طريقة مقترحة لتقدير القيم المفقودة في نموذج الانحدار الخطي المتعدد

A New Method for Estimating Missing Data in the Multiple Linear Regression Model

د. أشرف أحمد العتيبي *

* كلية التربية – جامعة الملك خالد – المملكة العربية السعودية

البريد الإلكتروني: batal9383@hotmail.com

ISSN: 2410-7727

ISSN: 2410-7727

طريقة مقترحة لتقدير القيم المفقودة في نموذج الانحدار الخطي المتعدد

الملخص

هدفت الدراسة الحالية إلى تقديم طريقة جديدة (متوسط الميول Slopes Mean) لمعالجة البيانات المفقودة في نموذج الانحدار الخطي المتعدد ومقارنتها مع طريقة التعويض بالمتوسط الحسابي وطريقة التعويض بالانحدار وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML). تعتمد الطريقة المقترحة على حساب الميل المطلق بين كل نقطتين معلومتين في نموذج الانحدار ومن ثم أخذ المتوسط الحسابي للميول والتعويض به بدلاً من القيم المفقودة ومقارنة نتائج هذه الطريقة المقترحة مع نتائج الطرق الأخرى لمعالجة البيانات المفقودة باستخدام المعايير(MSE, R^2, R^2 , VIF) مستخدماً أسلوب المحاكاة في المقارنة بين الطرق . وبينت النتائج ان الطريقة المقترحة (SM) تعد هي ثاني أفضل الطرق بعد طريقة EM وذلك لأنها تحتوي على MSE أقل من طريقة التعويض بالمتوسط، وطريقة التعويض بالانحدار، وطريقة المعلومات الكاملة. كما أنها تحتوي على معامل تحديد ومعامل تحديد معدل أكبر من هذه الطرق. وأشارت النتائج أيضا إلى أن الطريقة المقترحة أفضل من طريقة EM من حيث معيار VIF حيث أنها تحتوي على VIF أقل.

الكلمات المفتاحية: الانحدار المتعدد، البيانات المفقودة، أنماط البيانات المفقودة، آليات فقد البيانات. التعويض بالانحدار، التعويض بالمتوسط، المعلومات الكاملة للاحتمال الأعظم، تعظيم التوقع.

Summary

the current study aimed to introduce a new method (Slopes Mean) to handle the missing data in the multiple linear regression model and compare it with the Mean Imputation method, Regression Imputation, the (EM) method, and the (FIML) method. The proposed method is depend on the calculation of the absolute slope between each of the two points in the regression model and then taking the arithmetic mean and compensating it instead of the missing values and comparing the results of this proposed method with the results of other methods that used to handle the missing data using the criteria (MSE, R^2, Adjusted R^2, VIF). Simulation method have been done to compare these methods. The results indicated that the proposed method (SM) is the second best way after the EM method because it contains less MSE than the Mean Imputation method, the regression method, and FIML method. It also contains a coefficient of determination and a coefficient of adjusted determination the larger than these methods. The results also indicated that the proposed method

ISSN: 2410-7727

is better than the EM method in terms of the VIF where it contains less VIF.

Key words: Multiple Regression; Missing Data; Pattern of Missing Data; Mechanisms of Missing Data; Regression Imputation; Mean Imputation; EM; FIML.

مقدمة:

تعد مشكلة البيانات المفقودة أحد أبرز المشكلات التي تواجه الباحثين أثناء تحليلهم لبيانات بحوثهم مما يعيق عملية التحليل الإحصائي ويؤثر على جودة البيانات بشكل عام. وهو ما أشار إليه (Allison, 2001) بقوله "عاجلاً أم آجلاً: عادةً عاجلاً: كل من يعمل التحليل الإحصائي سيصطدم بمشكلة البيانات المفقودة ". ويظهر هذا الفقد في البيانات على شكل أنماط مختلفة وبسبب آليات مختلفة للفقد أيضاً. وقد حاول الإحصائيون تطوير وإيجاد أفضل الطرق في تقدير القيم المفقودة في شتى بيانات البحث العلمي, وكل طريقة من هذه الطرق لها شروطها وافتراضاتها الخاصة بها بحسب نمط الفقد وآليته (Mcknight, Sidani, Figueredo & Mcknight, 2007).

وعلى الرغم من التطور الكبير الذي تشهده مشكلة البيانات المفقودة وذلك فيما يتعلق بتطوير طرق تقديرها ومعالجتها, إلا أن الطرق التقليدية التي تعتمد على طريقة الحذف لاتزال مستخدمة وذلك لسهولة تنفيذها, ولأن الطرق الجديدة تحتاج للكثير من الوقت حتى يتم استيعابها وتطبيقها بدل الطرق القديمة (Enders, 2010).

الأمر الذي حذرت منه الجمعية الأمريكية للعلوم النفسية والتربوية وطالبت باستخدام طرق إحصائية متطورة تأخذ في الاعتبار الظروف التي يتم في ظلها فقد البيانات (Peng, Harwell, Liou & Ehman, 2006) . ويشير Tabachnick & Fidell (2007) إلى أن وجود البيانات المفقودة عند استخدام أساليب التحليل الإحصائي المتعددة المتغيرات يعد أكثر تأثيراً وخطورة على نتائج التحليل. لا سيّما لنماذج الانحدار المتعدد, حيث تؤثر البيانات المفقودة على كفاءة نموذج الانحدار وقدرته التنبؤية مما يترتب على ذلك عدم دقة ومصداقية النتائج التي يتوصل إليها الباحث. وقد أشارت الكثير من البحوث إلى مشكلة البيانات المفقودة وطرق معالجتها وبيان أنماط وآليات الفقد, حيث ركزت هذه البحوث بشكل أكبر على المقارنة بين طرق المعالجة المختلفة وتقديم طرق مقترحة لتقدير القيم المفقودة, مما دفع الباحث إلى تقديم طريقة مقترحة لمعالجة البيانات المفقودة ومقارنتها ببعض الطرق التقليدية والحديثة المتاحة لمعرفة فاعلية الطريقة المقترحة مقارنة ببقية الطرق المستخدمة في الدراسة.

ISSN: 2410-7727

1. **مشكلة الدراسة :**

إن وجود مشكلة البيانات المفقودة في بيانات نماذج الانحدار المتعدد بشكل أنماط مختلفة للفقد وبنسب فقد مختلفة ينشأ عنها استنتاجات خاطئة وتقديرات متحيزة وذلك نتيجة حذف جزء من بيانات العينة. وعلى الرغم من إسهام الرواد الأوائل في تطوير وإيجاد طرق لمعالجة البيانات المفقودة وكل طريقة لها افتراضاتها في الاستعمال وبحسب نمط الفقد وآليته, إلا أن الباحثين يجدون صعوبة في تحديد الطريقة المثلى للتعامل مع البيانات المفقودة فجميع طرق التعامل معها تتطلب إقحام نوع من الافتراضات الإضافية وفي الوقت نفسه هناك اتفاق عام حول نقاط قوة وضعف كل منها. ومن هنا أتت الحاجة إلى مثل هذه الدراسة والتي تعتمد على تقديم طريقة مقترحة لمعالجة البيانات المفقودة في نموذج الانحدار المتعدد ومقارنتها بنتائج الطرق الأخرى المستخدمة في هذه الدراسة. وعلى ذلك يمكن صياغة السؤال الرئيسي للدراسة كالتالي:

ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار (REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML))؟

ليتفرع من هذا السؤال أربعة أسئلة فرعية وهي:

- ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار (REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML) في ظل وجود معيار MSE للمقارنة بين الطرق؟
- ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار (REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML) في ظل وجود معيار R-Square للمقارنة بين الطرق؟
- ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار (REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML) في ظل وجود معيار Adjusted R-Square للمقارنة بين الطرق؟
- ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار (REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML) في ظل وجود معيار VIF للمقارنة بين الطرق؟

ISSN: 2410-7727

2. **أهداف الدراسة :**

هدفت الدراسة إلى معرفة كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة في نموذج الانحدار المتعدد, من خلال استقصاء كفاءة نماذج الانحدار المقدّرة بطرق المعالجة المختلفة (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار(REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)) ومقارنتها بالطريقة المقترحة (SM).

3. **أهمية الدراسة :**

تستمد الدراسة أهميتها النظرية من خلال تناولها موضوع البيانات المفقودة وطرق معالجتها وبيان أنماط وآليات فقد البيانات كذلك مساهمتها إلى جانب الدراسات العديدة في هذا المجال في تقديم طريقة مقترحة لمعالجة البيانات المفقودة . أما الجانب الثاني فهو الأهمية التطبيقية التي تتمثل في تزويد الباحثين بطريقة إحصائية لمعالجة البيانات المفقودة يأمل الباحث أن تساهم في إعطاء تقديرات أكثر دقة للبيانات المفقودة يمكن أن يعتمد عليها كي تحل محل البيانات المفقودة.

4. **الجانب النظري:**

4.1. **مفهوم البيانات المفقودة:**

هي البيانات التي تُركت دون إجابة بسبب الفشل في الحصول على بيانات من أجزاء من المجتمع المستهدف أو العينة (**عبد الرحمن**, 2013). كما يعرف ليتيل وروبن (Little & Rubin, 2002) البيانات المفقودة بأنها هي البيانات غير التامة والتي تعيق التحليل الإحصائي.

4.2. **أنماط البيانات المفقودة:**

إن معرفة نمط وآلية البيانات المفقودة يساعد الباحث على اختيار المعالجة الإحصائية المناسبة لتقدير معالم النموذج, خصوصاً أنّ هنالك معالجات إحصائية تكون مناسبة لأنماط خاصة من البيانات المفقودة وتكون واضحة الخطوات وسهلة التطبيق بينما هنالك معالجات إحصائية مناسبة للنمط العام وتكون أكثر تعقيداً من معالجات الأنماط الخاصة (Little, 1992). يرى إندرز (Enders, 2010) بأن نمط البيانات المفقودة يشير إلى استعراض القيم الملحوظة والمفقودة ضمن مجموعة البيانات. بمعنى أن نمط البيانات المفقودة يصف لنا ببساطة موقع تلك الفجوات في البيانات ولا يشرح لنا لماذا فقدت هذه البيانات. وعليه فإن البيانات المفقودة تأخذ شكل الأنماط التالية:

- **النمط الأول: النمط أحادي المتغير** (Univar ate pattern): في هذا النمط من البيانات المفقودة يكون فقد البيانات في متغير مستقل واحد, وباقي المتغيرات تكون تامة (Little & Rubin, 2002).
- **النمط الثاني: نمط البيانات المفقودة متعددة المتغيرات** (Multivariate two patterns): في هذا النمط من البيانات المفقودة يكون فقد البيانات في أكثر من متغير مستقل, بحيث تتساوى عدد حالات الفقد في جميع المتغيرات المستقلة التي تحتوي على بيانات مفقودة (Little & Rubin, 2002).

ISSN: 2410-7727

- **النمط الثالث: النمط ذو الوتيرة الواحدة** (Monotone pattern): في هذا النمط من البيانات المفقودة يكون فقد البيانات بشكل مرتب لبعض المتغيرات المستقلة, بحيث يكون المتغير الذي به أكبر عدد من حالات الفقد هو أول المتغيرات ثم يأتي المتغير الذي يحتوي الترتيب الثاني لأكبر عدد من حالات الفقد وهكذا بالنسبة لبقية المتغيرات المستقلة (Schafer, 1997).
- **النمط الرابع: النمط العام** (General pattern): في هذا النمط من البيانات المفقودة لا يأخذ فقد البيانات شكلاً معيناً بل تكون البيانات المفقودة بشكل عشوائي ومبعثرة. ويعد النمط العام هو النمط السائد من بين أنماط البيانات المفقودة الأخرى (Graham, 2012).
- **النمط الخامس: نمط تطابق الملف** (File matching): هذا النمط من البيانات المفقودة يحدث في متغيرين مستقلين فقط, بحيث لا توجد بيانات مفقودة مشتركة بين المتغيرين (Molenberghs, Fitzmaurice, Kenward, Tsiatisand & Verbeke, 2015).

4.3. آليات البيانات المفقودة:

يرى إندرز (Enders, 2010) بأن آلية البيانات المفقودة تشير إلى تلك العلاقة المحتملة بين المتغيرات المقاسة واحتمالية فقد البيانات. وعلى الرغم من أن آليات البيانات المفقودة لا توفر لنا تفسيراً سببياً لفقد البيانات إلا إنها تقدم لنا علاقة رياضية عامة بين البيانات وعملية فقدها. إن فهم هذه الآليات وتحديد طبيعتها يساعد في اختيار الطريقة المناسبة التي سيتم بها علاج البيانات المفقودة. وعليه فإن آليات فقد البيانات هي (العتيبي, 2018):

- **آلية الفقد بشكل عشوائي تام** MCAR (Missing Completely At Random): يقال أن البيانات فقدت بشكل عشوائي تام إذا كانت البيانات المفقودة لا تعتمد على قيم المتغير المستقل نفسها الذي يحتوي على هذه القيم المفقودة أو أي متغير مستقل آخر في قاعدة البيانات.
- **آلية الفقد بشكل عشوائي** MAR (Missing At Random): يقال أن البيانات فقدت بشكل عشوائي إذا كانت البيانات المفقودة لا تعتمد على قيم المتغير المستقل نفسها الذي يحتوي على هذه القيم المفقودة, ولكنها تعتمد على متغير (أو متغيرات) مستقلة أخرى.
- **آلية الفقد بشكل غير عشوائي** NMAR (Not Missing At Random): يقال أن البيانات فقدت بشكل غير عشوائي إذا كانت البيانات المفقودة في متغير مستقل معين تعتمد على قيم المتغير نفسها.

4.4. طرق معالجة البيانات المفقودة:

أولاً : طرق الحذف (Deletion):

1. **حذف الحالة** بالكامل ((Leastwise Deletion (LD): وتعتمد هذه الطريقة على حذف الحالة التي لديها بيانات مفقودة في متغير واحد على الأقل, وتستخدم فقط الحالات ذات البيانات الكاملة على جميع المتغيرات, ولذلك هي تسمى أيضاً تحليل الحالة الكاملة (Complete Case Analysis)

ISSN: 2410-7727

(دودين, 2013). وتصلح هذه الطريقة في حالة البيانات المفقودة القليلة وعندما يكون حجم العينة كبيراً مما يسمح بحذف تلك الحالات, كذلك تستخدم في حالة البيانات المفقودة عشوائياً بالكامل حيث تؤدي عدم العشوائية إلى تحيز النتائج (رزق الله, 2002). ومن أحد مشاكل هذه الطريقة هي أنها ربما تنتج تقديرات قياسية متحيزة أي ربما يحدث تحيزاً في نتائج هذه الطريقة لأن الحالة بأثرها لم تمثل كامل المجتمع (Nakai & Ke, 2011).

2. **الحذف بالأزواج** ((Pairwise Deletion (PD): في هذه الطريقة يتم حذف المفردة التي فيها بيانات مفقودة من المتغير المستخدم في التحليل فقط أما في بقية المتغيرات فلا تحذف هذه المفردة, هذا وتعد بديلاً أقل تطرفاً من طريقة حذف الحالة بالكامل وذلك كونها تحافظ نوعاً ما على البيانات المتجمعة ولم تحذفها على الإطلاق, ويعيب على هذه الطريقة أن العينات تتباين من تحليل إلى آخر نتيجة حذف بعض المفردات مرة وعدم حذفها مرات أخرى (هويت, كرمر, 2016).

ثانياً : طرق التعويض (Imputation Methods):

1. **التعويض بالمتوسط الحسابي**(Mean Imputation): في هذه الطريقة يتم التعويض عن القيمة المفقودة في متغير ما بمتوسط البيانات المتاحة في ذلك المتغير, وهنا يقوم الباحث بحساب المتوسط للمتغير الذي يحوي بيانات مفقودة مستخدماً جميع البيانات المتوفرة لديه ثم يضع قيمة المتوسط الناتجة في الخلايا الخالية لتحل محل كل قيمة مفقودة (القحطاني, 2015).
2. **التعويض بآخر قيمة ملاحظة** (Last Observation Carried Forward): هذه الطريقة مقتصرة على معالجة البيانات الطولية, وتستخدم هذه الطريقة لاستبدال كل قيمة مفقودة بآخر قيمة مرصودة في نفس الموضوع, وكلما افتقدت القيم يتم الاستبدال بآخر قيمة مرصودة. ويندر استخدام هذه الطريقة نسبياً في العلوم السلوكية والاجتماعية, وعلى الرغم من الاستخدام المتكرر لهذه الطريقة في الدراسات الطبية والتجارب السريرية, إلا أن هناك العديد من الدراسات التجريبية أثبتت ضعف هذه الطريقة في التعامل مع البيانات المفقودة الطولية (Molenberghs et al., 2015).
3. **التعويض بتقدير هوت دك** (Hot Deck Imputation): تقدير هوت دك للبيانات المفقودة هو مجموعة من التقنيات التي تعوض البيانات المفقودة بدرجات مماثلة من مستجيبين مماثلين للمستجيبين الناقصة بياناتهم, ويعتبر أكثر التطبيقات الشائعة لتقدير هوت دك أن تملأ القيم المفقودة لمستجيبين بقيم عشوائية من مستجيبين آخرين مشابهين لمجموعات المستجيبين الناقصة بياناتهم في نفس الأبعاد والبنود(Little & Rubin, 2002) .
4. **التعويض بتقدير كولد دك** (Cold Dec Imputation): وتعتمد هذه الطريقة على استبدال القيمة المفقودة بقيمة ثابتة يتم اختيارها من مصدر غير قاعدة البيانات الحالية, حيث يتم استبدال القيمة المفقودة بقيمة أخرى من مصدر آخر, كأخذ القيمة من إصدار سابق لنفس الاستقصاء, ويجب أن

ISSN: 2410-7727

يضمن الباحث بأن القيمة المشتقة من بحث سابق أصح وأدق من أي قيمة تشتق داخلياً. ولسوء الحظ فإن القيم التي يمكن استخدامها هي ليست دائماً متوفرة باستخدام طريقة كولد دك (الحمامي، 2007).

5. **التعويض باستخدام طريقة الانحدار** (Imputation Using Regression Analysis) : تعتمد هذه الطريقة على استخدام الانحدار لغرض التقدير والتنبؤ بقيمة متغير ما بدلالة متغير آخر، ولهذه الطريقة العديد من العيوب منها أنها قد تؤدي إلى تعديل في البيانات أكثر من اللازم، كذلك تفترض وجود علاقة ارتباط مع المتغير الذي يحتوي على بيانات مفقودة، ولذلك إذا كان الارتباط ضعيفاً فإنه لا يمكن التنبؤ بالبيانات المفقودة. وعلى الرغم من عيوبها إلا أنها تصلح في الحالات التي تنتشر فيها البيانات المفقودة باعتدال مع وجود علاقة ارتباط كافية بين المتغيرات (أبوعلام، 2007).
6. **طريقة التعويض المتعدد** Multiple Imputation (MI): إن مشكلة التعويض باستخدام المتوسط وباستخدام الانحدار تتمثل في أنه يبخس تقدير التباين في البيانات، والحل لهذه المشكلة هو استخدام عدة تعويضات لتقدير كل من المتغيرات التي تحتوي على قيم مفقودة، والتعويضات المختلفة تتمايز عن بعضها فقط عن طريق إضافة كمية صغيرة من الخطأ العشوائي وبهذا ينتهي الأمر مع العديد من مجموعات البيانات بتعويضات مختلفة عن القيم المفقودة، وكل من مجموعات البيانات هذه يتم تحليلها كما لو كانت مجموعات طبيعية كاملة البيانات وكخطوة أخيرة يتم الجمع بين تقديرات النموذج المختلفة مع الأخذ بعين الاعتبار خطأ الانحدار. لذلك يمر التعويض المتعدد بالخطوات الثلاثة التالية:
   - إنشاء عدة مجموعات بيانات كاملة باستخدام التعويض المتعدد.
   - القيام بإجراء التحليل المطلوب على كل مجموعة من مجموعات البيانات.
   - الجمع بين التحليلات المختلفة (بلنتش، 2017).
7. **خوارزمية تعظيم التوقع** (EM Algorithm) The Expectation Maximization: خوارزمية تعظيم التوقع والتي تسمى اختصاراً (EM) هي طريقة لإيجاد تقدير الإمكان الأعظم لمعلمات النموذج وهي خوارزمية تكرارية تنتج أقصى احتمالات التقديرات للبيانات المفقودة، وتمتاز هذه الطريقة ببساطتها وشموليتها للحالات التي يمكن معالجتها، وتعتمد هذه الطريقة على تعويض البيانات المفقودة من خلال تقديرات يتم الحصول عليها من خلال عملية تقدير متكررة للاستفادة من جميع المعلومات المتاحة من الحالات التامة وغير التامة، حيث يتم أولاً التنبؤ بالبيانات المفقودة بناءاً على تقديرات أولية لقيم معالم النموذج، ثم تستخدم هذه التنبؤات بعد ذلك لتعديل قيم المعالم، ويتم تكرار عملية التنبؤ باستخدام معالم النموذج المعدلة حتى تقترب المعالم من تقديرات الإمكان الأعظم (Davey & savla, 2010).

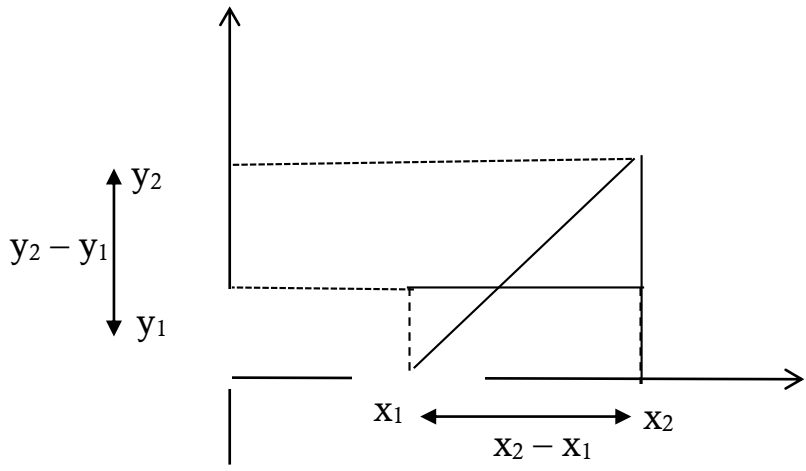
ISSN: 2410-7727

8. **طريقة المعلومات الكاملة للاحتمال الأعظم:**

Full Information Maximum Likelihood (FIML)

تعد طريقة FIML من أشهر طرق تقدير المعالم في النمذجة البنائية , وهي طريقة أساسية لتقدير المتوسطات والتغايرات على أساس البيانات غير المكتملة وعلى افتراض أنها فقدت بصورة عشوائية تامة أو بصورة عشوائية. في هذه الطريقة يتم تقدير البيانات المفقودة باستخدام كامل بيانات الحالة فقط دون الاعتماد على بيانات الحالات الأخرى. (Peng, Harwell, Liou & Ehman, 2006) كذلك تتيح هذه الطريقة الاستفادة من جميع البيانات المتاحة دون اللجوء إلى أي أسلوب من أساليب التعويض المفقودة (بلنتش, 2017).

9. **متوسط الميول المطلقة** Absolute Slopes Mean: تعتمد هذه الطريقة المقترحة على مفهوم ميل المستقيم والذي تعرفه سلامة (2003) بأنه " ظل الزاوية التي يصنعها المستقيم مع الاتجاه الموجب لمحور السينات ". يتحدد ميل المستقيم بمعلومية أي نقطتين عليه ولتكن $A(x_1, y_1)$, $B(x_2, y_2)$ كما هو مبين في الشكل التالي (الجوعي, السعيد, 2005):



ويعطى الميل **S** بالعلاقة :

$$S = \frac{y_2 - y_1}{x_2 - x_1}$$

وتستند فكرة هذه الطريقة المقترحة على أساس حساب الميل المطلق لكل نقطتين معلومتين من عناصر المتغير المستقل وعناصر المتغير التابع ثم حساب المتوسط الحسابي لتلك الميول والتعويض به بدل القيم المفقودة. وفيها يتم حذف جميع الحالات التي تحتوي على القيم المفقودة من جميع المتغيرات الموجودة

ISSN: 2410-7727

بالنموذج وذلك لكل متغير مستقل على حدة مع المتغير التابع. ثم بعد ذلك يتم حساب القيم المطلقة للميول وذلك للحالات الموجودة بعد الحذف مع تجاهل الحالة الأولى عن طريق القانون التالي:

$$\text{القيمة المطلقة للميل} = \left|\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right| \qquad ,i = 2,3,\dots,m$$

حيث أن:

y_i : هو قيمة المتغير التابع رقم i

x_i : هو قيمة المتغير المستقل رقم i

m : حجم العينة بعد حذف الحالات المفقودة منها

ثم يتم حساب المتوسط لجميع القيم التي أخذها في الاعتبار. ويتم التعويض بقيمة هذا المتوسط بجميع الحالات المفقودة بالمتغير المستقل. وأخيراً تكرر الخطوات السابقة لكل متغير من المتغيرات المستقلة الموجودة.

وتمتاز هذه الطريقة بسهولة حسابها كما أنها تعطى أفضل النتائج لتقدير كفاءة نموذج الانحدار للبيانات بعد طريقة EM ، حيث أنها تعطى أقل قيمة لمتوسط مربعات الأخطاء الخاص بالنموذج MSE بعد طريقة EM ، كما أنها تعطى أعلى قيمة لكلٍ من معامل التحديد الخاص بالنموذج ومعامل التحديد المعدل بعد طريقة EM. ولكن يعاب على هذه الطريقة أنه يصعب حسابها في حالة تساوي القيمتين المتتاليتين من المتغير المستقل ولكن يمكن معالجة هذا العيب وذلك عن طريق حذف إحدى الحالات المتشابهة من الحالتين المتتاليتين.

5. الدراسات السابقة:

من خلال اطلاع الباحث على الدراسات السابقة ذات الصلة بموضوع طرق التعامل مع البيانات المفقودة. تبين للباحث وجود عدد من الدراسات التي تناولت طرق معالجة البيانات المفقودة والمقارنة بين كفاءة تلك الطرق في تقدير معالم نماذج الانحدار (البسيط والمتعدد). إلا أن هناك ندرة في الدراسات السابقة التي قدّمت طرقاً مقترحة لتقدير البيانات المفقودة في نموذج الانحدار المتعدد.

وفيما يلي بعض الدراسات والبحوث المنشورة التي تناولت موضوع معالجة البيانات المفقودة والتي لها علاقة بموضوع الدراسة الحالية :

- **دراسة حسين** (2010) بعنوان تحليل البيانات غير التامة لنماذج الانحدار المتعدد باستخدام الخوارزميات EM, ECM, ECME. هدفت هذه الدراسة إلى المقارنة بين الطرق الثلاثة لمعالجة البيانات المفقودة. استخدمت هذه الدراسة أحجام عينات (25, 50, 100, 150) وتم افتراض أن البيانات فقدت بآلية MAR وبنسب فقد 5% و10% و15%. اعتمدت الباحثة معيار متوسط مربعات الخطأ للمعلمات للمقارنة بين طرق التقدير. وبتكرار التجربة 1000 مرة. توصلت الباحثة إلى تفوق طريقتي EM, ECM على طريقة ECME لجميع حجوم العينات ونسب الفقد وللنمطين الأول

ISSN: 2410-7727

والثالث عندما $\sigma_e^2 = 0.5$ فقط. ما عدا ذلك تتفوق طريقة ECME على بقية الطرق لجميع حجوم العينات ونسب الفقد وللنمطين الأول والثالث.

- دراسة حسين (2012) بعنوان تقدير القيم المفقودة لمتغير الاستجابة في نموذج الانحدار الخطي المتعدد. قام الباحث باقتراح طريقة لمعالجة البيانات غير التامة معتمداً على طريقتين وهما طريقة خوارزمية تعظيم التوقع (EM) وطريقة التعويض بالانحدار ومقارنة نتائج هذه الطريقة المقترحة مع نتائج بعض طرق معالجة البيانات المفقودة الأخرى. في هذه الدراسة تم استعمال أحجام عينات (50 , 100) وتم افتراض أن البيانات فقدت بآلية MCAR وبنسب فقد (5%, 10% , 15%) وبتكرار التجربة 1000 مرة. توصل الباحث إلى أن طريقة EM المقترحة تعطي مقدرات جيدة للقيم المفقودة مقارنة بطريقتي التعويض بالانحدار والتعويض بالمتوسط.
- **دراسة** Rubright, Nandakumar & Glutting (2014) بعنوان استخدام المحاكاة لدراسة البيانات المفقودة في حالة تعدد المتغيرات . هدفت هذه الدراسة إلى المقارنة بين طريقتين مشهورتين لمعالجة البيانات المفقودة وهما التعويض المتعدد وطريقة الحذف. وتم استخدام التحيز النسبي كمعيار مقارنة بين الطريقتين. وتم افتراض أن البيانات فقدت بآليتي MCAR و MAR وبنسب فقد (10%, 25%, 50%). توصل الباحثون إلى أن التحيز النسبي يزداد بارتفاع نسبة الفقد في البيانات وتحت آلية الفقد MCAR وجد الباحثون أن التحيز المطلق كان متساوي بالطريقتين . وعند استخدام أسلوب الفقد MAR كان التحيز المطلق أقل بطريقة التعويض المتعدد عندما يكون هناك متغير واحد فقط تعرض للفقد . ووجد الباحثين أيضاً أن طريقة الحذف جيدة عندما تكون آلية الفقد MCAR وذلك بحالة وجود أكثر من متغير تعرض للفقد . واستخدم الباحثين اختبار تحليل التباين ANOVA للكشف عن الفروق للتحيز المطلق بين الطرق المختلفة.
- **دراسة الوردي** (2015) بعنوان التعويض الجزئي للقيم المفقودة. هدفت هذه الدراسة إلى تقديم طريقة جديدة لمعالجة القيم المفقودة, تستند إلى تحديد المتغير المسبب للمشكلة أولاً, ثم التعويض الجزئي بإحدى طرق التعويض الأحادي البسيطة للقيمة المفقودة ثانياً, بحيث نحصل على حالة MCAR للفقدان, وأخيراً استخدام الحذف لمعالجة المشكلة. توصلت النتائج إلى أنه في حالة فقدان البيانات بشكل عشوائي تام MCAR فإنه يمكن معالجتها باستخدام طريقة الحذف, أما إذا تم الفقدان بشكل عشوائي MAR فنستطيع غالباً جعله عشوائي تام (MCAR) بالتعويض الجزئي.
- **دراسة اللصاصمة** (2016) والتي بعنوان أثر نسبية القيم المفقودة وطريقة معالجتها فى دقة تقدير معالم معادلة الانحداد البسيط. هدفت هذه الدراسة إلى الكشف عن أثر طرق معالجة البيانات المفقودة على دقة معاملات معادلة الانحدار البسيط. . استخدمت هذه الدراسة المحاكاة وبنسب فقد (5%, 10%, 15%, 20%) وتم تعويض القيم المفقودة بثلاث طرق (الحذف, الوسط الحسابي,

ISSN: 2410-7727

التعويض المتعدد). وأوضحت النتائج عدم كفاءة طريقة الحذف لوجود تأثير في تقدير معالم معادلة الانحدار. كما أظهرت النتائج كفاءة طريقة التعويض بالوسط الحسابي في حالة نسب الفقد (5%, 10%) وعلى العكس في حالة زيادة نسب الفقد إلى (15%, 20%). وأوضحت النتائج أن التعويض المتعدد هي الطريقة الأكثر كفاءة مقارنة بطريقتي الحذف والتعويض بالوسط الحسابي وذلك في حالة جميع نسب الفقد (5%, 10%, 15%, 20%).

- دراسة Mamun, Zubairi, Hussin & Rana (2016) بعنوان مقارنة بين طرق معالجة البيانات المفقودة بالنموذج الخطي . هدفت هذه الدراسة إلى تقييم أداء طرق التعويض عن البيانات المفقودة العديدة مثل التعويض بالمتوسط وخوارزمية تعظيم التوقع EM بالنموذج الخطي . وتم استخدام جذر متوسط مربعات الأخطاء والمتوسط المطلق للأخطاء كمعايير مقارنة بين هذه الطرق . وهذا بالنسبة لبيانات حقيقية وبيانات عشوائية باستخدام المحاكاة . وكان حجم العينة المستخدم بهذه الدراسة 50 و70 . وتم استخدام نسب الفقد 5%، 10%، 20%، 30% وكان نمط الفقد المستخدم هو MCAR . وأوضحت النتائج أن خوارزمية تعظيم التوقع هي الطريقة الأكثر فاعلية عن طريقة التعويض بالمتوسط وذلك لتحليل النموذج الخطي الذي يحتوي على البيانات المفقودة.
- دراسة **عبدالعاطي، وإسماعيل، وحراز، وأبو الخير** (2017) والتي بعنوان طرق التحليل الإحصائي للبيانات غير التامة في نماذج الانحدار الخطي المتعدد باستخدام أسلوب بيز. هدفت هذه الدراسة إلى معرفة الأسلوب المناسب لتقدير معالم نموذج الانحدار في وجود مشكلة البيانات المفقودة، ومتى يستخدم تقدير البيانات المفقودة والطرق المستخدمة في التقدير. في هذه الدراسة اعتمد الباحثون على بيانات طبية متاحة ومسجلة في مركز الكلى والمسالك البولية. حيث تم إجراء تحليل الانحدار على ثلاثة عشر متغير مستقل لبيان تأثيرها على المتغير التابع والذي هو نسبة الكرياتينين في الدم. وتم استخدام أسلوب بيز في تحليل البيانات غير التامة. وأوضحت النتائج أن طرق معالجة البيانات المفقودة تختلف باختلاف مشاكل الفقد وأن طريقة التعويض بالمتوسط غير الشرطي أسهل في التطبيق من بقية الطرق، كما أظهرت النتائج بأن متوسط مربعات الخطأ النسبي المطلق (MAPE) يتناقص مع زيادة حجم العينة.
- دراسة **العتيبي** (2018) والتي بعنوان تأثير البيانات المفقودة على كفاءة نموذج الانحدار المتعدد وطرق معالجتها. هدفت هذه الدراسة إلى معرفة تأثير البيانات المفقودة على كفاءة نموذج الانحدار المتعدد، كذلك طرق معالجة البيانات المفقودة. في هذه الدراسة تم استعمال أحجام عينات (25 , 71) وتم افتراض أن البيانات فقدت بآلية MCAR وبنسب فقد (10%, 20% , 30%) ثم تقدير البيانات المفقودة بأربعة طرق مختلفة وهي: (طريقة الحذف، طريقة تعظيم التوقع، طريقة التعويض بالمتوسط، طريقة المعلومات الكاملة للاحتمال الأعظم) وبتكرار التجربة 1000 مرة بغية التعرف على أفضل

ISSN: 2410-7727

طرق المعالجة, ومن ثم تتبع التغيرات التي تطرأ على كفاءة النموذج من خلال المعايير (مستوى الدلالة الإحصائية, عدد المتغيرات المستقلة الدالة إحصائياً, معامل التحديد ومعامل التحديد المعدل, معامل تضخم التباين, معاملات الانحدار), وتم التوصل إلى مجموعة من النتائج من أهمها: عدم كفاءة نماذج الانحدار المتعدد في حالة نسب الفقد المتوسطة والعالية بالنسبة لحجوم العينات الصغيرة, وفي حالة نسب الفقد العالية بالنسبة لحجوم العينات الكبيرة. كذلك طريقة تعظيم التوقع هي الأفضل في تقدير كل من (مستوى الدلالة الإحصائية وعدد المتغيرات المستقلة الدالة إحصائياً, معاملات نموذج الانحدار) في جميع حجوم العينات ولمختلف نسب الفقد مقارنة بالطرق الأخرى, وتعطي طريقة التعويض بالمتوسط وطريقة المعلومات الكاملة للاحتمال الأعظم نتائج جيدة ومتقاربة من بعضها البعض ولكنها أقل فاعلية من طريقة تعظيم التوقع, كما أن طريقة الحذف تعد طريقة جيدة في حالة حجوم العينات الكبيرة ونسب الفقد القليلة.

6. منهجية البحث:

تم استخدام المحاكاة لتوليد أرقام عشوائية تتبع التوزيع الطبيعي القياسي وذلك لعمل نموذج انحدار لثلاثة متغيرات مستقلة باستخدام برنامج **R** اصدار 3.4.4. بشرط توافر جميع شروط الانحدار المتعدد بالنموذج. ثم تم عمل فقد لهذه البيانات بنسب فقد 10%، 20%، 30% على التوالي. وأحجام عينات 25 و100 ، وذلك باستخدام أربعة طرق معروفة مضافاً إليهم الطريقة المقترحة. كما تم استخدام كلاً من متوسط مربعات الأخطاء الخاص بنموذج الانحدار MSE، معامل التحديد R^2 ومعامل التحديد المعدل Adjusted $\mathbf{R^2}$ ، و معامل تضخم التباين VIF.

7. خطوات إجراء البحث:

1. توليد أربعة متغيرات عشوائية تتبع التوزيع الطبيعي القياسي وبحجم العينة 25، 100 على أن يكون ثلاثة منهم متغيرات مستقلة (x_1, x_2, x_3) والأخر يمثل الخطأ العشوائي بنموذج الانحدار المتعدد (e)

$$y = x_1 + x_2 + x_3 + e$$

2. عمل فقد للمتغيرات المستقلة بنسب الفقد الثلاث المختلفة 10%، 20%، و30%
3. استخدام خمسة طرق تعويضية للتعويض عن القيم المفقودة (متوسط الميول (SM)، تعظيم التوقع (EM) ، التعويض بالانحدار (REG), والتعويض بالمتوسط (MEAN)، وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)).

ISSN: 2410-7727

4. تطبيق نموذج الانحدار المتعدد بعد التعويض بالطرق المذكورة بالخطوة (3) واستخراج أربعة معايير مختلفة للمقارنة بين الطرق التعويضية السابقة، وهم معامل التحديد R^2، معامل التحديد المعدل R^2 Adjusted، متوسط مربعات الأخطاء MSE، معامل تضخم التباين VIF.
5. تكرار الخطوات السابقة 1000 مرة.
6. الاعتماد على المتوسط الحسابي بمخرجات البرنامج (1000 مرة) في تقييم الطرق التعويضية.

8. النتائج

- نتائج السؤال الفرعي الأول والذي ينص على "ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار(REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)) في ظل وجود معيار MSE للمقارنة بين الطرق؟"

يوضح الجدول التالي النتائج التي تم الحصول عليها من إجراء المحاكاة لخمسة طرق تعويضية مختلفة كما يلي:

جدول (1): متوسط مربعات الأخطاء للطرق التعويضية الخمسة

حجم العينة	نسبة الفقدان	طريقة التقدير	MSE
25	0.1	SM	1.19
		EM	0.84
		FIML	1.31
		REG	1.67
		MEAN	1.33
	0.2	SM	1.20
		EM	0.71
		FIML	1.61
		REG	2.30
		MEAN	1.67
	0.3	SM	1.18
		EM	0.66
		FIML	1.90
		REG	2.76
		MEAN	1.98
100	0.1	SM	1.20

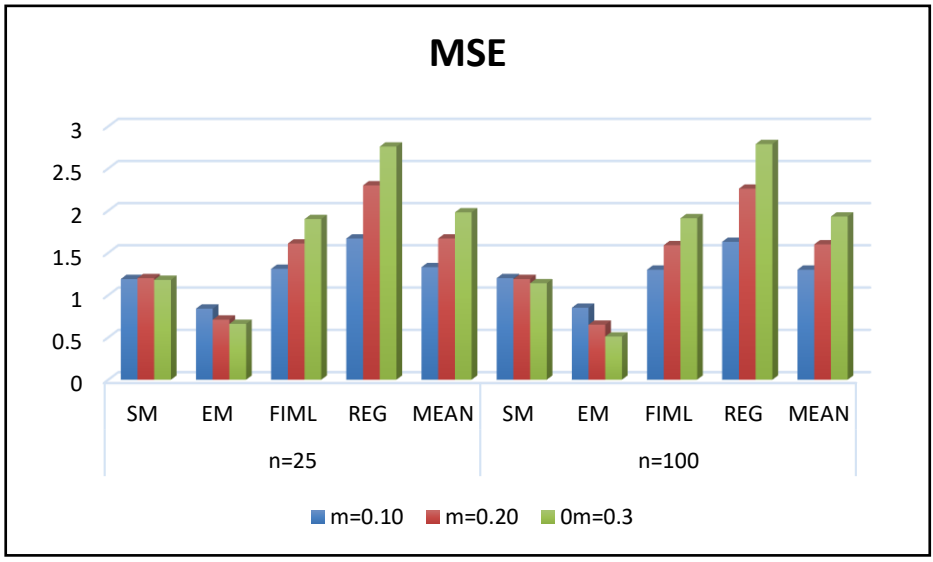
ISSN: 2410-7727

حجم العينة	نسبة الفقدان	طريقة التقدير	MSE
		EM	0.85
		FIML	1.30
		REG	1.63
		MEAN	1.30
	0.2	**SM**	1.19
		EM	0.65
		FIML	1.59
		REG	2.26
		MEAN	1.60
	0.3	**SM**	1.14
		EM	0.51
		FIML	1.91
		REG	2.79
		MEAN	1.93

حيث يظهر الجدول السابق ما يلي:

بالنسبة لمعيار المقارنة MSE بجميع أحجام العينات وجميع نسب الفقد المستخدمة تبين أن الطريقة المقترحة (SM) تعد هي ثاني أفضل الطرق بعد طريقة تعظيم التوقع (EM) وذلك لأنها تحتوي على MSE أقل من طريقة التعويض بالمتوسط (MEAN)، وطريقة التعويض بالانحدار (REG)، وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)، وذلك كما هو موضح بالشكل التالي:

ISSN: 2410-7727



شكل (2): متوسط مربعات الأخطاء للطرق التعويضية الخمسة

- نتائج السؤال الفرعي الثاني والذي ينص على "ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار(REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)) في ظل وجود معيار R Square للمقارنة بين الطرق؟"

يوضح الجدول التالي النتائج التي تم الحصول عليها من إجراء المحاكاة لخمسة طرق تعويضية مختلفة كما يلي:

جدول (2): معامل التحديد للطرق التعويضية الخمسة

حجم العينة	نسبة الفقدان	طريقة التقدير	R Square
25	0.1	SM	0.73
		EM	0.80
		FIML	0.70
		REG	0.62
		MEAN	0.70
	0.2	SM	0.74
		EM	0.84

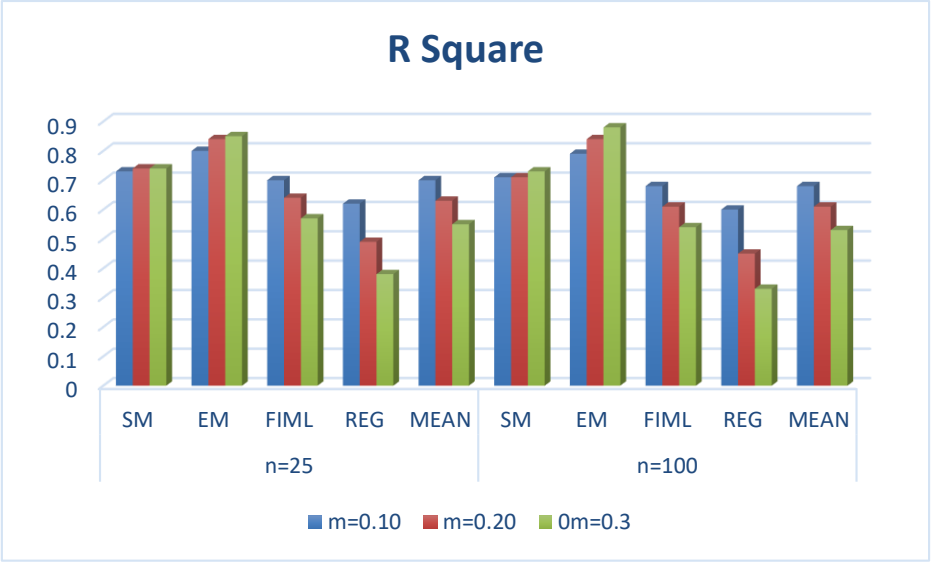
ISSN: 2410-7727

حجم العينة	نسبة الفقدان	طريقة التقدير	R Square
		FIML	0.64
		REG	0.49
		MEAN	0.63
		SM	0.74
		EM	0.85
	0.3	**FIML**	0.57
		REG	0.38
		MEAN	0.55
		SM	0.71
		EM	0.79
	0.1	**FIML**	0.68
		REG	0.60
		MEAN	0.68
		SM	0.71
		EM	0.84
100	**0.2**	**FIML**	0.61
		REG	0.45
		MEAN	0.61
		SM	0.73
		EM	0.88
	0.3	**FIML**	0.54
		REG	0.33
		MEAN	0.53

حيث يظهر الجدول السابق ما يلي:

بالنسبة لمعيار المقارنة $\mathbf{R^2}$ نجد أنه بجميع أحجام العينات وجميع نسب الفقد المستخدمة تبين أن الطريقة المقترحة (SM) تعد هي ثاني أفضل الطرق بعد طريقة تعظيم التوقع (EM)، وذلك لأنها تحتوي على R^2 أكبر من طريقة التعويض بالمتوسط (MEAN)، وطريقة التعويض بالانحدار (REG)، وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)، وذلك كما هو موضح بالشكل التالي:

ISSN: 2410-7727



شكل (3): معامل التحديد للطرق التعويضية الخمسة

- نتائج السؤال الفرعي الثالث والذي ينص على "ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار(REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)) في ظل وجود معيار Adjusted R Square للمقارنة بين الطرق؟"

يوضح الجدول التالي النتائج التي تم الحصول عليها من إجراء المحاكاة لخمسة طرق تعويضية مختلفة كما يلي :

ISSN: 2410-7727

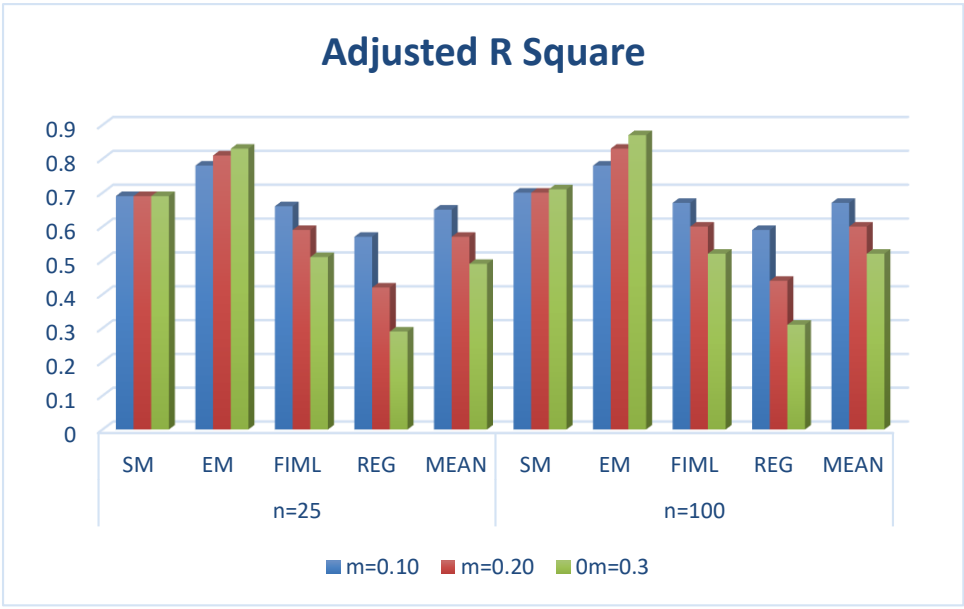
جدول (3): معامل التحديد المعدل للطرق التعويضية الخمسة

حجم العينة	نسبة الفقدان	طريقة التقدير	Adjusted R Square
25	0.1	SM	0.69
		EM	0.78
		FIML	0.66
		REG	0.57
		MEAN	0.65
	0.2	SM	0.69
		EM	0.81
		FIML	0.59
		REG	0.42
		MEAN	0.57
	0.3	SM	0.69
		EM	0.83
		FIML	0.51
		REG	0.29
		MEAN	0.49
100	0.1	SM	0.70
		EM	0.78
		FIML	0.67
		REG	0.59
		MEAN	0.67
	0.2	SM	0.70
		EM	0.83
		FIML	0.60
		REG	0.44
		MEAN	0.60
	0.3	SM	0.71
		EM	0.87
		FIML	0.52
		REG	0.31
		MEAN	0.52

ISSN: 2410-7727

حيث يظهر الجدول السابق ما يلي:

بالنسبة لمعيار المقارنة Adjusted $\mathbf{R^2}$ نجد أنه بجميع أحجام العينات وجميع نسب الفقد المستخدمة تبين أن الطريقة المقترحة (SM) تعد هي ثاني أفضل الطرق بعد طريقة تعظيم التوقع (EM), وذلك لأنها تحتوي على Adjusted $\mathbf{R^2}$ أكبر من طريقة التعويض بالمتوسط (MEAN)، وطريقة التعويض بالانحدار (REG)، وطريقة المعلومات الكاملة للإمكان الأعظم (FIML), وذلك كما هو موضح بالشكل التالي:



شكل (4): معامل التحديد المعدل للطرق التعويضية الخمسة

- نتائج السؤال الفرعي الرابع والذي ينص على "ما كفاءة الطريقة المقترحة (SM) في تقدير القيم المفقودة مقارنة بطرق المعالجات (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار(REG) وطريقة تعظيم التوقع (EM) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)) في ظل وجود معيار VIF للمقارنة بين الطرق؟"

ISSN: 2410-7727

يوضح الجدول التالي النتائج التي تم الحصول عليها من إجراء المحاكاة لخمسة طرق تعويضية مختلفة كما يلي:

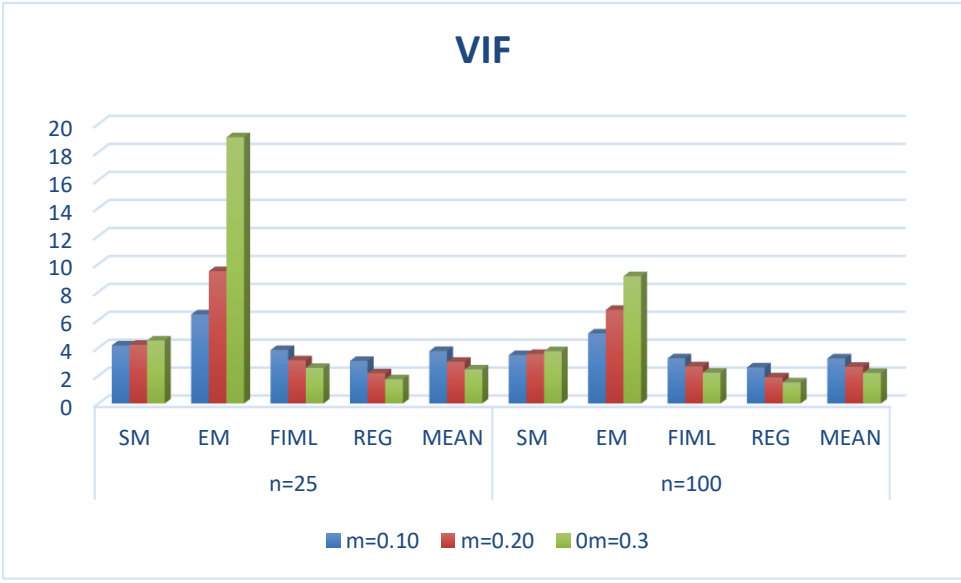
جدول (4): معامل تضخم التباين للطرق التعويضية الخمسة

حجم العينة	نسبة الفقدان	طريقة التقدير	VIF
25	0.1	SM	4.16
		EM	6.38
		FIML	3.83
		REG	3.05
		MEAN	3.76
	0.2	SM	4.21
		EM	9.48
		FIML	3.09
		REG	2.17
		MEAN	2.99
	0.3	SM	4.51
		EM	19.05
		FIML	2.55
		REG	1.74
		MEAN	2.44
100	0.1	SM	3.47
		EM	5.02
		FIML	3.24
		REG	2.59
		MEAN	3.23
	0.2	SM	3.55
		EM	6.71
		FIML	2.66
		REG	1.87
		MEAN	2.63
	0.3	SM	3.75
		EM	9.12
		FIML	2.21
		REG	1.51
		MEAN	2.19

ISSN: 2410-7727

حيث يظهر الجدول السابق ما يلي:

بالنسبة لمعيار معامل تضخم التباين (VIF) نجد أن نتيجة الطريقة المقترحة (SM) تتفوق على طريقة تعظيم التوقع (EM) بجميع أحجام العينات وجميع نسب الفقد حيث لها معامل تضخم تباين أقل من طريقة تعظيم التوقع. بينما تتفوق بقية الطرق الأخرى على الطريقة المقترحة (SM), حيث لها معامل تضخم تباين أقل من الطريقة المقترحة (SM), وذلك كما هو موضح بالشكل التالي:



شكل (5): معامل تضخم التباين للطرق التعويضية الخمسة

ISSN: 2410-7727

الخلاصة والتوصيات

تعد طريقة متوسط الميول المطلقة (SM) طريقة تعويضية جيدة لاستخدامها بالتعويض عن البيانات المفقودة بحالة الانحدار المتعدد. حيث أظهرت النتائج أنها أفضل من طريقة التعويض بالمتوسط (MEAN), وطريقة التعويض بالانحدار (REG), وطريقة المعلومات الكاملة للإمكان الأعظم (FIML), وذلك بجميع أحجام العينات (25 , 100) وجميع نسب الفقد 10%، 20%، 30%, عند استخدام المعايير (متوسط مربعات الأخطاء MSE, معامل التحديد $\mathbf{R^2}$, معامل التحديد المعدل $\mathbf{R^2}$ Adjusted) كطرق مقارنة. وعلى الرغم من تفوق الطريقة المقترحة (SM) على طريقة تعظيم التوقع (EM) بجميع أحجام العينات وجميع نسب الفقد عند استخدام معامل تضخم التباين VIF كمعيار للمقارنة إلا أن النتائج أظهرت بأنها أقل من الطرق الأخرى (طريقة التعويض بالمتوسط الحسابي (MEAN) وطريقة التعويض بالانحدار(REG) وطريقة المعلومات الكاملة للإمكان الأعظم (FIML)). في ظل وجود معيار VIF. وفي ضوء ما توصلت إليه نتائج الدراسة الحالية توصي الدراسة بما يلي:

1- وفقاً لمبدأ "الوقاية خير من العلاج", فإنه لا توجد طريقة مثلى لمعالجة البيانات المفقودة والحل الأفضل هو تقليل المشكلة أثناء مرحلة جمع البيانات.
2- يوصي الباحث باستخدام طريقتي تعظيم التوقع (EM) والطريقة المقترحة (SM) لمعالجة البيانات المفقودة في نموذج الانحدار الخطي المتعدد.
3- إجراء المزيد من الدراسات للتحقق من كفاءة طريقة (SM) في نماذج الانحدار الأخرى (غير الخطية).

ISSN: 2410-7727

المراجع العربية

أبو علام، رجاء محمود. (2007). *مناهج البحث في العلوم النفسية والتربوية*. ط6. القاهرة: دار النشر للجامعات .

بلنتش، نيلز جيه. (2017). *مقدمة في نمذجة المعادلات البنائية باستخدام* IBM SPSS STATISTICS AND AMOS. (ترجمة: سعد القحطاني). الرياض: معهد الإدارة العامة.

الجوعي، عبدالله محمد والسعيد، السيد أنور. (2005). *فصول في مبادئ الرياضيات. ط2*. الرياض: مكتبة الرشد.

حسين، أنعام عبود. (2010). *تحليل البيانات غير التامة لنماذج الانحدار المتعدد باستخدام الخوارزميات* ECME, ECM, EM [رسالة ماجستير غير منشورة]. جامعة بغداد.

حسين، علي ناصر. (2012). تقدير القيم المفقودة لمتغير الاستجابة في نموذج الانحدار الخطي المتعدد. *مجلة العلوم الاقتصادية*، 8(30) , 231-246 .

الحمامي، علاء حسين. (2007). *تنقيب البيانات*. ط1. عمّان: إثراء للنشر والتوزيع .

دودين، حمزة محمد. (2013). *التحليل الإحصائي المتقدم للبيانات باستخدام SPSS*. ط2. عمّان: دار المسيرة للنشر والتوزيع.

رزق الله، عايدة نخلة. (2002). *دليل الباحثين في التحليل الإحصائي الاختيار والتفسير*. القاهرة: دار الكتب.

سلامة، ميرفانا ياسر. (2003). *موسوعة التعريفات العلمية الرياضيات*. عمّان: دار صفاء للنشر والتوزيع.

عبد الرحمن، طارق عطية. (2013). *دليل تصميم وتنفيذ البحوث في العلوم الاجتماعية منهج تطبيقي لبناء المهارات البحثية*. الرياض: معهد الإدارة العامة.

عبد العاطي، فاطمة، إسماعيل، عائشة، حراز، أحمد وأبو الخير، أحمد. (2017). طرق التحليل الإحصائي للبيانات غير التامة في نماذج الانحدار الخطي المتعدد باستخدام أسلوب بيز. *المجلة المصرية للدراسات التجارية*، 41(2), 421-442 .

العتيبي، أشرف. (2018). *تأثير البيانات المفقودة على كفاءة نموذج الانحدار المتعدد وطرق معالجتها* [أطروحة دكتوراة غير منشورة]. جامعة أم القرى.

القحطاني، سعد. (2015). *الإحصاء التطبيقي – المفاهيم الأساسية وأدوات التحليل الإحصائي الأكثر استخداماً في الدراسات والبحوث الاجتماعية والإنسانية باستخدام* .SPSS الرياض: معهد

ISSN: 2410-7727

الإدارة العامة.

اللصاصمة، عمران إسماعيل. (2016). *أثر نسبية القيم المفقودة وطريقة معالجتها فى دقة تقدير معالم معادلة الانحداد البسيط* [رسالة ماجستير، جامعة مؤتة]. دار المنظومة http://search.mandumah.com/Record/784622

هويت، دنس وكرمر، دنكن. (2016). *مقدمة لحزمة البرامج الإحصائية SPSS في علم النفس*. (ترجمة: صلاح الدين محمود علام). عمّان: دار الفكر للنشر.

الوردي، هيثم عبدالأمير. (2015). التعويض الجزئي للقيم المفقودة. *مجلة الباحث، دون مجلد* (9)، 178-195.

ISSN: 2410-7727

المراجع الأجنبية

Allison, P. D. (2001) . Missing data. Thousand Oaks, CA: Sage Publications.

Davey, A. & Savla, J. (2010) . *Statistical powers analysis with missing data astructural equation modeling approach* . New York : Routledge .

Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.

Graham, J. W. (2012). *Missing data analysis and design***.** New York: springer.

Little, R . J. A. (1992) . Regression with missing X‘s : *Journal of the American Statistical Association,* 87, 1227 – 1237.

Little, R. J. A. & Rubin, D. B. (2002) **.** *Statistical analysis with missing data*. New York: Wiley.

Mamun, A., Zubairi, Y., Hussin, A., & Rana, S . (2016). A comparison of missing data handling methods in linear structural relationship model: evidence from BDHS2007 data. *Electronic Journal of Applied Statistical Analysis.* 122 - 133 .

Mcknight, P. E., Mcknight, K. M., Sidani, S. & Figuredo, A . J. (2007) **.** *Missingdata a gentle introduction***.** New York : The Guilford Press.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2015). *Handbook of missing data methodology*. New York: CRC Press.

Nakai, M. &Ke, W. (2011). Review of methods for handling missing data in longitudinal data analysis. *Journal of Math. Analysis*, 5, 1-13.

Peng, C., Harwell, M., Liou, S. &Ehman, L. (2006*). Advances in missing data methods and implications for educational research*. Real Data Analysis, PP. 31-78.

Rubright, J., Nandakumar, R., &Glutting, J . (2014). *A Simulation Study of Missing Data with Multiple Missing X’s*. Practical Assessment, Research & Evaluation .

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.