مقارنة لبعض طرائق تقدير القيم المفقودة في تصميم القطاعات العشوائية الكاملة

أ.م. مروان عبد العزيز دبدوب
جامعة الموصل / كلية علوم الحاسبات والرياضيات

م.م. وكاع علي هدبة
جامعة كركوك/ كلية الزراعة

تاريخ تسليم البحث : 2006/10/1 ؛ تاريخ قبول النشر : 2007/2/18

ملخص البحث :

استخدمت أربع طرائق لتقدير المشاهدات المفقودة في تصميم القطاعات العشوائية الكاملة، وهي: Yates وHarry وRubin وHaseman and Gaylar. وقد تمت المفاضلة بين هذه الطرائق وفقا للمعايير الإحصائية الآتية: قيمة متوسط مربعات الخطأ (MSe)، ومتوسط مربعات المعاملات (MSt) وقيمة الفرق المعنوي بين المعاملات (cal. F)، مع الأخذ بالاعتبار سهولة إجراء التحاليل الإحصائية.

وقد تبين أن أكفأ الطرائق هي طريقة Haseman and Gaylar، تليها طريقة Harry ومن ثم طريقة Yates. وقد عجزت طريقة Rubin من مجارات الطرائق الثلاث الأخرى.

Comparison of Some Methods of Estimation the Missing Values in the Randomized Complete Blocks Design

Assistant Prof.
Marwan A. Dabdob
University of Mosul
College of Computer & Mathematic Science

Assistant Lecturer
Wagaa Ali Hadba
University of Karkook
College of Agriculture

Abstract:

Four methods for tackling missing values were presented in randomized complete blocks design: Yates, Harry, Rubin, Haseman and Gaylar. To prefer between these methods some statistical measurements were used: error mean square (MSe), treatments mean square (MSt) and the value of significant differences between treatments (cal. F), with addition to the easiest path of statistical analysis.

It was found that the most preferable method was Haseman and Gaylar, followed by Harry method and the last one was Yates method. However, Rubin method has failed to go along with the others.

المقدمة:

في ضوء حركة التطور والتقدم العلمي المعتمد على أجراء التجارب المصممة على أسس علمية دقيقة، كان من المهم معالجة بعض المشاكل التي تواجه الباحث، ومنها فقدان بعض مشاهدات التجربة، بخاصة في موضوع تصميم وتحليل التجارب. وقد لاقت هذه المشكلة اهتماماً كبيراً من الباحثين لما تسببه من نقص في المعلومات وتأثيرها على دقة النتائج الإحصائية واتزان التصميم والخلل في توفر فروض التحليل.

إن فقدان قيمة واحدة أو أكثر في التصميم العشوائي الكامل (completely randomized design) لا يترتب عليه أية مشكلة ما دام أن تحليل التباين يمكن إجراؤه في حالة عدم تساوي عدد المشاهدات. ولكن عند فقدان قيمة في تحليل ذي اتجاهين مثل تصميم القطاعات العشوائية الكاملة (randomized complete block design) يعني فقدان التعامد والاتزان، ولحل هذه المشكلة وضعت عدة طرائق لإحلال قيم تقديرية بدلاً من قيم المشاهدات المفقودة، كل طريقة تسعى لإعطاء أفضل تقدير وإتباع ابسط أسلوب.

بدأت معالجة مشكلة القيم المفقودة عندما قام Allan and Wishart (1930) بتقدير قيمة مفقودة واحدة في تصميمي القطاعات العشوائية الكاملة وتصميم المربع اللاتيني، إذ وضعا معادلة تعتمد على جعل مجموع مربعات الخطأ اقل ما يمكن. وقد اقترح Yates (1933) أسلوب التعويض المتتالي iterative procedure لتقدير أكثر من قيمة مفقودة واحدة.

وقد قدر Harry (1964) القيم المفقودة في تصميم القطاعات العشوائية الكاملة من خلال حل عدد من المعادلات الآنية بعدد القيم المفقودة.

واستخدم Rubin (1972) الأسلوب غير المتتالي (non-iterative) لتقدير القيم المفقودة في تصميم القطاعات العشوائية الكاملة وذلك بالاستعانة بطريقة المربعات الصغرى (least squares method) وجعل مجموع مربعات خطأ الخلايا مساويا للصفر.

وقد أشار Haseman and Gaylar (1973) إلى أسلوب غير المتتالي (non-itreative) لتقدير القيم المفقودة في تصميم القطاعات العشوائية الكاملة.

مازالت الطرائق الأربع وهي: Yates (1933) وHarry (1964) و Rubin (1972) وHaseman and Gaylar (1973)، هي أكثر الطرائق استخداما في تقدير القيم المفقودة (هدبة، 2005).

قدم الجبوري (1998) دراسة تحليلية تعني بالقيم الشاذة والمفقودة في تصميمي العشوائي الكامل والمربع اللاتيني في حالة تكرار مشاهدات العينة. وأشار دبدوب وبدل (2004) إلى تأثر بعض طرائق معالجة القيم المفقودة في بيانات احتوت على قيم شاذة في تصميم العبور البسيط. وقد ناقش هدبة (2005) مشكلة القيم المفقودة في تصاميم التجارب الأساسية.

هدف البحث:

إن من أهم المعايير الإحصائية التي تشير إلى دقة تقدير القيم المفقودة هي انخفاض قيمة متوسط مربعات الخطأ error mean square (MSe)، فضلاً عن ارتفاع قيمة معيارين تعزيزيين هما: متوسط مربعات المعاملات treatments mean square (MSt) وقيمة الفرق المعنوي بين المعاملات value of significant difference between treatments والتي تدعى بقيمة F–المحسوبة (cal. F).

ستجرى المفاضلة بين أربع طرائق شائعة الاستخدام لتقدير قيم المشاهدات المفقودة وهي: Yates وHarry وRubin و(H & G) Haseman and Gayler، وسيتم ذلك في حالة تقديرها لأعداد مختلفة من قيم المشاهدات المفقودة وبمواقع مختلفة في التجربة، وذلك بالاستعانة بالمعايير الإحصائية المشار إليها، فضلا عن سهولة تطبيق الطريقة.

القيم المفقودة:

يحقق زيادة تكرار ظهور المعاملة في التجربة زيادة في دقة التجربة، ناتجة عن ارتفاع في درجات الحرية للخطأ التجريبي ومن ثم تقليل قيمة متوسط مربعات الخطأ (MSe)، لذا فإن فقدان بعض المشاهدات يعني انخفاض في عدد ظهور المعاملات فضلا عن تعقيد عمليات التحليل الإحصائي لعدم اتزان التصميم، وأفضل إجراء هو تقدير قيم المشاهدات المفقودة.

قد يحصل فقدان لبعض قيم مشاهدات التجربة بخاصة عندما تكون مرتبطة بالظروف الجوية والطبيعية التي غالباً ما تسبب تدميراً لبعض الوحدات التجريبية، فقد يحصل فقدان لبعض المشاهدات بسبب العوارض كالأمطار والعواصف والثلوج والأمراض. وقد يحصل فقدان للبيانات في مصنع ما نتيجة للتوقفات، وحتى في المختبرات العلمية التي تخضع للرقابة والسيطرة قد تقع عوارض مثل كسر إناء أو تلوث وسط نسجي أو انقطاع تيار كهربائي، وهكذا فإن فقدان بعض نتائج التجربة يحصل في شتى مجالات الحياة ولأسباب مختلفة.

في التصميم العشوائي الكامل تحلل البيانات في حالة فقدان بعض المشاهدات على أساس عدم تساوي التكرار وبذلك يمكن التغلب على هذه المشكلة، أما في التصاميم متعددة الاتجاهات ومنها تصميم القطاعات العشوائي الكاملة، فيتم إيجاد قيم تقديرية لتلك القيم المفقودة وذلك للمحافظة على التعامد واتزان التصميم ودقة النتائج. إن أفضل قيمة تقديريه هي التي تجعل الفرق بين مجموع مربعات الخطأ التجريبي بعد تقدير القيمة المفقودة والخطأ التجريبي في حالة عدم وجود القيمة المفقودة اصغر ما يمكن.

تحليل التباين:

هو الأسلوب الرياضي الذي يتم بموجبه تجزئة مجموع المربعات الكلي لبيانات ظاهرة ما إلى مصادره المختلفة (Keller and Warrack ,2003)، وتلخص نتائج التحليل في جدول يعرف بجدول تحليل التباين.

إن تصميم القطاعات العشوائية الكاملة هو ذلك التصميم الذي تجمع فيه الوحدات التجريبية المتجانسة في مجاميع عددها r تسمى قطاعات، في كل قطاع توزع عدد t من المعاملات على نفس العدد من الوحدات التجريبية توزيعاً عشوائياً ومستقلاً (المشهداني والمشهداني، 2002)، والنموذج الرياضي لهذا التصميم هو:

$$y_{ij} = \mu + \tau_i + \rho_j + \varepsilon_{ij}$$

$$i = 1,2,\ldots,t \ , \ j = 1,2,..,r$$

y_{ij}: قيمة مشاهدة الوحدة التجريبية التي أخذت المعاملة i والواقعة في القطاع j.

μ : الوسط العام.

τ_i : تأثير المعاملة i.

ρ_j : تأثير القطاع j.

ε_{ij} : تأثير الخطأ التجريبي للوحدة التجريبية التي أخذت المعاملة i والواقعة في القطاع j.

إذا احتوى التصميم على قيمة أو قيم مفقودة وجب تقديرها، ومن ثم إجراء تحليل التباين الذي تلخص نتائجه في الجدول (1)، عن كل قيمة مقدرة يجب أن يصحح مجموع مربعات المعاملات وأن تطرح درجة حرية واحدة من الخطأ التجريبي(الراوي وخلف الله، 2000).

الجدول 1: تحليل التباين لتصميم القطاعات العشوائية الكاملة.

S.O.V.	d.f.	S.S.	M.S.	F
Block	*r-1*	$t\sum_j (\bar{y}._j - \bar{y}..)^2$		
Treatments	*t-1*	$r\sum_{i=1} (\bar{y}_i. - \bar{y}..)^2$	$\frac{SSt}{t-1}$	$\frac{MSt}{MSe}$
Error	*(r-1)(t-1)*	$\sum_{j=1}^{r}\sum_{i=1}^{t}(Y_{ij} - \bar{y}._j - \bar{y}_i. + \bar{y}..)^2$	$\frac{SSe}{(t-1)(r-1)}$	
Total	rt-1	$SST = \sum_{j=1}^{r}\sum_{i=1}^{t}(y_{ij} - \bar{y}..)^2$		

تقدير القيم المفقودة:

إن أفضل طريقة للتغلب على القيمة المفقودة هي بتقدير قيمتها، سنشير إلى أربع طرائق شائعة الاستخدام لتقدير القيمة المفقودة في تصميم التجارب (هدبة، 2005). ولتوضيح فكرة هذه الطرائق وأسلوبها الرياضي ولتبسيط عملياتها الحسابية، فقد أجرينا تحويراً في معادلات

بعض هذه الطرائق واستبدلنا بعض رموزها حتى يسهل استخدامها وتكون ذات مدلول في مجال تصميم وتحليل التجارب.

1. طريقة Yates:

يتم تقدير عدة قيم مفقودة بتطبيق اسلوب التعويض المتتالي لجعل مجموع مربعات الخطأ اقل ما يمكن. لنفرض أن المشاهدة y_{11} قيمتها مفقودة، وبالاعتماد على نموذج التصميم تحسب قيمة الخطأ التجريبي ε_{ij} كالأتي:

$$\varepsilon_{ij} = y_{11}^2 + \sum_{i=2}^{t}\sum_{j=2}^{r} y_{ij}^2 - \frac{(Y_1. + y_{11})^2 + \sum_{i=2}^{t} Y_{1.}^2}{r} - \frac{(Y_{.1} + y_{11})^2 + \sum_{J=2}^{r} Y_{.j}^2}{t} + \frac{(Y.. + y_{11})^2}{tr}$$

نأخذ المشتقة بالنسبة إلى y_{11}:

$$\frac{\partial e_{ij}}{\partial y_{11}} = 2\hat{y}_{11} - \frac{2(Y_1. + \hat{y}_{11})}{r} - \frac{2(Y_{.1} + \hat{y}_{11})}{t} + \frac{2(Y.. + \hat{y}_{11})}{tr}$$

و بمساواة المعادلة بالصفر وحلها نحصل على:

$$\hat{y}_{11} = \frac{tY_1. + rY_{.1} - Y..}{(r-1)(t-1)} \quad \ldots \quad (1)$$

لتكون $\hat{y}_{11}$ هي القيمة التقديرية للمشاهدة المفقودة y_{11}، حيث أن:

$Y_i.$: مجموع قيم مشاهدات المعاملة التي تحتوي على القيمة المفقودة.

$Y_{.j}$: مجموع القطاع الذي يحتوي على المشاهدة المفقودة.

$Y..$: المجموع العام.

في حالة وجود مشاهدتين مفقودتين يتم تقدير إحدى القيمتين بالوسط الحسابي لبقية قيم مشاهدات المعاملة، ومن ثم تستخدم المعادلة (1) لتقدير القيمة الأخرى، أما القيمة المفقودة التي كان تقديرها الوسط الحسابي فيعاد تقديرها عن طريق التعويض المتتالي، وهكذا الحال عند وجود ثلاث قيم مفقودة أو أكثر (المشهداني والمشهداني، 2002).

2. طريقة Harry:

إن هذه الطريقة لها نفس أسلوب طريقة Yates في حالة وجود قيمة مفقودة واحدة، وتختلف عنها في الحالات الأخرى، وتعتمد على أسلوب التعويض غير المتتالي (non-iterative)، ولهذه الطريقة عدة حالات هي:

أ. فقدان قيمة مشاهدتين لنفس المعاملة: لنرمز إلى القيمة المفقودة الأولى y_{ij} والقيمة المفقودة الثانية $y_{ij'}$، عندئذ تكون قيمة مجموع مربعات الخطأ (SSe) كالآتي:

$$SSe = y_{ij}^2 + y_{ij'}^2 - \frac{(Y._j + y_{ij})^2 - (Y._{j'} + y_{ij'})^2}{t} - \frac{(Yi. + y_{ij} + y_{ij'})^2}{r} + \frac{(Y.. + y_{ij} + y_{ij'})^2}{rt} + \frac{(Y..)^2}{rt}$$

بأخذ المشتقة بالنسبة إلى $y_{ij'}, y_{ij}$ ومساواتها بالصفر نحصل على:

$$\frac{\partial SSe}{\partial y_{ij}} = 2\hat{y}_{ij} - \frac{2(Y._j + \hat{y}_{ij})}{t} - \frac{2(Yi. + \hat{y}_{ij} + \hat{y}_{ij'})}{r} + \frac{2(Y.. + \hat{y}_{ij} + \hat{y}_{ij'})}{rt} = 0$$

$$\frac{\partial SSe}{\partial y_{ij'}} = 2\hat{y}_{ij'} - \frac{2(Y._{j'} + \hat{y}_{ij'})}{t} - \frac{2(Yi. + \hat{y}_{ij} + \hat{y}_{ij'})}{r} + \frac{2(Y.. + \hat{y}_{ij} + \hat{y}_{ij'})}{rt} = 0$$

بعد التبسيط ينتج:

$$\hat{y}_{ij}(t-1)(r-2) = tY_{i.} + (r-1)Y_{.j} + Y_{.j'} - Y..$$

$$\hat{y}_{ij'}(t-1)(r-2) = tY_{i.} + Y_{.j} + (r-1)Y_{.j'} - Y..$$

ثم يتم الحصول على القيمتين التقديريتين كالآتي:

$$\hat{y}_{ij} = \frac{tY_{i.} + (r-1)Y_{.j} + Y_{.j'} - Y..}{(t-1)(r-2)} \quad \ldots \quad (2)$$

$$\hat{y}_{ij'} = \frac{tY_{i.} + Y_{.j} + (r-1)Y_{.j'} - Y..}{(t-1)(r-2)} \quad \ldots \quad (3)$$

ب. قيمة مشاهدتين مفقودتين في نفس القطاع: لنرمز إلى القيمة المفقودة الأولى y_{ij} والقيمة المفقودة الثانية $y_{i'j}$، لتكون قيمة مجموع مربعات الخطأ (SSe) كالآتي:

$$SSe = y_{ij}^2 + y_{i'j}^2 - \frac{(Yi. + y_{ij})^2 - (Yi'. + y_{i'j})^2}{r} - \frac{(Y._j + y_{ij} + y_{i'j})^2}{t} + \frac{(Y.. + y_{ij} + y_{i'j})^2}{r} + \frac{(Y..)^2}{tr}$$

بأخذ المشتقة بالنسبة إلى $y_{i'j}, y_{ij}$ ومساواتها بالصفر وتبسيط المعادلتين نحصل على:

$$\hat{y}_1(t-2)(r-1) = rY_{.j} + (r-1)Y_{i.} + Y_{i'.} - Y..$$

$$\hat{y}_1 = \frac{rY_{.j} + (t-1)Y_{i.} + Y_{i'.} - Y..}{(t-2)(r-1)} \quad \ldots \quad (4)$$

$$\hat{y}_{i'j}(t-2)(r-1) = rY_{.j} + Y_{i.} + (r-1)y_{i'.} - Y..$$

$$\hat{y}_{i'j} = \frac{rY_{.j} + Y_{i.} + (t-1)Y_{i'.} - Y..}{(t-2)(r-1)} \quad \ldots \quad (5)$$

هنالك تناظر في أسلوب تقدير قيمة المشاهدتين المفقودتين عند إنتمائهما إلى نفس المعاملة أو وقوعهما في نفس القطاع، ويمكن ملاحظة ذلك من المعادلتين 2 و4، والمعادلتين 3 و5.

ت. فقدان قيمة مشاهدتين تعودان لمعاملتين مختلفتين تقعان في قطاعين مختلفين: ليكن رمز القيمتين المفقودتين y_{ij} و$y_{i'j'}$، ولتبسيط الفكرة نضع المساواة الآتية:

$f = (t-1)(r-1)$

$w_1 = tY_{i.} + rY_{.j} - Y..$

$w_2 = tY_{i'.} + rY_{.j'} - Y..$

وقد أشار هدبة (2005) إلى العلاقتين الآتيتين:

$(f^2 - 1)y_{ij} = fw_1 - w_2$

$(f^2 - 1)y_{i'j'} = fw_2 - w_1$

وبتعويض عن f و w_1 و w_2 فيهما نحصل على:

$$\hat{y}_{ij} = \frac{(t-1)(r-1)(tY_{i.} + rY_{.j} - Y..) - (tY_{i'.} + rY_{.j'} - Y..)}{(t-1)^2(r-1)^2 - 1} \quad \ldots \quad (6)$$

$$\hat{y}_2 = \frac{(t-1)(r-1)(tY_{i'.} + rY_{.j'} - Y..) - (tY_{i.} + rY_{j.} - Y..)}{[(t-1)(r-1)]^2 - 1} \quad \ldots \quad (7)$$

ث. في حالة وجود اكثر من قيمتين مفقودتين :

تقدر إحدى القيم بالوسط الحسابي لبقية قيم مشاهدات المعاملة، وتستخدم إحدى صيغ تقدير القيمتين الأخريين، أما القيمة المفقودة التي كان تقديرها الوسط الحسابي فيعاد تقديرها.

3. طريقة Rubin:

تعتمد هذه الطريقة على الأسلوب غير المتكرر (non-iterative) لتقدير القيم المفقودة بإتباع المعادلة:

$$\hat{y}_{ij}' = -p'A^{-1} \quad \ldots \quad (8)$$

$\hat{y}_{ij}$, p : كل منهما عبارة عن متجه m×1، حيث أن: m يمثل عدد المشاهدات المفقودة، وأن: $k = 1,2,...,m$. وكل عنصر p_k في المتجه p مقداره:

$$p_k = e_{ij} = -\frac{Y_{i}.}{r} - \frac{Y_{.j}}{t} + \frac{Y..}{tr} \quad \ldots \quad (9)$$

A : مصفوفة m×m، وقيم عناصرها هي:

$$A_{kk} = 1 - \frac{1}{r} - \frac{1}{t} + \frac{1}{rt} \quad \ldots \quad (10)$$

$$A_{kk'} = \frac{1}{rt} \quad \ldots \quad (11)$$

4. طريقة Haseman & Gaylar (H&G):

بهذه الطريقة يتم تقدير m من القيم المفقودة بالاعتماد على مجموعة من المعادلات الآنية، وصيغتها العامة هي:

$$(r-1)(t-1)\hat{y}_h + \sum_{k \neq k'}^{m} [\hat{y}_{k'} - g(h\hat{y}_{k'})] = rY_{.j} + tY_{i.} - Y_{..} \quad \ldots \quad (12)$$

حيث أن:

$k \neq k' = 1,2,...,m$

$g = 1$ و $h = r$ إذا وقعت القيمتين المفقودتين y_k و $y_{k'}$ في نفس القطاع، وإذا كانتا y_k و $y_{k'}$ تعودان إلى نفس المعاملة فإن $g = 1$ و $h = t$. وغير ما ذكر فإن $g = 0$.

$Y_{i.}$: مجموع قيم مشاهدات المعاملة التي تعود إليها المشاهدة المفقودة y_k.

$Y_{.j}$: مجموع قيم مشاهدات القطاع الذي توجد فيه المشاهدة المفقودة y_k.

تطبيق:

أتاح البرنامج ألحاسوبي الجاهز Minitab 13.2 تكرار تطبيق كل من الطرائق التقديرية الأربع لمئة مرة، وذلك بإتباع أسلوب المحاكاة (Robinson, 2004) الذي ولَّد بيانات مختلفة تحاكي تصميم القطاعات العشوائية الكاملة. ولإعطاء فكرة عما توصل اليه أسلوب المحاكاة ولتوضيح التطبيقات العملية، ندرج أدناه تطبيق ذا بيانات واقعية تطابقت نتائجه مع ما أعطاه أسلوب المحاكاة. وقد اتبع السنجاري (2001) نفس هذا الأسلوب عند استخدامه للمحاكاة في معالجة القيم المفقودة وتأثرها بالقيم الشاذة في تصميم العبور البسيط.

أخذت البيانات من هدبة (2005)، وهي نتيجة تجربة أجريت في قسم البحوث الزراعية في موقع الرشيدية–محافظة نينوى، حيث أُقيمت التجربة لمقارنة إنتاج خمسة أصناف من القطن: الصنف الأول (نازلي 503) والصنف الثاني (1k347) والصنف الثالث (طاقة 2) والصنف الرابع (مرسومي 1) والصنف الخامس (دلتا بايه 5409)، احتوت التجربة على خمسة قطاعات كل قطاع قسم إلى خمس وحدات تجريبية وزعت عليها الأصناف عشوائياً. في كل وحدة تجريبية أنشأ أربعة مروز طول المرز خمسة أمتار والمسافة بين كل مرزين متتاليين 75 سنتيمتر، وزرع نبات واحد في الجورة بمسافة بين نبات وآخر 25 سنتيمتر. وقد تمت الزراعة بتاريخ 2001/5/3. كانت أوزان حاصل التجربة من القطن الزهر بالكيلوغرام لكل وحدة تجريبية كما في الجدول (2).

الجدول 2: نتائج التجربة لأوزان خمسة أصناف من القطن بالكيلوغرام لكل وحدة تجريبية.

القطاعات ← / الأصناف ↓	1	2	3	4	5	المجموع ($Y_{i.}$)
نازلي 503	4.970	5.030	4.890	4.850	4.840	24.580
1k347	4.580	4.490	4.490	4.350	4.330	22.240
طاقة 2	3.750	4.020	3.810	3.770	3.500	18.850
مرسومي	5.430	6.190	5.870	5.160	5.150	27.800
دلتاباية 5409	6.480	6.350	5.290	5.000	4.910	28.030
المجموع ($Y_{.j}$)	25.210	26.080	24.350	23.130	22.730	121.500

طبق تحليل التباين على بيانات الجدول (2) وكانت النتائج في الجدول (3).

الجدول 3: نتائج تحليل التباين عند عدم وجود قيم مفقودة.

S.O.V.	d.f.	S.S.	M.S	F
treatment	4	12.037	3.009	27.530
block	4	1.567	0.392	
Error	16	1.749	0.109	
Total	24	15.353		

عند فقدان مشاهدات من التجربة، يتم تقديرها بالطرائق الأربع المشار إليها، ومن ثم ملاحظة نتائج تحليل التباين، وكالآتي:

أ. **فقدان قيمة واحدة:** لنفرض فقدان المشاهدة التابعة للمعاملة الأولى في القطاع الثاني (y_{12}=5.030). لتقدير القيمة المفقودة ستتبع طرائق التقدير المشار إليها أنفا وهي:

طريقة Yates: باستخدام المعادلة (1) حسبت القيمة التقديرية، فكانت:

$$\hat{y}_{12} = \frac{5(19.550) + 5(21.050) - 116.470}{16} = 5.408$$

طريقة Harry: باستخدام المعادلة التي أشار إليها Yates (معادلة 1)، فإن $\hat{y}_{12} = 5.408$.

طريقة Rubin: من المعادلة (9) نحصل على:

$$p_1 = e_{12} - \frac{19.550}{5} - \frac{21.050}{5} + \frac{116.470}{25} = -3.461$$

لوجود قيمة مفقودة واحدة تستخدم المعادلة (10) فقط لتحسب قيمة $A_{1\times1}$ كالآتي:

$$A_{1\times1} = A_{kk} = 1 - \frac{1}{5} - \frac{1}{5} + \frac{1}{25} = 0.640$$

وبموجب العلاقة (8) نحصل على تقدير للقيمة المفقودة وهي:

$$\hat{y}_{12} = -(-3.460)\left(\frac{1}{0.640}\right) = 5.408$$

طريقة H&G: بإتباع المعادلة (12) تم تقدير قيمة المشاهدة المفقودة، فكانت:

$$16\,\hat{y}_{12} = 5(19.550) + 5(21.050) - 116.470 = 86.530$$

$$\hat{y}_{12} = 5.408$$

أعطت الطرائق الأربع نفس القيمة التقديرية للمشاهدة المفقودة ومقدارها 5.408، وهذا هو الحال عند فقدان قيمة مشاهدة واحدة. وبعد تصحيح قيمة مجموع مربعات المعاملات (SSt) وطرح درجة حرية واحدة من الخطأ (الراوي وخلف الله، 2000) نحصل على الجدول (4) لتحليل التباين الموحد للطرائق الأربع.

الجدول 4: تحليل التباين المصحح بعد تقدير قيمة المشاهدة المفقودة.

S.O.V.	d.f	S.S.	M.S.	Cal. F
treatment	4	12.103	3.026	29.210
block	4	1.859	0.465	
Error	15	1.657	0.104	
Total	23	15.619		

ب. فقدان قيمتين لنفس المعاملة: نفرض أن المشاهدتين العائدتين للمعاملة الخامسة في القطاعين الأول والثاني قد فقدتا، وقيمتهما: y_{51}=6.480 وy_{52}=6.350. وقد قدرت قيمة المشاهدتين المفقودتين كالآتي:

طريقة Yates: بإتباع أسلوب التعويض المتتالي والاستعانة بالمعادلة (1)، تم الحصول على القيمتين تقديريتين هما: $\hat{y}_{51} = 5.415$ و $\hat{y}_{52} = 5.165$.

طريقة Harry: عن طريق المعادلة (2) قدرت القيمة المفقودة y_{51} كالآتي:

$$\hat{y}_{51} = \frac{5(15.200) + 4(19.730) + 18.730 - 108.690}{12} = 5.413$$

وعن طريق المعادلة (3) قدرت القيمة المفقودة y_{52} كالآتي:

$$\hat{y}_{52} = \frac{5(15.200) + (19.730) + 4(18.730) - 108.670}{12} = 5.165$$

طريقة Rubin: باستخدام المعادلة (9) نحصل على المتجه p كالآتي:

$$p_1 = e_{51} = -\frac{15.200}{5} - \frac{19.730}{5} + \frac{108.670}{25} = -2.639$$

$$p_2 = e_{52} = -\frac{15.200}{5} - \frac{18.750}{5} + \frac{108.670}{25} = -2.443$$

$$p' = (-2.639 \quad -2.443)$$

وبالرجوع إلى المعادلتين (10) و(11) تكون عناصر المصفوفة $A_{2\times 2}$ كالآتي:

$$A = \begin{pmatrix} 0.640 & 0.040 \\ 0.040 & 0.640 \end{pmatrix}$$

وباستخدام العلاقة (8) تم الحصول على تقدير للقيمتين المفقودتين:

$$\hat{y}_{51} = 3.900 \quad \hat{y}_{52} = 3.573$$

طريقة H&G: بما أن المشاهدتين المفقودتين تعودان إلى نفس المعاملة فإن المعادلة (12) ستعطي المعادلتين الآنيتين الآتيتين:

$$16\hat{y}_{51} - 4\hat{y}_{52} = 65.980$$
$$16\hat{y}_{52} - 4\hat{y}_{51} = 60..980$$

بحل هاتين المعادلتين نحصل على قيمتين تقديريتين هما:

$$\hat{y}_{51} = 5.415 \quad \hat{y}_{52} = 5.165$$

اختلفت طريقة Rubin عن الطرائق الثلاث الأخرى في المقدار التقديري للمشاهدتين المفقودتين. وبعد تصحيح قيمة مجموع مربعات المعاملات (SSt) للقيمتين المقدرتين وطرح درجتين حرية من الخطأ (الراوي وخلف الله، 2000) تم الحصول على الجدول (5).

الجدول 5: نتائج تحليل التباين عند فقدان قيمة مشاهدتين عائدتين إلى نفس المعاملة.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	Cal. F بعد التقدير
Yates	5.415 5.165	0.037	2.234	60.378
Harry	5.415 5.165	0.037	2.234	60.378
Rubin	3.900 3.573	0.200	1.700	8.500
H&G	5.415 5.165	0.037	2.234	60.378

ت. فقدان قيمتين في نفس القطاع: لنفرض فقدان مشاهدتين في القطاع الخامس تابعتين للمعاملة الأولى (y_{15}=4.840) والمعاملة الخامسة (y_{55}=4.910). وقد قدرت قيمتاهما كالآتي:

طريقة Yates: بعد مساواة إحدى القيمتين بالوسط الحسابي لبقية قيم مشاهدات نفس المعاملة، واستخدام الأسلوب المتتالي والاستعانة بالمعادلة (1)، تم الحصول على قيمتين تقديريتين للمعاملتين المفقودتين ومقدارهما: $\hat{y}_{15} = 5.272$ و $\hat{y}_{55} = 5.804$.

طريقة Harry: أعطت المعادلة (4) القيمة التقديرية للمعاملة الأولى، فكانت:

$$\hat{y}_{15} = \frac{5(14.700) + 4(19.550) + (21.680) - 110.120}{12} = 5.272$$

وأعطت المعادلة (5) القيمة التقديرية للمعاملة الخامسة فكانت:

$$\hat{y}_{55} = \frac{5(14.700) + (19.550) + 4(21.680) - 110.120}{12} = 5.804$$

طريقة Rubin: عن طريق المعادلة (9)، تم الحصول على المتجه $p_{2\times1}$ والمصفوفة $A_{2\times2}$ كالآتي:

$$p_1 = e_{15} = -\frac{19.550}{5} - \frac{14.700}{5} + \frac{110.120}{25} = -2.445$$

$$p_2 = e_{55} = -\frac{21.680}{5} - \frac{14.700}{5} + \frac{110.120}{25} = -2.871$$

$$p' = \begin{pmatrix} -2.445 & -2.871 \end{pmatrix}$$

من المعادلتين (10) و(11)، نحصل على:

$$A_{kk} = 0.640$$

$$A_{kk'} = 0.040$$

علية فإن المصفوفة $A_{2\times2}$ ستكون على الشكل الآتي:

$$A = \begin{pmatrix} 0.640 & 0.040 \\ 0.040 & 0.640 \end{pmatrix}$$

وبعد الرجوع إلى المعادلة (8) تم تقدير قيمة المشاهدتين المفقودتين وقيمتهما التقديرية:

$\hat{y}_{15} = 3.554$ $\hat{y}_{55} = 4.264$,

طريقة H&G: من خلال المعادلة (12) تم الحصول على المعادلتين الآنيتين الآتيتين:

$$16\hat{y}_{15} - 4\hat{y}_{55} = 61.130$$

$$16\hat{y}_{55} - 4\hat{y}_{15} = 71.780$$

وبحل هاتين المعادلتين نحصل على القيمتين التقديريتين الآتيتين:

$\hat{y}_{15} = 5.272$ $\hat{y}_{55} = 5.804$,

أعطت الثلاث طرائق Yates وHarry وH&G نفس القيم التقديرية، ومرة أخرى اختلفت طريقة Rubin عن بقية الطرائق. وبعد تصحيح قيمة مجموع مربعات المعاملات (SSt) للقيمتين المقدرتين وطرح درجتين حرية من الخطأ تم الحصول على الجدول (6).

الجدول 6: نتائج تحليل التباين عند فقدان قيمة مشاهدتين تقعان في نفس القطاع.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	Cal. F بعد التقدير
Yates	5.272 5.804	0.106	1.412	13.321
Harry	5.272 5.804	0.106	1.412	13.321
Rubin	3.554 4.264	0.288	2.317	8.045
H&G	5.272 5.804	0.106	1.412	13.321

ث. فقدان ثلاث قيم: لنفرض فقدان قيم المشاهدات الآتية: المعاملة الخامسة في القطاع الأول (y_{51}=6.480) والمعاملة الخامسة في القطاع الثاني (y_{52}=6.350) والمعاملة الأولى في القطاع الثاني (y_{12}=5.030). عند استخدام طريقتي Harry وH&G يجب ملاحظ الآتي: هناك مشاهدتان مفقودتان تعودان إلى نفس المعاملة وهما: y_{51} وy_{52}، وهناك مشاهدتان مفقودتان موقعهما في نفس القطاع وهما: y_{52} وy_{12}، كما يوجد مشاهدتان مختلفتان في موقعهما ويعودان إلى معاملتين مختلفتين وهما: y_{51} وy_{12}.

وقد قدرت قيم المشاهدات المفقودة كالآتي:

طريقة Yates: بعد الاستعانة بالمعادلة (1) وإتباع أسلوب التعويض المتتالي تم الحصول على القيم التقديرية الآتية: $\hat{y}_{51} = 5.272$ و $\hat{y}_{52} = 5.475$ و $\hat{y}_{12} = 5.165$.

طريقة Harry: أعطت الصيغ (2 و3 و4 و5 و6 و7) القيم التقديرية الآتية: $\hat{y}_{51} = 5.276$ و $\hat{y}_{52} = 5.469$ و. $\hat{y}_{12} = 5.165$.

طريقة Rubin: بوساطة المعادلة (9) يتم الحصول على الآتي:

$$p_1 = e_{51} = -\frac{19.550}{5} - \frac{14.700}{5} + \frac{103.640}{25} = -2.704$$

$$p_2 = e_{52} = -\frac{15.200}{5} - \frac{14.700}{5} + \frac{103.640}{25} = -1.834$$

$$p_3 = e_{12} = -\frac{15.200}{5} - \frac{18.770}{5} + \frac{103.640}{25} = -2.648$$

علية فإن عناصر المتجه $p_{3\times1}$ هي:

$$p' = (-2.704 \quad -1.834 \quad -2.648)$$

ومن المعادلتين (10) و(11)، نحصل على:

$$A_{kk} = 0.640$$

$$A_{kk'} = 0.040$$

لتكون عناصر المصفوفة $A_{3\times3}$ كالآتي:

$$A=\begin{pmatrix}0.640 & 0.040 & 0.040\\ 0.040 & 0.640 & 0.040\\ 0.040 & 0.040 & 0.640\end{pmatrix}$$

وباستخدام العلاقة (8) تكون القيم التقديرية هي: $\hat{y}_{51}=3.841$ و $\hat{y}_{52}=2.391$ و $\hat{y}_{12}=3.748$.

طريقة H&G: بعد الأخذ بعين الاعتبار موقع القيم المفقودة وعائديتها إلى المعاملة أعطت المعادلة (10) المعدلات الآنية الآتية:

$$16\hat{y}_{51}-4\hat{y}_{52}+\hat{y}_{12}=67.610$$
$$16\hat{y}_{52}-4\hat{y}_{51}-4\hat{y}_{12}=45.860$$
$$16\hat{y}_{12}+\hat{y}_{51}-4\hat{y}_{52}=66.010$$

وبحل المعادلات الآنية قدرت القيم المفقودة، فكانت: $\hat{y}_{51}=5.272, \hat{y}_{52}=5.476, \hat{y}_{12}=5.165$.

يلاحظ توافق نتائج الطرائق الثلاث Yates وHarry وH&G في القيمة التقديرية للمشاهدة المفقودة واختلفت عنهم طريقة Rubin، وقد وضعت نتائج تحليل التباين للطرائق الأربع في الجدول (7).

الجدول 7: تحليل التباين عند فقدان قيم ثلاث مشاهدات.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	Cal. F بعد التقدير
Yates	5.272	0.035	1.205	34.429
	5.475			
	5.165			
Harry	5.276	0.034	1.201	35.324
	5.469			
	5.165			
Rubin	3.841	0.386	1.616	4.187
	2.391			
	3.748			
H&G	5.272	0.035	1.205	34.429
	5.476			
	5.165			

الاستنتاجات:

ندرج في أدناه ما تم ملاحظته من نتائج على شكل نقاط:

1. عند فقدان قيمة مشاهدة واحدة تتساوى الطرائق الأربع في مقدار القيمة التقديرية للمشاهدة المفقودة، ومن ثم الحصول على نفس جدول تحليل التباين.
2. إن الطرائق الثلاث وهي: Yates و Harry و(H&G) Haseman & Gaylar تقاربت في قيمها التقديرية للقيمة المفقودة، وكلما زاد عدد القيم المفقودة زاد الاختلاف بينها بعض الشيء. وقد اختلفت طريقة Rubin في مقدار القيمة التقديرية عن الطرائق الأخرى إذ أعطت قيم عالية لمتوسط مربعات الخطأ (MSe) وأقل قيمة للفرق المعنوي بين المعاملات (cal. F)، وقد تذبذبت قيمة متوسط مربعات المعاملات (MSt) فيها.
3. لم يكن لموقع القيم المفقودة تأثيرٌ في نتائج جدول تحليل التباين، وكان لقيم المشاهدات المفقودة علاقة مع قيم بقية المشاهدات، إذ لوحظ أن القيم التقديرية تنحدر لتكون قريبة من الوسط الحسابي للمعاملة العائدة لها، لذا قد تعطي عملية التقدير قيمة لمتوسط مربعات الخطأ (MSe) أقل من تلك القيمة الحقيقية عند عدم وجود قيم مفقودة وتحدث مثل هذه الحالة عندما تكون القيم الحقيقية للمشاهدات المفقودة ذات قيم تقع إلى أحد جانبي قيم المعاملة.
4. لتقارب نتائج تحليل التباين بين الطرائق الثلاث: Yates و Harry و Haseman & Gaylar (H&G)، لذا تتم المفاضلة بين الطرائق على أساس سهولة حساب القيمة التقديرية، وهنا عينت طريقة Haseman & Gaylar (H&G) كأكفأ طريقة إذ تعطي قيم تقديرية مباشرة من خلال حل عدد من المعادلات الآنية التي يمكن التوصل إليها بسهولة. وتأتي بالدرجة الثانية طريقة Harry التي تدخل في عدة مراحل تحليلية عند وجود أكثر من مشاهدتين مفقودتين. وأخيراً طريقة Yates التي تتميز بطول مراحلها التحليلية كلما زاد عدد المشاهدات المفقودة وذلك لإتباعها أسلوب التعويض المتتالي.

المصادر العربية

الجبوري، منى حسين، 1998، " دراسة تحليلية للقيم الشاذة والقيم المفقودة لتصميم المربع اللاتيني وتصميم تام التعشية في حالة تكرار مشاهدات المعاينة"، رسالة ماجستير، كلية الإدارة والاقتصاد، الجامعة المستنصرية.

دبدوب، مروان عبد العزيز وبدل، محمد عبد المجيد، 2004، "تقييم ثلاث طرائق لمعالجة القيم المفقودة وتأثرها بالقيم الشاذة في تصميم العبور البسيط"، مجلة علوم الرافدين، المجلد 15، العدد 2 خاص بالحاسبات والرياضيات والإحصاء، ص 17-26.

الراوي، خاشع محمود وخلف الله، عبد العزيز محمد، 2000، " تصميم وتحليل التجارب الزراعية"، الطبعة الثانية، دار الكتب للطباعة والنشر، جامعة الموصل.

السنجاري، محمد عبد المجيد بدل، 2001، " معالجة القيم المفقودة وتأثرها بالقيم الشاذة في تصميم العبور البسيط"، رسالة ماجستير، كلية علوم الحاسبات و الرياضيات-جامعة الموصل.

المشهداني، محمود حسن والمشهداني، كمال علوان، 2002، " تصميم وتحليل التجارب"، الدار الجامعية للطباعة والنشر والترجمة، جامعة بغداد.

هدبة، وكاع علي، 2005، " مقارنة لبعض طرائق تقدير ومعالجة القيم المفقودة في تصاميم التجارب الأساسية"، رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.

المصادر الأجنبية

Allan, F. and Wishart, J., 1930, "A method of estimating the yield of a missing plot in field experiment work", J. of Agric. Sci., 20, 399; citied by J. of App. Stat., 79,p 173-185.

Harry, C., 1964, " Introduction to experimental statistics", McGraw-Hill Book Company, USA .

Haseman , J.K and D.W Gaylor , 1973, "An algorithm for non–iterative estimation of multiple missing values for crossed classifications", Technometrics , 15, pp. 631-636.

Keller.G and warrack, 2003, "Statistics for management and conomics", Duxbury Press, USA.

Robinson, S., 2004, "Simulation: The Practice of Model Development and Use", John Wiley & Sons, USA.

Rubin, D., 1972, "A non-iterative algorithm for least square estimation of missing values in any analysis of variance design", J. of App. Stat., 21, P 136-141.

Yates, F., 1933, "The analysis of replicated experiments when the field results are incomplete", Emp. J. Exp. Agric., 1, p 129-142.