أسلوب بيز في تحليل البيانات غير التامة

أ. د. اموري هادي كاظم   د. قتيبة نبيل نايف
جامعة بغداد/ كلية الادارة والاقتصاد
قسم الاحصاء

المستخلص:

في هذا البحث سوف يتم توضيح كيفية توظيف اسلوب بيز في تحليل أنموذج الأنحدار الخطي المتعدد الذي يعاني من فقدان في بعض مشاهدات متغيراتة التوضيحية $X^{'s}$ كأسلوب مقترح جديد، إذ سيتم توضيح بعض من انماط الفقدان وتحت شرط الية فقدان من نوع MCAR، وكيفة توظيف هذا الأسلوب في ظل وجود هذه الأنماط ومقارنة مقدارت بيز مع الطريقة الكلاسيكية المتبعة في التحليل والمتمثلة بطرية الحالة التامة ، إذ تم استخدام أسلوب المحاكات في المقارنة بين الطريقتين .

Abstract:

In this paper we will explain ,how use Bayesian procedure in analysis multiple linear regression model with missing data in variables X's as the new method suggest , and explain some of missing Patterns under missing mechanism , missing complete at random MCAR and compare Bayesian estimator with complete case estimator by use simulation procedure .

1) المقدمة:

لقد شهدت مشكلة البيانات المفقودة أهتماماً ملحوظاً في السنوات الأخيرة ولاسيما لنماذج الانحدار الخطي، ومع التطور السريع لأجهزة الحاسوب في معالجة العمليات أصبح تطوير طرائق تحليل البيانات المفقودة ممكن نظرياً وعلى الرغم من ذلك ما زال العديد منها بحاجة للتطوير ويعاني من مشاكل عديدة.من هنا تأتي أهمية تسليط الضوء على طرائق تقدير معالم أنموذج الانحدار الخطي المتعدد بهدف أيجاد أفضل الطرائق التي تلائم هذا الأنموذج في تقدير معالمة بوجود القيم المفقودة أخذين بنظر الاعتبار نسب الفقدان و آلية الفقدان ونمط الفقدان بالإضافة لذلك إذ توافرت معلومات أولية لمعالم أنموذج الانحدار الخطي المراد تقديرها، كيفية توظيف هذه المعلومات باستخدام أسلوب بيز في ظل كون مشاهدات العينة قيد الدراسة تعاني من فقدان، لذلك يهدف هذا البحث الى توظيف اسلوب بيز في تقدير معالم أنموذج الأنحدار الخطي الطبيعي المتعدد الذي يعاني من فقدان في بعض مشاهدات متغيراته التوضيحية علماً أن متغير الأستجابة يكون تام المشاهدة وأستخدام أسلوب المحاكات لمقارنة أسلوب بيز في التقدير مع طريقة الحالة التامة.

2) أنماط البيانات المفقودة :[1], [3]

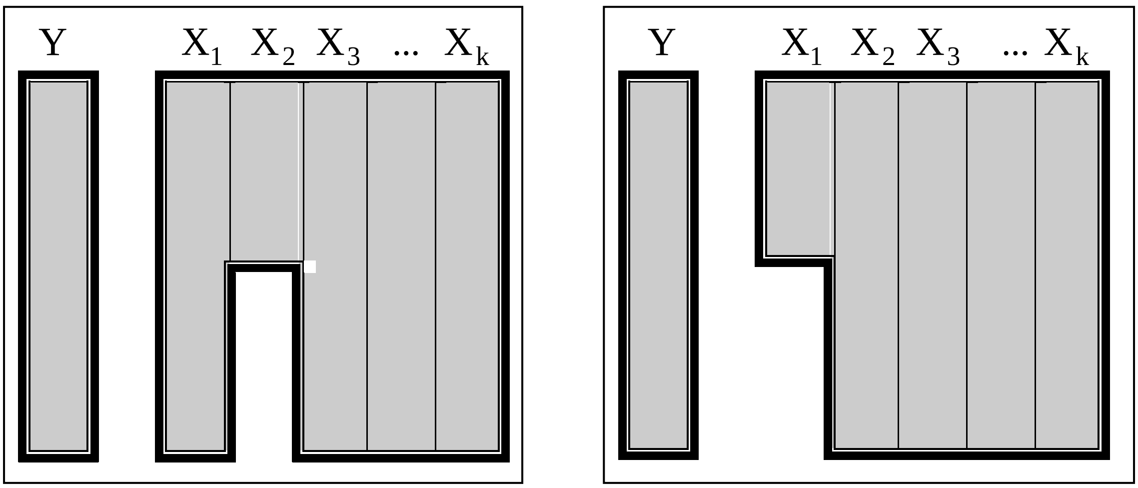
أن معرفة نمط والية الفقدان تساعد الباحث على تحديد الطريقة الإحصائية المناسبة لتقدير معالم الأنموذج، وبالأخص الأنماط، لأن هناك طرائق إحصائية تكون مناسبة لأنماط خاصة من البيانات غير التامة التي يمكن إن ترتب بشكل محدد، حيث تكون هذه الطرائق واضحة الخطوات وسهلة التطبيق. أما طرائق التحليل المناسبة للنمط العام لفقدان البيانات فتكون أكثر تعقيداً من طرائق الأنماط الخاصة، وهذا ما يدفع كثير من الباحثين إلى ترتيب بياناتهم بحسب نمط منتظم كلما أمكن ذلك تجنباً لاستخدام الطرائق المعقدة حسابياً. وعليه فأن أنماط البيانات المفقودة تقسم إلى قسمين الأولى منها يكون ضمن الأنماط الخاصة **Special Patterns** والثانية ضمن النمط العام **General Pattern**.

وفي أدناه عدد من أنماط البيانات غير التامة والمتمثلة بقيم المشاهدات لأنموذج الأنحدار الخطي:

$$Y \quad X_1\, X_2 \;\, X_3 \,\ldots\ldots. X_k$$

النمط الأول:

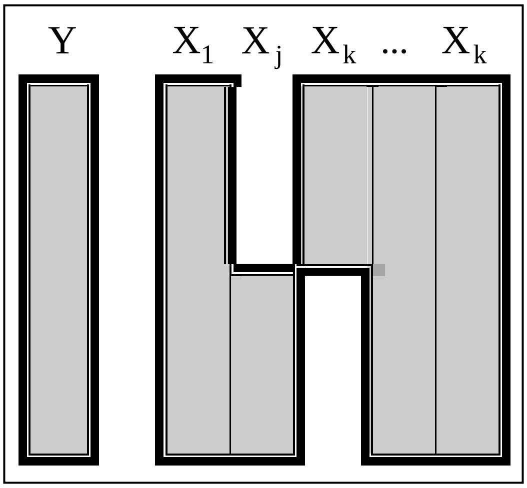
يعد هذا النوع من الأنماط الخاصة، وهو أبسط حالة من حالات البيانات غير التامة والتي يكون فيها جميع المتغيرات تامة المشاهدة عدا متغيراً واحداً يتضمن قيماً مفقودة في قسم من مشاهداته .



شكل رقم (1)
نمط فقدان البيانات لأحد المتغيرات

النمط الثاني: النمط المرتب أو المتداخل

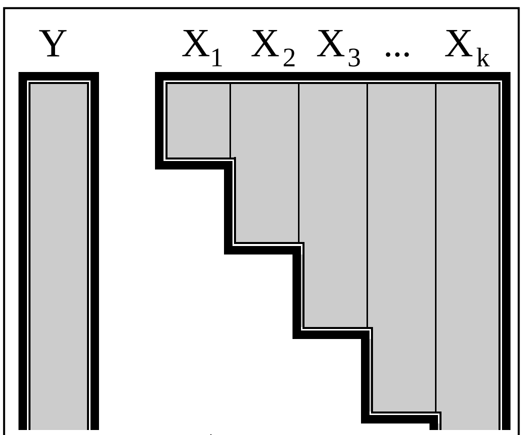
يعد هذا النمط من الأنماط الخاصة ايضاً، وفي هذا النمط ترتب البيانات حسب عدد القيم المفقودة ويمكن ترتيبها بشكل تصاعدي أو تنازلي، كذلك يتكون هذا النمط عادة عند اختيار عينة لحساب عدد من المتغيرات التوضيحية ومن ثم سحب عينة جزئية لحساب عدد من المتغيرات التوضيحية الأخرى والتي لم تسحب سابقاً كما في الشكل الأتي :



شكل رقم(2)
النمط المرتب أو المتداخل

النمــط الثالـث :

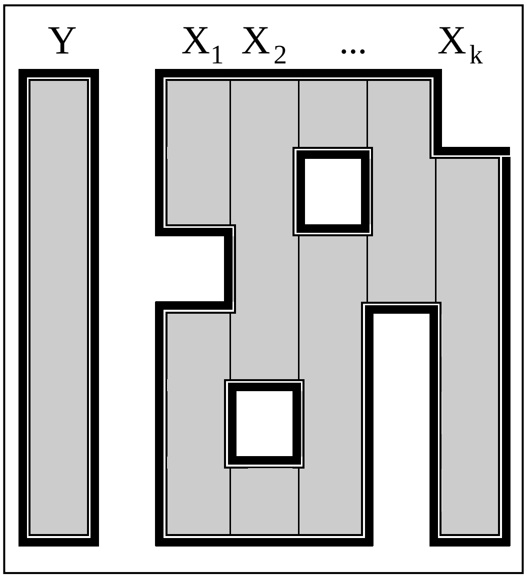
يعد هـذا النمط أخر الأنماط الخاصة، ويسمى بنمط البيانات المفقودة في حالة عدم تطابق المعالم، ويكون هذا النمط في حالة مشاهدات X_j و X_k غير مسجلة في مشاهدات واحدة، أي أن أي مشاهدة في المتغير X_j يقابلها مشاهدة مفقودة في المتغير X_k ، وتحدث هذه الحالة عند دمج أو توليف عينتين والشكل الأتي يبين هذا النمط من البيانات غير التامة.



شكل رقم(3)
نمط البيانات المفقودة في حالة عدم تطابق المعالم

النمـط الرابـع : النمـط العـام

هذا النمـط يوضـح فقدان البيانـات بشكل عشـوائي لأي قيمـه مـن قيم المتغيـرات قيـد الدراسـة والشكل التالي يبين هـذا النمط من البيانات غير التامة.



شكل رقم (4)
النمط العــام للبيانات المفقودة

3) اليات البيانات المفقودة :[1], [4] , [6]

تختلـف الطرائـق الإحصائيـة الخاصـة بتحليل البيانـات غيـر التامـة فـي فرضياتـها حـول الإليـة التـي تـؤدي إلـى فقـدان البيانـــات . وأن فهـم هـذه الآليـة وتحديـد طبيعتهـا يساعـد كثيـراً فـي اختيـار الطريقـة المناسبـة للتحليـل بـل يعد المـدخل لتشخيـص الطريقـة التـي تقتـرب نتائجها مـن الأمثليـة للبيانـات المدروسة. ويمكـن أن تلخـص علاقة فقـدان البيانـات لمتغيـر معيـن بقيـم المتغيـر نفسـه أو بقيـم المتغيرات الأخرى وكما يأتي :

1- أن فقـدان قيـم X_j يكـون مستقـلاً عـن قيـم المتغيـرات الأخـرى وعـن القيـم المفقـودة نفسهـا .

2- أن فقـدان قيـم X_j يعتمـد على القيمـة المفقودة نفسها .

3- اعتماد القيـم المفقـودة لـ X_j علـى قيـم المتغيـرات الأخـرى في العينـة .

وعليــة يمكن تقسيـم الاليــة كمـا يلـي :

1- فقدان البيانات تماماً بشكل عشوائي (Missing Complete At Random (MCAR

2- فقدان البيانات بشكل عشوائي (Missing At Random (MAR

3- فقدان البيانات بشكل غيـر عشوائي (Missing Not At Random (Not MAR

وبمـا أننـا سـوف نعتمـد علـى الأليه الأولـى سنوضـح كيفيـة حدوثهـا ، إذا كـان سبـب الفقـدان مستقلاً عن القيمـة المفقـودة نفسهـا وعن قيم المتغيرات الأخرى في العينـة عندهـا يمكـن القـول أن البيانـات تفقـد تمامـاً بشكـل عشوائـي (Missing Complete At Random (MCAR . ويمكـن

التعبير عن هذه الاليه رياضياً وذلك من خلال التوزيع الخاص بها والمقترحة من قبل Rubin(1976) [ ]والمتمثل بالتوزيع الشرطي لـ (X/R) وبمعالم مجهولة هي Ψ .

$$P\left(R/X\ ,\ \Psi\right)$$

حيث أن :

X : مصفوفة تمثل البيانات الحقيقية من مرتبة (n x p)

R : مصفوفة ثنائية تأخذ القيم (0 , 1) مناظرة للمصفوفة X

حيث أن:

$$r_{ij}=\begin{cases}1 & \text{if } X_{ij}\ \text{obs.}\\ 0 & \text{if } X_{ij}\ \text{miss.}\end{cases}$$

والمصفوفة **R** تدعى مصفوفة مؤشر البيانات المفقودة **Missing Data Indictor Matrix**

فإذا كان :

$$P\left(R/X,\Psi\right)=P\left(R/\Psi\right) \qquad \text{for all } X_{miss.} \qquad ...(1)$$

فأن البيانات تفقد تماماً بشكل عشوائي(MCAR) .

4) تحليل الحالة التامة :[1] , [3] , [4] Complete Case Analysis (CC)

تعد هذه الطريقة من الطرائق الأساسية التي تم استخدامها في تحليل البيانات التي تعاني من مشكلة الفقدان ويمكن القول أنها طريقة أساس في التحليل ،على فرض لدينا أنموذج الأنحدار الأتي :

$$y_I=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+...+\beta_k x_{ik}+\varepsilon_i \qquad ...(2)$$

أو حسب الصيغة العامة:

$$\underline{Y}=X\underline{\beta}+\underline{\varepsilon} \qquad ...(3)$$

حيث أن :-

$\underline{Y}$: متجه مشاهدات المتغير التابع (الاستجابة) n x 1.

X : مصفوفة التصميم بدرجة n x p وتمثل مشاهدات المتغيرات التوضيحية (المستقلة) فيما يخص نماذج الانحدار من رتبة p وأن p=k+1 ومحتوياتها كمية، وأن عدد هذه المتغيرات k حيث $x_j\ ,\ j=1,2,...,k$.

$\underline{\beta}$: متجه المعالم المجهولة ذات البعد p x 1 .

$\underline{\varepsilon}$: متجه الأخطاء العشوائية ذات البعد n x 1 .

فأن أسلوب هذه الطريقة وتحت شرط الية فقدان من نوع MCAR ، هو عند فقدان مشاهدات لمتغير ما ولتكن مثلاً المشاهدات $x_{i'1}$ نقوم بحذف جميع المشاهدات المقابلة للتسلسل i' ولجميع المتغيرات (المتغيرات ألتوضيحيه ومتغير الأستجابه) وبعد الانتهاء من عملية

الحذف يتكون أنموذج تام المشاهدات ولكن بعينه حجمها n_C (والرمز C هنا يشير إلى رمز التمام Complete) وكما يأتي :

$$y_j = \beta_0 + \beta_1 x_{i'1} + \beta_{i'} x_{i'2} + \dots + \beta_k x_{i'k} + \varepsilon_{i'} \qquad \dots (4)$$

إذ أن : $n_C < n \quad \text{and} \quad i' = 1, 2, \dots, n_C$

n هي عدد المشاهدات الأصلية في الأنموذج.

ومن ثم يتم استخدام طريقة المربعات الصغرى OLS في تقدير معالم الأنموذج الخطي العام التالي :

$$\underline{Y}_C = X_C \underline{\beta} + \underline{\varepsilon}_C \qquad \dots (5)$$

إذ إن :

$\underline{Y}_C$: هو متجه متغير الأستجابه للبيانات التامة ذو البعد $(n_C \times 1)$.

X_C : هي مصفوفة المشاهدات التامة للمتغيرات التوضيحية من الدرجة $(n_C \times p)$.

$\underline{\beta}$: هو متجه المعالم المطلوب تقديره ذو البعد $(p \times 1)$.

$\underline{\varepsilon}_C$: هو متجه الأخطاء العشوائية للبيانات التامة ذو البعد $(n_C \times 1)$.

علماً أن p هي عدد المعالم و k عدد المتغيرات التوضيحية في الأنموذج وأن $p = k + 1$

وباستخدام المصفوفات سوف يتم الحصول على:

$$\hat{\underline{\beta}}_{CC} = (X'_C X_C)^{-1} X'_C \underline{Y}_C \qquad \dots (6)$$

وبما أن آلية الفقدان هي MCAR فعلية فان تقديرات المعالم هي تقديرات غير متحيزة لتقديرات المربعات الصغرى الاعتيادية للبيانات التامة وأما التباين لهذه التقديرات فهو [1،3،66،91،59]:

$$\operatorname{var}(\hat{\underline{\beta}}_{CC}) = \hat{\sigma}^2_{CC} (X'_C X_C)^{-1} \qquad \dots (7)$$

$$\hat{\sigma}^2_{CC} = \frac{\underline{Y}'_C \underline{Y}_C - \hat{\underline{\beta}}'_{CC} X'_C \underline{Y}_C}{n_C - p}$$

وبصورة عامه يمكن أيجاز بعض النقاط المهمة والخصائص لهذه الطريقة :[1،3،66]

1- أن هذه الطريقة تستخدم في حالة وجود آلية فقدان بيانات من نوع MCAR لأنه بعد عملية الحذف سوف تتكون عينه ذات حجم n_C التي سوف تكون متحيزة عن العينة الأصلية ذات الحجم n .

2- ولاستخدام هذه الطريقة يجب أن تكون نسبة المشاهدات المحذوفة لأجمالي المشاهدات قليلة نسبياً أي لا تتجاوز نسبة الفقدان 40% عندما $(n > 50)$ ولا يتجاوز 20% عندما $n < 50$.

3- لا يمكن استخدام هذه الطريقة في حالة النمط الثالث لأنة سوف يتم حذف جميع المشاهدات.

5) أسلوب بيز في تحليل الحالة التامة: [1] , [2] , [5]

Bayes Procedure in Complete – Case Analysis (BCC)

كما وضحنا في الفقرة 4، أن طريقة الحالة التامة تستند في أساس عملها على حذف المشاهدات التي تقابل قيم مفقودة مما يؤدي إلى الحصول على عينة جزئية من العينة الأصلية قيد الدراسة، لذلك عندما يتم توظيف أسلوب بيز في التقدير وبخاصة في حالة استخدام دالة كثافة احتمالية مرفقة طبيعية والتي تم استنتاجها من تجارب سابقة، سوف نحتاج إلى عينة جزئية من هذه التجارب السابقة تتوافق مع العينة قيد الدراسة.

لذلك استندت الفكرة هنا على استخراج عينة جزئية من العينة المسبقة وذلك عن طريق حذف مشاهدات العينة المسبقة التي تقابلها مشاهدات مفقودة في العينة قيد الدراسة مما يؤدي إلى الحصول على عينة جزئية من العينة المسبقة تتوافق مع العينة الجزئية قيد الدراسة وكما يأتي:

بحسب الصيغة (2) فإن دالة الكثافة الاحتمالية المرافقة الطبيعية لـ $\underline{\beta}_{CC} / \sigma^2_{CC}$ بعد حذف المشاهدات للعينة المسبقة التي تقابل قيم مفقودة لنفس صف العينة الأصلية تكون بحسب الصيغة الآتية :

$$\pi\left(\underline{\beta}_{CC} / \sigma^2_{CC}\right) \propto \frac{1}{\left(\sigma^2_{CC}\right)^{\frac{n_{C(0)}}{2}}} \exp\left\{-\frac{1}{2\sigma^2_{CC}}\left(\underline{\beta}_{CC} - \hat{\underline{\beta}}_{CC(0)}\right)' Q_C \left(\underline{\beta}_{CC} - \hat{\underline{\beta}}_{CC(0)}\right)\right\} \ldots(8)$$

أما دالة الكثافة الاحتمالية المرفقة الطبيعية الحدية للمعلمة σ^2 تكون بالشكل التالي :

$$\pi\left(\sigma^2_{CC}\right) \propto \left(\sigma^2_{CC}\right)^{-\left(\frac{v_{C(0)}}{2}+1\right)} \exp\left\{-\frac{v_{C(0)}\sigma^2_{CC(0)}}{2\sigma^2_{CC}}\right\} \qquad \ldots(9)$$

وعلية فإن الصيغة (8) هي توزيع طبيعي متعدد المتغيرات لـ $\underline{\beta}_{CC}$ والصيغة (9) هي توزيع معكوس مربع كآي المقيس **Scaled inverse chi-square** للمعلمة σ^2_{CC}.

وبدمج الصيغة (8) و الصيغة (9) يتم الحصول على دالة مرافقة طبيعية مشتركة مسبقة للمعالم $\left(\underline{\beta}_{CC}, \sigma^2_{CC}\right)$ وكما يلي :

$$\pi\left(\underline{\beta}_{CC}, \sigma^2_{CC}\right) \propto \sigma^{-1}_{CC}\left(\sigma^2_{CC}\right)^{-\left(\frac{v_0}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2_{CC}}\left[v_{C(0)}\sigma^2_{CC(0)} + \left(\underline{\beta}_{CC} - \hat{\underline{\beta}}_{CC(0)}\right)' Q_C \left(\underline{\beta}_{CC} - \hat{\underline{\beta}}_{CC(0)}\right)\right]\right\} \ldots(10)$$

وعلية فأن الصيغة (10) ما هي إلا دالة الاحتمالية المسبقة المشتركة متعدد متغيرات طبيعي - معكوس مربع كآي **Multivariate Normal - Inverted – chi-square** أي أن :

$$\left(\underline{\beta}_{CC}, \sigma^2_{CC}\right) \sim \mathrm{MVN} - \mathrm{Inv} - \mathrm{Gamma}\left(\hat{\underline{\beta}}_{CC(0)}, Q_C ; \sigma^2_{CC(0)} / n_{C(0)}, v_{C(0)}\right)$$

أما دالة الكثافة الاحتمالية للصيغة (2) بعد حذف المشاهدات هي:

$$\left(\underline{Y}_C / X_C \underline{\beta}_{CC}, \sigma^2_{CC}\right) \sim MVN\left(X'_C \hat{\underline{\beta}}_{CC}, \sigma^2_{CC}(X'_C X_C)\right)$$

وعلية فان دالة الكثافة الاحتمالية ألاحقــة للمعالم $\left(\underline{\beta}_{CC}, \sigma^2_{CC}\right)$ هي :

$$\pi\left(\underline{\beta}_{CC}, \sigma^2_{CC} / \underline{Y}_C\right) \sim MVN - Inv - Gamma\left(\tilde{\underline{\beta}}_{CC(n)}, Q_{C(n)}; \tilde{\sigma}^2_{CC(n)}, v_{C(n)}\right) \quad ...(11)$$

وعلية فأن:

$$\left(\underline{\beta}_{CC} / \underline{Y}_C\right) \sim t_p\left(\tilde{\underline{\beta}}_{CC(n)}, Q_{C(n)}; \tilde{\sigma}^2_{CC(n)}, v_{C(n)}\right) \quad ...(12)$$

$$\left(\sigma^2_{CC} / \underline{Y}_C\right) \sim Inv - Gamma\left(v_{C(n)}, \tilde{\sigma}^2_{CC(n)}\right) \quad ...(13)$$

وعلية فأن مُقدر بيـز يكون :

$$\tilde{\underline{\beta}}_{BCC} = \tilde{\underline{\beta}}_{CC(n)} = \left(X'_C X_C + Q_C^{-1}\right)^{-1}\left(X'_C \underline{Y}_C + Q_C^{-1} \hat{\underline{\beta}}_{CC(0)}\right) \quad ...(14)$$

وان $\hat{\underline{\beta}}_{CC(0)}$ تمثل تقدير المربعات الصغرى LS التي تحسب من المعلومات الأولية .

أما مصفوفة التباين والتباين المشتركة لـ $\tilde{\underline{\beta}}_{BCC}$ هي :

$$var - cov\left(\tilde{\underline{\beta}}_{BCC}\right) = \frac{v_{C(n)}}{v_{C(n)} - 2} \tilde{\sigma}^2_{BCC} \left(X'_C X_C\right)^{-1} \quad ...(15)$$

وأن $\dfrac{\boldsymbol{v}_{C(n)}}{\boldsymbol{v}_{C(n)} - 2} \tilde{\boldsymbol{\sigma}}^2_{BCC}$ ما هي إلا الوسط الحسابي للصيغة (15) وتحسب كما يلي:

$$\tilde{\sigma}_{BCC} = s^2_{CC(n)} = \left(n_{C(0)} s^2_{CC(0)} + (n_C - 1) s^2_{CC} + \left(\hat{\underline{\beta}}_{CC} - \hat{\underline{\beta}}_{CC(0)}\right)' Q_{C(n)} \left(\hat{\underline{\beta}}_{CC} - \hat{\underline{\beta}}_{CC(0)}\right)\right) / v_{C(n)} \quad ...(16)$$

وأن :

$$n_{C(n)} = n_{C(0)} + n_C$$

$$Q_{C(n)} = \left(X'_C X_C + Q_C\right)^{-1}$$

$$v_{C(n)} = n_{C(n)} - p$$

أما $\underline{\hat{\beta}}_{CC}$ و $\underline{\hat{\beta}}_{CC(0)}$ يتم حسابيهما من الصيغة (6) من مشاهدات العينة و المعلومات الأولية وعلى التوالي وكذلك s^2_{CC} و $s^2_{CC(0)}$ تحسبا أيضاً من الصيغة (7) من مشاهدات العينة والمعلومات الأولية وعلى التوالي .

6) المحاكاة:

لغرض معرفة كفاءة المقدرات في تحليل البيانات غير التامة وبيان تأثير نسب الفقدان، وتغير التباين للأخطاء وكذلك حجوم العينات، سوف يتم الأنموذج الأفتراضي التالي:

$$y_i = 3.39 - 0.601x_{1i} + 0.05x_{2i} + 0.25x_{3i} + e_i \qquad ...(17)$$

وأن اقيم ألفتراضية لمتجة المعالم $\underline{\beta}$ تم أفتراضها بشكل ينسجم مع طبيعة الظاهرة المدروسة وذلك بالاعتماد على الخلفية النظرية للظاهرة .

أما حالة الفقدان وحسب إلية الفقدان MCAR فسيتم توليدها حسب الصيغة التالية [1]:

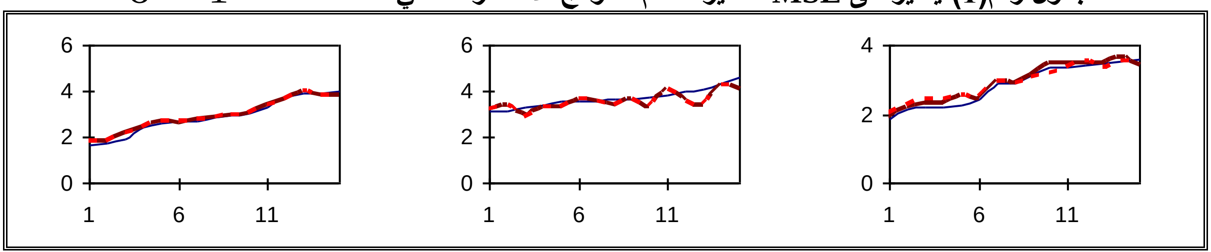
MCAR: $$p(x) = p(\delta = 1 / X = x) = 0.9 \quad , \forall x$$

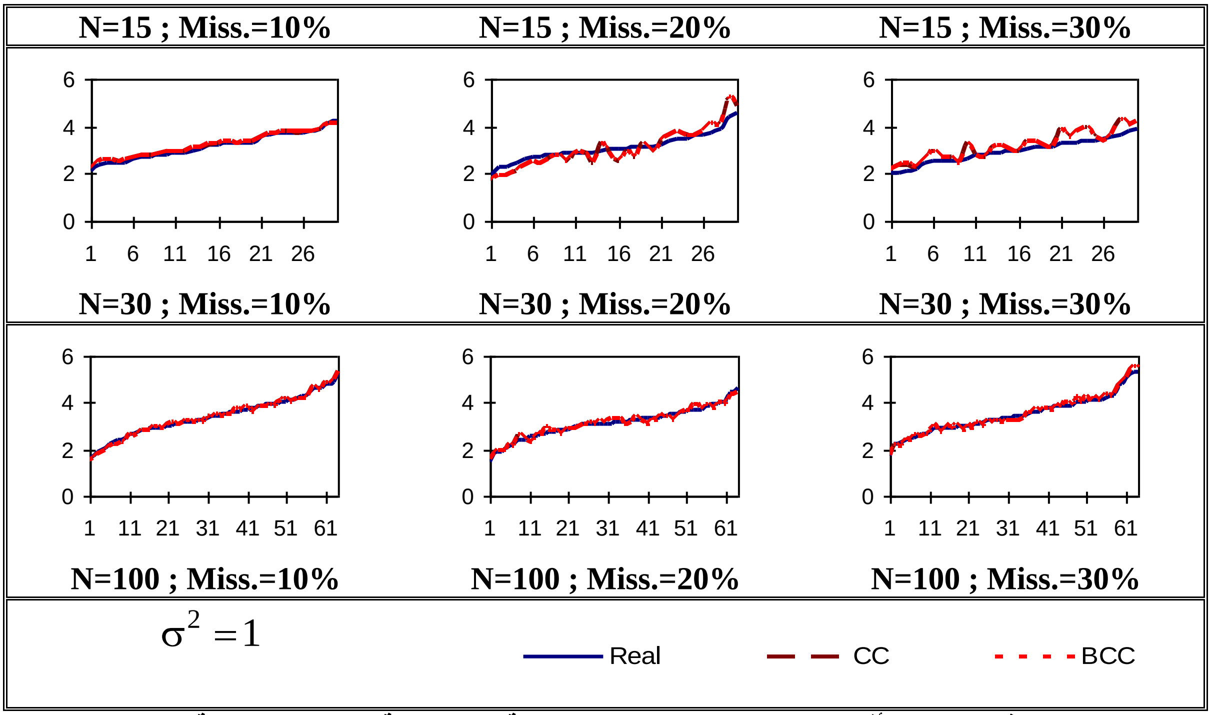
وتم افتراض قيم مختلفة لتباين الخطأ، وهي (2 ,1.5 ,1) .إما حجوم العينات المستخدمة فكانت (15, 30, 100) ، في حين كانت نسب فقدان (10%, 20%, 30%).

والجداول الأتية توضح قيم MSE لمقدرات دالة الانحدار المستعملة، التباينات، حجوم العينات ونسب الفقدان .

N	Methods / Parameters / Missing	CC β_0	CC β_1	CC β_2	CC β_3	BCC β_0	BCC β_1	BCC β_2	BCC β_3
15	10%	0.1087	0.1232	0.1262	0.1304	0.0976	0.1108	0.1138	0.1185
	20%	0.1215	0.1425	0.1465	0.1452	0.1077	0.1267	0.1314	0.1298
	30%	0.1590	0.1984	0.1905	0.1965	0.1223	0.1527	0.1450	0.1492
30	10%	0.0418	0.0449	0.0450	0.0451	0.0340	0.0365	0.0367	0.0367
	20%	0.0479	0.0518	0.0514	0.0524	0.0460	0.0497	0.0495	0.0503
	30%	0.0571	0.0617	0.0637	0.0637	0.0560	0.0602	0.0627	0.0623
100	10%	0.0116	0.0118	0.0119	0.0119	0.0123	0.0126	0.0126	0.0127
	20%	0.0131	0.0135	0.0134	0.0133	0.0132	0.0136	0.0135	0.0135
	30%	0.0150	0.0154	0.0154	0.0155	0.0132	0.0136	0.0136	0.0137

جدول رقم(1) يشير الى MSE لتقدير معالم أنموذج الانحدار الخطي المتعدد عندما $\sigma^2 = 1$



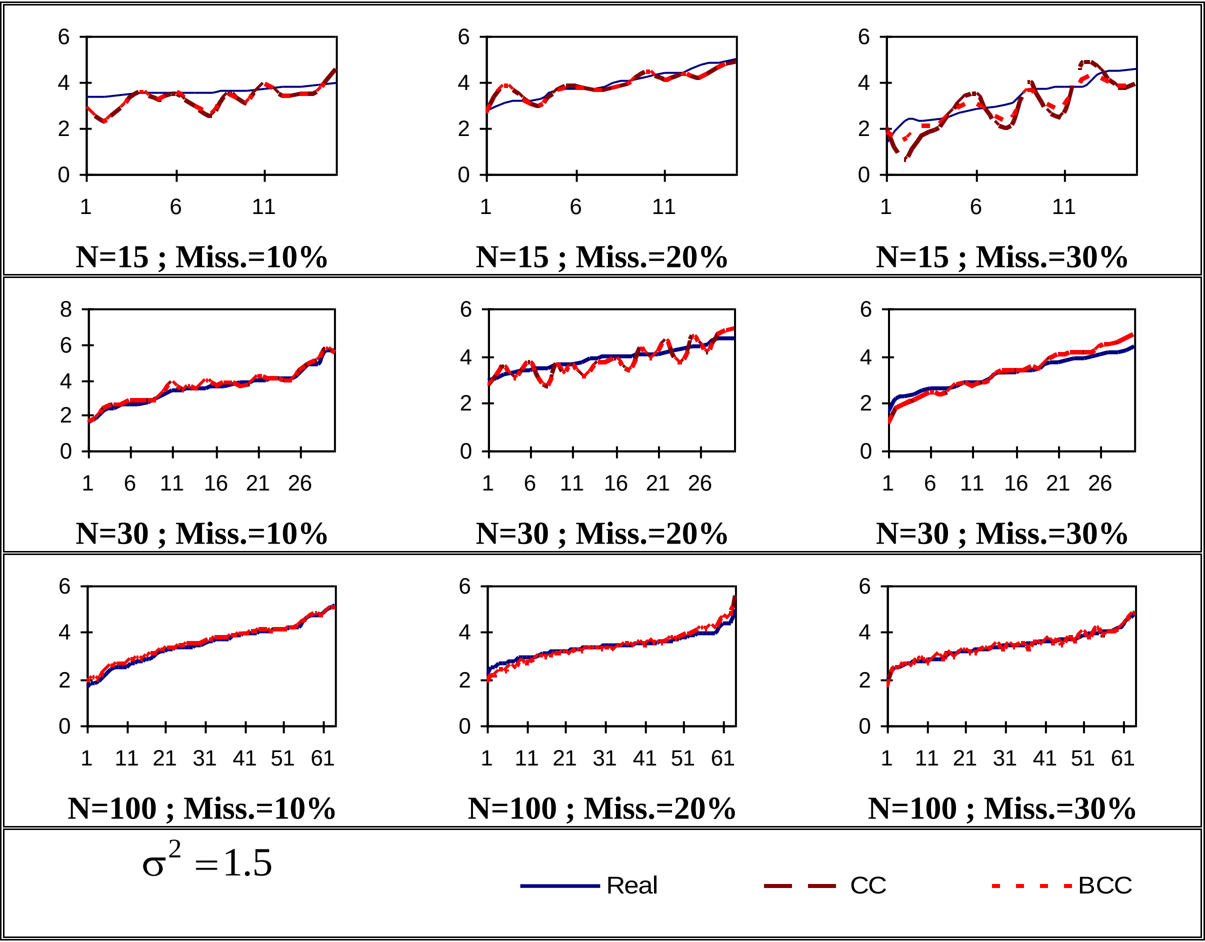


شكل رقم(5) يبين تأثير نسب الفقدان لجميع حجوم العينة على معادلة الانحدار التقديرية عندما

$\sigma^2 = 1$

Methods		CC				BCC			
N	Parameters / Missing	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
15	10%	0.2383	0.2810	0.2807	0.2846	0.1727	0.2036	0.2030	0.2054
	20%	0.2580	0.3154	0.3043	0.3151	0.1797	0.2195	0.2117	0.2182
	30%	0.3683	0.4695	0.4583	0.4513	0.2633	0.3432	0.3464	0.3236
30	10%	0.0957	0.1048	0.1030	0.1042	0.0871	0.0950	0.0936	0.0947
	20%	0.1082	0.1188	0.1180	0.1184	0.1070	0.1172	0.1167	0.1168
	30%	0.1279	0.1411	0.1389	0.1424	0.1259	0.1387	0.1370	0.1398
100	10%	0.0259	0.0265	0.0263	0.0265	0.0252	0.0258	0.0256	0.0258
	20%	0.0294	0.0302	0.0302	0.0302	0.0319	0.0327	0.0328	0.0328
	30%	0.0338	0.0350	0.0348	0.0348	0.0291	0.0300	0.0298	0.0299

جدول رقم(2) يشير الى MSE لتقدير معالم أنموذج الانحدار الخطي المتعدد عندما $\sigma^2 = 1.5$

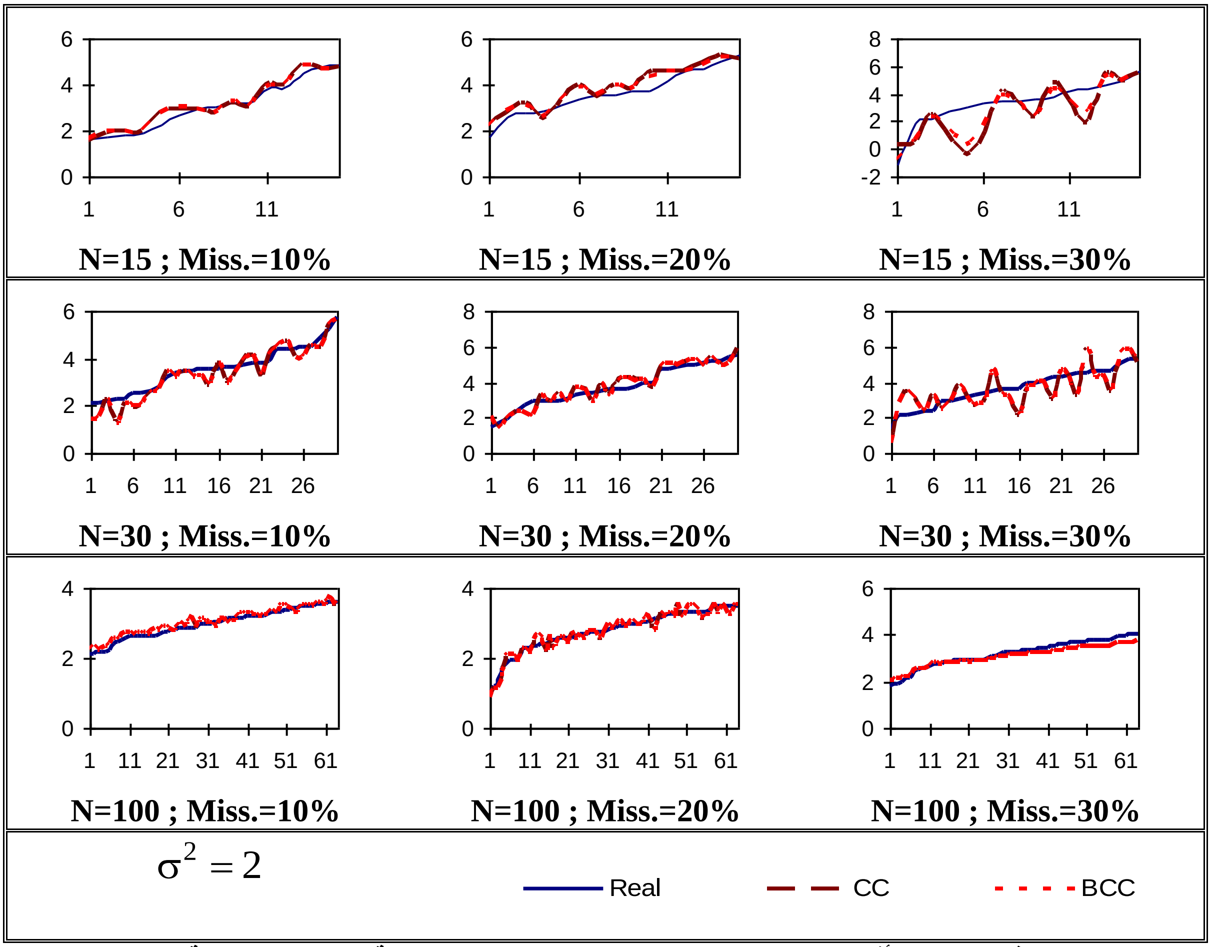


شكل رقم(6) يبين تأثير نسب الفقدان لجميع حجوم العينات على معادلة الانحدار التقديرية عندما

$\sigma^2 = 1.5$

Methods		CC				BCC			
N	Parameters / Missing	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
15	10%	0.4245	0.4948	0.4973	0.5105	0.3916	0.4546	0.4546	0.4689
	20%	0.4700	0.5754	0.5814	0.5969	0.3840	0.4718	0.4728	0.4860
	30%	0.6407	0.7891	0.7860	0.8213	0.5485	0.6895	0.6711	0.7015
30	10%	0.1714	0.1889	0.1833	0.1797	0.1576	0.1738	0.1683	0.1657
	20%	0.1920	0.2143	0.2103	0.2107	0.1603	0.1789	0.1763	0.1755
	30%	0.2256	0.2519	0.2521	0.2524	0.2147	0.2384	0.2401	0.2404
100	10%	0.0457	0.0466	0.0466	0.0468	0.0446	0.0455	0.0454	0.0457
	20%	0.0518	0.0532	0.0539	0.0533	0.0473	0.0485	0.0492	0.0487
	30%	0.0595	0.0611	0.0614	0.0614	0.0510	0.0523	0.0526	0.0526

جدول رقم(3) يشير الى MSE لتقدير معالم أنموذج الانحدار الخطي المتعدد عندما $\sigma^2 = 2$



شكل رقم(7) يبين تأثير نسب الفقدان لجميع حجوم العينات على معادلة الانحدار التقديرية عندما

$$\sigma^2 = 2$$

7) تفسير النتائج:

يتضح من النتائج في الجداول (1) ، (2) ، (3) المذكورة أنفاً الأتي :

- عند أحجام العينة الصغيرة نلاحظ أن اسلوب بيز والمتمثل بالطريقة المقترحة BCC كأن أفضل من طريقة CC في تقدير معالم الأنموذج .
- نلاحظ تناقص MSE عند زيادة حجوم العينة ولكنه يزداد عند زيادة نسب الفقدان خاصه عند استخدام طريقة CC .
- كذلك نلاحظ تزايد MSE عند زيادة التباينات إذ تكون الزيادة كبيره عند استخدام طريقة CC .
- نلاحظ تناقص كفاءة كلا الطريقتين في تقدير معالم الأنموذج وبشكل ملحوظ عند زيادة نسب الفقدان وخاصه عند احجام العينة الصغيرة ، إذ كان MSE لطريقة CC اكثر تأثراً بنسب الفقدان عند أحجام العينة الصغيرة من طريقة BCC .

8) الأستنتاجات:

نلاحظ من النتائج المذكورة في المبحث السابق أفضلية أسلوب بيز في تحليل البيانات المفقودة والمتمثل بالطريقة المقترحة BCC ولجميع الحالات وخاصه عند احجام العينة الصغيرة ويرجع ذلك الى توضيف المعلومات الأولية في التقدير ، بصورة اخرى ، أن المعلومات الأولية زادة من وفرة المعلومات حول المعلمة المراد تقديرها ، لذلك يوصى بأستخدام مقدر BCC في تحليل البيانات المفقودة تحت شرط الية فقدان MCAR وعند توفر معلومات أوليه حول الظاهرة المراد دراستها وخاصه كون هذه المعلومات تتوافق ومعلومات العينه قيد الدراسة .

8) المصادر:

[1] القزاز، قتيبه نبيل نايف ،(2007) "مقارنة أساليب بيز الحصين مع طرائق أخرى لتقدير معالم أنموذج الانحدار الخطي المتعدد في حالة البيانات غير التامة" أطروحة دكتوراه فلسفة في الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.

[2] Carlin , J.B ; Gelman , A. ; Rubin , D.B & Stern , H.S (2004) "Bayesian Data Analysis" 2nd ed. , Chapman & Hall, New York.

[3] Little, R.J.A. (1992) "Regression with Missing X's: A Review" JASA, vol. 87, p. 1227 – 1237.

[4] Little, R.J.A & Rubin, D.B (2003) "Statistical Analysis with Missing Data" 2nd ed., John Wile & Sons, New York.

[5] Rowe, D.R. (2003) "Multivariate Bayesian statistics" John Wile & Sons, New York.

[6] Schafer, J.L. (1997) "Analysis of Incomplete Multivariate Data" Chapman & Hall New York.