)

Missing		10%									20%								
n	$\rho$	0.3			0.5			0.8			0.3			0.5			0.8		
	$\sigma_1^2$	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0
	$\sigma_2^2$	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0
25	$\hat{f}_{UCM}$	٤٤	١٣	١.٧	٥٢	٢٦	٣.٤	٤٥٠	٢٠٤	٢٦	٣٧	١٩	٢.٦	٦٤	٣٣	٤.١	٣٤٥	١٨١	٢٤
	$\hat{f}_{CM}$	٤٤	١٧	٢.١	٨٣	٤١	٥.٥	١١٥٥	٥٣٧	٧١	٥٦	٢٩	٣.٩	١٣٤	٦٨	٨.٨	١٨٦١	٩٠١	١٢٠
	$\hat{f}_{Prop.}$	١٥	٨	١.١	٣٥	١٧	٢.٣	٣٢١	١٤٥	١٩	١٨	١٠	١.٢	٣٤	١٧	٢.١	٢٢٤	١١١	١٤
50	$\hat{f}_{UCM}$	١٢	٦.١	١.٦	٢٧	١٤	١.٨	٢٣٧	١٣٧	١٩	٢١	١١	١.٤	٣٩	١٩	٢.٥	٢٢٤	١١٠	١٥
	$\hat{f}_{CM}$	١٤	٧.٤	٢.١	٤١	٢٠	٢.٧	٦٢٤	٣١٩	٤٤	٣٠	١٥	١.٩	٨٠	٣٩	٥.١	١١٥٥	٥٧٢	٧٧
	$\hat{f}_{Prop.}$	٨	٤.٢	١.١	١٨	٩.٢	١.٢	٢٠١	٩٩	١٤	١٠	٥.٤	٠.٧	١٩	٩.٤	١.٢	١٢٤	٦٢	٨.٩
100	$\hat{f}_{UCM}$	٨.٣	٤	٠.٥	١٨	٩.٨	١.٣	٢٢٤	١١٣	١٥	١٥	٧.٦	٠.٩	٢٩	١٤	١.٩	١٧٩	٩١	١٢
	$\hat{f}_{CM}$	٩.٩	٤.٧	٠.٦	٢٨	١٥	١.٩	٥٣٩	٢٦٨	٣٥	٢٠	١٠	١.٣	٥٨	٢٨	٣.٧	٩٠٨	٤٥٧	٦٤
	$\hat{f}_{Prop.}$	٥.٤	٢.٦	٠.٣	١١	٦.٢	٠.٩	١٥٩	٨٠	١٠	٧.٧	٣.٨	٠.٥	١٣	٦.٦	٠.٩	٩٥	٤٧	٦.٧
Missing		30%									40%								
n	$\rho$	0.3			0.5			0.8			0.3			0.5			0.8		
	$\sigma_1^2$	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0
	$\sigma_2^2$	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0
25	$\hat{f}_{UCM}$	٦١	٣٠	٣.٨	٨٦	٤٨	٦.١	٣٠٧	١٥٥	٢٠	٨٢	٤٤	٥	١١٢	٥٩	٨	٣٠٢	١٤٦	١٨
	$\hat{f}_{CM}$	١١٧	٥٦	٧.٥	٢٧٩	١٥٤	٢٠	٣٩٣٩	١٨٩٥	٢٤٢	٢٠٧	١٢٠	١٥	٥٦٧	٣٣٣	٣٢	٧٦٤٤	٣٤٥١	٥٢٢
	$\hat{f}_{Prop.}$	٢٠	١١	١.٣	٣٣	١٥	٢.١	١٤٠	٧٢	٩.٣	٢٢	١١	١	٣٣	١٧	٢	١١٠	٥٦	٧
50	$\hat{f}_{UCM}$	٣٢	١٦	٢.٤	٥٥	٢٧	٣.٥	٢٠١	٩٥	١٣	٥٤	٢٦	٣.٦	٧٢	٣٦	٤.٨	١٨٥	٩٥	١٢
	$\hat{f}_{CM}$	٥١	٢٦	٣.٨	١٥٢	٧٣	٩.٥	٢١٤٣	٩٧٧	١٣٤	١٠٤	٤٨	٧.٨	٢٧٠	٢٣٤	١٨	٣٩٢١	١٨٨٦	٢٥٥
	$\hat{f}_{Prop.}$	١٢	٦.٣	٠.٩	٢٠	١٠	١.٣	٩٠	٤١	٦	١٥	٧.٤	٠.٩	٢١	١١	١.٤	٧٠	٣٦	٤.٦
100	$\hat{f}_{UCM}$	٢٦	١٣	١.٧	٤٠	٢١	٣	١٥٩	٨٠	١٠	٤٢	٢١	٢.٨	٥٧	٢١	٢.٧	١٤٢	٧٤	٩.٦
	$\hat{f}_{CM}$	٣٩	١٩	٢.٥	١٠٣	٥٣	٧	١٦٠٥	٨١٠	١٠٧	٧٣	٣٦	٤.٩	١٩٣	٣٥	٤.٦	٢٨٨١	١٤١١	١٩٠
	$\hat{f}_{Prop.}$	١٠	٥	٠.٧	١٥	٨	١	٦٨	٣٤	٥	١٣	٦.٢	٠.٨	١٧	٦.٢	٠.٨	٥٢	٢٧	٣.٦

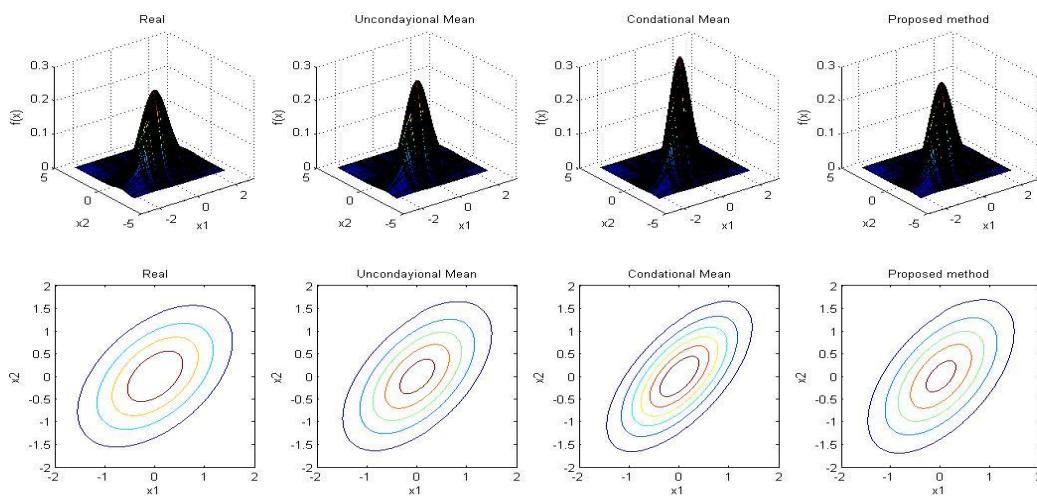


الطبيعي ثانوي المتغيرات

شكل رقم (١)

يشير الى القيم الحقيقية والتقديرات دالة الكثافة الإحتمالية وحسب النمط الأول عند حجم عينة  $n = 100$  وبيانات

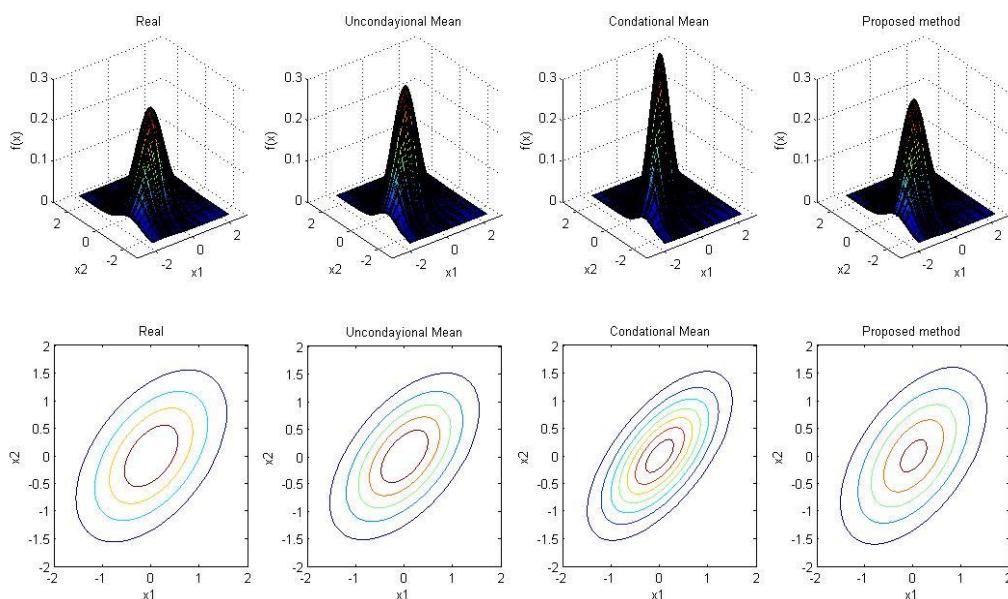
$$\rho = 0.5 \text{ وقيمة الارتباط } \sigma_1^2 = 1, \sigma_2^2 = 1 \text{ ونسبة فقدان}$$



شكل رقم (٢)

يشير الى القيم الحقيقية والتقديرات دالة الكثافة الإحتمالية وحسب النمط الأول عند حجم عينة  $n = 100$  وبيانات

$$\rho = 0.5 \text{ وقيمة الارتباط } \sigma_1^2 = 1, \sigma_2^2 = 1 \text{ ونسبة فقدان} 20\%$$



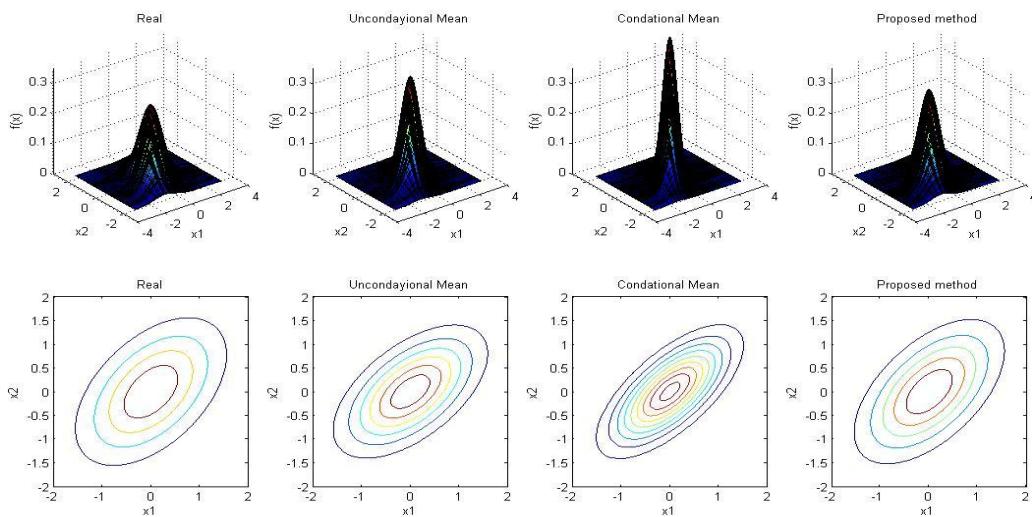


الطبيعي ثانوي المتغيرات

شكل رقم (٣)

يشير الى القيم الحقيقية والتقديرات دالة الكثافة الإحتمالية وحسب النمط الأول عند حجم عينة  $n = 100$  وبيانات

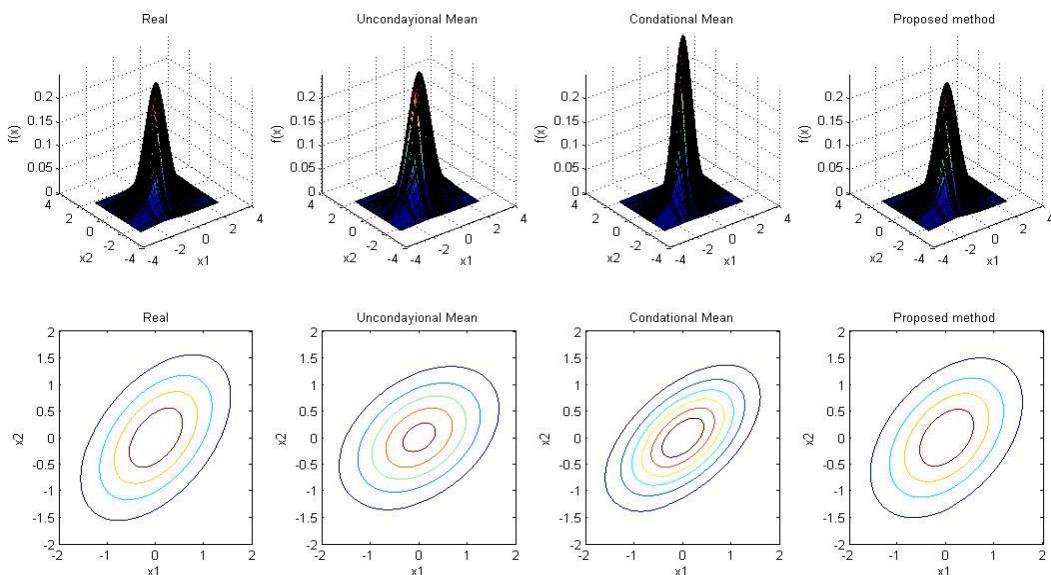
$$30\% \quad \rho = 0.5 \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 1$$



شكل رقم (٤)

يشير الى القيم الحقيقية والتقديرات دالة الكثافة الإحتمالية وحسب النمط الأول عند حجم عينة  $n = 100$  وبيانات

$$40\% \quad \rho = 0.5 \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 1$$





الطبيعي ثانوي المتغيرات

جدول (٢)

يبين قيمة متوسط مربعات الخطأ لتقدير دالة الكثافة للتوزيع الطبيعي ثانوي المتغيرات ( $MSE(\hat{f}(\underline{x}))$ ) وحسب النمط الثالث للقيم المفقودة وقيم البيانات وحجم العينات المستخدمة ونسب الفقدان (العدد الناتج مضروب في ١٠٠٠٠)

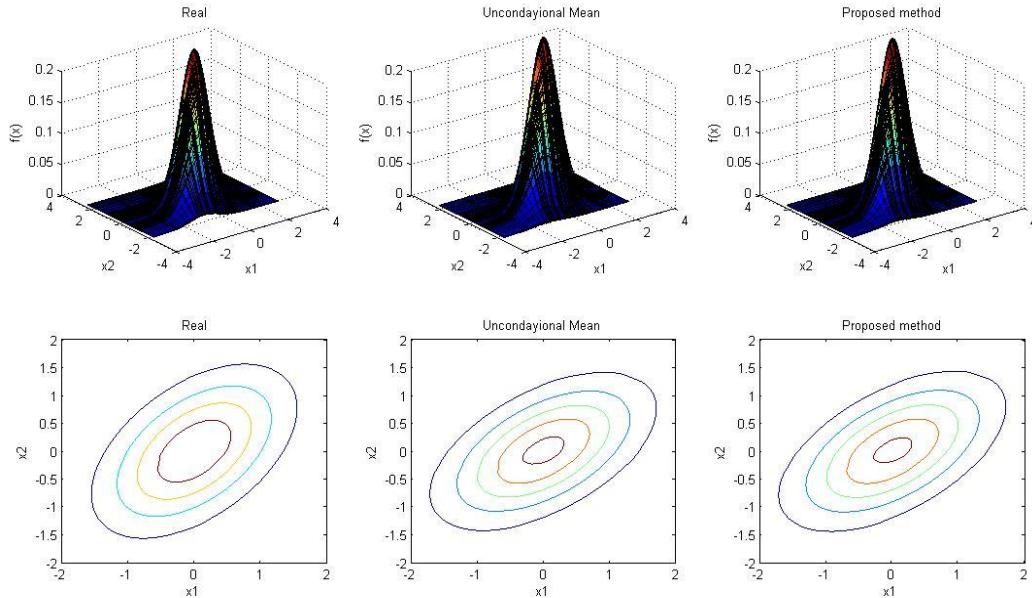
Missing		10%									20%								
n	$\rho$	0.3			0.5			0.8			0.3			0.5			0.8		
	$\sigma_1^2$	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0
	$\sigma_2^2$	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
25	$\hat{f}_{UCM}$	٤٦	١٢	١.٧	٤٠	١٩	٢.٦	١٨٥	١٠٠	١٣	٤٢	١٩	٢.٧	٥٩	٢٧	٤.١	١٩٢	٩٩	١٣
	$\hat{f}_{Prop.}$	٣٤	١٩	٥.٢	٤٣	٢٥	٨.٩	١١٠	٥٩	٣٠	٣٦	٢٠	٥.٨	٣٤	٢٦	٨.٢	١٠٣	٥٣	١٩
50	$\hat{f}_{UCM}$	١٤	٧.٢	٩.٩	٢٤	١٢	١.٧	١٤٦	٦٦	٩.٦	٢٧	١٤	١.٩	٣٨	٢١	٢.٦	١٢٧	٦٣	٨.٧
	$\hat{f}_{Prop.}$	٢٨	١٥	٤.٦	٣٤	٢٠	٧	٨٧	٤٦	٢٣	٣٠	١٧	٤.٧	٣٥	٢١	٦.٣	٨٣	٤١	١٤
100	$\hat{f}_{UCM}$	١١	٥.٥	٧.٢	٢٠	١٠	١.٣	١٢٤	٦٢	٨	٢١	١١	١.٤	٣٢	١٦	٢.١	١٠٧	٥٣	٧.١
	$\hat{f}_{Prop.}$	٢٦	١٤	٤.٢	٣١	١٨	٦.٤	٧٦	٣٩	٢٠	٢٧	١٥	٤.٢	٣٢	١٩	٥.٦	٧٣	٣٦	١١
Missing		30%									40%								
n	$\rho$	0.3			0.5			0.8			0.3			0.5			0.8		
	$\sigma_1^2$	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0
	$\sigma_2^2$	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
25	$\hat{f}_{UCM}$	٦٧	٣٤	٤.٧	٧٦	٤٠	٥.٤	١٦٨	٨٦	١٣	٧٧	٣٩	٥.٣	٩٦	٤٧	٦.٦	١٦٥	٨٩	١١
	$\hat{f}_{Prop.}$	٣٨	٢٣	٦.٥	٤٥	٢٧	٧.٦	٩٩	٥٠	١٤	٣٩	٢٥	٦.٨	٤٦	٢٨	٨.٣	٩٦	٤٨	١٢
50	$\hat{f}_{UCM}$	٤١	٢١	٢.٨	٥٥	٢٧	٣.٦	١٢١	٥٧	٧.٦	٥٣	٢٧	٣.٥	٦٥	٣٢	٤.٢	١١٢	٥٦	٧.٧
	$\hat{f}_{Prop.}$	٣١	١٨	٥.٢	٣٦	٢١	٦.٣	٨١	٣٩	١٠	٣٣	٢١	٥.٨	٣٧	٢٣	٦.٥	٧٨	٣٨	٩.٣
100	$\hat{f}_{UCM}$	٣٤	١٧	٢.٣	٤٣	٢١	٢.٩	٩٧	٤٥	٦.٤	٤١	٢١	٢.٧	٤٩	٢٥	٣.٣	٨٨	٤٤	٦
	$\hat{f}_{Prop.}$	٢٨	١٧	٤.٨	٣٣	١٩	٥.٧	٧٣	٣٥	٨.٨	٢٩	١٩	٥.٣	٣٣	٢١	٦	٧١	٣٥	٨



الطبيعي ثانوي المتغيرات

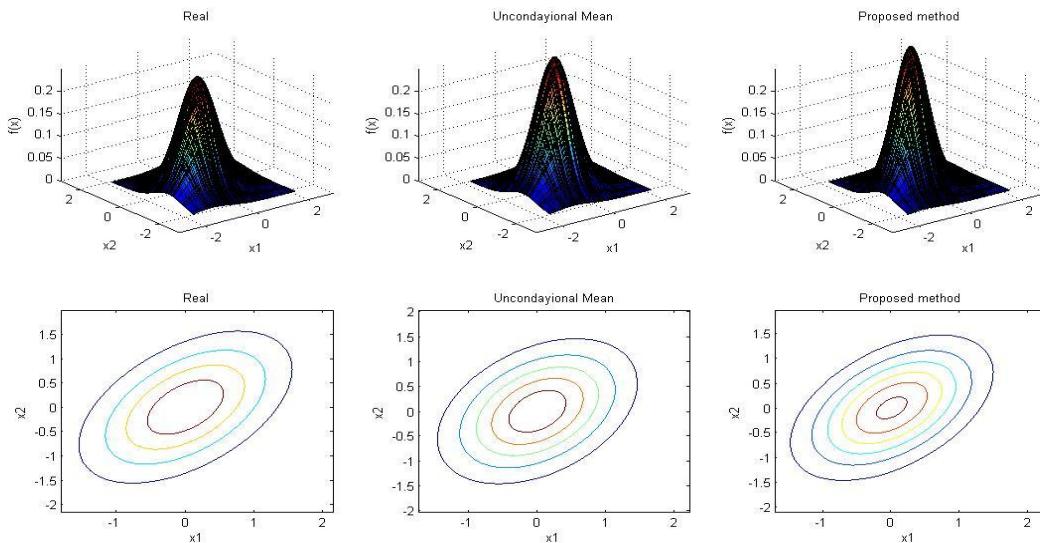
شكل رقم (٥)

يشير الى القيم الحقيقية والتقديرات لدالة الكثافة الإحتمالية وحسب النمط الثالث عند حجم عينة  $n = 100$  وتبينات  $\sigma_1^2 = 1$  ،  $\sigma_2^2 = 1$  وقيمة الارتباط  $\rho = 0.5$  ونسبة فقدان 10%



شكل رقم (٦)

يشير الى القيم الحقيقية والتقديرات لدالة الكثافة الإحتمالية وحسب النمط الثالث عند حجم عينة  $n = 100$  وتبينات  $\sigma_1^2 = 1$  ،  $\sigma_2^2 = 1$  وقيمة الارتباط  $\rho = 0.5$  ونسبة فقدان 20%

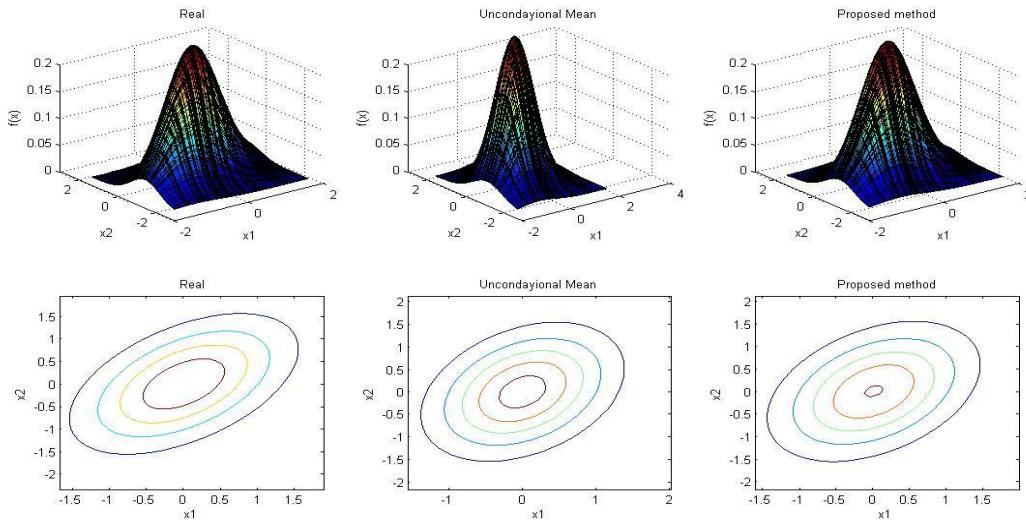




الطبيعي ثانوي المتغيرات

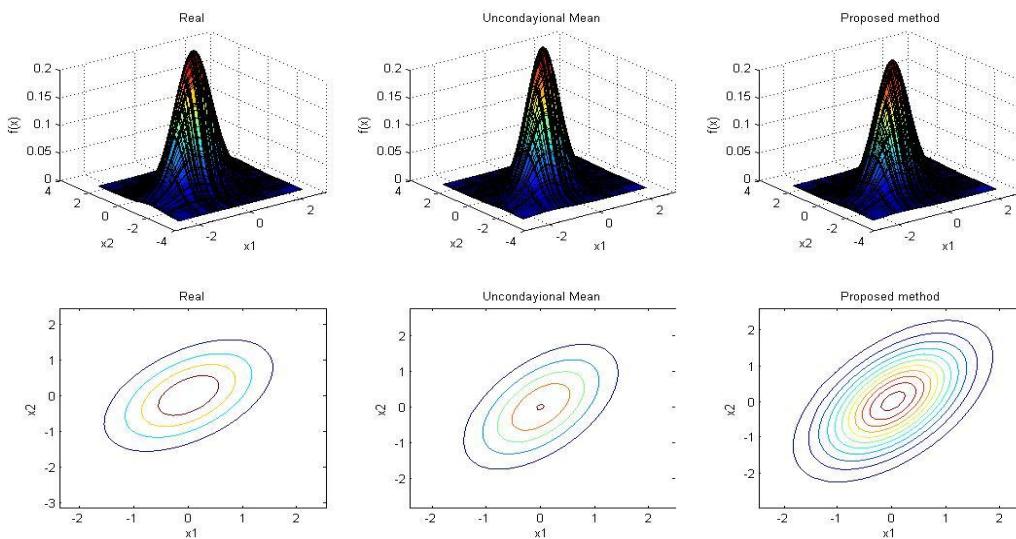
شكل رقم (٧)

يشير الى القيم الحقيقية والتقديرات لدالة الكثافة الاحتمالية وحسب النمط الثالث عند حجم عينة  $n = 100$  وتبينات  $\sigma_1^2 = 1$  ،  $\sigma_2^2 = 1$  وقيمة الارتباط  $\rho = 0.5$  ونسبة فقدان 30%



شكل رقم (٨)

يشير الى القيم الحقيقية والتقديرات لدالة الكثافة الاحتمالية وحسب النمط الثالث عند حجم عينة  $n = 100$  وتبينات  $\sigma_1^2 = 1$  ،  $\sigma_2^2 = 1$  وقيمة الارتباط  $\rho = 0.5$  ونسبة فقدان 40%



#### ٤. تفسير النتائج

من الجدول (١) نلاحظ النتائج الآتية: (النطء الأول للفقدان)

- لجميع نسب فقدان ولجميع حجوم العينات والتباينات وقيم الارتباطات المستخدمة أظهرت النتائج ان الطريقة المقترحة في التعويض عن القيمة المفقودة أفضل من طريقة التعويض بالمتوسط غير الشرطي UCM وطريقة التعويض بالمتوسط الشرطي CM وخاصة عند زيادة نسب فقدان ( 40% ، 30%)، إذ تمتع طريقة التعويض بالمتوسط غير الشرطي UCM والمتوسط الشرطي CM بعدم الكفاءة وخاصة عند نسب فقدان ( 30% ، 40%).
- نلاحظ عند زيادة قيم الارتباطات استقرار الطريقة المقترحة ولكن نلاحظ تأثير وبشكل كبير لكل من طريقة التعويض بالمتوسط غير الشرطي UCM وطريقة التعويض بالمتوسط الشرطي CM ، وهذا ما يعبّر على هاتين الطريقتين لأنه عند زيادة قيم الارتباطات تقل كفاءة هاتين الطريقتين.
- من الشكل ١ و ٢ و ٣ و ٤ نلاحظ قلة التحيز للطريقة المقترحة في التعويض عن القيمة المفقودة ، اي أن البيانات التي تم تقديرها تكون قريبة جداً من البيانات الحقيقة.
- أما ما يخص طريقة التعويض الأحادي ( UCM ، CM ) كانت طريقة UCM أفضل من طريقة CM لجميع الحالات .

من الجدول (٢) نلاحظ النتائج الآتية: (النطء الثالث للفقدان)

- عند نسبة فقدان 10% ولجميع التباينات وقيمة الارتباط ( 0.5 ، 0.3 ) وحجم عينة ٢٥ كفاءة طريقة التعويض UCM على الطريقة المقترحة أما عند قيمة ارتباط ٠.٨ كفاءة الطريقة المقترحة عند قيمة تباين  $\sigma^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$  ولكن ، أما في حالة حجم عينة ٥٠ و ١٠٠ نلاحظ كفاءة الطريقة UCM عند قيمة ارتباط ٠.٥ ولكن عند قيمة ارتباط ٠.٣ تكون الطريقة المقترحة أفضل عندما  $\sigma^2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$ .
- عند نسبة فقدان 20% وقيمة ارتباط ( 0.5 ، 0.3 ) ولجميع حجوم العينات تكون قيم متوازنات الخطأ التربيعي متقاربة جداً للطريقتين ولكن الأمثلية تكون لطريقة UCM ، أما عند قيمة ارتباط ٠.٨ تكون طريقة التعويض عن القيمة المفقودة المقترحة هي الأفضل و خاصة عندما  $\sigma^2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$ .
- عند نسبة فقدان ( 40% ، 30% ) ولجميع حجوم العينات وقيم الارتباطات وعندما  $\sigma^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$  تكون الطريقة المقترحة في التعويض عن القيمة المفقودة هي الأفضل.
- من الشكل ٥ و ٦ و ٧ و ٨ نلاحظ قلة التحيز للطريقة المقترحة في التعويض عن القيمة المفقودة ، اي أن البيانات التي تم تقديرها تكون قريبة جداً من البيانات الحقيقة.

## ٥. الاستنتاجات

- \* في حالة النمط الأول من أنماط فقدان أظهرت الطريقة المقترحة للتعويض عن القيمة المفقودة أفضلية في التقدير ولجميع نسب فقدان وحجوم العينات وقيم التباينات وقيم الارتباطات المستعملة.
- \* في حالة النمط الثالث من أنماط فقدان أظهرت النتائج تفاوت في كفاءة الطريقة المقترحة بالرغم من ان نتائج متوسط مربعات الخطأ للطريقتين ( المقترحة و UCM ) كانت متقاربة جدا وبصورة عامة كانت الطريقة المقترحة هي الافضل عند زيادة نسب فقدان ولجميع حجوم العينات وقيم الارتباط وعندما

$$\cdot \underline{\sigma}^2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \underline{\sigma}^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

## ٦. المصادر:

١. مناف يوسف حمود، تهاني مهدي عباس، قبيبة نبيل نايف، ٢٠٠٨ "تقدير لا معلمى لدالة كثافة احتمالية متعددة متغيرات" مجلة جامعة النهرین - العلوم ، المجلد ١١ ، العدد ٢ ، صفحة ٥٥ – ٦٣ .
2. Dagenais, M.G. (1973) "The use of Incomplete Observation in Multiple Regression Analysis: A generalized Least Squares Approach" J. of Econometrics, vol. 1, p. 317 – 328.
3. Heumann, C.; Nittner, T.; Scheid, S. & Toutenburg, H. (2002)"Parametric and Nonparametric Regression Missing X's: A Review" [www.Stat.uni-muenchen.de/sfb386/papers/dsp/paper286.ps](http://www.Stat.uni-muenchen.de/sfb386/papers/dsp/paper286.ps).
4. Kosobud, R. (1963) "A note on Problem Caused by Assignment of Missing Data in Sample Surveys" Econometrica, vol. 31, p. 562 – 563.
5. Little, R.J.A & Rubin, D.B (2003) "Statistical Analysis with Missing Data" 2<sup>nd</sup> ed., John Wile & Sons, New York.
6. Nittner, T. (2002) "The Additive with Missing values in the Independent Variable: Theory & Simulation" <http://www.pms.ifi.lmu.de/research-report/index.pdf>