

المقارنة بين طريقتي التعويض بآخر قيمة ملاحظة وهوت دك لمعالجة البيانات المفقودة في ضوء معيار قوة الاختبار الإحصائي "دراسة محاكاة"

د. أشرف أحمد العتيبي

أستاذ الإحصاء والبحوث المساعد بكلية التربية - جامعة الملك خالد - المملكة العربية السعودية

Doi10.55534/1320-009-003-005

المستخلص:

هدفت الدراسة الحالية إلى تحديد أكفأ طرق التعويض عن البيانات المفقودة، وذلك من خلال مقارنة طريقة التعويض بآخر قيمة ملاحظة، وطريقة هوت دك لتقدير البيانات المفقودة؛ ولتحقيق هدف الدراسة تم استخدام المحاكاة لتوليد أرقام عشوائية بأحجام عينات مختلفة (30، 60، 90)، حيث تم بناء نموذج محاكاة لأسلوب تحليل التباين الأحادي (في حالة تساوي أحجام العينات بين المجموعات، وفي حالة عدم تساوي أحجام العينات بين المجموعات) وتتوافر فيه الافتراضات الخاصة بأسلوب تحليل التباين الأحادي. وتم حساب القوة الإحصائية للنموذج التام للحالات المذكورة. ثم تم عمل فقد لهذه البيانات بنسب فقد بلغت (15%)، و(25%)، و(40%)، على التوالي وفق آلية الفقد (MCAR)، وتحت نمط فقد (النمط العام). وبعد ذلك تم تقدير القيم المفقودة باستخدام طريقتي التعويض. ثم إجراء اختبار تحليل التباين الأحادي للنماذج المقدرة، وحساب القوة الإحصائية لتلك النماذج للحالات المذكورة، ثم تتم المفاضلة بين طريقتي التعويض بآخر قيمة ملاحظة وهوت دك لتقدير البيانات المفقودة، وذلك من خلال مقارنة القوة الإحصائية لتلك النماذج المقدرة بقوة النموذج الإحصائي التام. وأظهرت نتائج الدراسة أن طريقة التعويض بآخر قيمة ملاحظة (LOCF) أكفأ من طريقة هوت دك (HOT DECK) بجميع حالات نسب الفقد وأحجام العينات المختلفة، إذ تعطي طريقة (LOCF) قيمة تقديرية تؤدي إلى نتائج أقرب إلى النموذج التام وفقاً لمعيار قوة الاختبار الإحصائي للنموذج.

الكلمات المفتاحية: البيانات المفقودة، التعويض بآخر قيمة ملاحظة، التعويض بطريقة هوت دك، قوة الاختبار الإحصائي.

**Comparison of the two methods of compensation with the Last Observation Carried Forward and hot deck to process missing data in light of the statistical test strength criterion
"Simulation study"**

Dr. Ashraf Ahmad Al otaibi

Assistant Professor of Statistics and Research
Saudi Arabia, King Khalid University, College of Education

Abstract

The study aimed to identify the most efficient ways to compensate for missing data, this is by comparing the two methods of compensation with the Last Observation Carried Forward and hot deck Imputation to estimate the missing data. To achieve the study goal, the simulation was used to generate random numbers of different sample sizes (30, 60, and 90). A simulation model of One-Way Analysis of Variance method was built (if sample sizes are equal between groups, and sample sizes are not equal between groups) and assumptions are available for One Way Analysis of Variance method. The statistical strength of the full model of these cases was calculated. This data was then lost by 15%, 25%, 40% respectively according to the MCAR mechanism, and under the pattern of loss (general pattern). The lost values were then estimated using the two methods of compensation. Then conduct the One-Way Analysis of Variance test for estimated models and calculate the statistical strength of those models for the cases mentioned. The two methods of compensation are then compared with the Last Observation Carried Forward and hot deck to estimate the missing data by comparing the statistical strength of those models estimated by the strength of the full statistical model. The results of the study showed that the compensation method for (LOCF) is more efficient than the hot deck. The LOCF method gives estimated values that lead to results closer to the full model according to the model's statistical test strength standard.

Keywords: Missing data, Last Observation Carried Forward Imputation, Hot Deck Imputation, statistical test power.

المقدمة:

تستند البحوث التربوية إلى أسسٍ متينةٍ نابعةٍ من حاجات المجتمع وملائمةٍ لظروفه وخصائصه، مستعينة بالأساليب العلمية الحديثة في معالجة المشكلات التربوية، وصنع القرارات المتعلقة بتطوير أو تعديل أو تغيير أي جانبٍ من جوانب العملية التربوية، وهذا من شأنه أن يعمل على تمكين التربويين من تجويد أساليب البحث التربوي و أدواته خلال سعيهم للكشف عن المعرفة الجديدة، عبر تقديم أفضل الحلول والبدائل التي تمكن من فهم العملية التعليمية بأبعادها المختلفة، وما يواجهها من مشكلاتٍ بأسلوبٍ موضوعي؛ في محاولة الوصول إلى نتائج بحثية موثوقة، واتخاذ قراراتٍ سليمة بصدد مشكلةٍ بحثيةٍ معينة.

ويتضمن البحث التربوي استخدام الاستدلال الإحصائي؛ بهدف الوصول إلى معرفةٍ قابلةٍ للتعميم من خلال الاعتماد على دراسة خصائص العينة، والاستدلال منها على خصائص المجتمع الذي أخذت منه تلك العينة، وبعد اختبار الفرضيات أحد الموضوعات الرئيسة للاستدلال الإحصائي التي تهتم بالتحقق من معرفة مدى الاتفاق أو الاختلاف بين بارامترات المجتمع الأصل وإحصاءات العينة (الشربجي، حسن، 2012). عندما يقوم الباحث باختبار صحة فرضياته يحاول تجنب نوعين من القرارات الإحصائية الخطأ: أولهما الخطأ من النوع الأول ويتمثل في رفض الفرضية الصفرية وهي صحيحة، ويرمز له بالرمز (α) ويسمى أحيانا بمستوى الدلالة الإحصائية، والآخر الخطأ من النوع الثاني ويتمثل في قبول الفرضية الصفرية وهي خطأ ويرمز له بالرمز (β) (عبد، 2017). ولتجنب الوقوع في خطأ النوع الثاني ظهر مفهوم قوة الاختبار الإحصائي ليعبر عن احتمالية الكشف عن حجم التأثير غير الصفري، أي احتمال قرار رفض الفرضية الصفرية وهي خطأ، وهذا ما يريده الباحث عادة، أي أن يؤدي بحثه في النهاية إلى قرارٍ صحيحٍ فيما يتعلق بالتظيرة التي يختبرها، ولذا من المفيد للباحث معرفة العوامل التي تؤثر على قوة الاختبار الإحصائي وكيفية تأثر كل منها (محمد، ٢٠١٣). وتتأثر قوة الاختبار الإحصائي بعوامل أبرزها حجم العينة، فكلما زاد حجم العينة زادت قوة الاختبار الإحصائي، وزادت بالتالي احتمالية الحصول على نتائج دالة إحصائية. وتعد مشكلة فقد البيانات إحدى أبرز المشكلات التي تعيق عملية التحليل الإحصائي، وتؤثر على جودة البيانات بشكل عام، وفي كثيرٍ من الحالات يتم إهمال البيانات المفقودة، أو لا تتم معالجتها بالطرق الإحصائية المناسبة، بل يتم أحيانا حذف الحالات التي تحتوي على بياناتٍ مفقودة في أثناء عملية التحليل مما يؤثر على القوة الإحصائية للاختبارات المستخدمة (دودين، 2013).

وبالرغم من الاهتمام المتزايد من الإحصائيين بمشكلة البيانات المفقودة من خلال تطوير طرق تقديرها ومعالجتها فإن الطرق التقليدية التي تعتمد على طريقة الحذف لاتزال مستخدمة وذلك لسهولة تنفيذها، ولأن الأساليب الإحصائية تفترض معلوماتٍ كاملةٍ عن المتغيرات المدرجة في التحليل، وبالتالي فإن معالجة هذا النوع من البيانات يتم بالطرق التقليدية وذلك بحذف الحالات التي تحتوي بياناتٍ مفقودة. الأمر الذي جعل كثيرا من الباحثين، يتساهلون في التعامل مع البيانات المفقودة دون إدراكٍ لما قد تسببه من آثارٍ على دقة نتائج التحليلات الإحصائية، وضعف في القوة الإحصائية للاختبارات والمقاييس المستخدمة (العتيبي، 2018).

مشكلة الدراسة:

تعد مشكلة فقد البيانات من المشكلات المزعجة في تحليل البيانات، وبخاصة إذا تجاهل الباحث التعامل معها نتيجة الميل للسهولة، وقلة الوعي بحجم المشكلة، والاعتماد على البرامج الإحصائية التي صممت على افتراض أن البيانات كاملة وبالتالي فإن معالجة هذا النوع من البيانات تتم بالطرق التقليدية وذلك بحذف الحالات التي تحتوي بياناتٍ مفقودة؛ مما قد

يؤدي إلى استنتاجاتٍ خطأ، وتقديراتٍ متحيزة، وضعفٍ في القوة الإحصائية للاختبارات الإحصائية المستخدمة في التحليل؛ وذلك نتيجة حذف جزءٍ من بيانات العينة، وبخاصة إذا كانت البيانات المفقودة كبيرة، وحجوم العينات متوسطة أو صغيرة.. وقد حذرت الجمعية الأمريكية للعلوم النفسية والتربوية من استخدام أساليب الحذف عند التعامل مع البيانات المفقودة، وطالب باستخدام طرق إحصائية متطورة تأخذ في الاعتبار الظروف التي يتم في ظلها فقد البيانات (Peng)، Harwell، Liou & Ehman (2006). وتتمثل مشكلة الدراسة في مقارنة طريقتين سهلتين للتعامل مع البيانات المفقودة، وهما طريقتان مبرجتان في أغلب البرامج الإحصائية المتقدمة، ولا يتطلب استخدامهما مهارات إحصائية كبيرة وتفاني بغرض الباحثين في مجال الأبحاث الإنسانية. ومن خلال اطلاع الباحث على عديد الدراسات النفسية والتربوية المنشورة في المجلات العلمية، التي تناولت موضوع البيانات المفقودة وطرق معالجتها، وعلى حد علم الباحث فإنه لا توجد أي من هذه الدراسات قامت بدراسة مقارنة بين طرق معالجة البيانات المفقودة اعتماداً على معيار قوة الاختبار الإحصائي. وتحاول هذه الدراسة تحديد كفاءة طرق معالجة البيانات المفقودة المستخدمة في تفسير قوة الاختبار الإحصائي مع محاولة الوصول إلى أفضل الطرق لتقدير البيانات المفقودة في ضوء: تغير حجم العينة، وفقدان جزءٍ من البيانات بنسبٍ مختلفة. ويمكن صياغة مشكلة الدراسة في السؤال الرئيس التالي:

ما كفاءة طريقتي التعويض بآخر قيمة ملاحظة وهوت دوك لتقدير البيانات المفقودة في ضوء معيار قوة الاختبار الإحصائي؟

ويتفرع من هذا السؤال الأسئلة التالية:

- ١- ما كفاءة طريقتي التعويض بآخر قيمة ملاحظة وهوت دوك لتقدير القيم المفقودة في ضوء معيار قوة اختبار تحليل التباين الأحادي في حالة تساوي أحجام العينات بين المجموعات؟
 - ٢- ما كفاءة طريقتي التعويض بآخر قيمة ملاحظة وهوت دوك لتقدير القيم المفقودة في ضوء معيار قوة اختبار تحليل التباين الأحادي في حالة عدم تساوي أحجام العينات بين المجموعات؟
- أهداف الدراسة:

- هدفت الدراسة الحالية إلى المفاضلة بين طريقتي التعويض بآخر قيمة ملاحظة وهوت دوك لتقدير القيم المفقودة وذلك من خلال تعرف دقتها في تقدير قوة الاختبار الإحصائي، وبناء على الهدف العام فقد تفرعت منه الأهداف التالية:
- ١- المفاضلة بين طريقتي التعويض بآخر قيمة ملاحظة وهوت دوك لتقدير القيم المفقودة في ضوء معيار قوة اختبار تحليل التباين الأحادي في حالة تساوي أحجام العينات بين المجموعات.
 - ٢- المفاضلة بين طريقتي التعويض بآخر قيمة ملاحظة وهوت دوك لتقدير القيم المفقودة في ضوء معيار قوة اختبار تحليل التباين الأحادي في حالة عدم تساوي أحجام العينات بين المجموعات.
- أهمية الدراسة:

تكمن أهمية هذه الدراسة - في حدود علم الباحث - في أنّ هناك ندرة في الدراسات التي أجريت حول موضوع طرق التعامل مع البيانات المفقودة ودقتها في تقدير القيم المفقودة؛ لذا تمثلت أهمية هذه الدراسة في جانبين: الأهمية النظرية، والأهمية التطبيقية. وتمثل الأهمية النظرية في:

- (١) معرفة البيانات المفقودة وطرق علاجها وبيان أنماط فقد البيانات وآلياتها.
- (٢) الكشف عن تأثير نسب فقد البيانات المختلفة على قوة اختبار تحليل التباين الأحادي.

(٣) تعدّ الدراسة الحالية من أوليات الدراسات - حسب اطلاع الباحث - التي اعتمدت قوة الاختبار الإحصائي معيارا للمقارنة بين مدى دقة طرائق معالجة البيانات المفقودة

(٤) كما تعدّ الدراسة الأولى التي قارنت بين كفاءة طريقتي التعويض بآخر قيمة ملاحظة وهوت دوك باعتبارهما طريقتين تعويضيتين بديلتين لتقليل تأثير القيم المفقودة على قوة الاختبار الإحصائي. أمّا الأهمية التطبيقية للدراسة الحالية، فيتوقع أن تحدّد نتائج الدراسة للباحثين أفضل الطريقتين لتقدير القيم المفقودة التي تعطي تقديرات دقيقة لقوة الاختبار الإحصائي لتحليل التباين الأحادي؛ ممّا يسهّل عليهم حلّ مشكلة البيانات المفقودة في الدراسات التربوية التي تستخدم تحليل التباين الأحادي.

مصطلحات الدراسة:

البيانات المفقودة (Missing Data):

تعرف البيانات المفقودة أنّها البيانات التي تركت دون إجابة بسبب الفشل في الحصول على بيانات من أجزاء من المجتمع المستهدف أو العينة (عبد الرحمن، 2013).

وتعرف إجرائيًا: فقد بيانات من البيانات التامة بنسب مختلفة (15%، 25%، 40%).

طرق التعامل مع البيانات المفقودة (Methods Handling Missing Data):

هي الطرق والأساليب الإحصائية المناسبة والمستخدم لغرض تقدير البيانات المفقودة وإيجادها، وتشمل طرائق الحذف وطرائق التعويض (النعيمي والعاني، 2008).

وتعرف إجرائيًا: أنّها الطرق المعتمدة لأغراض هذه الدراسة وتشمل: طريقتي التعويض بآخر قيمة ملاحظة، وهوت دوك لتقدير البيانات المفقودة.

طريقة التعويض بآخر قيمة ملاحظة (Last Observation Carried Forward):

وتستخدم هذه الطريقة لاستبدال كل قيمة مفقودة بآخر قيمة مرصودة في الموضوع نفسه (Molenberghs، Kenward، Fitzmaurice، Tsiatis and Verbeke، 2015).

طريقة هوت دك (Hot Deck Imputation):

تعتمد هذه الطريقة ملء القيم المفقودة لمستجيبين بقيم عشوائية من مستجيبين آخرين مشابهين لمجموعات المستجيبين الناقصة بياناتهم في نفس الأبعاد والبنود. (Little & Rubin 2002)

قوة الاختبار (Power of the Test):

ويقصد بها: قدرة الاختبار على رفض الفرضية الصفرية التي تكون في حقيقتها خاطئة نظرياً (الشوربجي، حسن، 2012). وتعرف إجرائيًا أنّها: القيمة التي قام الباحث بحسابها لاختبار تحليل التباين الأحادي في حالة تساوي حجوم العينات وفي حالة اختلاف حجوم العينات باستخدام برنامج حساب القوة الإحصائية G*Power 3.0.10.

حدود الدراسة:

تحدّث الدراسة الحالية بحدود موضوعية تمثّلت في: دراسة تحليل قوة الاختبار لتحليل التباين الأحادي في حالة تساوي حجوم العينات بين المجموعات، وفي حالة عدم تساوي حجوم العينات بين المجموعات، وذلك في ضوء أحجام عينات (30، 60، 90)، ونسب فقد (15%، 25%، 40%) على التوالي وفق آلية الفقد (MCAR)، وتحت نمط فقد (التمط العام). وطرق معالجة (طريقة التعويض بآخر قيمة ملاحظة، وطريقة هوت دك).

الجانِب النَّظري

أولاً: البيانات المفقودة

في كل بحثٍ علمي يوجد احتمال لبيانات ناقصة أو غير مكتملة، وهذه البيانات المفقودة تأتي نتيجة عدم إجابة المشاركين عن بعض الأسئلة إما بقصد أو بغير قصد، كما يمكن أن يكون ذلك نتيجة انسحاب المشاركين من البحث قبل استكمالهم، أو نتيجة خطأ في إدخال البيانات (أوزبورن، 2018/2015).

أنماط البيانات المفقودة

يعرف نمط فقد البيانات أنه الطريقة التي يمكن من خلالها معرفة موقع تلك الفجوات في البيانات، والتي تساعد في اكتشاف الأسباب التي أدت إلى فقد البيانات. وعليه فإنّ البيانات المفقودة تأخذ شكل الأنماط التالية:

(١) **النمط أحادي المتغير (Univariate pattern):** في هذا النمط من البيانات المفقودة يحدث فقد البيانات في متغيرٍ مستقل واحد، فإذا وجد (k) من المتغيرات فإنّ عدد المتغيرات التي تحتوي على بيانات تامة هو (k-1) في حين يكون هناك متغير واحد فقط يحتوي على بيانات غير تامة (حسين، 2012).

(٢) **النمط متعدد المتغيرات (Multivariate two patterns):** في هذا النمط من البيانات المفقودة يوجد أكثر من متغير يحتوي على بيانات مفقودة، بحيث تتساوى عدد حالات الفقد في جميع المتغيرات المستقلة التي تحتوي على بيانات مفقودة (Little & Rubin, 2002).

(٣) **النمط ذو الوتيرة الواحدة (Monotone pattern):** في هذا النمط من البيانات المفقودة يحدث فقد البيانات بشكلٍ مرتّب لبعض المتغيرات المستقلة، بحيث يكون المتغير الذي يحدث به أكبر عددٍ من حالات الفقد هو أول المتغيرات، ثم يأتي المتغير الذي يحتوي الترتيب الثاني لأكبر عددٍ من حالات الفقد، وهكذا بالنسبة لبقية المتغيرات المستقلة (Schafer, 1997).

(٤) **النمط العام (General pattern):** في هذا النمط من البيانات المفقودة ترتّب المتغيرات بحيث يكون المتغير الذي يحتوي على أكبر عددٍ من حالات الفقد هو أول المتغيرات، ثم يأتي المتغير الذي يحتوي الترتيب الثاني لأكبر عدد حالات فقد، وهكذا بالنسبة لبقية المتغيرات، كما يمكن ترتيب المتغيرات بطريقة عكسية (المريعي، 2014).

(٥) **نمط تطابق الملف (File matching):** يحدث هذا النمط من البيانات المفقودة في متغيرين مستقلين فقط، بحيث لا توجد بيانات مفقودة مشتركة بين المتغيرين (Molenberghs et al., 2015).

آليات فقد البيانات

تشير آلية فقد البيانات إلى تلك العلاقة المحتملة بين المتغيرات المقاسة واحتمالية فقد البيانات. وبالرغم من أنّ آليات الفقد لا توفر لنا تفسيراً سببياً لفقد البيانات فإنّها تقدّم لنا علاقة رياضية عامة بين البيانات وعملية فقدها. إن فهم هذه الآليات، وتحديد طبيعتها يساعد في اختيار الطريقة المناسبة التي سيتمّ بها علاج البيانات المفقودة. وعليه فإنّ آليات الفقد هي (Enders 2010):

(١) آلية الفقد بشكل عشوائي تام (Missing Completely At Random) MCAR:

وهي البيانات المفقودة بطريقة لا ترتبط بخصائص البيانات أو خصائص الأفراد، ولا تعتمد على قيم المتغير المستقل نفسها الذي يحتوي على هذه القيم المفقودة، أو أي متغير مستقل آخر في قاعدة البيانات.

٢) آلية الفقد بشكل عشوائي (Missing At Random) MAR:

يقال أنّ البيانات فقدت بشكل عشوائي عندما يكون سبب الفقد له علاقة بقيم المتغيرات الأخرى وغير مرتبط بالقيمة المفقودة.

٣) آلية الفقد بشكل غير عشوائي (Not Missing At Random) NMAR:

يقال أنّ البيانات فقدت بشكل غير عشوائي إذا كانت البيانات المفقودة في متغير مستقل معيّن تعتمد على قيم المتغير نفسها.

طرق معالجة البيانات المفقودة

تشير الأدبيات الإحصائية إلى تنوع طرق التعامل مع البيانات المفقودة، فمنها ما يعتمد على إجراءات الحذف، ومنها ما يعتمد على إجراءات التعويض، وفيما يلي عرض للطرق التي تندرج تحت كل منهما باختصار:

١) الطرق القائمة على الحذف:

توجد طريقتان للحذف هما: حذف الحالة بالكامل (Leastwise Deletion) والحذف المزدوج (Pairwise Deletion)، حيث تستند الطريقة الأولى على حذف كل بيانات الحالة التي لديها قيم مفقودة في متغير واحد على الأقل وتحليل الحالات الكاملة، وهذا الخيار هو الوضع الافتراضي لكثير من برامج التحليل الإحصائي، في حين تستند الطريقة الثانية على حذف الحالة من التحليل عندما توجد قيم مفقودة في المتغير قيد الدراسة، والاحتفاظ بها في حالة البيانات كاملة للمتغير قيد الدراسة (القحطاني، 2015).

٢) الطرق القائمة على التعويض

وهي طرق قائمة على استبدال البيانات المفقودة بالقيمة المقدرة، وقد يكون التعويض أحاديا (التعويض بقيمة واحدة) أو متعددا (التعويض بقيم متعددة). وتتضمن الطرق القائمة على التعويض: طريقة التعويض بالمتوسط الحسابي (Mean Imputation)، وطريقة التعويض بالانحدار (Imputation Using Regression Analysis)، وطريقة التعويض بآخر قيمة ملاحظة (Last Observation Carried Forward)، وطريقة هوت دك (Hot Deck Imputation)، وطريقة كولد دك (Cold Dec Imputation)، وطريقة التعويض المتعدد (Multiple Imputation)، وطريقة خوارزمية تعظيم التوقع (EM Algorithm)، وطريقة المعلومات الكاملة لاحتمال الأعظم (Full Information Maximum Likelihood). وتقوم طريقة التعويض بالمتوسط الحسابي (Mean Imputation) ، على استبدال البيانات المفقودة بمتوسط الدرجات المتوفرة على الفقرة، حيث يتم حساب الوسط لهذه الفقرة عبر جميع الاستجابات للمفحوصين، ويعوّض هذا الوسط جميع البيانات المفقودة على الفقرة، أو عن طريق حساب الوسط الحسابي للمفحوص عبر جميع الفقرات، ويعوّض هذا الوسط جميع القيم المفقودة للمفحوص (بني عواد والمومني، 2018).

أما طريقة التعويض بالانحدار (Imputation Using Regression Analysis) فتتضمن استخدام المتغيرات التي لها بيانات تامة، والتي لها علاقة بالمتغير الذي يحوي قيما مفقودة للتنبؤ بالقيم المفقودة (رزق الله، 2002). في حين أنّ طريقة التعويض بآخر قيمة ملاحظة (Last Observation Carried Forward) تعتمد في تحديدها للقيمة التعويضية للبيانات المفقودة على استبدال كل قيمة مفقودة بآخر قيمة مرصودة في الموضوع نفسه. وكلما اقتقدت القيم يتم الاستبدال بآخر قيمة مرصودة (Molenberghs et al. 2015). وتسهم هذه الطريقة في إبراز افتراض قوي حيث إنّ قيمة النتائج تظل كما هي بلا تغيير بعد فقدان، الأمر الذي يبدو غير واقعي في عديد من الأوضاع. في حين تتضمن طريقة هوت دك

(Hot Deck Imputation) لتقدير القيم المفقودة مجموعة من التقنيات التي تعوّض البيانات المفقودة بدرجاتٍ مماثلةٍ من مستجيبين مماثلين للمستجيبين الناقصة بياناتهم، ويعدّ أكثر التطبيقات الشائعة لتقدير هوت ذلك أن تملأ القيم المفقودة لمستجيبين بقيم عشوائية من مستجيبين آخرين مشابهي لمجموعات المستجيبين الناقصة بياناتهم في الأبعاد والبنود نفسها، وتسهم هذه الطريقة في الحفاظ بشكل عام على التوزيع أحادي المتغير للبيانات، ولا تضعف مستوى تباين أو تباين البيانات التي تم ملؤها. (Little & Rubin، 2002). وتعدّ معالجة البيانات المفقودة عبر استبدالها باستجابة من وحدةٍ مماثلةٍ طريقة شائعة الاستخدام في تطبيقات العلوم الاجتماعية، وهي مناسبة عند التعامل مع البيانات الفئوية. ودائما ما تستخدم هذه الطريقة في ترتيب البيانات، وتهدف إلى تجنب افتراضات التوزيع (الملاحظات). ويعدّ ذلك مهما في حالة ما إذا كان توزيع البيانات ملتويا، أو أنّها أظهرت خصائص محدّدة مثل إسقاط رقم من ضمن الأعداد، وذلك يحدث في كثير من بيانات العلوم الاجتماعية (العتيبي، 2018).

أما طريقة كولد دك ((Cold Dec Imputation فتقوم 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 للاستقصاء ذاته (الحمامي، 2007). في حين تستخدم طريقة التعويض المتعدّد (Multiple Imputation) أكثر من طريقة من الطرق المذكورة ثم الحصول على المتوسط الحسابي لهذه التقديرات المتعددة كقيمة بديلة للبيانات المفقودة (رزق الله، 2002). أما خوارزمية تعظيم التوقع (EM Algorithm) فتعتمد على تعويض البيانات المفقودة من خلال تقديرات يتم الحصول عليها عبر عملية تقدير متكررة تداول بين خطوة التوقع وخطوة التعظيم للاستفادة من جميع المعلومات المتاحة من الحالات التامة وغير التامة (Davey & savla, 2010). في حين تلخص طريقة المعلومات الكاملة لاحتمال الأعظم (Full Information Maximum Likelihood) في الاستفادة من البيانات المتاحة للحالة التي بها فقد في تقدير البيانات المفقودة دون الاعتماد على بيانات الحالات الأخرى (Liou & Ehman، Harwell، Peng، 2006).

ثانيا: قوة الاختبار:

تعبر قوة الاختبار عن قدرته على رفض الفرض الصفري عندما يكون خطأ، وتكون تلك القوة في صورة احتمال تعتمد قيمته على احتمال الوقوع في خطأ النوع الثاني، وتحدّد قوة الاختبار من العلاقة: $P = 1 - \beta$ (الشرييني، 2007). كما تشير قوة الاختبار إلى احتمال أن ينتج الاختبار الإحصائي نتائج دالة إحصائية، وبالرغم من العلاقة الوثيقة بين القوة والدلالة الإحصائية باعتبار أن القوة الإحصائية هي أحد المؤشرات المكتملة لاختبارات الدلالة الإحصائية وبالتالي يصبح من الضروري الاهتمام بالقوة الإحصائية مثل الاهتمام بالدلالة الإحصائية (Galarza-Hernanddez، 1993)، كما ورد في الضوي، (2006).

استخدامات قوة الاختبار

توجد عديد الاستخدامات لقوة الاختبار الإحصائي، منها:

أولا: تحليل القوة القبلي (Priori power analysis):

يستخدم تحليل القوة القبلي في مرحلة تخطيط دراسة بحثية معينة بهدف تقدير حجم العينة المناسب لاستخدامه في الدراسة، للوصول إلى المستوى المقبول من القوة الإحصائية مع الأخذ في الاعتبار مستوى الدلالة الإحصائي α وحجم التأثير في المجتمع المراد الكشف عنه (هويت، كرم، 2016).

ثانيا: تحليل القوة البعدي (Post hoc power analysis):

يستخدم تحليل القوة البعدي لتقييم مدى قدرة الاختبار الإحصائي المستخدم في رفض الفرضية الصفرية الخطأ بعد الانتهاء من الدراسة، ففي الحالات التي لا يتم فيها رفض الفرض الصفرية قد يكون ذلك بسبب عدم وجود تأثير فعلي، أو أنّ التأثير موجود، ولكن لضعف قوة الاختبار يتعدّد الكشف عن هذا التأثير (عبد الفتاح، 2011؛ المالكي، 2018).

العوامل المؤثرة على قوة الاختبار:

تعتمد قوة الاختبار الإحصائي على عددٍ من العوامل هي:

(١) حجم التأثير:

كلّما زاد حجم التأثير؛ ازدادت قوة الاختبار الإحصائي؛ وذلك لأنّ زيادة مستويات حجم التأثير تعني زيادة وضوح الفروق بين متوسطات مجموعات المقارنة، وكلّما زاد هذا الفرق زادت احتمالية رفض الفرضية الصفرية، أي زيادة قوة الاختبار الإحصائي؛ لأنّ قوة الاختبار ماهي إلا احتمال رفض الفرضية الصفرية، عندما يجب أن ترفض (المالكي، 2020).

(٢) حجم العينة:

تزداد قوة الاختبار الإحصائي بازدياد حجم العينة مع ثبات العوامل الأخرى المؤثرة على قوة الاختبار الإحصائي، لكنّ زيادة حجم العينة بشكل كبير قد تؤدي إلى أنّ تصبح الفروقات مهما كانت بسيطة أو تافهة؛ دالة إحصائيا بالرغم من عدم دلالتها العملية (الصياد، 1988).

مستوى الدلالة:

تنخفض قوة الاختبار الإحصائي عندما تكون مستويات الدلالة أكثر تحفظا وانخفاضاً، وترتفع قوة الاختبار الإحصائي بزيادة مستويات الدلالة، أي أنّ الاختبار الإحصائي عند مستوى دلالة إحصائية (0.05) يعد أقوى كفاءة من الاختبار عند مستوى دلالة إحصائية (0.01) وذلك بافتراض ثبات العوامل الأخرى (الشوربجي، 2012).

(٣) طبيعة الاختبار الإحصائي المستخدم لاختبار صحة الفرضيات:

تعدّ الاختبارات البارامترية أقوى كفاءة من الاختبارات اللابارامترية، فالخطأ من النوع الثاني أقل حدوثاً عند استخدام الاختبارات البارامترية في حالة توافر الافتراضات الخاصة بها، أمّا في حالة انتهاك الافتراضات الأساسية للاختبارات البارامترية، تكون الاختبارات اللابارامترية أقوى كفاءة (النجار، 2006).

(٤) مقدار التباين (σ^2):

تناسب قوة الاختبار الإحصائي تناسباً عكسياً مع مقدار التباين، حيث تقلّ قوة الاختبار الإحصائي بارتفاع مقدار التباين، وتزداد بانخفاض مقدار التباين، وهذا يعني أنّ قوة الاختبار الإحصائي تقلّ بانخفاض معامل التباين (بابطين، 2002).

(٥) طبيعة الفرض البديل:

عندما يكون الفرض الصفرى زائفاً في حقيقة الأمر، فإنّ قوة الاختبار تزداد في حالة الفرض الموجه عنه بالنسبة لحالة الفرض غير الموجه، أي أنّ الاختبار بديل واحد أقوى من الاختبار بذيلين عند ثبوت قيمة (α) (شينار و بولجبال، 2021).

(٦) عوامل أخرى:

هناك عوامل أخرى تؤثر على قوة الاختبار الإحصائي تختص بتصميم البحث، حيث حدّد Baguley (2004؛ Sink & Orlikowsky، 2007؛ Mvududu، 1989؛ 2010، كما ورد في عامر، 2014)

عوامل تصميم البحث المؤثرة في قوة الاختبار ومنها: اختيار المتغيرات، فاختبار أهم المتغيرات المستقلة المؤثرة في الظاهرة يزيد من قوة الاختبار الإحصائي، وكذلك الثبات، فالثبات المرتفع للقياسات يؤدي إلى قوة إحصائية عالية. كما أن توزيع مفردات العينة بشكل متساوٍ على المجموعات المتعددة، أو مستويات العامل يزيد من قوة الاختبار، في حين أن التوزيع غير المتساوي على مجموعات التصميم العاملي يؤدي إلى تقليص الارتباطات أو الفروق، ومن ثم تقليص حجم التأثير وهذا بدوره يؤدي إلى قوة منخفضة. كما أن إجراء معالجات تجريبية ذات قياسات متكررة يؤدي إلى قوة عالية.

الدراسات والبحوث السابقة

من خلال اطلاع الباحث على الدراسات السابقة ذات الصلة بموضوع طرق التعامل مع البيانات المفقودة، تبين للباحث وجود عددٍ 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 قدمت 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 (Bokossa, 2000) و (Huang & Cohen, 2000) التي هدفت إلى المقارنة بين دقة طريقتي هوت دك (Hot Deck Imputation)، وطريقة النموذج المستند إلى التعويض العشوائي (Model-based random imputation method)، حيث استخدموا لهذه الغاية بيانات عينة عشوائية طبقية تكونت من (8252) طالبا من طلبة المرحلة الأساسية العليا والثانوية الذين تقدموا لاختبارات وطنية في القراءة والحساب في مدينة أورلنجتون بولاية فرجينيا، حيث تم استخدام النموذج ثلاثي المعالم لتقدير قدرات المفحوصين، وقد تم حذف ما نسبته (20%) من البيانات، والتعويض عنها بالطريقتين السابقتين، ومن ثم استخدموا البيانات الناتجة في تقدير قدرة المفحوصين، واعتمدوا في الحكم على دقة تقدير القدرة على ثلاثة معايير هي: معدل خطأ التعويض، وتحيز التباين، ومتوسط التحيز. توصلت نتائج الدراسة إلى أن طريقة النموذج المستند إلى التعويض العشوائي أكثر دقة في تقدير القيم المفقودة من طريقة هوت دك.

– دراسة (Langkamp, 2010) و (Lehman & Lemeshow, 2010) التي هدفت إلى المقارنة بين أربع طرق لمعالجة البيانات المفقودة: طريقة إعادة الوزن (Reweighting)، وطريقة التعويض المتعدد (Multiple imputation)، وطريقة حذف الحالة (Case-Deletion)، وطريقة هوت دك (Hot Deck Imputation) في دقة تقديرها لمعاملات نموذج الانحدار، وذلك في ضوء نسب فقد مختلفة (10%، 20%، 30%، 40%)، توصلت نتائج الدراسة إلى كفاءة طريقتي التعويض المتعدد (Multiple imputation)، وطريقة إعادة الوزن (Reweighting) مقارنة بطريقتي حذف الحالة (Case-Deletion)، وطريقة هوت دك (Hot Deck Imputation)، وذلك لنسب الفقد التي تزيد عن (10%). وأوصت الدراسة باستخدام طريقة إعادة الوزن أو طريقة التعويض المتعدد في حالة نسب الفقد الكبيرة.

– دراسة (Nakai, 2011) بعنوان دراسة محاكاة: مقدمة بين طرق التعويض عن البيانات المفقودة بالتحليل الطولي هدفت هذه الدراسة المقارنة بين أربع طرق مختلفة للتعويض عن البيانات المفقودة. حيث تم استخدام حجم عينة مكون من (200)، وكانت نسب الفقد المستخدمة 0%، 10%، 20%، 30%، و 50% بآلية فقد للبيانات من النوع MCAR. وكانت الطرق الأربعة هي: التعويض المتعدد (Multiple imputation)، والتعويض بالمتوسط (Mean Imputation)، والتعويض بالمشاهدة الأخيرة (Last Observation Carried Forward)، وطريقة الحذف

(Case-Deletion). وتوصلت الدراسة إلى أنّ طريقة الحذف تعدّ طريقة جيدة إذا كانت نسبة الفقد بالبيانات قليلة وحجم العينة كبير. ورشح الباحث استخدام طريقة التعويض بالمتوسط عندما يتمّ فقد جزء من البيانات تتشابه مع بعضها بنفس ظروف الفقد. وتستخدم طريقة التعويض بالمشاهدة الأخيرة إذا كانت قيم البيانات صغيرة. ولكن تعاب طريقة التعويض المتعدد في أنّها تبعد عن افتراضات اتباع البيانات للتوزيع الطبيعي.

– دراسة (Zhu 2014) بعنوان: مقارنة بين أربعة طرق للتعامل مع البيانات المفقودة في تحليل البيانات الطولية باستخدام أسلوب المحاكاة. هدفت هذه الدراسة إلى تقييم أربعة طرق لمعالجة البيانات المفقودة وهي: طريقة الحذف (Case-Deletion)، والتعويض بالمتوسط (Mean Imputation)، والتعويض بالمشاهدة الأخيرة (Last Observation Carried Forward)، والتعويض المتعدد (Multiple imputation) وكان حجم العينة المستخدم بالدراسة (100). وتمّ استخدام ثلاث آليات مختلفة للفقد MCAR، MAR، NMAR. تمّ استخدام نسب فقد مختلفة (5%، 10%، 15%، 20%، 30%). وتمّ استخدام معياري: التحيز، ومتوسط مربعات الأخطاء للمفاضلة بين الطرق الأربعة. وتوضح من النتائج أنّ التعويض بالمشاهدة الأخيرة تحتوي على أعلى قيمة للتحيز وذلك تحت نمطي الفقد MCAR و MAR. ووجد الباحث أنّ طريقة الحذف وطريقة التعويض المتعدد تحت آلية الفقد MCAR لهما أقل متوسط مربعات أخطاء

– دراسة عيسوي (2015) بعنوان: معالجة البيانات المفقودة لبعض التوزيعات المتتوية. هدفت هذه الدراسة إلى معالجة البيانات المفقودة لحالة التوزيعات المتتوية، وتمّ استخدام حالة توزيع جاما، والتوزيع الهندسي بوصفها حالات خاصة من التوزيعات المتتوية بالدراسة، وذلك من خلال توليد بيانات عشوائية باستخدام نظام المحاكاة؛ للمقارنة بين طريقة تقليدية وأربع طرق حديثة مختلفة لمعالجة البيانات المفقودة بهذه الدراسة وهي: التعويض بالمشاهدة الأخيرة (Last Observation Carried Forward)، وخوارزمية تعظيم التوقع (EM The Expectation Maximization Algorithm) وطريقة تحليل العناصر الرئيسية للمتغير البيزي Variational Bayesian Principle Component Analysis (VBPCA) وطريقة خريطة التنظيم الذاتي (Self-Organizing MAP (SOM) وطريقة مختلطة مكونة من طريقتي: تحليل العناصر الرئيسية للمتغير البيزي، وخريطة التنظيم الذاتي. واستخدم الباحث ثلاثة طرق مختلفة لتقييم أداء طرق المعالجات وهي: الجذر التربيعي لمتوسط مربعات الأخطاء (RMSE) ومتوسط الخطأ المطلق (MAE) ومتوسط الخطأ النسبي (MMRE). كما استخدم الباحث بالدراسة نمط الفقد (MCAR) ونسب الفقد المختلفة الآتية (5%، 10%، 15%، 25%، 40%)، وأوضح النتائج أنّ الطريقة التقليدية المستخدمة بالدراسة (التعويض بالمشاهدة الأخيرة) هي أكثر فاعلية من الطرق الأربعة الحديثة الأخرى؛ وذلك لجميع نسب الفقد المختلفة. ففي حالة أحجام العينات الكبيرة لتوزيع جاما تعدّ طريقة التعويض بالمشاهدة الأخيرة لها أقل (MAE) و (MMRE) كما أنّها تأخذ الترتيب الثاني بالنسبة لمعيار RMSE وذلك بعد خوارزمية تعظيم التوقع (EM). في حين أنّ خوارزمية تعظيم التوقع (EM) لها أكبر قيم وذلك لجميع المعايير المختلفة خلاف معيار RMSE. أمّا طرق المعالجات الثلاثة الأخرى (SOM-VBPCA، SOM، VBPCA) فقد احتوت على نتائج قريبة جدا من البيانات الأصلية، حيث كانت قيم معياري (MAE) و (MMRE) للطرق الثلاثة تقع بين طريقتي التعويض بالمشاهدة الأخيرة وخوارزمية (EM). بينما احتوت هذه الطرق الثلاثة على قيم أعلى من طريقتي التعويض بالمشاهدة الأخيرة وخوارزمية تعظيم التوقع (EM) وذلك بالنسبة لمعيار (RMSE). كما يلاحظ أنّ طرق المعالجات (VBPCA، SOM and SOM-VBPCA) أعطت نتائج مقاربة جدا فيما بينها، وذلك لجميع نسب الفقد المستخدمة بالدراسة؛ ممّا يعني أنّ استخدام طريقة المزج

(SOM-VBPCA) بين طريقتي (SOM & VBPCA) غير فعال. كما يلاحظ أنه في حالة ارتفاع نسبة الفقد بالبيانات فإن جميع طرق المعالجات المستخدمة تفقد دقتها بالنسبة لمعيار (RMSE) في حين أوضحت الطرق الأربعة الحديثة عدم الاستقرار بالنسبة لمعاري (MAE) و (MMR). وأوضح اختبار t للعينات المرتبطة أنه لا يوجد فرق معنوي بين طريقة التعويض بالمشاهدة الأخيرة والبيانات الأصلية، لكن ظهرت اختلافات معنوية بين الطرق الحديثة الأربعة والبيانات الأصلية عند مستوى الفقد بالبيانات (5%).

– دراسة Vink & Schoot (2015)، Zondervan-Zwijnenburg، Peeters بعنوان: كيفية التعامل مع البيانات المفقودة: مقارنة بين بعض الطرق. هدفت الدراسة إيجاد الطريقة المثلى للتعامل مع البيانات المفقودة. ولهذا تم تقييم طرق مختلفة للتعامل مع البيانات المفقودة. وتم استخدام المحاكاة لتحقيق هدف الدراسة. وتم استخدام نسب الفقد (2%، 7.8%، 9.4%) لحجم العينة (374)، والنسب (25%، 29.9%) لحجم العينة (279)، والنسب (39%، 39.9%) للحجم (230)، والنسب (48%، 52%) للحجم (196) تحت آلية فقد (MAR). وكانت طرق معالجة البيانات هي: طريقة الحذف (Case-Deletion)، وطريقة التعويض بالمشاهدة الأخيرة (Last Observation Carried Forward)، وطريقة التعويض بالمتوسط الشرطي (Conditional Mean (CM) وطريقة التعويض المتعدد (Multiple imputation) باستخدام الانحدار اللوجستي، وطريقة (EM). وأوضحت نتائج الدراسة أن التعويض المتعدد هو الطريقة المثلى للتعامل مع البيانات المفقودة؛ وذلك لأنه يحقق أقل تحيز وأصغر فترة ثقة. أما طريقة الحذف وطريقة التعويض بالمشاهدة الأخيرة فأعطيتا تحيزاً أكبر، وفترة ثقة أكبر من طريقة التعويض المتعدد.

التعليق على الدراسات السابقة

في ضوء ما تقدم من دراسات سابقة حول طرق معالجة البيانات المفقودة، تبين أن بعض تلك الدراسات تم إجراؤها باستخدام بيانات افتراضية تم توليدها باستخدام البرامج الإحصائية المختلفة، والبعض الآخر تم إجراؤها باستخدام بيانات حقيقية. كما أن معظم هذه الدراسات تناولت المقارنة بين طرق معالجة البيانات المفقودة في دقة التقدير في ضوء المعايير: التحيز في تقدير المعالم، والجذر التربيعي لمتوسط مربعات الأخطاء، ومتوسط الخطأ المطلق، ومتوسط الخطأ النسبي. ولا توجد دراسات في حدود علم الباحث اعتمدت على قوة الاختبار الإحصائي بوصفه معياراً للمقارنة بين دقة طرائق معالجة البيانات المفقودة. وكذلك لا توجد دراسات عربية أو أجنبية جمعت بين طريقتي: التعويض بآخر قيمة ملاحظة، وهوت دوك باعتبارهما طريقتين تعويضيتين بديلتين لتقليل تأثير القيم المفقودة على قوة الاختبار الإحصائي؛ لذلك جاءت هذه الدراسة لتوضيح تأثير البيانات المفقودة على قوة الاختبار الإحصائي، في ضوء نسب فقد، وأحجام عينات مختلفة. وكذلك تبين كفاءة طريقتي التعويض المستخدمتين في تقدير البيانات المفقودة.

منهجية الدراسة وإجراءاتها

منهج الدراسة:

بناء على طبيعة مشكلة الدراسة، وفي ضوء الدراسات السابقة، حدّد الباحث المنهج الملائم للدراسة الحالية وهو المنهج الوصفي (المقارن)، وتمّ استخدام هذا المنهج لتحقيق أهداف الدراسة المتعلقة بتحديد كفاءة طرق معالجة البيانات المفقودة المستخدمة في تفسير قوة الاختبار الإحصائي، مع محاولة الوصول إلى أفضل الطرق لتقدير البيانات المفقودة في ضوء تعيّر حجم العينة، وفقدان جزء من البيانات بنسب مختلفة.

مجتمع الدراسة وعينتها:

مجتمع الدراسة أو المجتمع الإحصائي، هو جميع الأفراد أو الحالات التي تخضع لحالة البحث. ونجد هنا أنّ المجتمع الإحصائي الخاص بالدراسة الحالية هو جميع البيانات الإحصائية التي تمّ توليدها باستخدام برنامج (R) إصدار (3.4.4). وتتوافر فيها الافتراضات الخاصة بأسلوب تحليل التباين الأحادي.

إجراءات الدراسة:

تمّ استخدام المحاكاة لتوليد أرقام عشوائية بأحجام عينات مختلفة (30، 60، 90)، حيث تمّ بناء نموذج محاكاة لأسلوب تحليل التباين الأحادي (في حالة تساوي أحجام العينات بين المجموعات، وفي حالة عدم تساوي أحجام العينات بين المجموعات) باستخدام برنامج (R) إصدار (3.4.4). بشرط توافر جميع الافتراضات اللازمة لاستخدام أسلوب تحليل التباين الأحادي، سواء فيما يتعلق بالاختيار العشوائي، أم بمستوى القياس الفئوي، أم بالتوزيع الاعتدالي، أم بتجانس التباين. وتمّ حساب القوة الإحصائية لاختبار تحليل التباين الأحادي للحالات المذكورة، وذلك في حالة البيانات التامة باستخدام برنامج حساب القوة الإحصائية (G*Power 3.0.10). ثمّ تم عمل فقد لهذه البيانات بنسب فقد (15%، 25%، 40%) على التوالي وفق آلية الفقد (MCAR)، وتحت نمط فقد (النمط العام). بعد ذلك تمّ تقدير القيم المفقودة باستخدام طريقتي: التعويض بآخر قيمة ملاحظة (Last Observation Carried Forward)، وطريقة هوت دك (Hot Deck) (Imputation). ثمّ إجراء اختبار تحليل التباين الأحادي (في حالة تساوي أحجام العينات بين المجموعات، وفي حالة عدم تساوي أحجام العينات بين المجموعات) وحساب القوة الإحصائية لاختبار تحليل التباين الأحادي للحالات المذكورة في حالة التعويض عن البيانات المفقودة؛ وذلك للمفاضلة بين طريقتي التعويض بآخر قيمة ملاحظة وهوت دك لتقدير البيانات المفقودة، وذلك من خلال تعرّف دقّتها في تقدير قوة الاختبار الإحصائي. وتمّ تطبيق هذا البرنامج على السيناريوهات المختلفة المذكورة سابقاً، وأخيراً تمّ تكرار هذا البرنامج (1000) مره لكل سيناريو، وتمّ أخذ المتوسطات الحسابية لاعتمادها نتيجة نهائية لإجراء المقارنة والتحليل بين النماذج المختلفة.

عرض النتائج ومناقشتها

نتائج السؤال الأول:

ما كفاءة طريقتي التعويض بآخر قيمة ملاحظة وهوت دك لتقدير القيم المفقودة في ضوء معيار قوة اختبار تحليل التباين الأحادي في حالة تساوي أحجام العينات بين المجموعات؟

• أولاً: في حالة حجم العينة (30):

جدول ١

قوة اختبار تحليل التباين في حالة تساوي أحجام العينات بين المجموعات وحجم العينة 30

التمودج	نسبة الفقد	alpha	Beta	Mean1	Mean2	Mean3	قوة التمودج
FULL DATA	-	0.05	0.2419	1	-2	2	0.7581
LOCF	M= 0.15	0.05	0.2870	1	-2	2	0.7130
	M= 0.25	0.05	0.3203	1	-2	2	0.6797
	M= 0.4	0.05	0.3831	1	-2	2	0.6169
HOT DEC	M= 0.15	0.05	0.4948	1	-2	2	0.5052
	M= 0.25	0.05	0.6400	1	-2	2	0.3600
	M= 0.4	0.05	0.7893	1	-2	2	0.2107

من الجدول السابق يتضح تفوق طريقة التعويض بالملاحظة الأخيرة LOCF عن طريقة الهوت دك HOT DEC

بجميع نسب الفقد المختلفة للبيانات التامة كما يلي:

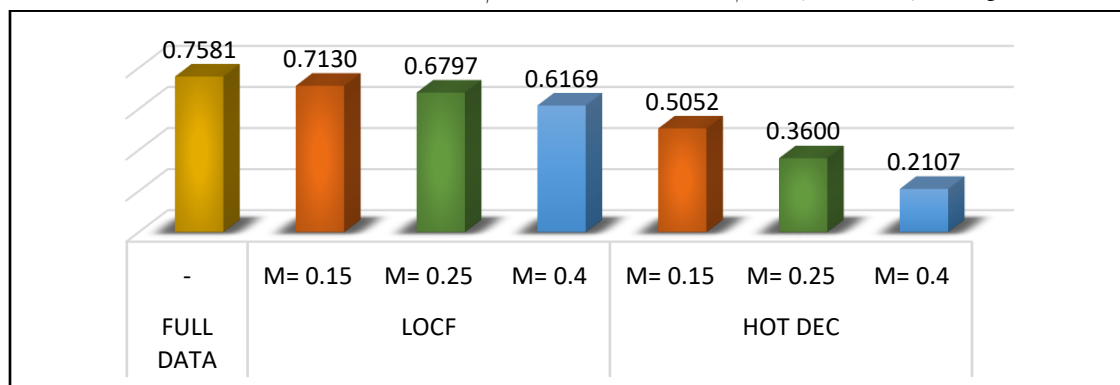
(١) عند نسبة الفقد بالبيانات (15%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (71.3%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (50.5%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (75.8%)

(٢) عند نسبة الفقد بالبيانات (25%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (68%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (36%)، وبالتالي تعدّ طريقة LOCF هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (75.8%).

(٣) عند نسبة الفقد بالبيانات (40%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (61.7%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (21.1%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب للبيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (75.8%)، وذلك كما هو موضح بالشكل التالي:

شكل ١

قوة اختبار تحليل التباين في حالة تساوي أحجام العينات بين المجموعات وحجم العينة 30



• ثانياً: في حالة حجم العينة 60:

جدول ٢

قوة اختبار تحليل التباين في حالة تساوي أحجام العينات بين المجموعات وحجم العينة 60

التمودج	نسبة الفقد	alpha	beta	Mean1	Mean2	Mean3	قوة التمودج
FULL DATA	-	0.05	0.1579	1	-2	2	0.8421
LOCF	M= 0.15	0.05	0.1857	1	-2	2	0.8143
	M= 0.25	0.05	0.2019	1	-2	2	0.7981
	M= 0.4	0.05	0.2394	1	-2	2	0.7606
HOT DEC	M= 0.15	0.05	0.4240	1	-2	2	0.5760
	M= 0.25	0.05	0.5756	1	-2	2	0.4244
	M= 0.4	0.05	0.7636	1	-2	2	0.2364

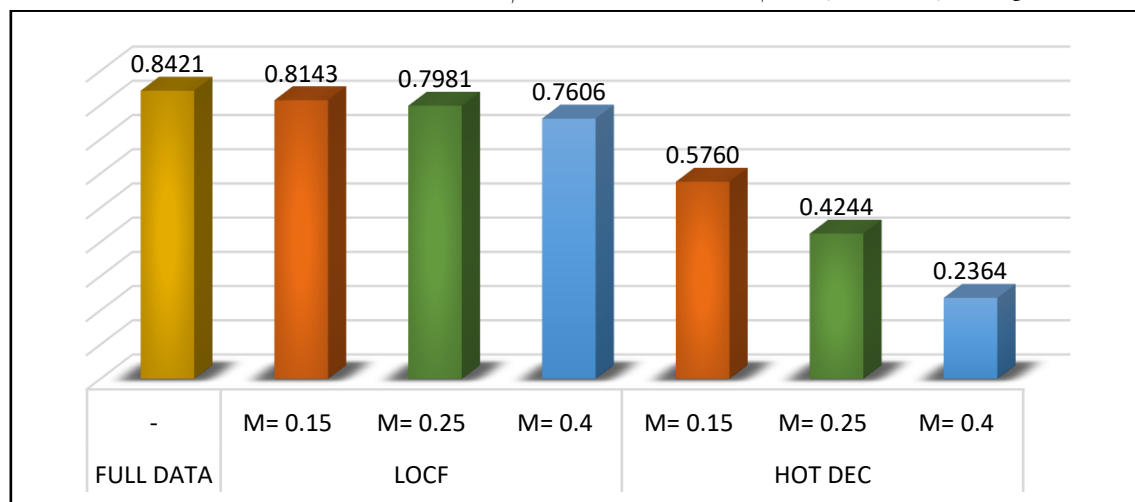
من الجدول السابق يتضح تفوق طريقة التعويض بالملاحظة الأخيرة LOCF عن طريقة الهوت دك HOT DEC بجميع

نسب الفقد المختلفة للبيانات التامة كما يلي:

- (١) عند نسبة الفقد بالبيانات (15%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (81.4%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (57.6%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها 84.2%
- (٢) عند نسبة الفقد بالبيانات (25%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (79.8%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (42.4%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (84.2%)
- (٣) عند نسبة الفقد بالبيانات (40%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (76.1%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (23.6%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (84.2%)، وذلك كما هو موضّح بالشكل التالي:

شكل ٢

قوة اختبار تحليل التباين في حالة تساوي أحجام العينات بين المجموعات وحجم العينة 60



● ثالثاً: في حالة حجم العينة 90:

جدول ٣

قوة اختبار تحليل التباين في حالة تساوي أحجام العينات بين المجموعات وحجم العينة 90

التمودج	نسبة الفقد	alpha	beta	Mean1	Mean2	Mean3	قوة التمودج
FULL DATA	-	0.05	0.0855	1	-2	2	0.9145
LOCF	M= 0.15	0.05	0.1046	1	-2	2	0.8954
	M= 0.25	0.05	0.1136	1	-2	2	0.8864
	M= 0.4	0.05	0.1497	1	-2	2	0.8503
HOT DEC	M= 0.15	0.05	0.3262	1	-2	2	0.6738
	M= 0.25	0.05	0.5047	1	-2	2	0.4953
	M= 0.4	0.05	0.7328	1	-2	2	0.2672

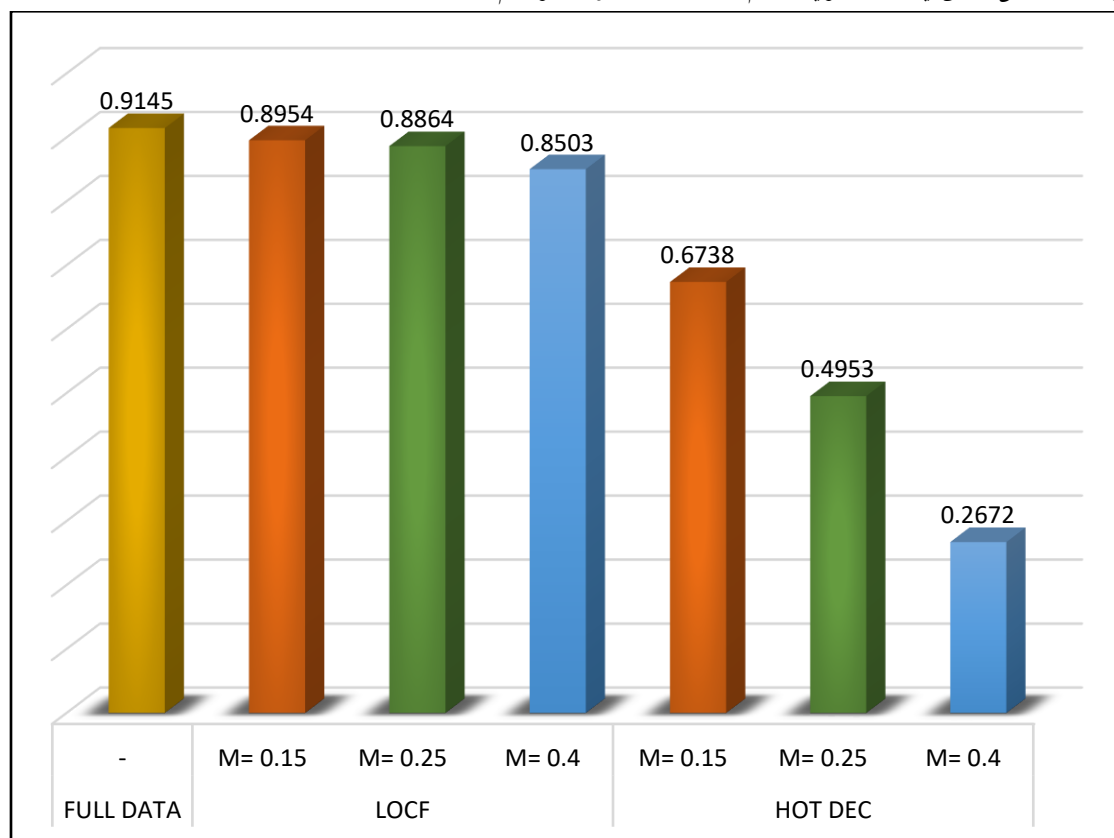
من الجدول السابق يتضح تفوق طريقة التعويض بالملاحظة الأخيرة LOCF عن طريقة الهوت دك HOT DEC بجميع نسب الفقد المختلفة للبيانات التامة كما يلي:

- (١) عند نسبة الفقد بالبيانات (15%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالمشاهدة الأخيرة (LOCF) (89.5%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (67.4%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (91.4%)
- (٢) عند نسبة الفقد بالبيانات (25%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالمشاهدة الأخيرة (LOCF) (88.6%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (49.5%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (91.4%)
- (٣) عند نسبة الفقد بالبيانات (40%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالمشاهدة الأخيرة (LOCF) (85%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (26.7%)، وبالتالي تعدّ طريقة LOCF

هي الأقرب لنموذج البيانات التامة من حيث قوة النموذج الإحصائي وقد بلغت قيمتها (91.4%)، وذلك كما هو موضح بالشكل التالي:

شكل ٣

قوة اختبار تحليل التباين في حالة تساوي أحجام العينات بين المجموعات وحجم العينة 90



نتائج السؤال الثاني:

ما كفاءة طريقتي التعويض بآخر قيمة ملاحظة وهوت دك لتقدير القيم المفقودة في ضوء معيار قوة اختبار تحليل التباين الأحادي في حالة عدم تساوي أحجام العينات بين المجموعات؟

● أولاً: في حالة حجم العينة (30):

جدول ٤

قوة اختبار تحليل التباين في حالة عدم تساوي أحجام العينات بين المجموعات وحجم العينة 30

قوة النموذج	Mean3	Mean2	Mean1	beta	alpha	نسبة الفقد	النموذج
0.5336	3.3	2.2	0.8	0.4664	0.05	-	FULL DATA
0.5172	3.3	2.2	0.8	0.4828	0.05	M= 0.15	LOCF
0.5006	3.3	2.2	0.8	0.4994	0.05	M= 0.25	
0.4600	3.3	2.2	0.8	0.5400	0.05	M= 0.4	
0.3670	3.3	2.2	0.8	0.6330	0.05	M= 0.15	HOT DEC
0.2735	3.3	2.2	0.8	0.7265	0.05	M= 0.25	

0.1686	3.3	2.2	0.8	0.8314	0.05	M= 0.4	HOT DEC
--------	-----	-----	-----	--------	------	--------	---------

من الجدول السابق يتضح تفوق طريقة التعويض بالملاحظة الأخيرة LOCF عن طريقة الهوت دك HOT DEC بجميع نسب الفقد المختلفة للبيانات التامة كما يلي:

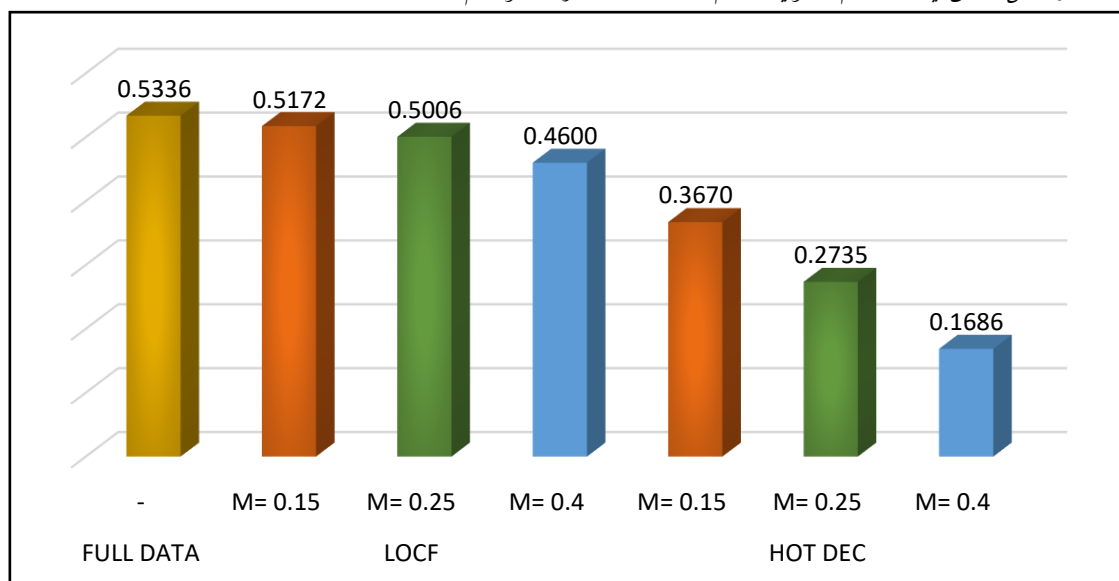
(١) عند نسبة الفقد بالبيانات (15%) بلغت قوة النموذج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) 51.7%، في حين كانت قوة النموذج الإحصائي بطريقة الهوت دك (HOT DEC) 36.7%، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة النموذج الإحصائي وقد بلغت قيمتها (53.4%)

(٢) عند نسبة الفقد بالبيانات (25%) بلغت قوة النموذج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (50.1%)، في حين كانت قوة النموذج الإحصائي بطريقة الهوت دك (HOT DEC) (27.4%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة النموذج الإحصائي وقد بلغت قيمتها (53.4%)

(٣) عند نسبة الفقد بالبيانات (40%) بلغت قوة النموذج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) 46%، في حين كانت قوة النموذج الإحصائي بطريقة الهوت دك (HOT DEC) 36.9%، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة النموذج الإحصائي وقد بلغت قيمتها (53.4%)، وذلك كما هو موضح بالشكل التالي:

شكل ٤

قوة اختبار تحليل التباين في حالة عدم تساوي أحجام العينات بين المجموعات وحجم العينة 30



● ثانياً: في حالة حجم العينة 60:

جدول ٥

قوة اختبار تحليل التباين في حالة عدم تساوي أحجام العينات بين المجموعات وحجم العينة 60

التّموذج	نسبة الفقد	alpha	beta	Mean1	Mean2	Mean3	قوة التّموذج
FULL DATA	-	0.05	0.0987	0.8	2.2	3.3	0.9013
	M= 0.15	0.05	0.1097	0.8	2.2	3.3	0.8903
	M= 0.25	0.05	0.1230	0.8	2.2	3.3	0.8770

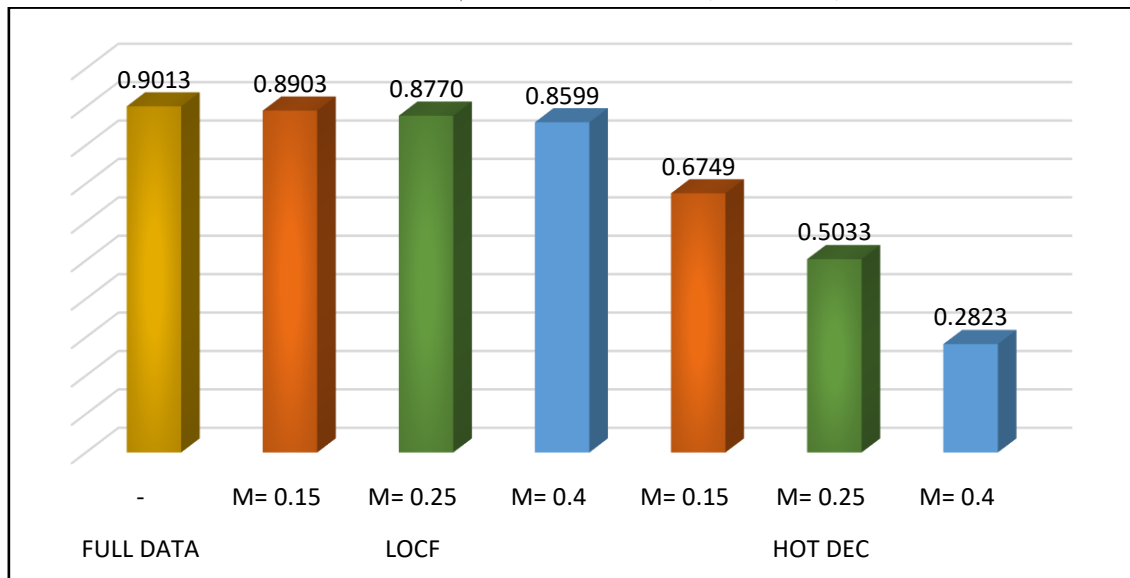
0.8599	3.3	2.2	0.8	0.1401	0.05	M= 0.4	LOCF
0.6749	3.3	2.2	0.8	0.3251	0.05	M= 0.15	
0.5033	3.3	2.2	0.8	0.4967	0.05	M= 0.25	
0.2823	3.3	2.2	0.8	0.7177	0.05	M= 0.4	HOT DEC

من الجدول السابق يتضح تفوق طريقة التعويض بالملاحظة الأخيرة LOCF عن طريقة الهوت دك HOT DEC بجميع نسب الفقد المختلفة للبيانات التامة كما يلي:

- (١) عند نسبة الفقد بالبيانات (15%) بلغت قوة التّموذج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (89%)، في حين كانت قوة التّموذج الإحصائي بطريقة الهوت دك (HOT DEC) (67.5%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التّموذج الإحصائي وقد بلغت قيمتها (90.1%)
- (٢) عند نسبة الفقد بالبيانات (25%) بلغت قوة التّموذج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (87.7%)، في حين كانت قوة التّموذج الإحصائي بطريقة الهوت دك (HOT DEC) (50.3%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التّموذج الإحصائي وقد بلغت قيمتها (90.1%)
- (٣) عند نسبة الفقد بالبيانات (40%) بلغت قوة التّموذج الإحصائي في طريقة التعويض بالملاحظة الأخيرة (LOCF) (86%)، في حين كانت قوة التّموذج الإحصائي بطريقة الهوت دك (HOT DEC) (28.2%)، وبالتالي تعدّ طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التّموذج الإحصائي وقد بلغت قيمتها 90.1%، وذلك كما هو موضّح بالشكل التالي:

شكله

قوة اختبار تحليل التباين في حالة عدم تساوي أحجام العينات بين المجموعات وحجم العينة 60



●ثالثاً: في حالة حجم العينة (90):

جدول ٦

قوة اختبار تحليل التباين في حالة عدم تساوي أحجام العينات بين المجموعات وحجم العينة 90

التمودج	نسبة الفقد	alpha	beta	Mean1	Mean2	Mean3	قوة التمودج
FULL DATA	-	0.05	0.0461	0.8	2.2	3.3	0.9539
LOCF	M= 0.15	0.05	0.0555	0.8	2.2	3.3	0.9445
	M= 0.25	0.05	0.0604	0.8	2.2	3.3	0.9396
	M= 0.4	0.05	0.0793	0.8	2.2	3.3	0.9207
HOT DEC	M= 0.15	0.05	0.2212	0.8	2.2	3.3	0.7788
	M= 0.25	0.05	0.3895	0.8	2.2	3.3	0.6105
	M= 0.4	0.05	0.6492	0.8	2.2	3.3	0.3508

من الجدول السابق يتضح تفوق طريقة التعويض بالملاحظة الأخيرة (LOCF) عن طريقة الهوت دك HOT DEC بجميع نسب الفقد المختلفة للبيانات التامة كما يلي:

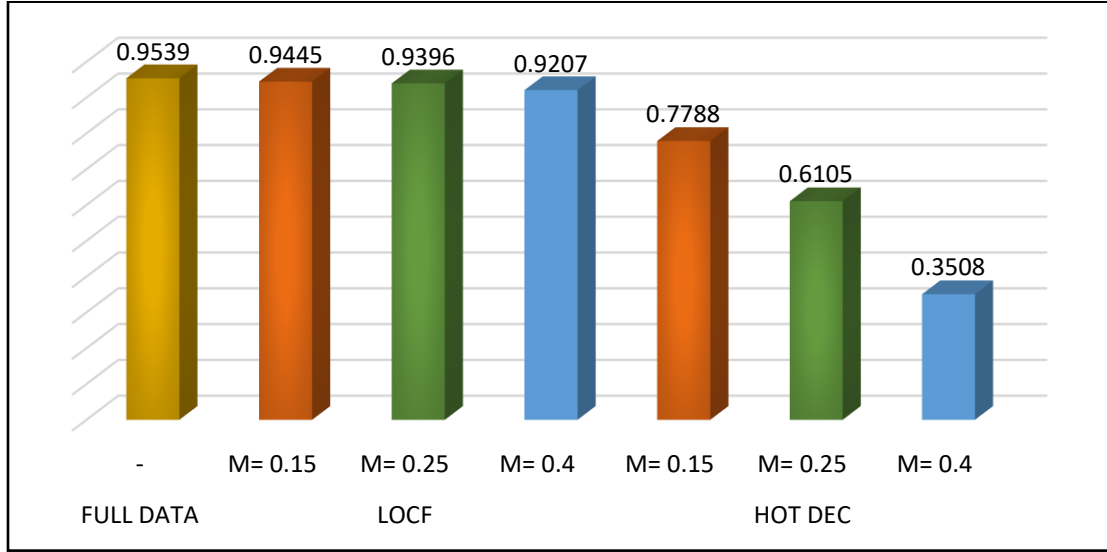
(١) عند نسبة الفقد بالبيانات (15%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالمشاهدة الأخيرة (LOCF) (94.5%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (77.9%)، وبالتالي تعد طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (95.4%)

(٢) عند نسبة الفقد بالبيانات (25%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالمشاهدة الأخيرة (LOCF) (94%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (61.1%)، وبالتالي تعد طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (95.4%)

(٣) عند نسبة الفقد بالبيانات (40%) بلغت قوة التمودج الإحصائي في طريقة التعويض بالمشاهدة الأخيرة (LOCF) (92.1%)، في حين كانت قوة التمودج الإحصائي بطريقة الهوت دك (HOT DEC) (35.1%)، وبالتالي تعد طريقة (LOCF) هي الأقرب لنموذج البيانات التامة من حيث قوة التمودج الإحصائي وقد بلغت قيمتها (95.4%) وذلك كما هو موضح بالشكل التالي:

شكل ٦

قوة اختبار تحليل التباين في حالة عدم تساوي أحجام العينات بين المجموعات وحجم العينة 90



التوصيات:

- في ضوء ما توصلت إليه نتائج الدراسة الحالية فإنها توصي بما يلي:
 - لا توجد طريقة مثالية لمعالجة البيانات المفقودة تؤدي إلى الحصول على نتائج دقيقة ماثلة لتلك النتائج فيما لو كانت البيانات تامة؛ لذا على الباحث أن يحرص على جمع بيانات تامة قدر الإمكان.
 - على الباحث فحص نموذج البيانات المفقودة ومعرفة نمط فقد البيانات وآلياتها حتى يتمكن من اختيار أسلوب التعويض المناسب.
 - النموذج المقدر بطريقة التعويض بالمشاهدة الأخيرة (LOCF) أكفأ من النموذج المقدر بطريقة الهوت دك (HOT DEC) بجميع حالات نسب الفقد وأحجام العينات المختلفة، إذ تعطي طريقة (LOCF) قيمة تقديرية تؤدي إلى نتائج أقرب إلى النموذج التام وفقا لمعيار قوة الاختبار الإحصائي.

المقترحات

- إجراء دراسات تقارن بين كفاءة طرق أخرى لمعالجة البيانات المفقودة اعتمادا على معيار القوة الإحصائية.
- إجراء دراسات لبحث أثر فقد البيانات ومعالجتها على القوة الإحصائية للأساليب الإحصائية اللابارامترية.
- التوسع في استخدام الطرق الحديثة لمعالجة البيانات المفقودة بدلا من الاعتماد على الطرق التقليدية.

المراجع

أولاً: المراجع العربية

- أوزبورن، جايسون. (2018). أفضل الممارسات في الانحدار اللوجستي (حصة العقيل وسامي العنزي، مترجم). معهد الإدارة العامة. (العمل الأصلي نشر في 2015).
- بابطين، عادل أحمد حسن (2002). مشكلات الدلالة الإحصائية في البحث التربوي وحلول بديلة [رسالة ماجستير غير منشورة]. جامعة أم القرى.
- بني عواد، علي والمومني، خالد. (2018). إدراك أعضاء هيئة التدريس في جامعة الملك فيصل لأهمية طرق التعامل مع البيانات المفقودة في الاستبانات البحثية واختبارات الطلبة. مجلة جامعة النجاح، 32(6)، 1033-1062.
- 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. دار إثراء للنشر والتوزيع.
- دودين، حمزة محمد. (2013). التحليل الإحصائي المتقدم للبيانات باستخدام SPSS. دار المسيرة للنشر والتوزيع.
- رزق الله، عائدة نخله. (2002). دليل الباحثين في التحليل الإحصائي الاختيار والتفسير. دار الكتب.
- شينار، سامية وبولجبال، آية. (2021). الدلالة الإحصائية وقوة الاختبار الإحصائي. مجلة العلوم الإسلامية والحضارة، 6(2)، 283-298.
- الصياد، عبد العاطي أحمد. (1988، يوليو 2-4). الدلالة العملية وحجم العينة المصاحبتين لاختبار الدلالة الإحصائية لاختبار (ت) في البحث التربوي والنفسي العربي. بحوث مؤتمر البحث التربوي الواقع والمستقبل، المركز العربي للبحوث التربوية. القاهرة، 2، 197-233.
- الضوي، محسوب عبد القادر. (2006). الإحصاء الاستدلالي المتقدم في التربية وعلم النفس. مكتبة الأنجلو المصرية.
- عامر، عبدا لناصر السيد. (2014). تقدير القوة الإحصائية للدراسات النفسية في المجلات العربية 2000-2012. مجلة دراسات تربوية ونفسية، دون مجلد (84)، 213-244.
- عبد الرحمن، طارق عطية. (2013). دليل تصميم وتنفيذ البحوث في العلوم الاجتماعية منهج تطبيقي لبناء المهارات البحثية. معهد الإدارة العامة.
- عبد الفتاح، فيصل أحمد. (2011). قوة الاختبارات الإحصائية وحجم الأثر في البحوث التربوية والنفسية. مجلة جامعة الملك سعود للعلوم التربوية والإسلامية، 23(2)، 195-214.
- عبده، شحادة مصطفى. (2017). تصميم البحث التربوي وأساليبه الإحصائية. دار نشر نور.
- العتيبي، أشرف. (2018). تأثير البيانات المفقودة على كفاءة نموذج الانحدار المتعدد وطرق معالجتها [أطروحة دكتوراة غير منشورة]. جامعة أم القرى.
- عيسوي، محمد معتمد. (2015). معالجة البيانات المفقودة لبعض التوزيعات الملتوية [رسالة ماجستير غير منشورة]. جامعة القاهرة.
- القحطاني، سعد. (2015). الإحصاء التطبيقي - المفاهيم الأساسية وأدوات التحليل الإحصائي الأكثر استخداما في الدراسات والبحوث الاجتماعية والإنسانية باستخدام SPSS. معهد الإدارة العامة.

المالكي، فهد عبد الله. (2018). واقع الدلالة الإحصائية والدلالة العملية وقوة الاختبار الإحصائي دراسة تطبيقية على بحوث المجالات التربوية بالجامعات السعودية [أطروحة دكتوراة غير منشورة]. جامعة أم القرى.

المالكي، فهد عبد الله. (2020). القوة الإحصائية للبحوث التربوية في مجالات الخليج العربي بين الماضي والحاضر. رسالة الخليج العربي، 41(157)، 37-54.

الريمي، مهيبه أحمد إمام. (2014). معالجة البيانات المفقودة باستخدام طريقتي تحليل المكونات الرئيسية وخريطة التنظيم الذاتي [رسالة ماجستير غير منشورة]. جامعة القاهرة.

محمد، محمد إبراهيم محمد. (2013). تحليل قوة الاختبار الإحصائي وعلاقتها بمستوى الدلالة وحجم التأثير في البحوث التربوية. مجلة دراسات عربية في التربية وعلم النفس، 3(41)، 100-125.

النجار، عبد الله عمر. (2006). دراسة تحليلية لقوة الاختبار الإحصائي في البحوث الإدارية المنشورة. المجلة العلمية لجامعة الملك فيصل، العلوم الإنسانية والإدارية، 7(2)، 261-293.

النعيمي، محمد عبدالعال والعاني، مزهر شعبان. (2008). الأساليب الإحصائية باستخدام حزمة MATLAB. دار وائل للنشر والتوزيع.

هويت، دنس وكمر، دنكن. (2016). مقدمة الإحصاء في علم النفس (صلاح الدين محمود علام، مترجم). دار الفكر للنشر (العمل الأصلي نشر في 2014).

المراجع الأجنبية

- AbdelFattah، F. (2011). Statistical power and effect size in educational and sychological research (in Arabic). King Saud University Journal of Educational and Islamic Sciences، 23(2)، 195-214.
- AbdelRahman، T. (2013). The guide to the design and implementation of research in social sciences is an applied approach to building research skills. Institute of Public Administration.
- Abdo، S. (2017). Educational sciences research design and statistical procedures. Noor Publishing.
- Aldhowey، M. (2002). Advanced inference statistics in education and psychology. Anglo Egyptian Bookshop.
- Al-hammami، A. (2007). Data Mining. Ethraa Publishing and Distribution.
- Almaliki، F. (2018). Reality of statistical significance and the practical significance and power of statistical test applied study on the researches of educational journals in Saudi universities [unpublished doctoral dissertation]. Umm al-Qura University.
- Almaliki، F. (2020). The statistical power of educational research in The Arab Gulf journals between the past and the present (in Arabic). Risālat al-Khalīj al-‘Arabī، 41 (157)، 37-54.
- Almureemi، M. (2014). Processing missing data using the principal components analysis method and the self-organizing map [Unpublished master thesis]. Cairo University.
- Alnajjar، A. (2006). An analytical study of statistical power in published administrative research (in Arabic). Scientific Journal of King Faisal University، Humanities and Management sciences، 7(2)، 261-293.
- Alnuaimi، M.، &Alani، M. (2008). Programming using matlab package with statistical applications. Wael Publishing and Distribution.

- Alotaibi, A. (2018). The impact of missing data on the efficiency of the multiple regression model and its Treatment methods [unpublished doctoral dissertation]. Umm al-Qura University.
- Alqahtani, S. (2015). Applied statistics - basic concepts and statistical analysis tools most commonly used in social and humanitarian studies and research using SPSS. Institute of Public Administration.
- Alsayyad, A. (1988, Jul2-4). Practical significance and sample size associated with the test of statistical significance of the (t) test in Arab educational and psychological research. Research conference on real and future educational research, Arab Center for Educational Research. Cairo, 2, 197-233.
- Amer, A. (2014). Estimating the statistical strength of psychological studies in Arab journals 2000-2012 (in Arabic). Journal of Educational and Psychological Studies, (84), 213-244.
- Babtain, A. (2002). Problems of statistical significance in educational research and alternative solutions [Unpublished master thesis]. Umm al-Qura University.
- Bany Awad, Ali, and Momani, Khaled. (2018). Recognizing the Importance of Dealing with Missing Data in Research Questionnaires and Test Students of the Faculty Members at King Faisal University (in Arabic). An-Najah University Journal, 32(6), 1033-1062.
- Bokossa, M., Huang, G., and Cohen, M. (2000). Imputation of test scores in The National Education Longitudinal Study. National Center for Educational Statistics (NELS:88), 191-196.
- Chinar, S., and Boulahbal, A. (2021). Statistical significance and the strength of the statistical test (in Arabic). Journal of Islamic Sciences and Civilization, 6(2), 283-298.
- Davey, A. & Savla, J. (2010). Statistical powers analysis with missing data a structural equation modeling approach. New York : Routledge.
- Dodeen, Hamzeh Mohammed. (2013). Advanced statistical analysis of data using SPSS. Almassira Publishing and Distribution.
- Enders, C. K. (2010). Applied missing data analysis. The Guilford Press.
- Howitte, D., & Crammer, D. (2016). Introduction to Statistics in Psychology. Translated by: Salaheddine Mahmoud Allam, dar alfiker publishing. (The original work was published in 2014).
- Hussein, Ali Nasser. (2012). Estimating of lost value of Responding variable in the Multi Regression Model (in Arabic). Journal of Economics, 8(30), 231-246.
- Issawi, M. (2015). Processing missing data for some twisted distributions [Unpublished master thesis]. Cairo University.
- Langkamp, D., Leman, A. & Lemeshow, S. (2010). Techniques for Handling Missing Data in Secondary Analyses of Large Surveys. Academic Pediatrics, 10 (3), 205 - 211.
- Little, R. J. A. & Rubin, D. B. (2002). Statistical analysis with missing data. Wiley publishing.
- Mohammad, M. (2013). analyzing of the statistical power and its relation to the significance level and the effect size in educational research (in Arabic). Journal of Arabic Studies in Education and Psychology, 3(41), 100-125.
- Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2015). Handbook of missing data methodology. 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.

- Osborne, J. (2018). Best practices in logistics regression. Translated by: Hessah al-Aqeel and Sami al-Anzi, Institute of Public Administration. (The original work was published in 2015).
- Peeters, M., Zondervan-Zwijnenburg, M., Vink, G., & Schoot, R. (2015). How to handle missing data: A comparison of different approaches. *European Journal of Developmental Psychology*, 377-394.
- 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.
- Rizkallah, Aida Nakhla. (2002). *Researchers' guide in statistical analysis selection and interpretation*. DarAlkotob.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall.
- Zhu, X. (2014). Comparison of Four Methods for Handling Missing Data in Longitudinal Data Analysis through a Simulation Study. *Open Journal of Statistics* . 933-944.