تقدير القيم المفقودة لمتغير الاستجابة في نموذج الانحدار الخطي المتعدد

Estimating of lost value of Responding variable in the Multi Regression Model

م.م. علي ناصر حسين
Assist. L. A. N. Hussein
قسم الإحصاء/ كلية الإدارة والاقتصاد
جامعة البصرة

المستخلص

تعد مشكلة فقدان البيانات من المشاكل التي تعيق التحليل الإحصائي، ان وجود هذه المشكلة في نماذج الانحدار المتعدد يسهم بان تكون نتائج هذه الوسيلة الإحصائية نتائج غير دقيقة كما تعتبر هذه الوسيلة في ظل وجود مشكلة البيانات غير التامة وسيلة غير كفؤه. ان وضع حل لهذه المشكلة يتطلب معرفة سبب الفقدان و الية الفقدان. و قد تم في هذه الدراسة اقتراح طريقة لمعالجة البيانات غير التامة معتمدا على طريقتين و هما طريقة EM Algorithm و طريقة التعويض بالانحدار و مقارنة نتائج هذه الطريقة المقترحة مع نتائج بعض الطرق معالجة البيانات المفقودة الأخرى .

تم استلام البحث بتاريخ٢٠١١/٧/١٣ تم قبول البحث بتاريخ٢٠١١/٨/٧

١- المقدمة

إن أهم المشاكل التي تواجه الباحث تتمثل في فقدان جزء من بيانات العينة ،ويعود هذا الفقـدان إلى أسباب عديدة مثل الحروب و الإهمال و التلف و كلفة او مخاطرة جمع البيانات أو أي سبب أخر وقد وضعت عدة آليات لمعالجة مشكلة البيانات غير التامة تختلف هذه الإلية باختلاف آليـة فقدان البيانات، إن فهم إلية الفقدان و تحديد طبيعتها يسهم في تحديد الطريقة المناسبة للتحليـل. غير إن اغلب الطرق الإحصائية التقليدية و البرامج الإحصائية وضـعت علـى افتـراض إن البيانات هي بيانات تامة و بالتالي فان استخدام هذه الطرق في تحليل البيانات غير التامـة يـتم بحذف الحالات التي تحتوي بيانات مفقودة و تعرف هذه الطرق بطريقة الحذف و هـي احـدى طرق معالجة مشكلة البيانات المفقودة. ان نماذج الانحدار المتعدد التي تستخدم لقيـاس تـأثير مجموعة من المتغيرات تسمى المتغيرات المستقلة على المتغير المعتمد وحسب المعادلة أدناه.

$$\underline{Y} = X\underline{B} + \epsilon \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

غير ان وجود مشكلة البيانات المفقودة في بيانات النماذج الانحدار المتعدد تؤدي الـى اعطـاء نتائج غير دقيقة في حالة و جود قيم مفقودة نتيجة حذف جزء من بيانات العينة.

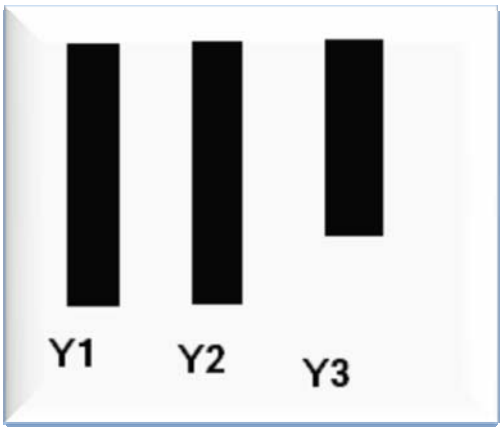
2-أنماط البيانات المفقودة

ان الطرق الإحصائية المستخدمة لمعالجة مشكلة فقدان البيانات تعتمد على نمط البيانات المفقودة و أدناه عرض لأهم أنماط البيانات غير التامة .

٢-١ Univariate missing data

في هذا النمط من البيانات غير التامة يكون فقدان البيانات في متغير واحد فقط و يعتبر هذا النمط من البيانات غير التامة من ابسط أنواع البيانات المفقودة فإذا وجد k من المتغيـرات فان عدد المتغيرات التي تحتوي على بيانات تامة هو k-1 بينما يوجد متغير واحد يحتـوي على حالات غير تامة ،و الشكل أدناه يوضح هذا النمط من البيانات غير التامة .

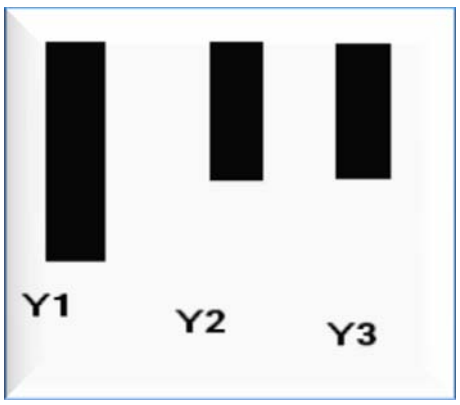
شكل (١) النمط الأول من البيانات غير التامة



Multivariate two patterns ٢-2

في هذا النمط من البيانات المفقودة يوجد اكثر من متغير يحتوي على بيانات مفقودة بحيـ ث ان عدد حالات البيانات المفقودة تكون متساوية في جميع المتغيرات التي توجد فيها بيانات مفقودة. و الشكل يوضح هذا النمط من البيانات المفقودة.

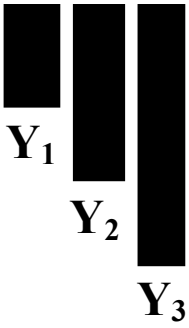
شكل (٢) النمط الثاني من البيانات غير التامة



Monotone ٣-٢

في هذا النمط من البيانات غير التامة ترتب المتغيرات بحيث يكون المتغير الذي يحتـوي على اكبر عدد من الحالات البيانات هو أول المتغيرات ثم يأتي المتغير الذي يحتوي العـدد الثاني اكبر عدد من الحالات المفقودة وهكذا بالنسبة لبقية المتغيرات . و من الممكن ترتيب المتغيرات بالاتجاه المعاكس.

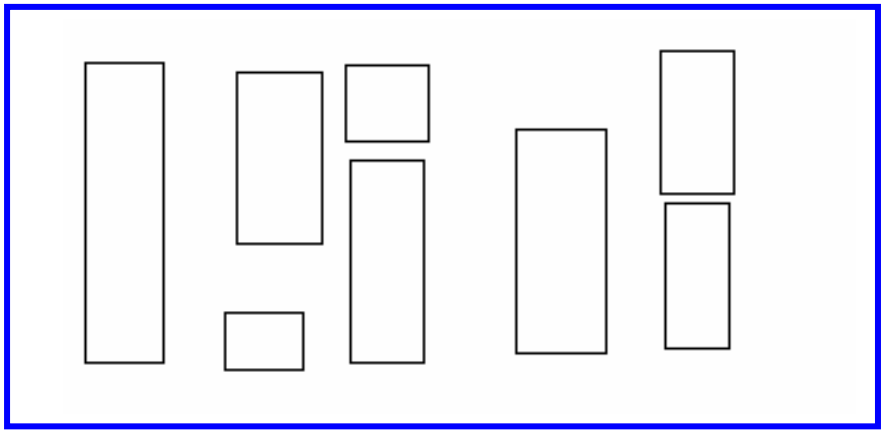
شكل رقم (٣) النمط الثالث من البيانات غير التامة



General ٤-٢

ان فقدان البيانات في هذا النمط هو فقدان عشوائي أي انه لا يوجـد نمـط معـين لفقـدان البيانات، و الشكل ادناه يوضح هذا النمط من البيانات المفقودة

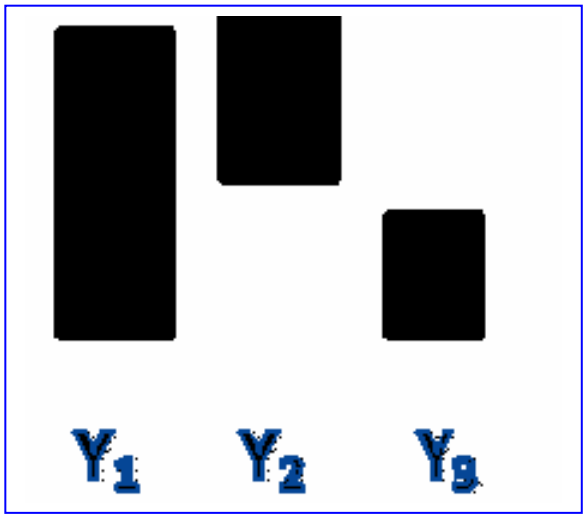
شكل رقم (٤) النمط الرابع من البيانات غير التامة



٢-٥ File matching

ان هذا النمط من البيانات المفقودة يظهر في حالات وجود متغيرين يحتويا على مشاهدات او بيانات مفقودة بحيث لا توجد بيانات مفقودة مشتركة بين المتغيرين

الشكل (٥) النمط الثالث من البيانات غير التامة



3-آليات الفقدان (1,3,4,6)

لتوضيح آليات فقد البيانات نفترض الاتي ان نموذج الانحدار المتعدد في المعادلة (١) و الـذي يحتوي على بيانات مفقودة حيث يفترض الاتي:

R: مصفوفة من درجة (n*p) و تدعى مؤشر البيانات المفقودة (Missing Data Indictor matrix) و تمثل القيم المشاهدة و القيم المفقودة لجميع المتغيرات

M تمثل مصفوفة لها نفس درجة المصفوفة R بحيث تكون عناصر المصفوفة m معرفة كالأتي

$$m_{ij} = \begin{cases} 1 & if\ r_{ij}\ is\ observed \\ o & if\ r_{ij}\ is\ missing \end{cases}$$

و بالاعتماد على المصفوفة R و المصفوفة M فان اليات الفقدان هي :

٣-١ آلية الفقدان بشكل عشوائي تام Missing Completely at Random (MCAR)

يقال ان البيانات فقدت بشكل عشوائي تماما اذا كانت احتمالية فقدان أي قيمة لا تعتمد علـى أي قيمة من القيم المشاهد و كذلك لا تعتمد على أي قيمة من القيمة المفقودة نفسها .اي ان

$$p_r(M_m/R,\emptyset) = p_r(M_m/\emptyset) \quad \forall M_m \quad \text{..... (2)}$$

حيث ان

M_m : مصفوفة القيم المفقودة .

ϕ معلمات مجهولة .

٣-٢ آلية الفقدان بشكل عشوائي Missing at Random (MAR)

يقال ان البيانات فقدت بشكل عشوائي اذا كانت احتمالية فقدان أي قيمة يعتمد فقط على القيم المشاهدة أي ان

$$p_r(M_m/R,\emptyset) = p_r(M_m/R_o,\emptyset) \quad \forall M_m \text{ ..... (3)}$$

حيث ان

٣-٣ آلية الفقدان بشكل غير عشوائي Missing not at Random (NMAR)

: MNAR

يقال ان البيانات فقدت بشكل غير عشوائي اذا كانت احتمالية فقدان اية قيمة يعتمـد علـى القيمة المفقودة فقط أي ان

$$p_r(M_m/M_oR,\emptyset) = p_r(M_m/R_m,\emptyset) \quad \forall M_m \text{.....(4)}$$

حيث ان : R_0 تمثل مؤشر مصفوفة القيم المفقودة

4-**طريقة المربعات الصغرى بالتعويض عن البيانات المفقودة**:

تعتبر هذه الطريقة من أكثر الطرق استخداما لتقدير القيمة المفقودة لما لها من مرونة في معالجة مشكلة البيانات غير التامة و تعتمد هذه الطريقة على أساس تعويض البيانات المفقودة ثم تحليـل البيانات كبيانات تامة .

توجد عدة من الآليات التي يمكن استخدامها لتعويض المفقودة لتعويض القـيم المفقـودة مثـل التعويض عن البيانات المفقودة باستخدام المتوسطات ،القيم المناسبة ، القيم او المعـدلات مـن بيانات مشابها او من بيانات سابقة.و من اهم طرق التعويض هي :

4-1- **طريقة المتوسط غير الشرطي**:

يتم تعويض بالمتوسط المشاهدات للمتغير الذي يحتوي بيانات مفقودة بدل من القيمة المفقودة و كالأتي

$$\hat{y}_i = \sum_{n_j} \frac{y_{obs}}{n_j} \quad \text{..... (5)}$$

n_j: تمثل عدد المتغيرات المشاهدة للمتغير الذي فقدت جزء من بياناته (y).

غير إن هذه الطريقة محدودة الاستخدام حيث أنها تعطي مقدرات غير متسقة حيث نلاحـظ إن مقدار التحيز في التباين في هذه الحالة هو

$$n_j - 1/n - 1 \quad \text{..... (6)}$$

بينما يكون مقدار التحيز في التباين المشترك بين أي متغيرين

$$n_{jk} - 1/n - 1 \quad \text{..... (7)}$$

التعويض بالمتوسط الشرطي :

يتم هنا استخدام المتوسط الشرطي المحسوب على أساس قيم المتغير (X) للتعويض عن القـيم المفقودة و كالاتي :

$$E(y_i/x_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \text{..... (8)}$$

$$E(y_i/x_{i2} \dots . x_{ip}) = \beta_0 + \beta_1 x^*_{i1} + \sum_{j=1}^{p} \beta_j x_{ij} \text{..... (9)}$$

Where $\beta_1 x^*_{1j} = E(x_{i1}/x_{i2} \dots \ x_{ip})$

4-2-**طريقة التعويض بالانحدار**:

في حالة معينة يمكن النظر إلى البيانات المفقودة على أنها معالم إضافية يتوجب تقديرها و على هذا الأساس يمكن تقدير البيانات في المتغير المعتمد و كالآتي:

فاذا فرضنا ان نموذج الخطي الانحدار هو

$$y = x\beta + \epsilon \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

حيث إن y متجة ذات رتبه (n*1) يمثل مشاهدات المتغير المعتمد،

x مصفوفة ذات رتبه (n*P) يمثل مشاهدات المتغير المستقل حيث توجـد p-1 مـن المتغيرات المستقلة

β متجه ذات رتبه (p*1) يمثل معالم الانحدار بما فيها المقطع الصادي،

ϵ الخطاء العشوائي حيث ان $\epsilon \sim N(\underline{0}, \sigma^2 I_n)$

فإذا كانت m تمثل عدد المشاهدات المفقودة من n من المشاهدات الكلية للمتغير المعتمد فان النموذج (10) يجزئ كالأتي

$$\begin{bmatrix} y_{obs} \\ y_{mis} \end{bmatrix} = \begin{bmatrix} x_{obs} \\ x_{mis} \end{bmatrix} \beta + \begin{bmatrix} \epsilon_{obs} \\ \epsilon_{mis} \end{bmatrix} \ldots\ldots\ldots\ldots\ldots\ldots (11)$$

حيث ان :

y_{obs} : متجه ذات رتبه (r=n-m,1) يمثل القيم ألمشاهده للمتغير المعتمد .

y_{mis}:متجه ذات رتبه (m,1) يمثل القيم المفقودة للمتغير المعتمد.

x_{obs}:مصفوفة ذات رتبه (r,p) تمثل قيم المتغيرات المستقلة المقابلة للقيم المشاهدة من المتغير المعتمد.

x_{mis}مصفوفة ذات رتبه (m,p) تمثل قيم المتغيرات المستقلة المقابلة للقيم المفقودة من المتغير المعتمد.

ϵ_{obs}:الخطاء العشوائي حيث ان $\epsilon_{obs} \sim N(\underline{0}, \sigma^2 I_r)$

ϵ_{mis}: الخطاء العشوائي حيث ان $\epsilon_{mis} \sim N(\underline{0}, \sigma^2 I_m)$

ان القيمة المفقودة للمتغير المعتمد يمكن تقديرها من خلال المعادلة

$$y_{mis} = x_{mis}\hat{\beta} \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

حيث ان

$$\hat{\beta} = (\acute{x}_{obs}x_{obs})^{-1}\acute{x}_{obs}y_{obs} \quad \text{..... (13)}$$

E M Algorithm -5

خوارزمية تعظيم التوقع يطلق عليها اختصارا (EM) وهي أسلوب لإيجاد تقدير الإمكان الأعظم لمعلمات الأنموذج وهي خوارزمية تكرارية لتقدير (Maximum Likelihood Ml) للبيانات غير التامة تمتاز هذه الخوارزمية بالبساطة و اتساع الحالات التي يمكن معالجتها ولتوضيح هذه الطريقة كالأتي:

إن هذه الخوارزمية تتألف من خطوتان لتقدير البيانات المفقودة الخطوة الأولى تسمى (E-step) و فيه يتم حساب المقدرات الكافية للمشاهدات التي يوجد فيها فقدان وكالاتي

$$E\left(\sum_{i=1}^{n} y_{ij} / Y_{obs}\theta^{t}\right) = \sum_{i=1}^{n} y_{ij} \quad \text{..... (14)}$$

حيث ان j=1,2,…k

θ^{t}هي تقدير اولى للمعلمة المجهولة θ

$$E\left(\sum_{i=1}^{n} y_{ij}y_{ik} / Y_{obs}, \theta^{t}\right) = \sum_{i=1}^{n}\left(y_{ij}^{t}y_{ik}^{t} + c^{t}{}_{jki}\right) \quad \text{.....} \quad (15)$$

Where I,j=1,2 …k , i≠k

$$y_{ij}^{t} = \begin{cases} y_{ij} & if\ y_{ij}\ is\ observed \\ E\left(y_{ij}/y_{obs}, \theta^{t}\right) & if\ y_{ij}\ is\ missing \end{cases} \quad \text{..... (16)}$$

$$c_{jki}^{t} = \begin{cases} 0 & if\ y_{ij} or\ y_{ik}\ is\ observed \\ cov\left(y_{ij}, y_{ik} \mathrm{I} y_{obs}, \theta^{t}\right) & if\ y_{ij}\ and\ y_{ik}\ are\ missing \end{cases} \quad \text{.....(17)}$$

اما في الخطوة M : في هذه الخطوة يتم استعمال تقديرات جديدة ل Ɵ و المتمثلة بمعلمات التوزيع الطبيعي المتعدد و هذه التقديرات هي θ^{t+1} و بالاعتماد على الخطوة E و كما يلي :

$$\hat{\mu}_{j} = \frac{\sum_{i=1}^{n} y_{ij}^{t}}{n} \quad \text{.....(18)}$$

$$\sigma_{ij}^{t+1} = \frac{1}{n}\sum_{i=1}^{n}\left[\left(x_{ij}^{t} - \mu_{j}^{t}\right)\left(x_{ik}^{t} - \mu_{k}^{t}\right) + c_{jki}^{t}\right] \quad \text{..... (19)}$$

6- **الطريقة المقترحة** :

تعتمد هذه الطريقة على مفهوم طريقة (E M Algorithm) حيث سيتم في الخطوة E step حساب المقدرات الكافية للمشاهدات التي يوجد فيها فقدان حيث تم اقتراح الاتي لحساب هذه المقدرات:

$$\hat{y}_{ij} = \left[1 - \lambda\sigma^{2}tr(w)/(\hat{\beta}'x'_{m}\Sigma^{-1}\chi_{m}\hat{\beta}\right]\chi_{m}\hat{\beta} \quad \text{..... (20)}$$

حيث ان :

$\lambda = (n-p)(m-2)/((n-p+2)tr(w))$.....(21)

$\Sigma = x_{mis}(x'_{obs}x_{obs})^{-1}x'_{mis}$..... (22)

$w = m'\Sigma^{-1}m$ ..... (23)

$m = x_{mis}(x'_{obs}x_{obs})^{-1}$..... (24)

الخطوة M يتم في هذه الخطوة استعمال مقدرات جديدة للمعلمة و كما تم توضيحه في الفقرة ويستمر تكرار الخطوات السابقة E و M إلى ان يتم الحصول على مقدرات مستقرة و حـسب قاعدة التوقف التالية:

$\left|\hat{\mu}^{t+1} - \hat{\mu}^{t}\right| < \delta \underline{I}$ .....(25)

$\left|\hat{\underline{\Sigma}}^{t+1} - \hat{\underline{\Sigma}}^{t}\right| < \delta \underline{J}$..... (26)

7-الجانب التطبيقي:

لغرض توضيح العلاقات الواردة في الجانب النظري من البحث فقد تم استعمال اسلوب المحاكاة للحصول على بيانات نموذج الانحدار المتعدد ذي متغيرات عشوائية تتبـع التوزيـع الطبيعـي المتعدد المتغيرات . حيث تم استخدم برنامج ال *matlab* لتوليد بيانات تتبع التوزيـع الطبيعـي بالمعالم (μ,σ^2) .ان خطوات الحصول على بيانات تتبع التوزيع الطبيعي المتعـدد المتغيـرات بمتوسط $\mu = (\mu_1,\mu_2,\dots\mu_3)'$ و مصفوفة تباين مشترك (Σ) يكون كالاتي

I.توليد $x_1,x_2,\dots x_d$.*as iid N(0,1)*

II. لكل $x_t = \mu_t + \sum_{j=1}^{i} c_{tj}z_j$

حيث ان *i=1,2 ...d*

ولان (Σ)هي مصفوفة متماثلة وموجبة يمكـن تجزئـة المـصفوفة باسـتعمال *(cholesky decomposition)* حيث ان مرتبة المصفوفة *c* هي *d*d* و كالاتي:

بسبب كون المصفوفة (∑) هي مصفوفة مربعة و موجبة نجد مصفوفة مثل المصفوفة (*c*)بحيث ان

$cc' = \Sigma$ ..... (27)

نحسب عناصر المصفوفة c من العلاقة الآتية:

$$c_{ii} = \left(\Sigma_{ii} - \sum_{k=1}^{i-1} c_{ik}^2\right)^{\frac{1}{2}} \quad \ldots\ldots (28)$$

$$c_{jt} = \frac{1}{c_{ii}}\left(\Sigma_{ij} - \sum_{k=1}^{i-1} c_{ik} c_{jk}\right) \quad \ldots\ldots (29)$$

بعد ان تم توليد بيانات تتبع التوزيع الطبيعي متعدد المتغيرات لبيانات المتغيرات التوضـيحية و الخطاء العشوائي يتم توليد قيم متغير الاستجابة من العلاقة الخطية الآتية :

$$\underline{y} = x\underline{\beta} + \underline{e} \quad \ldots\ldots (30)$$

8-**العينة**:

تم استعمال إحجام لعينات بحجمين هي كـالأتي (١٠٠، ٥٠) و تبـاين الخطـاء (*0.5*) و تـم الافتراض إن البيانات فقدت بالية من نوع (*MCAR*) و بنـسب فقـدان (*5%,10%,15%*) و بتكرار التجربة ١٠٠٠ بغية التعرف على الطرق أفضل في حالة الفقدان من النـوع *MCAR* حيث تم استخدام مقياس متوسط مربعات الخطاء للمقارنة بين طرق التقدير حيث يعتبـر هـذا المقياس من المقايس الجيدة للمفاضلة بين تقديرات الطرائق المستخدمة و قد اعتمـدت الـصيغة الاتية

$$\boldsymbol{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p} \quad \ldots\ldots (31)$$

9-**تحليل النتائج**:

تم تحليل نموذج انحدار خطي متعدد حسب الصيغة الاتية

$$y_i = 2 + 1.5x_{i1} + .5x_{i2} + .2x_{i3} + e_i$$

و الذي يعاني من فقدان في مشاهدات المتغير المعتمد y و تحت فرضـية فقـدان MCAR و للنمط الاول حيث كانت كالاتي

اولا : في حالة حجم العينة ٥٠ نلاحظ الاتي

١- في حالة الفقدان 5% ان طريقة EM المقترحة هي افضل الطرق ثم تأتي طريقة التقدير من خطا الانحدار و طريقة التعويض بالمتوسط غير الشرطي حيث بلغ MSE للطـرق الثلاثة و على الترتيب كالاتي (0.821)، (0.872)، (0.92).

٢- في حالة الفقدان ١٠% ان طريقة EM المقترحة هي افضل الطرق ثـم تـأتي طريقـة التقدير من خطا الانحدار و طريقة التعويض بالمتوسط غير الشرطي حيث بلـغ MSE للطرق الثلاثة و على الترتيب كالاتي (2.451)، (2.492)، (2.491)

٣- في حالة الفقدان ١٥% ان طريقة EM المقترحة هي افضل الطرق ثـم تـأتي طريقـة التقدير من خطا الانحدار و طريقة التعويض بالمتوسط غير الشرطي حيث بلـغ MSE للطرق الثلاثة و على الترتيب كالاتي (0.695)، (0.973)، (0.952)

جدول (١) نتائج تقدير البيانات المفقودة لطرائق التقدير

حجم العينة	نسبة الفقدان	طريقة التقدير		B0	B1	B2	B3	MSE
50	5%	طريقة EM المقترحة	قيمة المعلمة	0.123	1.032	0.584	0.316	0.821
			الانحراف المعياري	0,094	0.102	0.092	0.102	
		طريقة التعويض بالانحدار	قيمة المعلمة	0.127	1.05	0.591	1.314	0.872
			الانحراف المعياري	1	0.11	0.097	0.107	
		طريقة المتوسط غير الشرطي	قيمة المعلمة	0.15	0.957	0.549	0.317	0.92
			الانحراف المعياري	0.099	0.108	0.097	0.108	
	10%	طريقة EM المقترحة	قيمة المعلمة	0.2204	0.3846	0.0602	0.2337	2.4٥1
			الانحراف المعياري	0.1608	0.1384	0.1614	0.1605	
		طريقة التعويض بالانحدار	قيمة المعلمة	0.2462	0.4173	0.0610	0.2879	2.4٩2
			الانحراف المعياري	0.1596	0.1373	0.1601	0.1593	
		طريقة المتوسط غير الشرطي	قيمة المعلمة	0.2204	0.3846	0.0602	0.2337	2.491
			الانحراف المعياري	0.1608	0.1384	0.1614	0.1605	
	15%	طريقة EM المقترحة	قيمة المعلمة	0.248	0.982	0.503	0.1806	
			الانحراف المعياري	0.084	0.092	0.074	0.0819	0.695
		طريقة التعويض بالانحدار	قيمة المعلمة	0.1832	0.7658	0.4376	0.1589	
			الانحراف المعياري	0.099	0.109	0.088	0.0969	0.973
		طريقة المتوسط غير الشرطي	قيمة المعلمة	0.221	0.776	0.436	0.1566	
			الانحراف المعياري	0.098	0.108	0.087	0.0958	0.952

ثانيا : **في حالة حجم العينة** 100 **نلاحظ الاتي**

١- في حالة الفقدان ٥% ان طريقة EM المقترحة هي افضل الطرق ثم تأتي طريقة التقدير من خط الانحدار و طريقة التعويض بالمتوسط غير الشرطي حيث بلغ MSE للطـرق الثلاثة و على الترتيب كالاتي (0.967)، (0.995)، (1.026).

٢- في حالة الفقدان ١٠% ان طريقة EM المقترحة هي افضل الطرق ثـم تـأتي طريقـة التقدير من خطا الانحدار و طريقة التعويض بالمتوسط غير الشرطي حيث بلـغ MSE للطرق الثلاثة و على الترتيب كالاتي (0.158)، (0.168)، (1.042)

٣- في حالة الفقدان ١٥% ان طريقة EM المقترحة هي افضل الطرق ثـم تـأتي طريقـة التقدير من خطا الانحدار و طريقة التعويض بالمتوسط غير الشرطي حيث بلـغ MSE للطرق الثلاثة و على الترتيب كالاتي (0.942)، (0.961)، (1.576)

جدول (٢) نتائج تقدير البيانات المفقودة لطرق التقدير عند حجم عينة (١٠٠)

	نسبة الفقدان	طريقة التقدير	طريقة التقدير					
١٠٠	5%	طريقة EM المقترحة	قيمة المعلمة الانحراف المعياري	0.302	1.037	0.61	0.209	0.967
				0.082	0.083	0.081	0.077	
		طريقة التعويض بالانحدار	قيمة المعلمة الانحراف المعياري	0.309	1.007	0.601	0.189	0.995
				0.083	0.084	0.082	0.078	
		طريقة المتوسط غير الشرطي	قيمة المعلمة الانحراف المعياري	0.317	0.994	0.594	0.177	1.026
				0.084	0.086	0.083	0.079	
	10%	طريقة EM المقترحة	قيمة المعلمة الانحراف المعياري	0.6250	0.0730	0.3320	0.1070	0.158
				0.0220	0.0720	0.0750	0.0770	
		طريقة التعويض بالانحدار	قيمة المعلمة الانحراف المعياري	0.6360	0.0810	0.3410	0.1160	0.168
				0.0250	0.0780	0.0760	0.0775	
		طريقة المتوسط غير الشرطي	قيمة المعلمة الانحراف المعياري	0.6720	0.0830	0.3250	0.1240	1.042
				0.0290	0.0820	0.0861	0.0768	
	15%	طريقة EM المقترحة	قيمة المعلمة الانحراف المعياري	0.523	0.568	0.246	0.127	0.942
				0.622	0.006	0.047	0.0.757	
		طريقة التعويض بالانحدار	قيمة المعلمة الانحراف المعياري	0.628	0.582	0.285	0.088	0.961
				0.733	0.014	0.053	0.078	
		طريقة المتوسط غير الشرطي	قيمة المعلمة	0.736	0.812	0.254	0.092	1.576
				0.963	0.123	0.099	0.096	

10-**الاستنتاجات** :

من الجانب العملي للدراسة تم استنتاج الاتي:

١- ان الطريقة المقترحة تعطي مقدرات جيدة للقيم المفقودة بحيث تكـون MSE لنمـوذج الانحدار الذي قدرت قيمة المفقودة بهذه الطريقة اقل من بقية الطرق التي تـم المقارنـة معها.

٢- ان طريقة التعويض بالانحدار كانت افضل من طريقة المتوسط غير الشرطي.

٣- ان طريقة المتوسط غير الشرطي و رغم عدم أفضليتها الا انها ابسط في التطبيق من بقية الطرق التي تم مناقشتها في الدراسة

المصادر :

أولاً : **المصادر العربية**

١- حسين، انعام عبود ٢٠١٠ " تحليل البيانات غير التامة لنماذج الانحدار المتعدد باستخدام الخوارزميات ECME, ECM , EM مع تطبيق" رسالة ماجستير في الاحصاء، كلية الادارة و الاقتصاد، جامعة بغداد. غير منشور.

٢- عبد الاحد، باسم منصور " طريقة (James – stein) لتقدير المشاهدات المفقودة في نموذج الانحدار الخطية " مجلة الادارة و الاقتصاد، العدد ٥٤، ٢٠٠٥، ص ص ٢٦-٤٢.

٣- النعيمي، اسوان محمد طيب " معالجة البيانات غير التامة و تقديرها بطريقة المكونات الرئيسية " المجلة العراقية للعلوم الإحصائية، العدد ١٧ ، المجلد ١٠، ٢٠١٠ ص ص ٢٢٧- ٢٣٨.

ثانيا : **المصادر الانكليزية** .

3- C.Yang Yuan " Multiple Imputation for Missing Data : concepts and New Development http://support.sas.com/rnd/app/papers/multipleimputation.pdf

4- Dempster .A.p & N.M Laird "Maximum Likelihood from Incomplete Data Via the EM Algorithm . (1977) " Journal of the Royal Statistical Society Vol.39.No.1http://people.csail.mit.edu/jrennie/trg/papers/rubin-missing-76.pdf

5- D.paul .Allison (2001) "Missing Data " http://www.research.ibm.com/dar/papers/pdf/rc21783.pdf

6- E.Trivellore .Raghunathan & James M. Lepkowski "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models http://www.statcan.gc.ca/ads-annonces/12-001-x/5857-eng.pdf

7- J.Skyler.Cranmer (2005)" Methods Exam Review " http://jmlr.csail.mit.edu/papers/volume8/saar-tsechansky07a/saar-tsechansky07a.pdf

8- Law.M Averill & W. David Kelton (2000) " Simulation Modeling And Analysis " third edition ,New York.

9- Nittner " The Addivtive Model with missing Values in the independent variable – theory and simulation"; http://epub.ub.uni-muenchen.de/1653/1/paper_272.pdf

10- Roche .Alexis EM algorithm and variants : an information tutorial http://webcache.googleusercontent.com/search?hl=ar&lr=&spell=1&q=cache:0tIu1vBmI4gJ:ftp://ftp.cea.fr/pub/dsv/madic/publis/Roche_em.pdf+e+m+algorithm+and+variants&ct=clnk

11- Sundberg. Rolf (1974) " Maximum Likelihood Theory for Incomplete Data from an Exponential Family http://www2.math.su.se/~rolfs/Publikationer/SJS1974.pdf