مقارنة بعض طرائق التعويض الأحادي للبيانات المفقودة لدالة الكثافة الاحتمالية للتوزيع الطبيعي ثنائي 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 ثنائي المتغيرات ومقارنة كفاءة هذا الطريقة مع طريقتي التعويض الأحادي، هما طريقة تعويض المتوسط الشرطي وطريقة تعويض المتوسط غير الشرطي وباستخدام المحاكاة التي اظهرت ان الطريقة المقترحة 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.

1. 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 ثنائي المتغيرات، لذلك كان لابد من التفكير في حل لمعالجة مثل هذه المشاكل وتعددت الطرائق المستخدمة في تحليل هكذا نوع من المشاكل وحسب آلية الفقدان ونمط الفقدان 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 تقل كفاءة هذه الطرائق بالتقدير كلما زادة نسبة الفقدان وخاصة عندما تتجاوز نسبة الفقدان للبيانات 20% ، لذلك في هذا البحث سوف نقدم طريقة مقترحة جديدة في تعويض البيانات المفقودة وذات كفاءة عالية وخاصة عندما تتجاوز نسبة الفقدان 20% من البيانات، وتم استخدام المحاكاة لمقارنة هذه الطريقة مع طريقتي التعويض الاحادي وباعتماد معيار متوسط مربعات الخطأ.

2. الجانب النظري

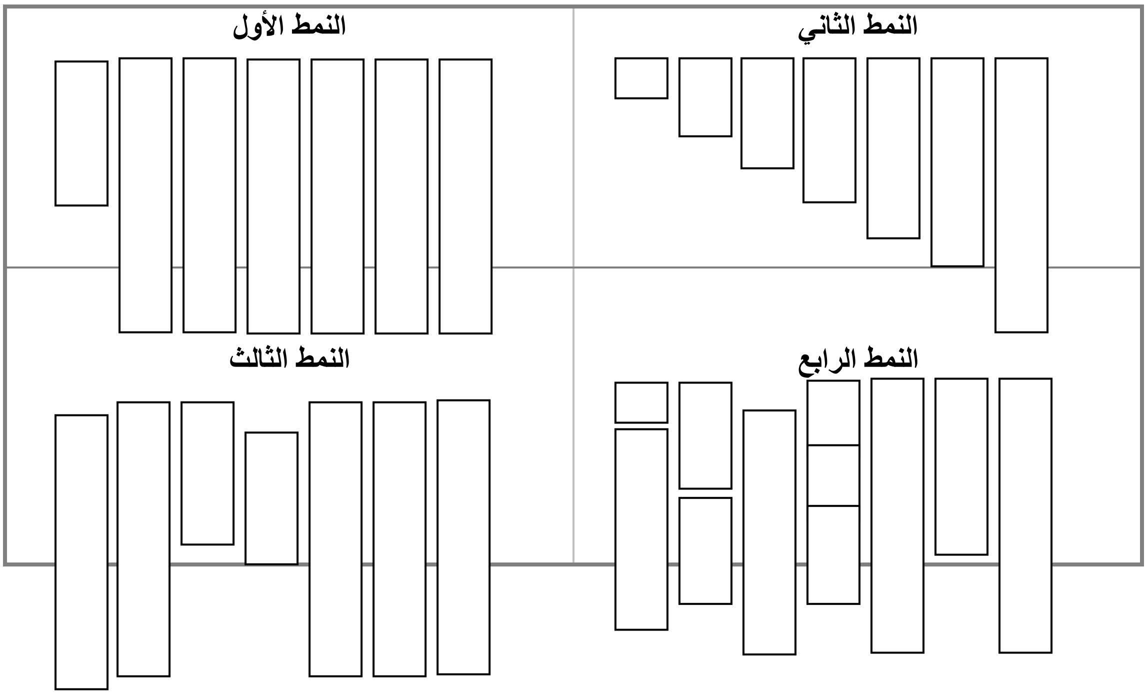
2.1 آليات وأنماط البيانات المفقودة

ان مشكلة المشاهدات المفقودة تنتج من أحدى آليات الفقدان الأتية[5]:

1. فقدان البيانات تماماً بشكل عشوائي **Missing Complete at Random (MCAR)**
2. فقدان البيانات بشكل عشوائي **Missing at Random (MAR)**
3. فقدان لبيانات بشكل غير عشوائي **Missing not at Random (Not MAR)**

أن لأي آلية من الآليات المبينة أنفا مسببات في الظهور وعلى ضوءها تظهر مشكلة البيانات المفقودة على شكل أربعة أنماط مختلفه هي[5]:

1. النمط الأول، فقدان في أحد المتغيرات **Pattern of univariate Missing Data**
2. النمط الثاني، النمط المرتب او المتداخل **Monotone or Nested Missing Data**
3. النمط الثالث، عدم تطابق المعالم **Missing Data with Unidentified Parameters**
4. النمط العام **General Pattern**



شكل (1)
أنمــاط البيانـــات المفقــودة

<u>2.2 طرائق تقدير القيم المفقودة للتوزيع الطبيعي ثنائي المتغيرات</u> :

اذا كان لدينا متجه المتغيرات $\underline{x}$ ذو البعد **p** يتبع التوزيع الطبيعي ثنائي المتغيرات [3]:

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

وأن

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

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

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

حيث ان

$\underline{\mu}'$ يمثل متجه المتوسطات

Σ مصفوفة التباينات والتباينات المشتركة **Variance – Covariance Matrix**، وهي مصفوفة متماثلة **Symmetric**.

وعلى فرض أن متجة المتغيرات $\underline{x}$ يعاني من وجد مشاهدات مفقودة وتحت شرط آلية فقدان **MCAR** ، ولتعويض قيم بدل عن القيم المفقودة سوف نستعمل طرائق التعويض الأحادي الأتية:

2.2.1 التعويض بالمتوسط غير الشرطي[6]: Unconditional Mean (UCM)

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

$$\tilde{x}_j = \frac{\sum\limits_{n_j} x_{obs.}}{n_j} \quad \cdots(2)$$

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 فأن المتوسط العام $\bar{x}_j$ لهذا المتغير يكون مساوياً الى $\tilde{x}_j$ ، وبما أن آلية الفقدان **MCAR** فأن التباين $\tilde{\sigma}_{jj}$ للمتغير $\tilde{x}_j$ هو (تقدير متسق للتباين الحقيقي)[5] .

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

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

تعتمد هذا الطريقة على أستخدام أنموذج الانحدار الخطي، ولفهم آلية هذه الطريقة نفرض لدينا متجه المتغيرات $\underline{x}' = [x_1, x_2]$ وأن المتغير x_1 متغير تام المشاهدات والمتغير x_2 يعاني من وجود فقدان في بعض مشاهداته، وعلية فأن خطوات طريقة **CM** ستكون كما يلي:

1. يتم فرض المتغير x_2 الذي يعاني من فقدان في بعض مشاهداته متغير أستجابة (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 الخطي التالية:

$$x_{i'2} = \alpha + \beta x_{i'1} + \varepsilon_{i'} \quad \cdots(3)$$
$$i' = 1, 2, ...n_c \ ; \ n_c < n$$

حيث أن n_c تمثل عدد قيم البيانات المتاحة

2. تستعمل طريقة المربعات الصغرى **OLS** للحصول على تقدير لمعلمات معادلة الانحدار α, β وكما يلي:

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

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

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

حيث أن:

$$\left.\begin{aligned}\tilde{x}_1 &= \frac{\sum_{i'=1}^{n_c} x_{i'1}}{n_c}\\ \tilde{x}_2 &= \frac{\sum_{i'=1}^{n_c} x_{i'2}}{n_c}\end{aligned}\right\} \cdots(6)$$

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

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

أن هذا الأسلوب في التعويض لا يمكن أستعمالة لجميع انماط الفقدان 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 ثنائي المتغيرات ذو معالم محدد يتم الحصول عليها من بيانات العينة المتاحة، وتم اقتراح أسلوبان للحصولعلى المعالم وحسب آلية الفقدان وهذان الأسلوبان هما:

1.3.3.1 الأسلوب الأول

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

1. يتم حذف صف المشاهدات الذي يحوي قيمة مفقودة ومن كل المتغيرات لنحصل على عينة جزئية تامه من العينة الكلية .
2. استعمال العينة الجزئية التي تم الحصول عليها في الخطوة (1) للحصول على تقديرات لمعلمات التوزيع الطبيعي ثنائي المتغيرات $(\hat{\mu}, \hat{\Sigma})$ باستخدام طريقة الإمكان الأعظم .
3. استعمال تقديرات معلمات التوزيع الطبيعي ثنائي المتغيرات $(\hat{\mu}, \hat{\Sigma})$ التي تم الحصول عليها في الخطوة (2) كمعلومات أولية في توليد مشاهدات تتبع التوزيع الطبيعي ثنائي المتغيرات وبحجم عينة مساوي الى حجم العينة الأصلية .
4. يتم مقارنة البيانات الأصلية مع البيانات المولدة باستخدام الخطوة (3) ويتم أحلال القيمة المولدة بدلا 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 الأعتماد على حدود الثقة **Confidence Intervals** لمعالم كل متغير وكما يلي:

1.3.3.2 الأسلوب الثاني

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

1. يتم الحصول على عينة جزئية لكل متغير.
2. استعمال العينة الجزئية للحصول على تقديرات لمعلمات التوزيع الطبيعي ولكل متغير باستخدام طريقة الإمكان الأعظم .
3. يتم الحصول على معلمات التوزيع الطبيعي ثنائي المتغيرات كما يأتي:

i. يتم الاعتماد على حدود الثقة Confidence Intervals لمتوسط كل متغير للحصول على قيمة متجه المتوسطات .

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

حيث أن $i=1\,,2\,,\cdots,\ p$ وأن $\bar{x}_i$ و s_i^2 تمثلان الوسط الحسابي والتباين المحسوبان من المشاهدات المتاحة لكل متغير

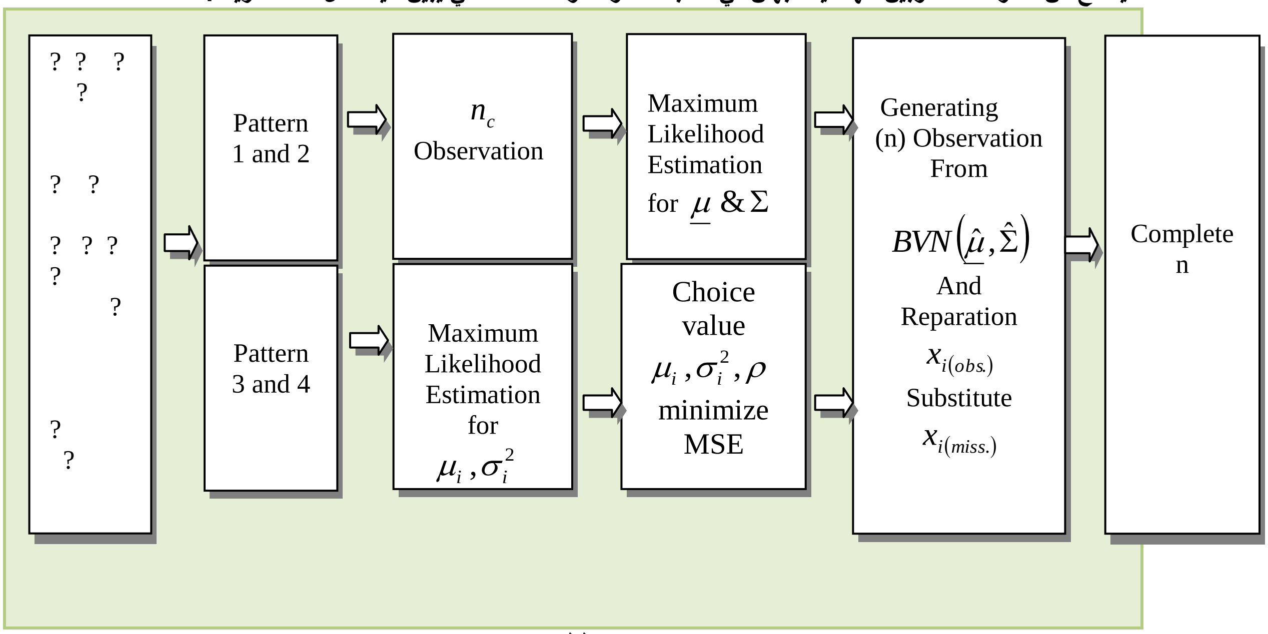
ii. يتم الاعتماد على حدود الثقة Confidence Intervals لتباين كل متغير للحصول على التباين.

$$\frac{(n-1)\, s_i^2}{\chi^2_{\frac{\alpha}{2}}} \le \sigma_i^2 \le \frac{(n-1)\, s_i^2}{\chi^2_{\left(1-\frac{\alpha}{2}\right)}} \quad ...(8)$$

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

4. استعمال المعلمات المقدرة والتي تم الحصول عليها في الخطوة السابقة في توليد مشاهدات تتبع التوزيع الطبيعي ثنائي المتغيرات.
5. يتم التعويض عن كل قيمة مفقودة بقيمة يتم توليده كما في الخطوة السابقة.

يتضح من خطوات الأسلوبين أنهما يتشابهان في اغلب الخطوات والمخطط الآتي يبين آلية عمل هذه الطريقة:



مخطط (1)
يبين آلية عمل الطريقة المقترحة

3. الجانب التجريبي

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

- **توليد توزيع طبيعي ثنائي 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(*) $\rho = 0.3\ ,\ 0.5\ ,\ 0.8$ كذلك تم استخدام أحجام مختلفة للعينات $n = 25\ ,\ 50\ ,\ 100$

- **تم توليد مصفوفة الفقدان تحت شرط الية فقدان من نوع MCAR [5] وللنمط الاول والثاني وبنسب فقدان $(10\%\ ,\ 20\%\ ,\ 30\%\ ,\ 40\%)$.**
- **ولغرض المقارنة تم استعمال معيار متوسط مربعات الخطأ MSE.**

والجداول الاتية توضح نتائج المحاكاة وفق المعطيات المذكورة انفاً ، كذلك تم عرض بعض الاشكال لعدد من الحالات المستخدمة.

(*) تم اخذ القيم الموجبة للارتباطات وأهمال القيم السالبة وذلك لعدم تأثر قيمة متوسط مربعات الخطأ باشارة الارتباط لمزيد من التفاصيل راجع المصدر [1].

الطبيعي ثنائي المتغيرات

جدول (1)

يبين قيمة متوسط مربعات الخطأ لتقدير دالة الكثافة للتوزيع الطبيعي ثنائي المتغيرات $MSE(\hat{f}(\underline{x}))$
وحسب النمط الأول للقيم المفقودة وقيم التباينات وحجوم العينات المستخدمة و نسب الفقدان
(العدد الناتج مضروب في 10000)

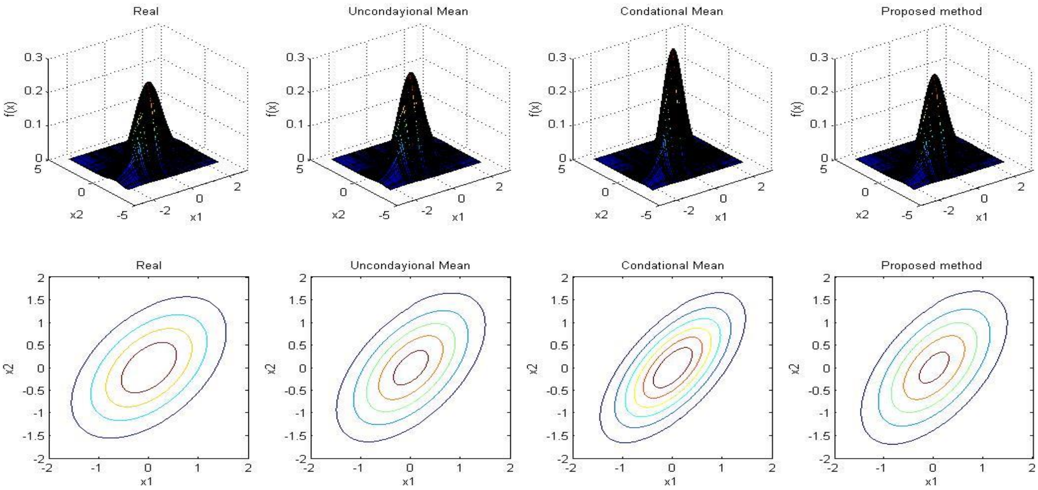
Missing		10%									20%								
n	ρ	0.3			0.5			0.8			0.3			0.5			0.8		
	σ_1^2	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5
	σ_2^2	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
25	$\hat{f}_{UCM}$	24	13	1.7	52	26	3.4	450	204	26	37	19	2.6	64	33	4.1	345	181	24
	$\hat{f}_{CM}$	32	17	2.1	83	41	5.5	1155	537	71	56	29	3.9	134	68	8.8	1861	901	120
	$\hat{f}_{Prop.}$	15	8	1.1	35	17	2.3	321	145	19	18	10	1.2	34	17	2.1	224	111	14
50	$\hat{f}_{UCM}$	12	6.1	1.6	27	14	1.8	237	137	19	21	11	1.4	39	19	2.5	224	110	15
	$\hat{f}_{CM}$	14	7.4	2.1	41	20	2.7	624	319	44	30	15	1.9	80	39	5.1	1105	572	77
	$\hat{f}_{Prop.}$	8	4.2	1.1	18	9.2	1.2	201	99	14	10	5.4	0.7	19	9.4	1.2	124	62	8.9
100	$\hat{f}_{UCM}$	8.3	4	0.5	18	9.8	1.3	224	113	15	15	7.6	0.9	29	14	1.9	179	91	12
	$\hat{f}_{CM}$	9.9	4.7	0.6	28	15	1.9	539	268	35	20	10	1.3	58	28	3.7	908	457	64
	$\hat{f}_{Prop.}$	5.4	2.6	0.3	11	6.2	0.9	159	80	10	7.7	3.8	0.5	13	6.6	0.9	95	47	6.7

Missing		30%									40%								
n	ρ	0.3			0.5			0.8			0.3			0.5			0.8		
	σ_1^2	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5
	σ_2^2	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
25	$\hat{f}_{UCM}$	61	30	3.8	86	48	6.1	307	155	20	82	44	5	112	59	8	302	146	18
	$\hat{f}_{CM}$	117	56	7.5	279	154	20	3939	1895	242	207	120	15	567	333	32	7644	3451	522
	$\hat{f}_{Prop.}$	20	11	1.3	33	15	2.1	140	72	9.3	22	11	1	33	17	2	110	56	7
50	$\hat{f}_{UCM}$	32	16	2.4	55	27	3.5	201	95	13	54	26	3.6	72	36	4.8	185	95	12
	$\hat{f}_{CM}$	51	26	3.8	152	73	9.5	2143	977	134	104	48	7.8	270	234	18	3921	1886	255
	$\hat{f}_{Prop.}$	12	6.3	0.9	20	10	1.3	90	41	6	15	7.4	0.9	21	11	1.4	70	36	4.6
100	$\hat{f}_{UCM}$	26	13	1.7	40	21	3	159	80	10	42	21	2.8	57	21	2.7	142	74	9.6
	$\hat{f}_{CM}$	39	19	2.5	103	53	7	1605	810	107	73	36	4.9	193	35	4.6	2881	1411	190
	$\hat{f}_{Prop.}$	10	5	0.7	15	8	1	68	34	5	13	6.2	0.8	17	6.2	0.8	52	27	3.6

الطبيعي ثنائي المتغيرات

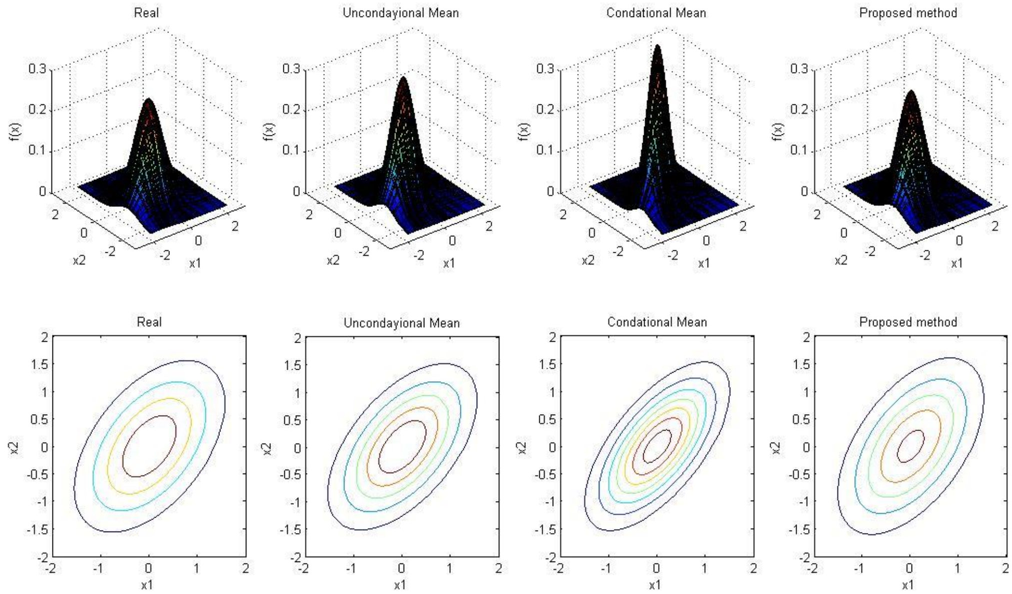
شكل رقم (1)

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



شكل رقم (2)

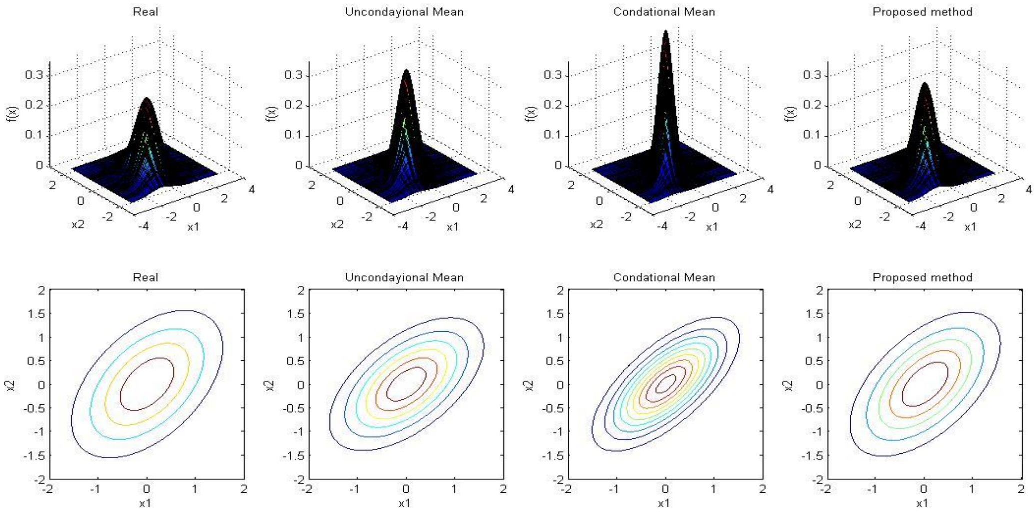
يشير الى القيم الحقيقية والتقديرات لدالة الكثافة الإحتمالية وحسب النمط الأول عند حجم عينة $n = 100$ وتباينات $\sigma_1^2 = 1 \;,\; \sigma_2^2 = 1$ وقيمة الارتباط $\rho = 0.5$ ونسبة فقدان 20%



شكل رقم (3)

يشير الى القيم الحقيقية والتقديرات لدالة الكثافة الإحتمالية وحسب النمط الأول عند حجم عينة $n = 100$ **وتباينات**

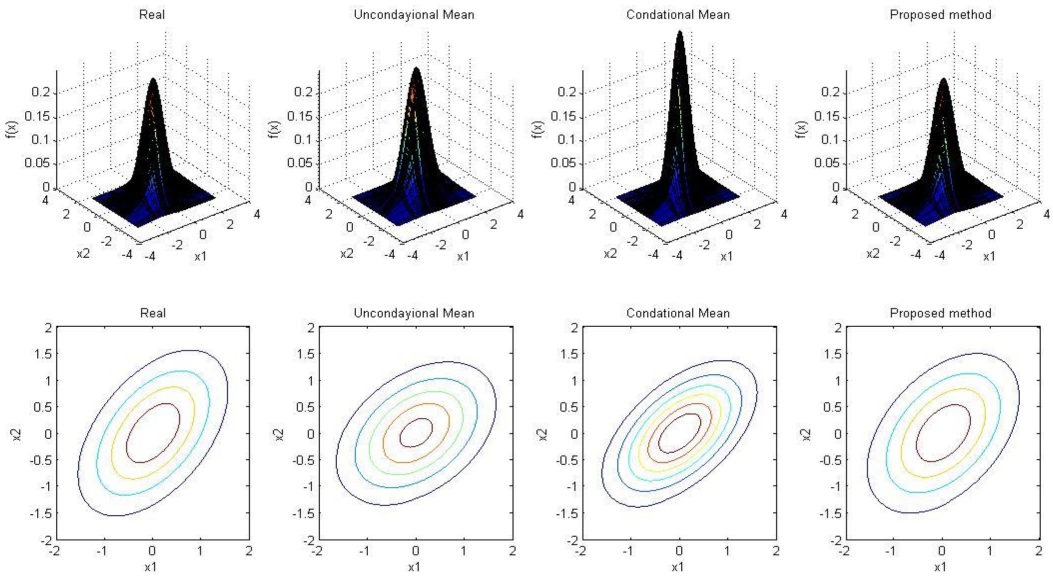
$\sigma_1^2 = 1 \;,\; \sigma_2^2 = 1$ **وقيمة الارتباط** $\rho = 0.5$ **ونسبة فقدان** 30%



شكل رقم (4)

يشير الى القيم الحقيقية والتقديرات لدالة الكثافة الإحتمالية وحسب النمط الأول عند حجم عينة $n = 100$ **وتباينات**

$\sigma_1^2 = 1 \;,\; \sigma_2^2 = 1$ **وقيمة الارتباط** $\rho = 0.5$ **ونسبة فقدان** 40%



جدول (2)

يبين قيمة متوسط مربعات الخطأ لتقدير دالة الكثافة للتوزيع الطبيعي ثنائي المتغيرات $MSE(\hat{f}(\underline{x}))$ وحسب النمط الثالث للقيم المفقودة وقيم التباينات وحجوم العينات المستخدمة و نسب الفقدان

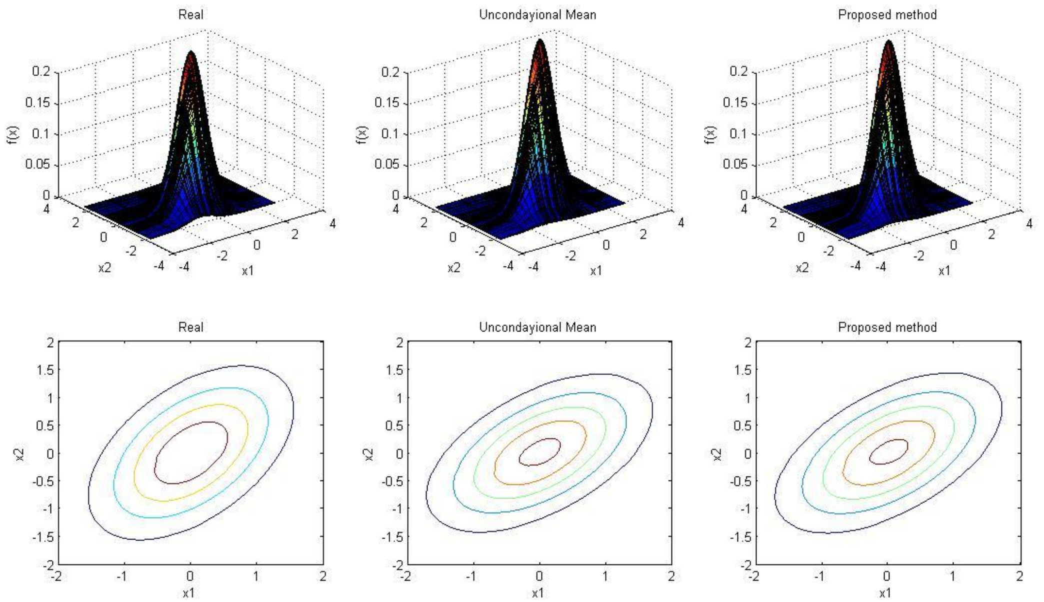
(العدد الناتج مضروب في 10000)

Missing		10%									20%								
n	ρ	0.3			0.5			0.8			0.3			0.5			0.8		
	σ_1^2	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5
	σ_2^2	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
25	$\hat{f}_{UCM}$	26	12	1.7	40	19	2.6	185	100	13	42	19	2.7	59	27	4.1	192	99	13
	$\hat{f}_{Prop.}$	34	19	5.2	43	25	8.9	110	59	30	36	20	5.8	34	26	8.2	103	53	19
50	$\hat{f}_{UCM}$	14	7.2	9.9	24	12	1.7	146	66	9.6	27	14	1.9	38	21	2.6	127	63	8.7
	$\hat{f}_{Prop.}$	28	15	4.6	34	20	7	87	46	23	30	17	4.7	35	21	6.3	83	41	14
100	$\hat{f}_{UCM}$	11	5.5	7.2	20	10	1.3	124	62	8	21	11	1.4	32	16	2.1	107	53	7.1
	$\hat{f}_{Prop.}$	26	14	4.2	31	18	6.4	76	39	20	27	15	4.2	32	19	5.6	73	36	11

Missing		30%									40%								
n	ρ	0.3			0.5			0.8			0.3			0.5			0.8		
	σ_1^2	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5	1	1	5
	σ_2^2	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
25	$\hat{f}_{UCM}$	67	34	4.7	76	40	5.4	168	86	13	77	39	5.3	96	47	6.6	165	89	11
	$\hat{f}_{Prop.}$	38	23	6.5	45	27	7.6	99	50	14	39	25	6.8	46	28	8.3	96	48	12
50	$\hat{f}_{UCM}$	41	21	2.8	55	27	3.6	121	57	7.6	53	27	3.5	65	32	4.2	112	56	7.7
	$\hat{f}_{Prop.}$	31	18	5.2	36	21	6.3	81	39	10	33	21	5.8	37	23	6.5	78	38	9.3
100	$\hat{f}_{UCM}$	34	17	2.3	43	21	2.9	97	45	6.4	41	21	2.7	49	25	3.3	88	44	6
	$\hat{f}_{Prop.}$	28	17	4.8	33	19	5.7	73	35	8.8	29	19	5.3	33	21	6	71	35	8

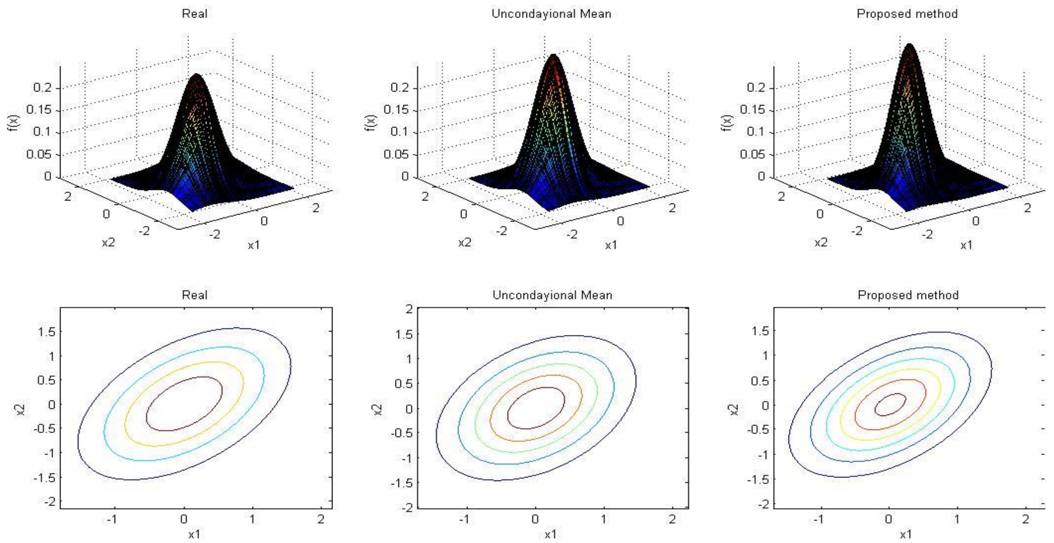
شكل رقم (5)

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



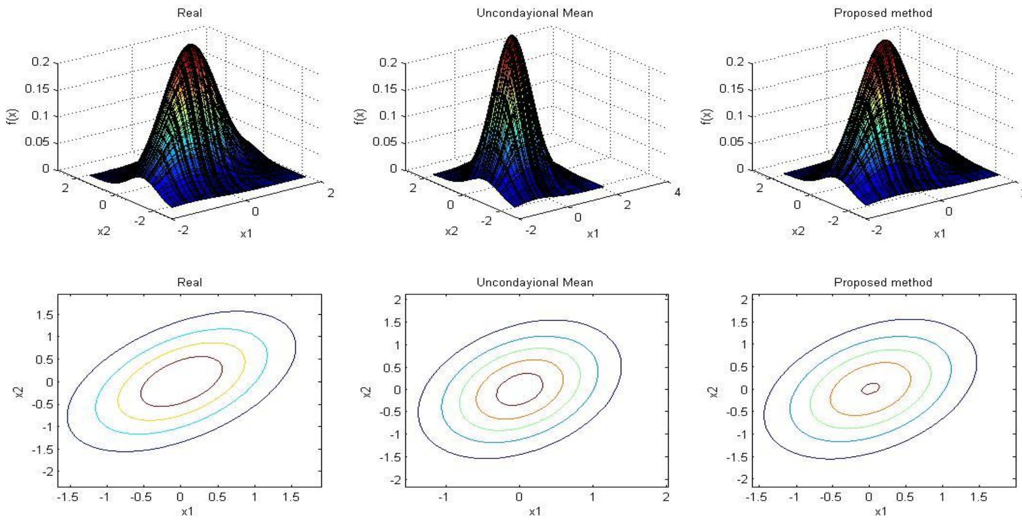
شكل رقم (6)

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



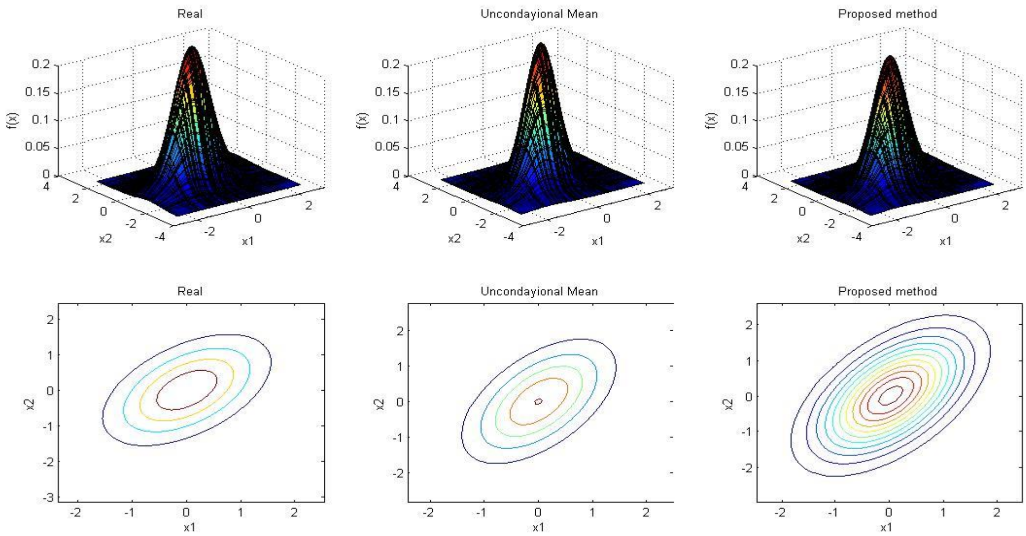
شكل رقم (7)

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



شكل رقم (8)

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



4. تفسير النتائج

من الجدول (1) نلاحظ النتائج الاتية: (النمط الأول للفقدان)

- **لجميع نسب الفقدان ولجميع حجوم العينات والتباينات وقيم الأرتباطات 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 الفقدان (40% , 30%)، إذ تمتاز طريقتي 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 تأثر وبشكل كبير لكل من طريقة التعويض بالمتوسط غير الشرطي UCM وطريقة التعويض بالمتوسط الشرطي CM ، وهذا ما يعاب على هاتين الطريقتين لأنه عند زيادة قيم الارتباطات تقل كفاءة هاتين الطريقتين.**
- **من الشكل 1 و 2 و 3 و 4 نلاحظ قلة التحيز للطريقة المقترحة في التعويض عن القيمة المفقودة ، اي أن البيانات التي تم تقديرها تكون قريبة جداً من البيانات الحقيقية.**
- **أما ما يخص طريقتي التعويض الأحادي (CM , UCM) كانت طريقة UCM أفضل من طريقة CM لجميع الحالات .**

من الجدول (2) نلاحظ النتائج الاتية: (النمط الثالث للفقدان)

- **عند نسبة فقدان 10% ولجميع التبايانات وقيمة الارتباط (0.5 , 0.3) وحجم عينة 25 كفاءة طريقة التعويض UCM على الطريقة المقترحة أما عند قيمة ارتباط 0.8 كفاءة الطريقة المقترحة عند قيمة تباين $\underline{\sigma}^2=\begin{pmatrix}1\\2\end{pmatrix}$, $\underline{\sigma}^2=\begin{pmatrix}1\\1\end{pmatrix}$ ولكن ، أما في حالة حجم عينة 50 و 100 نلاحظ كفاءة الطريقة UCM عند قيمة ارتباط 0.5 ولكن عند قيمة أرتباط 0.3 تكون الطريقة المقترحة أفضل عندما $\underline{\sigma}^2=\begin{pmatrix}5\\3\end{pmatrix}$.**
- **عند نسبة فقدان 20% وقيمة ارتباط (0.5 , 0.3) ولجميع حجوم العينات تكون قيم متوسطات الخطأ التربيعي متقاربة جداً للطريقتين ولكن الأمثلية تكون لطريقة UCM ، أما عند قيمة أرتباط 0.8 تكون طريقة التعويض عن القيمة المفقودة المقترحة هي الأفضل وخاصة عندما $\underline{\sigma}^2=\begin{pmatrix}5\\3\end{pmatrix}$.**
- **عند نسبة فقدان (40% , 30%) ولجميع حجوم العينات وقيم الارتباطات وعندما $\underline{\sigma}^2=\begin{pmatrix}1\\2\end{pmatrix}$, $\underline{\sigma}^2=\begin{pmatrix}1\\1\end{pmatrix}$ تكون الطريقة المقترحة في التعويض عن القيمة المفقودة هي الأفضل.**
- **من الشكل 5 و 6 و 7 و 8 نلاحظ قلة التحيز للطريقة المقترحة في التعويض عن القيمة المفقودة ، اي أن البيانات التي تم تقديرها تكون قريبة جداً من البيانات الحقيقية.**

5. الأستنتاجات

* في حالة النمط الأول من أنماط الفقدان أظهرت الطريقة المقترحة للتعويض عن القيمة المفقودة أفضلية في التقدير ولجميع نسب الفقدان وحجوم العينات وقيم التباينات وقيم الارتباطات المستعملة.
* في حالة النمط الثالث من انماط الفقدان أظهرت النتائج تفاوت في كفاءة الطريقة المقترحة بالرغم من ان نتائج متوسط مربعات الخطأ للطريقتين (المقترحة و UCM) كانت متقاربة جداً وبصورة عامة كانت الطريقة المقترحة هي الافضل عند زيادة نسب الفقدان ولجميع حجوم العينات وقيم الارتباط وعندما $\underline{\sigma}^2 = \begin{pmatrix}1\\2\end{pmatrix}$ ، $\underline{\sigma}^2 = \begin{pmatrix}1\\1\end{pmatrix}$.

6. المصادر:

1. مناف يوسف حمود، تهاني مهدي عباس، قتيبة نبيل نايف، 2008 " تقدير لا معلمي لدالة كثافة احتمالية متعددة متغيرات " مجلة جامعة النهرين – العلوم ، المجلد 11 ، العدد2 ، صفحة 55 – 63 .

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.

4. Kosobud, R. (1963) "*A note on Problem Caused by Assignment of Missing Data in Sample Surveys*" Econometric a, vol. 31, p. 562 – 563.

5. Little, R.J.A & Rubin, D.B (2003) "*Statistical Analysis with Missing Data*" 2nd 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