

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 : ٢٠٠٦/١٠/١ ؛ تاريخ قبول النشر : ٢٠٠٧/٢/١٨

ملخص البحث :

استخدمت أربع طرائق لتقدير المشاهدات المفقودة في تصميم القطاعات العشوائية الكاملة، وهي: Yates و Harry و Rubin و Haseman and Gaylar. وقد تمت المفاضلة بين هذه الطرائق وفقاً 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 أكفاً 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، خاصة 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 أبسط 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 الآتية 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 مساوياً للصفر.

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

مازالت الطرائق الأربع وهي: Yates (1933) و Harry (1964) و Rubin (1972) و Haseman and Gaylar (1973)، هي أكثر الطرائق استخداماً في تقدير القيم المفقودة (هدبة، ٢٠٠٥).

قدم الجبوري (1998) دراسة تحليلية تعني بالقيم الشاذة والمفقودة في تصميمي العشوائي الكامل والمربع اللاتيني في حالة تكرار مشاهدات العينة. وأشار بدوب وبدل (٢٠٠٤) إلى تأثر بعض طرائق معالجة القيم المفقودة في بيانات احتوت على قيم شاذة في تصميم العبور البسيط. وقد ناقش هدبة (٢٠٠٥) 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 (MSe) error mean square، فضلاً عن ارتفاع قيمة معيارين تعزيزيين هما: متوسط مربعات المعاملات (MSt) treatments mean square 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 و Harry و Rubin و Haseman and Gayler (H & G)، وسيتم ذلك في حالة تقديرها لأعداد مختلفة من قيم المشاهدات المفقودة وبمواقع مختلفة في التجربة، وذلك بالاستعانة بالمعايير الإحصائية المشار إليها، فضلاً 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 فضلاً عن تعقيد عمليات التحليل الإحصائي لعدم اتزان التصميم، وأفضل إجراء هو تقدير قيم المشاهدات المفقودة. قد يحصل فقدان لبعض قيم المشاهدات التجربة بخاصة عندما تكون مرتبطة بالظروف الجوية والطبيعية التي غالباً ما تسبب تدميراً لبعض الوحدات التجريبية، فقد يحصل فقدان لبعض المشاهدات بسبب العواصف والأمطار والعواصف والتلوج والأمراض. وقد يحصل فقدان للبيانات في مصنع ما نتيجة للتوقفات، وحتى في المختبرات العلمية التي تخضع للرقابة والسيطرة قد تقع عواصف 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 التصميمات 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 تقديرية 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، ٢٠٠٢)، والنموذج الرياضي لهذا التصميم هو:

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

$$i = 1, 2, \dots, t, \quad j = 1, 2, \dots, r$$

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 (١)، عن كل قيمة مقدرة يجب أن يصحح مجموع مربعات المعاملات وأن تطرح درجة حرية واحدة من الخطأ التجريبي (الراوي وخلف الله، ٢٠٠٠).

الجدول ١: 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	$t-1$	$r \sum_{i=1}^t (\bar{y}_{i.} - \bar{y}_{..})^2$	$\frac{SSt}{t-1}$	$\frac{MSt}{MSe}$
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، ٢٠٠٥). ولتوضيح فكرة هذه الطرائق وأسلوبها الرياضي ولتبسيط عملياتها الحسابية، فقد أجرينا تحويلاً في معادلات

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

١. طريقة Yates:

يتم تقدير عدة قيم مفقودة بتطبيق أسلوب التعويض المتتالي لجعل مجموع مربعات الخطأ أقل ما يمكن. لنفرض أن الملاحظة y_{11} قيمتها مفقودة، وبالاكتفاء على نموذج التصميم تحسب قيمة الخطأ التجريبي ε_{ij} كالآتي:

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

نأخذ المشتقة بالنسبة إلى y_{11} :

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

و بمساواة المعادلة بالصفر وحلها نحصل على:

$$\hat{y}_{11} = \frac{tY_{1.} + rY_{.1} - Y_{..}}{(r-1)(t-1)} \quad (1) \quad \dots$$

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

$Y_{i.}$: مجموع قيم مشاهدات المعاملة التي تحتوي على القيمة المفقودة.

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

$Y_{..}$: المجموع العام.

في حالة وجود مشاهدين مفقودتين يتم تقدير إحدى القيمتين بالوسط الحسابي لبقية قيم مشاهدات المعاملة، ومن ثم تستخدم المعادلة (١) 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، ٢٠٠٢).

٢. 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 مشاهدين لنفس المعاملة: لنرمز إلى القيمة المفقودة الأولى y_{ij} والقيمة المفقودة الثانية $y_{ij'}$ ، عندئذ تكون قيمة مجموع مربعات الخطأ (Sse) كالآتي:

$$SSe = y_{ij}^2 + y_{ij'}^2 - \frac{(Y_{.j} + y_{ij})^2 - (Y_{.j'} + y_{ij'})^2}{t} - \frac{(Y_{i.} + y_{ij} + y_{ij'})^2}{r} + \frac{(Y_{..} + y_{ij} + y_{ij'})^2}{rt} + \frac{(Y_{..})^2}{rt}$$

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

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

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

بعد التبسيط ينتج:

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

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

ثم يتم الحصول على القيمتين التقديريتين كآلاتي:

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

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

ب. قيمة مشاهدين مفقودتين في نفس القطاع: لنرمز إلى القيمة المفقودة الأولى y_{ij} والقيمة

المفقودة الثانية $y_{ij'}$ ، لتكون قيمة مجموع مربعات الخطأ (SSe) كآلاتي:

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

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

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

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

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

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

هناك تناظر في أسلوب تقدير قيمة المشاهدين المفقودتين عند إنتمائهما إلى نفس المعاملة

أو وقوعهما في نفس القطاع، ويمكن ملاحظة ذلك من المعادلتين ٢ و ٤، والمعادلتين ٣ و ٥.

ت. فقدان قيمة مشاهدين 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 هدية (٢٠٠٥) 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:

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

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

ث. في حالة وجود أكثر من قيمتين مفقودتين :

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

٣. طريقة Rubin:

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

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

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

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

A : مصفوفة $m \times m$ ، وقيم عناصرها هي:

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

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

٤. طريقة Haseman & Gaylar (H&G):

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

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

حيث أن:

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

$g = 1$ 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 الملاحظة المفقودة Y_k .
 Y_j : مجموع قيم مشاهدات القطاع الذي توجد فيه الملاحظة المفقودة Y_k .

تطبيق:

أتاح البرنامج الحاسوبي الجاهز Minitab 13.2 تكرار تطبيق كل من الطرائق التقديرية الأربع لمئة مرة، وذلك بإتباع أسلوب المحاكاة (Robinson, 2004) الذي ولّد بيانات مختلفة تحاكي تصميم القطاعات العشوائية الكاملة. ولإعطاء فكرة عما توصل إليه أسلوب المحاكاة ولتوضيح التطبيقات العملية، ندرج أدناه تطبيق ذا بيانات واقعية تطابقت نتائجه مع ما أعطاه أسلوب المحاكاة. وقد اتبع السنجاري (٢٠٠١) نفس هذا الأسلوب عند استخدامه للمحاكاة في معالجة القيم المفقودة وتأثرها بالقيم الشاذة في تصميم العبور البسيط.

أخذت البيانات من هدبة (٢٠٠٥)، 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 ٥٠٣) والصنف الثاني (١k٣٤٧) والصنف الثالث (طاقة ٢) والصنف الرابع (مرسومي ١) والصنف الخامس (دلتا بايه ٥٤٠٩)، 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 ٧٥ سنتيمتر، وزرع نبات واحد في الجورة بمسافة بين نبات وآخر ٢٥ سنتيمتر. وقد تمت الزراعة بتاريخ ٢٠٠١/٥/٣. كانت أوزان حاصل التجربة من القطن الزهر بالكيلوغرام لكل وحدة تجريبية كما في الجدول (٢).

الجدول ٢: نتائج التجربة لأوزان خمسة أصناف من القطن بالكيلوغرام لكل وحدة تجريبية.

المجموع (Y _{i.})	٥	٤	٣	٢	١	→ القطاعات ↓ الأصناف
٢٤.٥٨٠	٤.٨٤٠	٤.٨٥٠	٤.٨٩٠	٥.٠٣٠	٤.٩٧٠	نازلي ٥.٣
٢٢.٢٤٠	٤.٣٣٠	٤.٣٥٠	٤.٤٩٠	٤.٤٩٠	٤.٥٨٠	1k347
١٨.٨٥٠	٣.٥٠٠	٣.٧٧٠	٣.٨١٠	٤.٠٢٠	٣.٧٥٠	طاقة ٢
٢٧.٨٠٠	٥.١٥٠	٥.١٦٠	٥.٨٧٠	٦.١٩٠	٥.٤٣٠	مرسومي
٢٨.٠٣٠	٤.٩١٠	٥.٠٠٠	٥.٢٩٠	٦.٣٥٠	٦.٤٨٠	دلتاباية ٥٤.٠٩
١٢١.٥٠٠	٢٢.٧٣٠	٢٣.١٣٠	٢٤.٣٥٠	٢٦.٠٨٠	٢٥.٢١٠	المجموع (Y _{.j})

طبق تحليل التباين على بيانات الجدول (٢) وكانت النتائج في الجدول (٣).

الجدول ٣: نتائج تحليل التباين عند عدم وجود قيم مفقودة.

S.O.V.	d.f.	S.S.	M.S	F
treatment	4	١٢.٠٣٧	3.009	27.530
block	4	1.567	0.392	
Error	16	١.٧٤٩	0.109	
Total	٢٤	١٥.٣٥٣		

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

أ. فقدان قيمة واحدة: لنفرض فقدان المشاهدة التابعة للمعاملة الأولى في القطاع الثاني

(y₁₂=5.030). لتقدير القيمة المفقودة سنتبع طرائق التقدير المشار إليها أنفا وهي:

طريقة Yates: باستخدام المعادلة (١) حسب 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 ١)، فإن $\hat{y}_{12} = 5.408$.

طريقة Rubin: من المعادلة (٩) 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 (١٠) فقط لتحسب قيمة $A_{1 \times 1}$ كالآتي:

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

وبموجب العلاقة (٨) نحصل على تقدير للقيمة المفقودة وهي:

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

طريقة H&G: بإتباع المعادلة (١٢) 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 ٥.٤٠٨، وهذا هو الحال عند فقدان قيمة مشاهدة واحدة. ويعد تصحيح قيمة مجموع مربعات المعاملات (SSt) وطرح درجة حرية واحدة من الخطأ (الراوي وخلف الله، ٢٠٠٠) نحصل على الجدول (٤) لتحليل التباين الموحد للطرائق الأربع.

الجدول ٤: 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	٢٩.٢١٠
block	4	1.859	٠.٤٦٥	
Error	15	١.٦٥٧	٠.١٠٤	
Total	٢٣	١٥.٦١٩		

ب. فقدان قيمتين لنفس المعاملة: نفرض أن المشاهدين العائدين للمعاملة الخامسة في القطاعين الأول والثاني قد فقدتا، وقيمتها: $y_{51} = ٦.٤٨0$ و $y_{52} = ٦.٣٥0$. وقد قدرت قيمة المشاهدين المفقودتين كالآتي:

طريقة Yates: بإتباع أسلوب التعويض المتتالي والاستعانة بالمعادلة (١)، تم الحصول على القيمتين تقديريتين هما: $\hat{y}_{51} = 5.415$ و $\hat{y}_{52} = 5.165$.

طريقة Harry: عن طريق المعادلة (٢) 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 (٣) 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 (٩) 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 (١٠) و (١١) تكون عناصر المصفوفة $A_{2 \times 2}$ كالآتي:

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

وباستخدام العلاقة (٨) تم الحصول على تقدير للقيمتين المفقودتين:

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

طريقة H&G: بما أن المشاهدين المفقودتين تعودان إلى نفس المعاملة فإن المعادلة (١٢) ستعطي المعادلتين 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 للمشاهدين المفقودتين. وبعد تصحيح قيمة مجموع مربعات المعاملات (SSt) للقيمتين المقدرتين وطرح درجتين حرية من الخطأ (الراوي وخلف الله، ٢٠٠٠) تم الحصول على الجدول (٥).

الجدول ٥: نتائج تحليل التباين عند فقدان قيمة مشاهدين عائدين إلى نفس المعاملة.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	Cal. F بعد التقدير
Yates	٥.٤١٥	٠.٠٣٧	٢.٢٣٤	٦٠.٣٧٨
	٥.١٦٥			
Harry	٥.٤١٥	٠.٠٣٧	٢.٢٣٤	٦٠.٣٧٨
	٥.١٦٥			
Rubin	٣.٩٠٠	٠.٢٠٠	١.٧٠٠	٨.٥٠٠
	٣.٥٧٣			
H&G	٥.٤١٥	٠.٠٣٧	٢.٢٣٤	٦٠.٣٧٨
	٥.١٦٥			

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

طريقة Yates: بعد مساواة إحدى القيمتين بالوسط الحسابي لبقية قيم مشاهدات نفس المعاملة، واستخدام الأسلوب المتتالي والاستعانة بالمعادلة (١)، تم الحصول على قيمتين تقديريتين للمعاملتين المفقودتين ومقدارهما: $\hat{y}_{15} = 5.272$ و $\hat{y}_{55} = 5.804$.

طريقة Harry: أعطت المعادلة (٤) القيمة التقديرية للمعاملة الأولى، فكانت:

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

وأعطت المعادلة (٥) 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 (٩)، تم الحصول على المتجه $p_{2 \times 1}$ والمصفوفة $A_{2 \times 2}$ كالآتي:

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

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

$$p' = (-2.445 \quad -2.871)$$

من المعادلتين (١٠) و (١١)، نحصل على:

$$A_{kk} = 0.640$$

$$A_{kk'} = 0.040$$

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

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

وبعد الرجوع إلى المعادلة (٨) تم تقدير قيمة المشاهدين المفقودتين وقيمتها التقديرية:

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

طريقة H&G: من خلال المعادلة (١٢) تم الحصول على المعادلتين 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:

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

أعطت الثلاث طرائق Yates و Harry و H&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 (٦).

الجدول ٦: نتائج تحليل التباين عند فقدان قيمة مشاهدين تقعان في نفس القطاع.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	Cal. F بعد التقدير
Yates	٥.٢٧٢	٠.١٠٦	١.٤١٢	١٣.٣٢١
	٥.٨٠٤			
Harry	٥.٢٧٢	٠.١٠٦	١.٤١٢	١٣.٣٢١
	٥.٨٠٤			
Rubin	٣.٥٥٤	٠.٢٨٨	٢.٣١٧	٨.٠٤٥
	٤.٢٦٤			
H&G	٥.٢٧٢	٠.١٠٦	١.٤١٢	١٣.٣٢١
	٥.٨٠٤			

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 و H&G يجب ملاحظ الآتي: هناك مشاهدتان مفقودتان تعودان إلى نفس المعاملة وهما: y_{51} و y_{52} ، وهناك مشاهدتان مفقودتان موقعهما في نفس القطاع وهما: y_{12} و y_{52} ، كما يوجد مشاهدتان مختلفتان في موقعهما ويعودان إلى معاملتين مختلفتين وهما: y_{12} و y_{51} .

وقد قدرت قيم المشاهدات المفقودة كالآتي:

طريقة Yates: بعد الاستعانة بالمعادلة (١) 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 (٢ و ٣ و ٤ و ٥ و ٦ و ٧) القيم التقديرية الآتية: $\hat{y}_{51} = 5.276$ و $\hat{y}_{52} = 5.469$ و $\hat{y}_{12} = 5.165$.

طريقة Rubin: بوساطة المعادلة (٩) 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 $p_{3 \times 1}$ هي:

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

ومن المعادلتين (١٠) و (١١)، نحصل على:

$$A_{kk} = 0.640$$

$$A_{kk'} = 0.040$$

لتكون عناصر المصفوفة $A_{3 \times 3}$ 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 (٨) 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 (١٠) المعدلات الآتية 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 الآتية قدرت القيم المفقودة، فكانت: $\hat{y}_{12} = 5.165$, $\hat{y}_{52} = 5.476$, $\hat{y}_{51} = 5.272$.
يلاحظ توافق نتائج الطرائق الثلاث Yates و Harry و H&G في القيمة التقديرية للملاحظة المفقودة واختلفت عنهم طريقة Rubin، وقد وضعت نتائج تحليل التباين للطرائق الأربع في الجدول (٧).

الجدول ٧: تحليل التباين عند فقدان قيم ثلاث مشاهدات.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	Cal. F بعد التقدير
Yates	٥.٢٧٢	٠.٠٣٥	١.٢٠٥	٣٤.٤٢٩
	٥.٤٧٥			
	٥.١٦٥			
Harry	٥.٢٧٦	٠.٠٣٤	١.٢٠١	٣٥.٣٢٤
	٥.٤٦٩			
	٥.١٦٥			
Rubin	٣.٨٤١	٠.٣٨٦	١.٦١٦	٤.١٨٧
	٢.٣٩١			
	٣.٧٤٨			
H&G	٥.٢٧٢	٠.٠٣٥	١.٢٠٥	٣٤.٤٢٩
	٥.٤٧٦			
	٥.١٦٥			

الاستنتاجات:

- ندرج في أدناه ما تم ملاحظته من نتائج على شكل نقاط:
١. عند فقدان قيمة مشاهدة واحدة تتساوى الطرائق الأربع في مقدار القيمة التقديرية للمشاهدة المفقودة، ومن ثم الحصول على نفس جدول تحليل التباين.
 ٢. إن الطرائق الثلاث وهي: Yates و Harry و Haseman & Gaylar (H&G) 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.
 ٣. لم يكن لموقع القيم المفقودة تأثير 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.
 ٤. 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) كأفضل طريقة إذ تعطي قيم تقديرية مباشرة من خلال حل عدد من المعادلات الآتية 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، ١٩٩٨، "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، ٢٠٠٤، "تقييم ثلاث طرائق لمعالجة القيم المفقودة وتأثرها بالقيم الشاذة في تصميم العبور البسيط"، مجلة علوم الرافدين، المجلد ١٥، العدد ٢ خاص بالحاسبات والرياضيات والإحصاء، ص ١٧-٢٦.
- الراوي، خاشع محمود وخلف الله، عبد العزيز محمد، ٢٠٠٠، "تصميم وتحليل التجارب الزراعية"، الطبعة الثانية، دار الكتب للطباعة والنشر، جامعة الموصل.
- السنجاري، محمد عبد المجيد بدل، ٢٠٠١، "معالجة القيم المفقودة وتأثرها بالقيم الشاذة في تصميم العبور البسيط"، رسالة ماجستير، كلية علوم الحاسبات والرياضيات-جامعة الموصل.
- المشهداني، محمود حسن والمشهداني، كمال علوان، ٢٠٠٢، "تصميم وتحليل التجارب"، الدار الجامعية للطباعة والنشر والترجمة، جامعة بغداد.
- هدبة، وكاع علي، ٢٠٠٥، "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; cited 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 economics", 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.