بحوث مستقبلية (١٩) ٢٠٠٧
ص ص [١٦٣-١٨٤]

مقارنة أربع طرائق لتقدير القيم المفقودة في تصميم المربع اللاتيني

مروان عبد العزيز دبدوب[١]
وكاع علي هدبة[٢]

الملخص

استخدمت أربع طرائق لتقدير القيم المفقودة في تصميم المربع اللاتيني، وهي: Yates وHarry وRubin وHaseman and Gaylar. وقد تمت المفاضلة بين هذه الطرائق على وفق المعايير الاحصائية الآتية: انخفاض قيمة متوسط مربعات الخطأ (MSe)، وارتفاع قيمتي: متوسط مربعات المعاملات (MSt) والفرق المعنوي بين المعاملات (F cal.)، مع الأخذ بالاعتبار سهولة إجراء التحاليل الاحصائية. وقد لوحظ أن طريق Yates هي أكفأ الطرائق ولكنها تزداد تعقيداً كلما زاد عدد القيم المفقودة، تليها طريقة Harry المعتدلة التطبيق، ومن ثم طريقة Haseman and Gaylar، وكان لطريقة Rubin أقل كفاءة. وقد اتصفت الطريقتان الأخيرتان بسهولة التطبيق.

Abstract

Four methods for tackling missing values in Latin square design have been presented: Yates, Harry, Rubin, and the method of Haseman and Gaylar. To make preference among these methods some statistical measurements have been used, which are: lowest value of the mean square error (MSe), highest value of the mean square treatments (MSt) and the value of significant differences between treatments (F cal.). The easiest path of statistical analysis has been taken into account. It has been found that the most preferable method is Yates' method which has the most complicated application whenever the number of missing values are increased, followed by Harry method which has a moderate application difficulty, then Haseman and Gaylar method, and finally Rubin method. The last two methods have the easiest applications.

---

(١) أستاذ مساعد–قسم الإحصاء والمعلوماتية-كلية علوم الحاسبات والرياضيات–جامعة الموصل.

(٢) مدرس مساعد – كلية الزراعة–جامعة كركوك.

تاريخ الاستلام: ٢٠٠٧/٠١/١٦
تاريخ قبول النشر: ٢٠٠٧/٠٦/٠٧

المقدمة

قد يحصل فقدان لبعض مشاهدات التجربة عندما لا يمكن السيطرة بشكل كامل على المؤثرات المختلفة خاصة إذا أقيمت التجربة تحت الظروف الطبيعية كالتجارب الزراعية. وقد يحدث فقدان لبعض المشاهدات حتى في التجارب المختبرية المسيطر على مؤثراتها.

إن فقدان قيم بعض المشاهدات يؤدي إلى عدم اتزان التصميم وتعقيد عمليات التحليل الإحصائي، فضلاً عن أن انخفاضا في عدد ظهور المعاملات يسبب نقصان في درجات الحرية للخطأ التجريبي الذي يؤدي بدوره إلى ارتفاع قيمة متوسط مربعات الخطأ. ويتم التغلب على هذه المشكلة من خلال اختيار قيم تقديرية تعطي أقل فرق بين مجموع مربعات الخطأ الناتج من تقدير قيم المشاهدات المفقودة والخطأ التجريبي في حالة عدم وجود قيم مفقودة.

قد يعتقد البعض أن إعادة التجربة بتطبيق ذات التصميم أفضل من معالجة القيم المفقودة، ولكن اتباع هذا الاعتقاد يعني مضاعفة الكلفة وتأخير نتائج التجربة وهدر جهد ليس بالهين، ومن ثم قد يحصل الباحث على نتائج أسوأ من سابقتها، لذا عني الكثير من الباحثين بكيفية معالجة القيم المفقودة، وأول من حاول ذلك هو Allan and Wishart (1930) إذ أشارا إلى طريقة لتقدير قيمة مفقودة واحدة، تلاهما Yates (1933) الذي وضع طريقة التعويض المتوالي لتقدير أكثر من قيمة مفقودة واحدة.

من خلال تحليل عدد من المعادلات الآنية بعدد القيم المفقودة توصل Harry (1964) إلى تقدير للقيم المفقودة في تصميمي القطاعات العشوائية الكاملة والمربع اللاتيني. وقد استخدم Rubin (1972) أسلوب المصفوفات وهو أسلوب غير متتالٍ (non-Itrative) لتقدير القيم المفقودة وذلك بالاستعانة بطريقة المربعات الصغرى وجعل مجموع مربعات خطأ الخلايا مساويا للصفر. وعالج Haseman and Gaylar (1973) مشكلة القيم المفقودة باستخدام صيغة مباشرة تعتمد على عدد القيم المفقودة وموقعها في التجربة.

قدم الجبوري (1998) دراسة تحليلية تعني بالقيم الشاذة والمفقودة في تصميمي العشوائي الكامل والمربع اللاتيني في حالة تكرار مشاهدات العينة. وقارن هدبة (2005) بين خمس طرائق لمعالجة القيم المفقودة في تصميمي القطاعات العشوائية الكاملة والمربع اللاتيني.

هدف البحث

سيتم اتباع أربع طرائق لتقدير قيم المشاهدات المفقودة وهي: Yates وHarry وRubin و(Haseman and Gayler (H & G، وستجرى المفاضلة بين هذه الطرائق في حالة تقديرها لأعداد مختلفة من قيم المشاهدات المفقودة ذات المواقع المختلفة في التجربة، وذلك بالاستعانة بالمعايير الاحصائية الآتية: انخفاض قيمة متوسط مربعات الخطأ mean square error (MSe)، وارتفاع قيمة كل من: متوسط مربعات المعاملات mean square treatments (MSt) وقيمة الفرق المعنوي بين المعاملات value of significant difference between treatments (F cal.). فضلا عن سهولة تطبيق الطريقة.

تصميم المربع اللاتيني

هو التصميم الذي يجمع بين تصميمين عشوائيين كاملين في اتجاهين متعامدين، يدعى أحدهما صفوف بعدد r والآخر أعمدة بعدد c، وتوزع المعاملات بعدد t على الوحدات التجريبية لتظهر مرة واحدة في كل صف ومرة في كل عمود، وبذلك يكون $r = c = t$ (المشهداني والمشهداني، ٢٠٠٢).

يعد تحليل التباين في تصميم المربع اللاتيني الأسلوب الرياضي الذي يختبر الفرضية الخاصة بتساوي الأوساط الحسابية لعدد من المجتمعات، والفرضية هي:

$$H_o : \mu_1 = \mu_2 = ... = \mu_t \quad vs \quad H_1 : at\,least\,one\,of\;\mu_k\;is\,differ$$

وتتم بموجب تحليل التباين تجزئة مجموع المربعات الكلي إلى مصادره المختلفة (Keller and Warrock ,2003)، وتلخص نتائج التحليل في الجدول (1).

الجدول 1: تحليل التباين لتصميم المربع اللاتيني.

S.O.V.	D.F.	S.S.	M.S.	F cal.
Rows	r-1	$SS_r = r\sum_i (\bar{y}_i. - \bar{y}...)^2$	MSr	
Columns	c-1	$SS_c = r\sum_j (\bar{y}._j - \bar{y}...)^2$	MSc	
Treatment	t-1	$SS_t = r\sum_k (\bar{y}_k - \bar{y}...)^2$	MSt	$\frac{MSt}{MSe}$
Error	(r-1)(c-2)	$SS_e = SS_T - SS_r - SS_C - SS_t$	MSe	
Total	rc-1	$SS_T = \sum_i \sum_j (y_{ij} - \bar{y}...)^2$		

والنموذج الرياضي لهذا التصميم هو:

$$y_{ij(k)} = \mu + \rho_i + \gamma_j + \tau_{(k)} + \varepsilon_{ij(k)}$$
$$k = 1,2,\cdots,t \text{ ، } j = 1,2,\cdots,c \text{ ، } i = 1,2,\cdots,r$$

μ : المتوسط العام.

ρ_i : تأثير الصف i.

γ_j : تأثير العمود j.

$\tau_{(k)}$: تأثير المعاملة k.

$\varepsilon_{ij(k)}$: قيمة الخطأ التجريبي في الصف i والعمود j ، إذ طبقت المعاملة k.

طرائق تقدير القيم المفقودة

وضعت أربع طرائق شائعة الاستخدام لتقدير القيم المفقود في تصميم المربع اللاتيني (الراوي وخلف الله، ٢٠٠٠)، (المشهداني والمشهداني، ٢٠٠٢)، (هدبة، ٢٠٠٥)، وهي:

أ- طريقة Yates

نفرض فقدان قيمة المشاهدة الناتجة من تطبيق المعاملة الأولى على الوحدة التجريبية الواقعة في الصف الأول والعمود الأول، إن قيمة مجموع مربعات الخطأ في هذه الحالة هي:

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{t}e^2{}_{ij(k)} = y_{11(1)}^2 + \sum_{i=2}\sum_{j=2}\sum_{k=2} y_{ij(k)}^2 - \frac{1}{r}\Big[\ \left(Y_{1}.+y_{11(1)}\right)^2 + \sum_{i=2} Y_{i.}^2 + \left(Y_{\cdot 1}+y_{11(1)}\right)^2$$

$$+\sum_{j=2} Y_{.1}^2 + \left(Y_{\cdot(1)}+y_{11(1)}\right)^2 + \sum_{k=2} Y_{.(k)}^2\ \Big] + \frac{2\left(Y..+y_{11(1)}\right)^2}{r^2}$$

بعد أخذ المشتقة بالنسبة إلى $y_{11(1)}$ ينتج :

$$\frac{\partial\sum\sum e_{ij(k)}^2}{\partial y_{11(1)}} = 2\hat{y}_{11(1)} - \frac{2}{r}\left(y_{1.}+\hat{y}_{11(1)}+y_{.1}+\hat{y}_{11(1)}+\hat{y}_{11(1)}+y_{.1}\right) + \frac{4\left(y_{..}+\hat{y}_{11(1)}\right)}{r^2}$$

وبعد مساواة المعادلة بالصفر وتبسيط المقدار نحصل على:

$$\hat{y}_{11(1)} = \frac{r\left(y_{1.}+y_{.1}+y_{.1}\right)-2y_{..}}{(r-1)(r-2)} \qquad \cdots \qquad (1)$$

وعلية فإن $\hat{y}_{11(1)}$ هي القيمة التقديرية للقيمة المفقودة.

ب- طريق Harry

اعتمد Harry (1964) على أسلوب التعويض غير المتتالي، ولكل قيمة مفقودة وضع لها معادلة، وقد أخذ بنظر الاعتبار مواقع القيم المفقودة في التجربة، وقد استخدم أسلوب Yates نفسه في حالة فقدان مشاهدة واحدة. وللطريقة الأسلوب الآتي:

١- في حالة فقدان قيمة واحدة: تتبع طريقة Yates نفسها ، وتستخدم المعادلة (١) لإيجاد قيمة مفقودة واحدة.

٢ – في حالة فقدان قيمتين من المعاملة نفسها: نفرض أن y_1 وy_2 القيمتان المفقودتان، وأن:

R1: مجموع الصف الذي يحتوي على القيمة المفقودة الأولى.

R2: مجموع الصف الذي يحتوي على القيمة المفقودة الثانية.

C1: مجموع العمود الذي يحتوي على القيمة المفقودة الأولى.

C2: مجموع العمود الذي يحتوي على القيمة المفقودة الثانية.

T12: مجموع قيم مشاهدات المعاملة التي تضم القيمتين المفقودتين.

$Y..$: المجموع الكلي.

تحسب قيمة مجموع مربعات الخطأ (SSe) في هذه الحالة كالآتي:

$$SS_e = y_1^2 + y_2^2 - \frac{(R_1 + y_1)^2}{r} + \frac{(R_2 + y_1)^2}{r} - \frac{(c_1 + y_1)^2 + (c_2 + y_2)^2}{r}$$
$$- \frac{(T_{12} + y_1 + y_2)}{r} + \frac{2(Y... + y_1 + y_2)^2}{r^2}$$

بعد اخذ المشتقة بالنسبة لـ y2, y1:

$$Q_1 = \frac{\partial SSe}{\partial y_1} \quad , \quad Q_2 = \frac{\partial SSe}{\partial y_2}$$

نحصل على:

$$Q_1 = \hat{y}_1 - \frac{(R_1 + \hat{y}_1)}{r} - \frac{(C_1 + \hat{y}_1)}{r} - \frac{(T_{12} + \hat{y}_1 + \hat{y}_2)}{r} + \frac{2(Y... + \hat{y}_1 + \hat{y}_2)}{r^2}$$
$$Q_2 = \hat{y}_2 - \frac{(R_2 + \hat{y}_2)}{r} - \frac{(C_2 + \hat{y}_2)}{r} - \frac{(T_{12} + \hat{y}_1 + \hat{y}_2)}{r} + \frac{2(Y... + \hat{y}_1 + \hat{y}_2)}{r^2}$$

$$Q_1 = (r-1)(r-2)\hat{y}_1 - (r-2)\hat{y}_2 = r(R_1 + C_1 + T_{12}) - 2Y..$$
$$Q_2 = (r-1)(r-2)\hat{y}_2 - (r-2)\hat{y}_1 = r(R_2 + C_2 + T_{12}) - 2Y..$$
$$r(r-2)^2\,\hat{y}_1 = (r-1)Q_1 + Q_2$$
$$r(r-2)^2\,\hat{y}_2 = (r-1)Q_2 + Q_1$$

$$\hat{y}_1 = \frac{(r-1)Q_1 + Q_2}{r(r-2)^2} \qquad ... \qquad (2)$$

$$\hat{y}_2 = \frac{(r-1)Q_2 + Q_1}{r(r-2)^2} \quad \ldots \quad (3)$$

حيث أن $\hat{y}_1$ و $\hat{y}_2$ هما القيمتين التقديريتين للقيمتين المفقودتين y_1 وy_2 على التوالي، وأن:

$$Q_1 = r(R_1 + C_1 + T_{12}) - 2Y.. \quad , \quad Q_2 = r(R_2 + C_2 + T_{12}) - 2Y..$$

٣- في حالة فقدان قيمتان في نفس العمود: نفرض أن y_1 وy_2 القيمتان المفقودتان، وأن:

R1: مجموع الصف الذي يحتوي على القيمة المفقودة الأولى.

R2: مجموع الصف الذي يحتوي على القيمة المفقودة الثانية.

T1: مجموع المعاملة التي تحتوي على القيمة المفقودة الأولى.

T2: مجموع المعاملة التي تحتوي على القيمة المفقودة الثانية.

C12: مجموع قيم مشاهدات العمود الذي يضم القيمتين المفقودتين.

$Y..$: المجموع الكلي.

وبعد اخذ المشتقة لمجموع مربعات الخطأ (SSe) بالنسبة لـ y_1 و y_2 :

$$Q_1 = \frac{\partial SSe}{\partial y_1} \quad , \quad Q_2 = \frac{\partial SSe}{\partial y_2}$$

نحصل على الآتي:

$$Q_1 = \hat{y}_1 - \frac{(R_1 + \hat{y}_1)}{r} - \frac{(T_1 + \hat{y}_1)}{r} - \frac{(C_{12} + \hat{y}_1 + \hat{y}_2)}{r} + \frac{2(Y... + \hat{y}_1 + \hat{y}_2)}{r^2}$$

$$Q_2 = \hat{y}_2 - \frac{(R_2 + \hat{y}_2)}{r} - \frac{(T_2 + \hat{y}_2)}{r} - \frac{(C_{12} + \hat{y}_1 + \hat{y}_2)}{r} + \frac{2(Y... + \hat{y}_1 + \hat{y}_2)}{r^2}$$

$$Q_1 = (r-1)(r-2)\hat{y}_1 - (r-2)\hat{y}_2 = r(T_1 + R_1 + C_{12}) - 2Y..$$

$$Q_2 = (r-1)(r-2)\hat{y}_2 - (r-2)\hat{y}_1 = r(T_2 + R_2 + C_{12}) - 2Y..$$

وبعد التبسيط للحصول على $\hat{y}_1$ و $\hat{y}_2$:

$$\hat{y}_1 = \frac{(r-1)Q_1 + Q_2}{r(r-2)^2} \quad \ldots \quad (4)$$

$$\hat{y}_2 = \frac{(r-1)Q_2 + Q_1}{r(r-2)^2} \quad \ldots \quad (5)$$

إذ إن $\hat{y}_1$ و $\hat{y}_2$ هما القيمتان التقديريتان للقيمتين المفقودتين y_1 وy_2 على التوالي، وأن:

$$Q_1 = r(T_1 + R_1 + C_{12}) - 2Y.. \quad , \quad Q_2 = r(T_2 + R_2 + C_{12}) - 2Y..$$

٤- في حالة فقدان قيمتين في الصف نفسه: نفرض القيمتين المفقودتين y_1 , y_2. وأن:

C1: **مجموع العمود الذي يحتوي على القيمة المفقودة الأولى**

C2: **مجموع العمود الذي يحتوي على القيمة المفقودة الثانية**

T1: **مجموع المعاملة التي تحتوي على القيمة المفقودة الأولى**

T2: **مجموع المعاملة التي تحتوي على القيمة المفقودة الثانية**

R12: **مجموع قيم مشاهدات الصف الذي يضم القيمتين المفقودتين.**

$Y..$: **المجموع الكلي.**

بعد اخذ المشتقة لمجموع مربعات الخطأ (SSe) **بالنسبة لـ y_2 وy_1: تنتج:**

$$Q_1 = \frac{\partial SSe}{\partial y_1} \quad , \quad Q_2 = \frac{\partial SSe}{\partial y_2}$$

نحصل على المعادلتين الآتيتين:

$$Q_1 = \hat{y}_1 - \frac{(C_1 + \hat{y}_1)}{r} - \frac{(T_1 + \hat{y}_1)}{r} - \frac{(R_{12} + \hat{y}_1 + \hat{y}_2)}{r} + \frac{2(Y... + \hat{y}_1 + \hat{y}_2)}{r^2}.$$

$$Q_2 = \hat{y}_2 - \frac{(C_2 + \hat{y}_2)}{r} - \frac{(T_2 + \hat{y}_2)}{r} - \frac{(R_{12} + \hat{y}_1 + \hat{y}_2)}{r} + \frac{2(Y... + \hat{y}_1 + \hat{y}_2)}{r^2}$$

وبعد التبسيط للحصول على $\hat{y}_1$ و $\hat{y}_2$:

$$\hat{y}_1 = \frac{(r-1)Q_1 + Q_2}{r(r-2)^2} \quad \ldots \quad (6)$$

$$\hat{y}_2 = \frac{(r-1)Q_2 + Q_1}{r(r-2)^2} \qquad \ldots \qquad (7)$$

حيث إن:

$Q_1 = r(R_{12} + C_1 + T_1) - 2Y.. \quad , \quad Q_2 = r(R_{12} + C_1 + T_1) - 2Y..$

يلاحظ تناظر الأسلوب المتبع لإيجاد قيمتين تقديريتين عندما تعودان المشاهدتين المفقودتين إلى المعاملة نفسها (المعادلة ٢ و٣) أو تقعان في العمود نفسها (المعادلتين ٤ و٥) أو تقعان في الصف نفسها (المعادلتين ٦ و٧)، والاختلاف فيما بينها في احتساب قيمتي Q_1 و Q_2.

٥- في حالة فقدان قيمتي مشاهدتين تعودان إلى معاملتين مختلفتين وتقعان في صفين مختلفين وعمودين مختلفين. لنفرض الرموز الآتية:

C_1: مجموع العمود الذي يحتوي على القيمة المفقودة الأولى.

C_2: مجموع العمود الذي يحتوي على القيمة المفقودة الثانية.

T_1: مجموع المعاملة التي تحتوي على القيمة المفقودة الأولى.

T_2: مجموع المعاملة التي تحتوي على القيمة المفقودة الثانية.

R_1: مجموع الصف الذي يحتوي على القيمة المفقودة الأولى.

R_2: مجموع الصف الذي يحتوي على القيمة المفقودة الثانية.

$Y..$: المجموع الكلي.

f : درجات الحرية للخطأ، حيث إن: $f = (r-1)(r-2)$.

وبتبسيط المعادلة التي أشار إليها Harry (1964)، نحصل على:

$$fy_1 + 2y_2 = r(R_1 + C_1 + T_1) - 2Y.. = Q_1$$
$$2y_2 + fy_2 = r(R_2 + C_2 + T_2) - 2Y.. = Q_2$$
$$(f^2 - 4)y_1 = fQ_1 - 2Q_2 \qquad \ldots \qquad (8)$$
$$(f^2 - 4)y_2 = fQ_2 - 2Q_2 \qquad \ldots \qquad (9)$$

حيث إن :

$$Q_1 = r(R_1 + C_1 + T_1) - 2Y..$$
$$Q_2 = r(R_2 + C_2 + T_2) - 2Y..$$

٦- في حالة فقدان أكثر من قيمتين:

حتى تقدر القيم المفقودة يتبع الآتي: تترك قيمتان مفقودتان وتعوض كلٌ من بقية القيم المفقودة بالوسط الحسابي للمعاملة العائدة لها، ومن ثم يتبع أسلوب إيجاد قيمتين مفقودتين، أما القيم المفقودة التي عوضت بالوسط الحسابي فيعاد تقديرها بالأسلوب نفسه.

ت- طريقة Rubin

استخدم Rubin (1972) أسلوب المصفوفات لتقدير m من القيم المفقودة بتطبيق الآتي:

$$\hat{y}' = -p' A^{-1} \qquad (10)$$

$\hat{y}$: متجه القيم التقديرية $m\times 1$، حيث إن: $k = 1,2,\ldots,m$.

p : متجه $m\times 1$ قيم وتحسب عناصره كالآتي:

$$p_k = -\frac{Y_i..+Y._j+Y._{(k)}}{r} - \frac{2Y..}{r^2} \qquad \ldots \qquad (11)$$

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

$$a_{kk} = 1-\frac{1}{r}-\frac{1}{t}+\frac{1}{rt} \qquad \ldots \qquad (12)$$

$$a_{kk'} = \frac{1}{rt} \qquad \ldots \qquad (13)$$

ث- طريقة Haseman & Gaylar (H & G)

عند وجود m من القيم المفقودة أشار Haseman and Gaylar (1973) إلى تكوين m من المعادلات الآنية وبتحليلها يتم الحصول على تقدير للقيم المفقودة، والصيغة العامة للمعادلات الآنية هي:

$$(r-1)(r-2)y_h + \sum_{g\neq h}^{m} y_g\left[2-r\sum_{i=1}^{3} y_{gh}(Ai)\right] = r\sum_{i=1}^{3} T_h(Ai) - 2Y.. \qquad \ldots \qquad (14)$$

$$y_{gh}(A_1) = \begin{cases} 1 & \text{إذا كانت القيمتان المفقودتان } y_h \text{ و } y_g \text{ واقعتين في الصف نفسه} \\ 0 & \text{غير ذلك} \end{cases}$$

$$Y_{gh}(A_2) = \begin{cases} 1 & \text{إذا كانت القيمتان المفقودتان } Y_h \text{ و } Y_g \text{ واقعتين في العمود نفسه} \\ 0 & \text{غير ذلك} \end{cases}$$

$$Y_{g\,h}(A_3) = \begin{cases} 1 & \text{إذا كانت القيمتان المفقودتان } Y_h \text{ و } Y_g \text{ واقعتين في المعاملة نفسها} \\ 0 & \text{غير ذلك} \end{cases}$$

مجموع قيم المشاهدات الواقعة في الصف الذي يتضمن القيمة المفقودة: $T_{g\,h}(A_1)$.

مجموع قيم المشاهدات الواقعة في العمود الذي يتضمن القيمة المفقودة: $T_{g\,h}(A_2)$.

مجموع قيم المشاهدات الواقعة في المعاملة التي تتضمن القيمة المفقودة: $T_{g\,h}(A_3)$.

المجموع الكلي للمشاهدات: $Y..$.

بعد تقدير القيم المفقودة توضع في محلها بين البيانات ومن ثم تتبع خطوات تحليل التباين، مع مراعاة الآتي: طرح درجة حرية واحدة من الخطأ عن كل قيمة مفقودة، ويعدل مجموع مربعات المعاملات لكل قيمة مفقودة باستخدام معادلة التصحيح الآتية (هدبة، ٢٠٠٥):

$$ss't = sst - \left[\frac{Y.. - Y_i. - Y._j - (t-1)Y._{(k)}}{(r-1)(r-2)}\right]^2 \quad \cdots \quad (15)$$

حيث أن:

SSt: مجموع مربعات المعاملات قبل عملية تقدير القيمة المفقودة.

$Y..$: المجموع العام للتجربة عدا قيمة المشاهدة المفقودة.

$Y_i.$: مجموع مشاهدات الصف عدا قيمة المشاهدة المفقودة.

$Y._j$: مجموع مشاهدات العمود عدا قيمة المشاهدة المفقودة.

$Y._{(k)}$: مجموع مشاهدات المعاملات عدا قيمة المشاهدة المفقودة.

تطبيق: سيتم تطبيق الطرائق الأربع لتقدير القيم المفقودة على بيانات ذكرها هدبة (٢٠٠٥)، ويشير إليها الجدول (٢). وفي أدناه وصف للتصميم والتجربة والبيانات.

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

أعمدة → صفوف ↓	C1	C2	C3	C4	C5	$Y_{i\cdot}$
R1	4.97 t1	6.35 t5	5.87 t4	3.77 t3	4.33 t2	25.29
R2	4.58 t2	5.03 t1	5.29 t5	5.16 t4	3.50 t3	23.56
R3	3.75 t3	4.49 t2	4.89 t1	5.00 t5	5.15 t4	23.28
R4	5.43 t4	4.02 t3	4.49 t2	4.85 t1	4.91 t5	23.7
R5	6.48 t5	6.19 t4	3.81 t3	4.35 t2	4.84 t1	25.67
$Y_{\cdot j}$	25.21	26.08	24.35	23.13	22.73	121.5

$Y_{\cdot(1)} = 24.58$	$Y_{\cdot(2)} = 22.24$	$Y_{\cdot(3)} = 18.85$	$Y_{\cdot(4)} = 27.80$	$Y_{\cdot(5)} = 28.03$

أشار هدبة (٢٠٠٥) أن البيانات هي نتيجة تجربة أجراها قسم البحوث الزراعية في موقع الرشيدية–محافظة نينوى، إذ أقيمت التجربة لمقارنة خمسة أصناف من القطن (خمس معاملات: t_k, $k=1,2,3,4,5$) وهي: الصنف الأول (t_1) (نازلي ٥٠٣) والصنف الثاني (t_2) (k٣٤٧١) والصنف الثالث (t_3) (طاقة ٢) والصنف الرابع (t_4) (مرسومي ١) والصنف الخامس (t_5) (دلتا بايه ٥٤٠٩). لاحتواء التجربة على خمس معاملات فقد استنبط تصميم المربع اللاتيني (لملاءمته لمتطلبات البحث) بخمسة صفوف وخمسة أعمدة، وبذلك يكون التصميم من ٢٥ وحدة تجريبية وزعت عليها المعاملات عشوائيا بحيث تظهر كل معاملة مرة

واحدة في كل صف وكل عمود. في كل وحدة تجريبية أنشأ أربعة مروز طول المرز خمسة أمتار والمسافة بين كل مرزين متتاليين ٧٥ سنتيميتراً، وزرع نبات واحد في الجورة بمسافة بين نبات وآخر ٢٥ سنتيميتراً، وقد تمت الزراعة بتاريخ ٢٠٠١/٥/٣. وكان أوزان حاصل التجربة من القطن الزهر بالكيلوغرام لكل وحدة تجريبية كما في الجدول (٢).

تحليل التجربة

طبق تحليل التباين على نتائج التجربة وهي كاملة المشاهدات فأعطى الجدول (٣).

الجدول ٣: تحليل التباين لنتائج تجربة خمسة أصناف من القطن عند عدم وجود قيم مفقودة.

S.O.V.	D.F.	S.S.	M.S.	F cal.
treatment	4	12.037	3.009	45.591
rows	4	0.961	0.240	
columns	4	1.567	0.392	
error	12	0.788	0.066	
Total	24	15.353		

فقدان بعض قيم التجربة وطرائق تقديرها:

أ- فقدان قيمة مشاهدة واحدة: نفرض أن قيمة المشاهدة العائدة للمعاملة الأولى الواقعة في الصف الثاني والعمود الثاني ($y_{22(1)} = 5.03$) قد فقدت. أعطت الطرائق التقديرية الأربع النتائج الآتية:

طريقة Yates: أعطت المعادلة (١) القيمة التقديرية الآتية:

$$\hat{y}_{22(1)} = \frac{5(18.53 + 21.05 + 19.55) - 2(116.47)}{12} = 5.226$$

طريقة Harry: كما هو الحال في طريقة Yates، قد استخدمت المعادلة (١) فأعطت القيمة التقديرية ٥,٢٢٦ .

طريقة Rubin: لوجود قيمة مفقودة واحدة فإن المعادلة (١١) تعطي عنصر واحد للمتجه $p_{1\times1}$ مقداره:

$$p_1 = e_{22(1)} = -\frac{18.530}{5} - \frac{21.050}{5} - \frac{19.550}{5} + \frac{2(116.470)}{25} = -2.5084$$

$$\therefore\ p_{1\times1} = (-2.5084)$$

وبالرجوع إلى المعادلة (١٢) نحصل على عنصر واحد للمصفوفة $A_{1\times1}$ **هو:**

$$a_{11} = 1 - \frac{1}{5} - \frac{1}{5} - \frac{1}{5} + \frac{2(1)}{25} = 0.480$$

وبذلك تعطي المعادلة (١٠) تقديراً للقيمة المفقودة مقداره:

$$\hat{y}_{22(1)} = -2.5084(\frac{1}{0.480}) = 5.226$$

طريقة (H&G)**: بتطبيق المعادلة (١٤):**

$$(5-1)(5-2)\hat{y}_{22(1)} = 5(18.530 + 21.050 + 19.550) - 2(116.470)$$

$$\therefore \hat{y}_{22(1)} = 5.226$$

يلاحظ أن الطرائق الأربع أعطت القيمة التقديرية نفسها وهي: $\hat{y}_{22(1)} = 5.226$ **، وبعد تصحيح قيمة مجموع مربعات المعاملات** (SSt) **وطرح درجة حرية واحدة من الخطأ (المعادلة ١٥) نحصل على جدول تحليل التباين (الجدول ٤) الذي يبين نتائج الطرائق الأربع.**

الجدول ٤: تحليل التباين بعد تقدير القيمة المفقودة.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	F cal. بعد التقدير
Yates	5.226	0.070	3.013	43.043
Harry	5.226	0.070	3.013	43.043
Rubin	5.226	0.070	3.013	43.043
H&G	5.226	0.070	3.013	43.043

ب- فقدان قيمة مشاهدتين في المعاملة نفسها: لنفرض أن المشاهدة المفقودة هي التي أخذت المعاملة الخامسة والواقعة في الصف الأول والعمود الثاني ($y_{12(5)}$=6.35) **والمشاهدة الأخرى المفقودة قد أخذت المعاملة الخامسة والواقعة في الصف الخامس والعمود الأول** ($y_{51(5)}$=6.48)**، ويتم تقدير قيمتهما كالآتي:**

طريقة Yates**: باستخدام طريقة التعويض المتتالي والمعادلة (١) تم الحصول على:**

$\hat{y}_{51(5)} = 5.445$ ، $\hat{y}_{12(5)} = 5.695$

طريقة Harry: باستخدام المعادلتين (٢ و٣) تم الحصول على قيمتين تقديريتين:

$$\hat{y}_{12(5)} = \frac{4(52.000) + 48.260}{45} = 5.695$$

$$\hat{y}_{51(5)} = \frac{4(48.260) + 52.000}{45} = 5.445$$

طريقة Rubin: أعطت المعادلة (١١) الآتي:

$$p_1 = e_{12(5)} = -\frac{18.940}{5} - \frac{19.730}{5} - \frac{15.200}{5} + \frac{2(108.670)}{25} = -2.080$$

$$p_2 = e_{51(5)} = -\frac{19.190}{5} - \frac{18.730}{5} - \frac{15.200}{5} + \frac{2(108.670)}{25} = -1.930$$

$$p^{'} = (-2.080 \quad -1.930)$$

وقد أعطت المعادلتين (١٢ و١٣) قيمة عناصر المصفوفة $A_{2\times2}$ وكالآتي:

$$A = \begin{pmatrix} 0.480 & 0.080 \\ 0.080 & 0.480 \end{pmatrix}$$

ومن خلال المعادلة (١٠) تم الحصول على القيمتين المفقودتين وهما:

$\hat{y}_{51(5)} = 3.393$ ، $\hat{y}_{12(5)} = 3.768$

طريقة H&G: أعطت المعادلة (١٤) المعادلتين الآنيتين الآتيتين:

$$12\hat{y}_{12(5)} - 3\hat{y}_{51(5)} = 52.000$$

$$12\hat{y}_{51(5)} - 3\hat{y}_{12(5)} = 48.260$$

وبتحليلهما تم الحصول على القيمتين التقديريتين: $\hat{y}_{12(5)} = 5.695$ و $\hat{y}_{51(5)}$ 5.445 .

أدخلت القيم المقدرة إلى البيانات، ويوضح الجدول (٥) نتائج تحليل التباين بعد تصحيح قيمة مجموع مربعات المعاملات (SSt) للقيمتين المقدرتين (المعادلة ١٥) وطرح درجتي حرية من الخطأ.

الجدول ٥: بعض نتائج تحليل التباين عند فقدان قيمتين في المعاملة نفسها.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	F cal. بعد التقدير
Yates	٥,٦٩٥	٠,٠٢٣	٢,٣٢٧	١٠١,١٧٤
	٥,٤٤٥			
Harry	٥,٦٩٥	٠,٠٢٣	٢,٣٢٧	١٠١,١٧٤
	٥,٤٤٥			
Rubin	٣,٧٦٨	٠,٣٠٩	١,٨٦١	٦,٠٢٣
	٣,٣٩٣			
H&G	٥,٦٩٥	٠,٠٢٣	٢,٣٢٧	١٠١,١٧٤
	٥,٤٤٥			

ت- فقدان قيمة مشاهدتين للعمود نفسه: لنفرض أن المشاهدتين $y_{12(5)}=6.35$ **و**$y_{22(1)}=5.03$ **والواقعتين في العمود الثاني قد فقدتا. وسيتم تقديرهما كالآتي:**

طريقة Yates: **تعوض أولاً إحدى قيم المشاهدات المفقودة بالوسط الحسابي لبقية قيم مشاهدات المعاملة نفسها ، ثم الرجوع إلى المعادلة (١) لتقدير المشاهدة المفقودة الأخرى، وهكذا باتباع أسلوب التعويض المتتالي كان التقدير الآتي:** $\hat{y}_{12(5)} = 5.980$ **و** $\hat{y}_{22(1)} = 5.133$.

طريقة Harry: **بتطبيق المعادلتين (٤ و٥) تم الحصول على القيمتين التقديريتين وهما:**

$$\hat{y}_{12(5)} = \frac{4(56.360)+43.660}{45} = 5.980$$

$$\hat{y}_{22(1))} = \frac{4(43.660)+56.360}{45} = 5.133$$

طريقة Rubin: باستخدام المعادلة رقم (١١) نحصل على المعلومات الآتية:

$$p_1 = e_{12(5)} = -\frac{18.550}{5} - \frac{14.700}{5} - \frac{19.550}{5} + \frac{2(110.120)}{25} = -1.750$$

$$p_2 = e_{22(1)} = -\frac{18.940}{5} - \frac{14.700}{5} - \frac{21.680}{5} + \frac{2(110.120)}{25} = -2.254$$

$$\therefore p' = (-1.750 \quad -2.254)$$

وبتطبيق المعادلتين (١٢ و١٣) نحصل على عناصر المصفوفة $A_{2\times2}$:

$$A = \begin{pmatrix} 0.480 & 0.080 \\ 0.080 & 0.480 \end{pmatrix}$$

ومن المعادلة (١٠) يكون مقدار القيمتين التقديريتين كالآتي:

$\hat{y}_{12(5)} = 4.205$ ، $\hat{y}_{22(1)} = 2.945$.

طريقة H&G: أعطت المعادلة (١٢) المعادلتين الآنيتين الآتيتين:

$$12\hat{y}_{12(5)} - 3\hat{y}_{22(1))} = 56.360$$

$$12\hat{y}_{22(1)} - 3\hat{y}_{12(5)} = 43.660$$

وبتحليلهما تم الحصول على القيمتين التقديريتين: $\hat{y}_{12(5)} = 5.980$ و $\hat{y}_{22(1)} = 5.133$.

بعد إدخال القيم المقدرة إلى البيانات اجري تحليل التباين، ويوضح الجدول (٦) نتائج التحليل بعد تصحيح قيمة مجموع مربعات المعاملات (SSt) للقيمتين المقدرتين (المعادلة ١٥) وطرح درجتي حرية من الخطأ.

الجدول ٦: بعض نتائج تحليل التباين عند فقدان قيمتين في العمود نفسه.

الطرائق التقديرية	القيمة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	F cal. بعد التقدير
Yates	٥,٩٨٠	٠,٠٧١	٢,٦٧١	٣٧,٦٢٠
	٥,١٣٣			
Harry	٥,٩٨٠	٠,٠٧١	٢,٦٧١	٣٧,٦٢٠
	٥,١٣٣			
Rubin	٤,٢٠٥	٠,٠٧٩	٢,٧٥٧	٣٤,٨٩٩
	٢,٩٤٥			
H&G	٥,٩٨٠	٠,٠٧١	٢,٦٧١	٣٧,٦٢٠
	٥,١٣٣			

ث- فقدان ثلاث مشاهدات: لنفرض أن قيم المشاهدات الآتية قد فقدت: المعاملة الأولى الواقعة في الصف الثاني والعمود الثاني ($y_{22(1)}=5.03$) والمعاملة الخامسة الواقعة في الصف الأول والعمود الثاني ($y_{12(5)}=6.35$) والمعاملة الخامسة الواقعة في الصف الخامس والعمود الأول ($y_{51(5)}=6.48$). يلاحظ وقوع كل زوج من المشاهدتين المفقودتين كالآتي: في المعاملة نفسها $y_{12(5)}$ و$y_{51(5)}$، وفي العمود نفسه $y_{22(1)}$ و$y_{12(5)}$، ومشاهدتين تختلفان في المعاملة والصف والعمود وهما: $y_{22(1)}$ و$y_{51(5)}$. وستتبع في عملية التقدير الطرائق الآتية:

طريقة Yates: تعوض قيمتين مفقودتين بالوسط الحسابي لقيم مشاهدات المعاملات العائدة لكل منهما، وبالرجوع إلى المعادلة (١) يتم تقدير المشاهدة الثالثة، ثم نعود إلى المعادلة (١) لتقدير إحدى المشاهدتين السابقتين، وهكذا يتبع أسلوب التعويض المتتالي حتى تقارب القيم المقدرة من الثبات. وقد تم التوصل إلى القيم التقديرية الآتية: $\hat{y}_{22(1)} = 5.251$ و$\hat{y}_{12(5)} = 5.745$ و$\hat{y}_{51(5)} = 5.421$.

طريقة Harry: تعوض إحدى القيم المفقودة بالوسط الحسابي لبقية قيم مشاهدات المعاملة العائدة لها، يلاحظ بأن زوج من القيم المفقودة تعودان إلى المعاملة نفسها (المعاملة الخامسة)

وفي هذه الحالة تطبق المعادلتين (٢ و٣)، كما يوجد زوج يقعان في العمود نفسه (العمود الثاني) وهنا تطبق المعادلتان (٤ و٥)، كما يوجد زوجان مختلفان في المعاملة والعمود والصف تطبق عليهما المعادلتان (٨ و٩). وقد تم الحصول على القيم التقديرية الآتية: $\hat{y}_{22(1)} = 5.225$ و $\hat{y}_{12(5)} = 5.757$ و $\hat{y}_{51(5)} = 5.424$.

طريقة Rubin:

باستخدام المعادلة (١١) تم الحصول على الآتي:

$$p_1 = e_{22(1)} = -2.265 \qquad p_2 = e_{12(5)} = -1.477 \qquad p_3 = e_{51(5)} = -2.233$$

$$p' = \begin{pmatrix} -2.265 & -1.477 & -2.233 \end{pmatrix}$$

وبتطبيق المعادلتين (١٢ و١٣) نحصل على عناصر المصفوفة $A_{3\times3}$ وهي:

$$A = \begin{pmatrix} 0.480 & 0.080 & 0.080 \\ 0.080 & 0.480 & 0.080 \\ 0.080 & 0.080 & 0.480 \end{pmatrix}$$

وبالاستعانة بالمعادلة (١٠) تم الحصول على القيم التقديرية الآتية: $\hat{y}_{22(1)} = 3.795$ و $\hat{y}_{12(5)} = 1.825$ و $\hat{y}_{51(5)} = 3.715$.

طريقة H&G: أعطت المعادلة في (١٤) المعادلات الآنية الآتية:

$$12\hat{y}_{22(1)} - 3\hat{y}_{12(5)} + 2\hat{y}_{51(5)} = 56.620$$

$$12\hat{y}_{12(5)} - 3\hat{y}_{22(1)} - 3\hat{y}_{51(5)} = 36.920$$

$$12\hat{y}_{51(5)} - 3\hat{y}_{12(5)} + 2\hat{y}_{22(1)} = 58.320$$

وبتحليل المعادلات الآنية كانت القيم التقديرية الآتية: $\hat{y}_{22(1)} = 5.251$ و $\hat{y}_{12(5)} = 5.743$ و $\hat{y}_{51(5)} = 5.421$.

بعد إدخال القيم التقديرية إلى البيانات اجري تحليل التباين وصححت قيم مجموع مربعات المعاملات حسب المعادلة (١٥)، وتم طرح درجة حرية واحدة عن كل قيمة مقدرة، فكان الجدول (٧).

الجدول ٧: نتائج تحليل التباين عند فقدان ثلاث قيم.

الطرائق التقديرية	القيمـــة بعد التقدير	MSe بعد التقدير	MSt بعد التقدير	F cal. بعد التقدير
Yates	٥,٢٥١	٠,٠٢٣	٢,٣٢٧	١٠١,١٧٤
	٥,٧٤٥			
	٥,٤٢١			
Harry	٥,٢٢٥	٠,٠٢٥	٢,٢١٨	٨٨,٧٢٠
	٥,٧٥٧			
	٥,٤٢٤			
Rubin	٣,٧٩٥	٠,٨٣٥	٢,٠٠٧	٢,٤٠٤
	1.825			
	3.715			
H&G	٥,٢٥١	٠,٠٣٥	٢,٢٨٠	٦٥,١٤٣
	٥,٧٤٣			
	٥,٤٢١			

لقد ولدت لعشرات المرات بيانات مختلفة لذات التجربة معتمدين في ذلك على مسار البرنامج الحاسوبي الجاهز Minitab v 13.2 **وهو:**

Cal → Random Data → Normal → Generate $\boxed{5}$ rows of data + mean $\boxed{}$ + standard deviation $\boxed{}$

وأعيد تطبيق الطرائق الأربع بالأسلوب نفسه عند كل مجموعة بيانات مولده لتصميم المربع اللاتيني، وقد تم التوصل إلى النتائج نفسها المشار إليها.

المناقشة والاستنتاجات

ندرج في أدناه ما تم التوصل إليه من نتائج:

١- عند فقدان قيمة مشاهدة واحدة تساوت الطرائق الأربع: Yates وHarry وRubin وHaseman and Gaylar. في مقدار القيمة التقديرية للمشاهدة المفقودة، وبالتالي الحصول على جدول تحليل التباين نفسه (الجدول ٤).

٢- عند فقدان قيمتين تساوت نتائج ثلاث طرائق هي: Yates وHarry و Haseman and Gaylar. أما طريقة Rubin فقد سببت ارتفاعا في قيمة MSe وانخفاضا في قيمة F cal. (الجدول ٥).

٣- زيادة عدد القيم المفقودة يؤدي إلى زيادة الفروقات بين القيم التقديرية التي تعطيها الطرائق المختلفة. وعلى وفق معايير المفاضلة يلاحظ من الجدول (٧) الآتي: كانت طريقة Yates هي الأفضل، وقد جاءت بالمرتبة الثانية طريقة Harry، ومن ثم طريقة Haseman & Gaylar. واختلفت طريقة Rubin بشكل ملحوظ عن الطرائق الأخرى، فقد أعطت أعلى قيمة لمتوسط مربعات الخطأ (MSe)، وأقل قيمة للفرق المعنوي بين المعاملات (F cal.).

٤- لم تتأثر القيم التقديرية بموقع القيمة المفقودة، أو بعائديتها إلى أية معاملة.

٥- إن القيمة التقديرية تنحدر لتكون قريبة من الوسط الحسابي العام، وهذا بدوره يؤثر على قيمتين رئيستين في جدول تحليل التباين هما: مجموع مربعات المعاملات (SSt)، خاصة عند فقدان قيم تكون مسؤولة عن إظهار معنوية الفروقات بين المعاملات. القيمة الأخرى: مجموع مربعات الخطأ (SSe). وهنا يأتي دور درجات الحرية لتلعب دوراً في تحديد قيمة F cal.. هذا ما يشير إليه تحليل التباين عند مقارنة محتوى الجدول (٣) مع الجدولين (٥ و٧).

٦- عند المفاضلة بين الطرائق على أساس سهولة حساب القيم التقديرية: فقد عينت طريقة Haseman & Gaylar (H&G) وطريقة Rubin كأسهل طريقتين في التطبيق. ومن ثم 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 لاتباعها أسلوب التعويض المتتالي.

المصادر العربية

١- 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.

٢- الراوي، خاشع محمود وخلف الله، عبد العزيز محمد، ٢٠٠٠، " تصميم وتحليل التجارب الزراعية"، الطبعة الثانية، دار الكتب للطباعة والنشر، جامعة الموصل.

٣- المشهداني، محمود حسن والمشهداني، كمال علوان، ٢٠٠٢، "تصميم وتحليل التجارب"، الدار الجامعية للطباعة والنشر والترجمة، جامعة بغداد.

٤- 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

5- 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.

6- Harry, C., 1964, “Introduction to experimental statistics”, McGraw-Hill Book Company, USA .

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

8- Keller.G and warrack, 2003,“Statistics for management and conomics”, Duxbury Press, USA.

9- 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.

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