المتغيرات الوهمية بديلة عن تقدير القيم المفقودة في تجارب التصاميم الأساسية

أ.م.مروان عبدالعزيز دبدوب
كلية علوم الحاسبات والرياضيات -جامعة الموصل

الملخص

عولجت القيم المفقودة في نتائج التصاميم الأساسية وهي: العشوائي الكامل، والقطاعات العشوائية الكاملة، والمربع اللاتيني من دون اللجوء إلى تقديرها، والحصول على جدول تحليل التباين لتلك التصاميم باستخدام تحليل انحدار تلك النتائج على المتغيرات الوهمية وقد أُستعين بالبرنامج ألحاسوبي الجاهز Minitab v 13.12. وقد أظهر الأسلوب المقترح لتحليل للانحدار كفاءة عند مقارنة نتائجه مع طريقة Yates التقديرية.

Abstract

The missing values were tackled in three basic designs: complete randomized, randomized complete blocks and Latin square. The analysis of variance tables were constructed by using the regression analysis of the experimental data (response) on the dummy variables (predictors). The results were obtained by the aided of the software computer program-Minitab v 13.12. The suggest method of regression analysis appeared efficiency compare with the estimating method of Yates.

المقدمة

كأي تطور في مجال العلم كان لموضوع تصميم التجارب وتحليلها نصيب من ذلك، وقد ساعدت البرمجيات الحاسوبية الجاهزة في التغلب على الكثير من المشاكل والعقبات التي يواجهها الباحث في إجراء التجارب ولاسيما في الجانب التحليلي، ومن أهم هذه المشاكل: فقدان بعض نتائج التجربة أو الوحدات التجريبية بأكملها مما يؤثر في اتزان التصميم وكذلك ظهور خلل في فروض التحليل وبالتالي الحصول على نتائج تحليلية لا يعول عليها.

وقد يعتقد بعض الدارسين أن إعادة التجربة بتطبيق ذات التصميم أفضل من معالجة القيم المفقودة، ولكن إتباع 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 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، كما أشار Yates and Hale (1939) إلى كيفية التعامل مع طرائق التحليل في المربع اللاتيني عند فقدان عدة صفوف أو أعمدة أو معاملات.

وقد اعتمد Coons (1957) على القيم التقديرية لمعالم الانحدار بالاستعانة بالمتغيرات الوهمية لتقدير القيم المفقودة في تصميمي القطاعات العشوائية الكاملة والمربع اللاتيني، وكذلك استخدم Rubin (1972) طريقة التقدير غير المتوالي عن طريق المربعات الصغرى في التعامل مع المشكلة نفسها في التصميمين. وقد عالج John and Prescott (1975) القيم المفقودة باستخدام مصفوفات المتغيرات الوهمية، كما استعان Hunt and Triggs (1989) بالمصفوفات لتطبيق طريقة التقدير المتوالي، وفي العام التالي ذكر Hunt and Bell (1990) طرائق مختلفة للتعويض شبه المتوالي semi-iterative لمقدرات الانحدار بطريقة المربعات الصغرى لتفادي تأثير القيم المفقودة في نتائج التحليل في بعض التصاميم الشائعة الاستخدام.

إن لموضوع تحليل الانحدار أهمية كبيرة في محاولة التغلب على مشكلة وجود قيم مفقودة في نتائج التجارب المصممة، وقد أشار Little and Rubin (2002) إلى هذه الأهمية بتطبيقهما لتحليل الانحدار في معالجة القيم المفقودة من دون ذكرهما للمتغيرات الوهمية.

وناقش دبدوب وبدل (٢٠٠٤) تأثر بعض طرائق معالجة القيم المفقودة في بيانات احتوت على قيم شاذة في تصميم العبور البسيط. وقارن هدبة (2005) بين خمس طرائق لمعالجة القيم المفقودة إحداهن طريقة Yates في تصميمي القطاعات العشوائية الكاملة والمربع اللاتيني.

وتعطي كل واحدة من طرائق التقدير نتائج متباينة تحت ظروف وتصاميم مختلفة، وعليه اقترحنا أسلوباً بديلاً عن تقدير القيم المفقودة وذلك باستخدام المتغيرات الوهمية في تحليل الانحدار والاستعانة بالبرنامج الجاهز Minitab v 13.12 للتوصل إلى جدول تحليل التباين.

ولأهمية التصاميم الأساسية: العشوائي الكامل والقطاعات العشوائية الكاملة والمربع اللاتيني في تنفيذ التجارب ذوات التصاميم الأكثر تعقيداً، اخترنا هذه التصاميم لتطبيق الأسلوب المقترح لمعالجة القيم المفقودة في نتائج تلك التصاميم.

هدف البحث

مناقشة أسلوب مقترح لإيجاد جدول تحليل التباين للتصاميم الأساسية عند احتواء نتائج تجاربها على قيم مفقودة من دون اللجوء إلى إحدى طرائق التقدير، وذلك باستخدام تحليل انحدار تلك النتائج على المتغيرات الوهمية ذوات الرموز التأثيرية مع الاستعانة بالبرنامج الحاسوبي الجاهز Minitab v 13.12. وستقوّم كفاءة الأسلوب المقترح بمقارنة نتائجه مع طريقة Yates التقديرية.

القيم المفقودة وتقديرها

تعدُّ القيم مفقودة إذا فقدت بعضُ نتائج التجربة، أو إذا قرر الباحثُ إلغاء القيم الشاذة، أو التي حدث خطأ في تسجيلها، أو لأيّ سبب كان، وقد أشار Little and Rubin (2002) إلى الآلية المسببة لفقد البيانات. وبما أن فقدان القيمة التي تعطيها الوحدة التجريبية مستقلة عن تأثيرات المعاملات فلا يمكن تعويض القيمة المفقودة، وإنما يمكن حساب تقدير لها ويستعان في ذلك بترابطين: الأول: بين الوحدة التجريبية ذات القيمة المفقودة مع الوحدات التجريبية الأخرى في القطاع نفسه، والآخر: ترابط ذات الوحدة التجريبية مع الوحدات التجريبية الأخرى التي أخذت المعاملة نفسها.

وأكثر الطرائق شيوعاً لتقدير القيم المفقودة هي طريقة Yates للتعويض المتوالي، التي تحتاج إلى تصحيح مجموع المربعات العائد إلى كل اتجاه، ودرجات الحرية للخطأ. وأساس الطريقة في التصاميم الأساسية هي:

أ. **التصميم العشوائي الكامل:** هو مخطط ذو اتجاه واحد، ويتم التوزيع بين المعاملات والوحدات التجريبية المتجانسة عشوائيا. ومن ميزات هذا التصميم إمكانية إنشاء جدول تحليل تباين لنتائج تجربة من دون اللجوء إلى عملية تقدير للقيم المفقودة، إذ يكفي الحصول على قيمتين فقط لإجراء التحليل، وقد أشار كل من الراوي وخلف الله (2000) والمشهداني والمشهداني (2002) إلى طريقة التحليل المباشر، كما ذكر فهمي (2004) بعض الطرائق لإيجاد جدول تحليل التباين حتى عندما تكون نتائج التجربة بدون تكرار.

ب. **تصميم القطاعات العشوائية الكاملة:** هو تصميم باتجاهين: المعاملات والقطاعات، وفي كل قطاع تجمع وحدات تجريبية متجانسة توزع عليها المعاملات بصورة عشوائية.

ويتم تقدير القيمة المفقود y_{ij}، كالآتي: (المشهداني والمشهداني، ٢٠٠٢)

$$\hat{y}_{ij} = (tY_{i.} + rY_{.j} - Y_{..})/(r-1)(t-1) \quad \dots \quad (1)$$

عدد المعاملات= t وعدد القطاعات= r ; $i=1,2,\dots,t$ و $j=1,2,\dots,r$.

$Y_{i.}$: مجموع قيم المعاملة i التي تحتوي على القيمة المفقودة.
$Y_{.j}$: مجموع قيم القطاع j الذي يحتوي على القيمة المفقودة.
$Y_{..}$:المجموع الكلي لنتائج التجربة.

إن عملية التقدير هذه توجب إجراء تصحيح لمجموع المربعات للاتجاه q، وعادة ما يكون للمعاملات وقد يجرى للقطاعات أيضاً، وقيمة معامل التصحيح هي:

$$CF(q)=[A-(df)\ \hat{y}_{ij}]^2/[df(df+1)] \quad \dots \quad (2)$$

A: مجموع المشاهدات في الاتجاه الآخر الحاوي على القيمة المفقودة.
df: درجات الحرية للاتجاه q المراد إجراء تصحيح لمجموع مربعاته.

ت. تصميم المربع اللاتيني: هو تصميم قطاعين متعامدين يطلق على أحدهما الصفوف (r_j) rows والآخر الأعمدة (c_k) columns، لتظهر المعاملات (t_i) مرة واحدة في كل صف وكل عمود، وبذلك يحتوي هذا التصميم على ثلاثة اتجاهات هي: الصفوف والأعمدة والمعاملات، حيث إن:($j=1,2,\dots,r$; $k=1,2,\dots,c$; $i=1,2,\dots,t$).
وقد أشار المشهداني والمشهداني (٢٠٠٢) إلى تقدير القيمة المفقودة $y_{jk(i)}$، كالآتي:

$$\hat{Y}_{jk(i)}=r(Y_{j.}+Y_{.k}+Y_{(i).}-2Y_{..})/(r-1)(r-2) \quad \dots \quad (3)$$

Y_j: مجموع قيم الصف j الذي يحتوي على القيمة المفقودة.
$Y_{.k}$: مجموع قيم العمود k الذي يحتوي على القيمة المفقودة.
$Y_{(i)}$: مجموع قيم المعاملة i التي تحتوي على القيمة المفقودة.
إن معامل تصحيح مجموع المربعات للاتجاه q هو:

$$CF(q)=[Y_{..}-B-C-(df)A\ ]^2/[df(df-1)]^2 \quad \dots \quad (4)$$

عن كل قيمة مقدرة في التصاميم المختلفة تطرح درجة حرية واحدة من الخطأ، إن هذا الإجراء يؤدي إلى ارتفاع في مستوى الدلالة الإحصائية (α) اللازمة لرفض فرضية العدم.

المتغيرات الوهمية

يحتوي كل تصميم على اتجاه واحد أو أكثر (q عدد الاتجاهات في التصميم، $h=1,2,\dots,q$)، وكل اتجاه يحتوي على أكثر من مستوى واحد (ليكن p_h عدد المستويات في الاتجاه h, $g=1,2,\dots,p_h$)، ومن خلال هذه المستويات يمكن توليد متغيرات جديدة تدعى بالمتغيرات الوهمية وذلك باستخدام رموز معينة للمستوى المناظر لنتائج التجربة المتراكمة (stack) التي تمثل متغير الاستجابة (y_i) Response.
وسنذكر بعض أنواع الرموز لتوليد المتغيرات الوهمية التي تم استخدامها وهي:

أ. الرمز الرقمي Digital code: يعطى لكل مستوى رقم تسلسلي وبذلك يكون رمز المستوى الأول ١ والثاني ٢ وهكذا حتى آخر مستوى يرمز له p_h.
ونحصل عن طريق البرنامج ألحاسوبي ومن خلال المسار: Manip → Stack، على عمود يحتوي على القيم المتراكمة لنتيجة التجربة (y_i) Response ويسمى أيضاً Factor ويقابله الرمز الرقمي تحت عمود عنوانه Subscript.
يستخدم الرمز الرقمي في ثلاثة مسارات حاسوبيه هي: المسار (5) الذي يتبع عند تحليل نتائج التصميم العشوائي الكامل بمشاهدات كاملة أو احتوائه على قيم مفقودة.

Stat→ANOVA→One way→Response□+Factor□ … (5)

المسار الثاني (٦) خاص بتصميم القطاعات العشوائية الكاملة، ولا يمكن استخدامه عند احتواء البيانات على قيم مفقودة.

Stat→ANOVA→Two way→Response□+Factor1□+Factor2□ … (6)

Factor1 موقع رموز للاتجاه الأول وFactor2 موقع رموز للاتجاه الثاني.

المسار الثالث (٧)، يمكن اتباعه عند احتواء البيانات على قيم مفقودة أو كاملة وفي أيٍّ من التصاميم الأساسية الثلاثة.

Stat→ANOVA→General Linear Model→Response□+Model□… (7)

حيث يشار في الموقع Model إلى مواقع الاتجاهات المرمزة.

ب. الرمز ألتأثيري Effect Code: يتم توليد متغيرات وهمية لكل اتجاه، ويكون عددها بعدد مستويات الاتجاه مطروحاًمنها واحد (p_h-1)، ويثبت رمز المستوى الأخير ليكون ١- في مواقع المتغيرات الوهمية المولدة كافة، إن مجموعة قيم المتغير المعتمد المقابلة للمستوى الثابت (الذي اخذ الرمز ١- في المواقع كافة) تسمى بالعينة الرقابية، إذ إنه:

$$X_{hg} = \begin{cases} ١ & \text{عند تعاملنا مع المستوى g في الاتجاه h} \\ ٠ & \text{المستويات الأخرى عدا آخر مستوى} \\ -١ & \text{المستوى الأخير } p_h \text{ في الاتجاه h} \end{cases}$$

ت. الرمز الوهمي Dummy Code: يقوم على فكرة الرموز التأثيرية نفسها عدا أن المستوى الأخير يثبت بالرمز ٠ بدلاً من ١- .

إن الرمزين ألتأثيري والوهمي كليهما يُستخدمان في تحليل الانحدار، ويثبت في الموضع Predictors في مساره ألحاسوبي (٨) إلى موقع عمود المتغيرات الوهمية.

Stat→Regression→Regression→Response□+Predictors□ … (8)

المتغيرات الوهمية وتحليل الانحدار

يعطي تحليل الانحدار إنموذجاً يصف العلاقة بين متغيرات توضيحية (مستويات الاتجاهات) ومتغير معتمد (نتائج التجربة)، وجدول تحليل التباين لاختبار الفرضية العامة والجزئية لمعالم الأنموذج (كاظم ومسلم، ٢٠٠٢)، ويعطي أيضاً مقاييس يُفاد منها في تفسير الظاهرة وتوضيح العلاقات بين متغيرات التجربة، منها: معامل التحديد (coefficient of determination, R^2) الذي يقيس نسبة ما تفسره المتغيرات التوضيحية من الاختلافات في المتغير المعتمد، ومعامل تضخم التباين (variance inflation factor, VIF) وهو أحد مقاييس التعرف على تعدد العلاقة الخطية، وغيرهما من المقاييس التي ترتبط بالأنموذج أو بجدول تحليل التباين.

وبما أن هدف البحث اختبار الفرضية الخاصة بالتصاميم ألأساسية عند احتواء نتائج تجاربها على قيم مفقودة، فلذلك سنهتم بنتائج تحليل انحدار المتغيرات الوهمية ذات العلاقة بجدول تحليل التباين الخاص باختبار فرضيات ذلك التصميم.

وتختلف قيم معامل الانحدار (regression coefficients) الناتجة عن المتغيرات الوهمية ذات الرمز ألتأثيري وتلك ذات الرمز الوهمي لاختلاف طريقة حسابهما (الجدول ١)، وتختلف النتائج الثانوية ذات العلاقة المباشرة بمعامل الانحدار. ويبقى جدول تحليل التباين والنواتج المرتبطة به كمعامل التحديد (R^2) والانحراف القياسي (S) متساوية عند استخدام الرمزين.

الجدول (١)

حساب قيم معامل الانحدار عند استخدام الرمز التأثيري والرمز الوهمي

معامل الانحدار	الرمز التأثيري	الرمز الوهمي
$\hat{\beta}_o$	الوسط الحسابي العام.	الوسط الحسابي للعينة الرقابية.
$\hat{\beta}_j$	(الوسط الحسابي لقيم المتغير المعتمد المقابل للمتغير الوهمي j) – (الوسط الحسابي العام).	(الوسط الحسابي لقيم المتغير المعتمد المقابل للمتغير الوهمي j) – (الوسط الحسابي للعينة الرقابية).

وتختبر الأهمية الإحصائية لمعادلة الانحدار الحاوية على m من المتغيرات من خلال مصدر التباين العائد للانحدار ((Due to regression=R(all $X_{j's}=X_1...X_m$)، ويتم ذلك باختبار الفرضية الآتية: (الراوي، ١٩٨٧)

H_o: ($\beta_1=\beta_2=\ldots=\beta_m$)=(equality of β_j's)=0 vs H_A: at least one $\beta_j\neq 0$ ... (9)

ويمكن تجزئة مصدر التباين $R(X_1\ldots X_m)$ ليكون أحد أجزائه محتوياً على t من المتغيرات المختارة (selected $X_{j's}$) و r من المتغيرات الباقية (remainder $X_{j's}$) إذ أن t=m-r، وتختبر أهمية المتغيرات المختارة في الأنموذج من خلال الفرضية الآتية:

H_o: equality of the selected $\beta_{j's}=0$ vs H_A: at least one of them$_j\neq$٠ ... (10)

وعند تعاملنا مع البيانات الأصلية (غير الوهمية) يتطلب اختبار الفرضية (١٠) إيجاد مجموع المربعات العائد للانحدار للمتغيرات المختارة كالآتي:

$$SSR(\text{selected } X_{j's})=SSR(\text{all } X_{j's})-SSR(\text{remainder } X_{j's} \mid \text{selected } X_{j's}) \quad \ldots (11)$$

وقد لوحظ أن إدخال المتغيرات الوهمية في تحليل الانحدار يسهل التعامل مع الحد الثاني الأيمن في المساواة (11) لتصبح كالآتي:

$$SSR(\text{selected } X_{j's})=SSR(\text{all } X_{j's})-SSR(\text{remainder } X_{j's}) \quad \ldots (١٢)$$

فمثلاً عندما نتعامل مع نتائج تصميم القطاعات العشوائية الكاملة تشير المساواة (١٢) إلى اتجاهين: الأول: مجموعة المتغيرات المختارة (selected $X_{j's}$) التي تمثل المتغيرات الوهمية العائدة للمعاملات (t_i)، والاتجاه الآخر: بقية المتغيرات (remainder $X_{j's}$) التي تمثل القطاعات (r_j)، أي:

$$SSR(t_i)=SSR(\text{all } X_{j's}=t_i+r_j)-SSR(r_j)$$

في تصميم المربع اللاتيني يكون عندنا ثلاثة اتجاهات هي: المعاملات (t_i) والصفوف (r_j) والأعمدة (c_k)، لتصبح المساواة (12) للمعاملات كالأتي:

$$SSR(t_i)=SSR(\text{all } X_{j's}=t_i+r_j+c_k)-(SSR(r_j)+SSR(c_k)) \quad \ldots (13)$$

كما لوحظ أن:

$$SSR(r_j)+SSR(c_k)=SSR(r_j+c_k)$$

ويمكن تعميم المساواة في (13) لتشمل الصفوف والأعمدة أيضاً.

القيم المفقودة وتحليل الانحدار

عند تكوين جدول تحليل التباين لتصميم (عدا التصميم العشوائي الكامل)، لوحظ اختلاف في أسلوب تحليل الانحدار بين وجود قيم مفقودة وانعدامها، كما موضح في أدناه:

أ. نتائج التجربة ذات المشاهدات الكاملة: إن كل حد في المساواة (١٢) يكون مساوياً لمجموع مربعات (SS) المناظر له في تحليل التباين لتصميم القطاعات العشوائية الكاملة، أي:

$$SSR(\text{selected } X_{j's}=t_i)=SS(t) \text{ ، } SSR(\text{remainder } X_{j's}=r_j)=SS(r_j) \quad \dots (١٤)$$

وكذلك الحال عند تحليل نتائج تصميم المربع اللاتيني إذ يكون كل حد في المساواة (١٣) مساوياً لنظيرة في التصميم، أي إن:

$$SS(t_i)=SSR(t_i) \text{ , } SS(r_j)=SSR(r_j) \text{ , } SS(c_k)=SSR(c_k) \quad \dots (١٥)$$

إن ما لوحظ من علاقات بين نتائج تحليل الانحدار لمجاميع المتغيرات الوهمية في الحالتين المجتمعة أو المجزأة، يسري أيضاً على درجات الحرية والنتائج الثانوية المستخلصة من جدول تحليل التباين مثل: الانحراف القياسي (S) ومعامل التحديد (R^2)، وتبقى قيم معامل الانحدار ومعامل تضخم التباين (VIF) ثابتة في كافة الحالات ومختلفة بين الرمزين ألتأثيري والوهمي.

ب. عند احتواء نتائج التجارب على قيم مفقودة: لا تتحقق المساواة في(١٢) و(١٣) و (١٤) و(١٥). ففي تصميم القطاعات العشوائية الكاملة، تحسب مجموع المربعات للمعاملات $SS(t_i)$ والقطاعات $SS(r_j)$، كالآتي:

$$SS(t_i)=SSR(t_i+r_j)-SSR(r_j) \quad \text{و} \quad SS(r_j)=SSR(t_i+r_j)-SSR(t_i) \quad \dots (١٦)$$

كما تحسب مجموع المربعات للمعاملات $SS(t_i)$ في تصميم المربع اللاتيني، كالآتي:

$$SS(t_i)=SSR(r_j+c_k+t_i)-SSR(r_j+c_k) \quad \dots (١٧)$$

وتعمم العلاقة في (١٧) لتشمل حساب مجموع المربعات للصفوف $SS(r_j)$ والأعمدة $SS(c_k)$. ولوحظ أن:

$$SSR(r_j)+SSR(c_k) \neq SSR(r_j+c_k)$$

إن قيم المتغير المعتمد والمتمثلة بنتائج التجربة المصممة لا تتأثر بعملية توليد المتغيرات الوهمية، ومنها نجد مجموع المربعات الكلي لتحليل الانحدار (SSRT) الذي يكون مساوياً لمجموع المربعات الكلي للتصميم (SST)، وكذلك الحال في مجموع مربعات الخطأ (SSe)، وبالتالي متوسط مربعات الخطأ (MSe).

القيم المفقودة وجدول تحليل التباين

يهتم تحليل التباين للتصاميم باختبار الفرضية الخاصة بتساوي الأوساط الحسابية لمجتمعات مستويات كل اتجاه (Kuehl, 2000)، أي اختبار الفرضية الآتية:

$$H_o: \mu_1=\mu_2=\dots=\mu p_h \text{ vs } H_1: \text{at least two of them are not equal.} \quad \dots (18)$$

فضلاً عن ما تم طرحة في الجانب النظري، ولكي تتحقق الثقة بالأسلوب المقترح لاختبار الفرضية (١٨) عن طريق جدول تحليل التباين الخاص بالتصميم، فقد طبق في عدة مجالات منها: تحليل نتائج عدة تجارب ولاسيما تلك التي أقيمت في كلية الزراعة والغابات- جامعة الموصل، والتي وردت في بعض رسائل وأطاريح الدراسات العليا، وحيثما وجدنا مثالاً قد أُشير إلية في كتاب أو مصدر. إن أهم جزء في تطبيق الأسلوب المقترح هو استخدام المحاكاة simulation لتلك التجارب والأمثلة وقد تجاوز عدد مرات تكرارها المئتي مرة في كل

تصميم. وقد أشارت النتائج كافة إلى ملاءَمة الأسلوب المقترح، بالمقارنة مع طريقة Yates التقديرية، لمعالجة القيم المفقودة والحصول على جدول تحليل التباين المصحح.

وسنكتفي بإعطاء فكرة عن الأسلوب المقترحة من خلال ذكر مثال واحد لكل تصميم لاختبار الفرضية (١٨)، وسنستعين في ذلك بأمثلة أشار إليها الراوي وخلف الله (٢٠٠٠) في كتابهما بوصفة أحد كتب تصميم التجارب باللغة العربية ذات الانتشار الواسع.

أ. **التصميم العشوائي الكامل:** إن هذا التصميم له اتجاه واحد هو مستويات المعاملات، وكافة طرائق التحليل تعالج عدم تساوي تكرار المعاملات وتعطي النتائج نفسها، لذا يمكن الحصول على جدول تحليل التباين مباشرة باستخدام الرمز الرقمي في المسارين:(٥) و(٧)، أو استخدام الرمز التأثيري أو الوهمي في المسار (٨).

تجربة:

في دراسة لتأثير خمس طرائق تخزين في مستوى الرطوبة لألواح خشب الصنوبر الأبيض، استخدمت أعداد مختلفة من الألواح في كل طريقة، وكانت نتائج مستوى الرطوبة كما في الجدول (٢) (الراوي وخلف الله، ٢٠٠٠، ٨٤). ووضعت النتائج المتراكمة للتجربة مع المتغيرات الوهمية بثلاثة رموز: الرقمي والتأثيري والوهمي، في الجدول (٣) استعداداً لتحليلها باتباع المسارات الحاسوبية الخاصة بالتحليل.

الجدول (٢)

مستوى الرطوبة لألواح خشب الصنوبر الأبيض في خمس طرائق للتخزين

تكرار المشاهدات	خمس طرائق للتخزين (المعاملات t_i)				
	t_1	t_2	t_3	t_4	T_5
r_1	٨.٥	٨.٣	٧.٦	١٠.٧	٩.٢
r_2	٨.٨	٨.٦	٩.٣	١٠.١	٨.٣
r_3	٩.٥	٨.٦		١١.٢	
r_4	٩.٦				
r_5	٩.٥				

الجدول (٣)

المتغيرات الوهمية المستخدمة في تحليل نتائج تجربة التصميم العشوائي الكامل

نتائج التجربة (y_i)	الرموز الرقمية (subsc.)	الرموز التأثيرية				الرموز الوهمية			
		te_1	te_2	te_3	te_4	td_1	td_2	td_3	td_4
٨.٥	١	١	٠	٠	٠	١	٠	٠	٠
٨.٨	١	١	٠	٠	٠	١	٠	٠	٠
٩.٥	١	١	٠	٠	٠	١	٠	٠	٠
٩.٦	١	١	٠	٠	٠	١	٠	٠	٠
٩.٥	١	١	٠	٠	٠	١	٠	٠	٠
٨.٣	٢	٠	١	٠	٠	٠	١	٠	٠
٨.٦	٢	٠	١	٠	٠	٠	١	٠	٠

نتائج التجربة (y_i)	الرموز الرقمية (subsc.)	الرموز التأثيرية				الرموز الوهمية			
		te_1	te_2	te_3	te_4	td_1	td_2	td_3	td_4
٨.٦	٢	٠	١	٠	٠	٠	١	٠	٠
٧.٦	٣	٠	٠	١	٠	٠	٠	١	٠
٩.٣	٣	٠	٠	١	٠	٠	٠	١	٠
١٠.٧	٤	٠	٠	٠	١	٠	٠	٠	١
١٠.١	٤	٠	٠	٠	١	٠	٠	٠	١
١١.٢	٤	٠	٠	٠	١	٠	٠	٠	١
٩.٢	٥	-١	-١	-١	-١	٠	٠	٠	٠
٨.٣	٥	-١	-١	-١	-١	٠	٠	٠	٠

إن المسارات الحاسوبية كافة المتبعة في هذا التصميم أعطت محتويات الجدول (٤) نفسها.

الجدول(٤)
تحليل التباين للمتغيرات الوهمية

Source	F	SS	MS	F	P
treatments	4	9.453	2.363	6.74	0.007
Error	10	3.505	0.350		
Total	14	12.957			

ب. تصميم القطاعات العشوائية الكاملة: إننا نسعى للحصول على مجموع المربعات المصحح لاتجاهي التصميم (المعاملات والقطاعات) عند احتواء نتائجة على قيم مفقودة.

ومن خلال نتائج تجربة لتسمين الأغنام، سنقارن بين محتوى جدولين لتحليل التباين، وسنحصل على أحدهما بطريقة Yates والآخر بالأسلوب المقترح لتحليل المتغيرات الوهمية ذات الرمز ألتأثيري بإتباع المسار (٨). وقد اشتملت التجربة على ثلاث علائق غذائية مختلفة التراكيب (المعاملات) وطبقت في أربع حظائر (القطاعات)، وضعت نتائج التجربة (كغم زيادة في الوزن) في الجدول (٥) وفقدت منها مشاهدتان (الراوي وخلف الله، ٢٠٠٠، ١٢٧).

الجدول (٥)

نتائج تجربة لتغذية الحيوان بثلاث معاملات وأربعة قطاعات

المعاملات (t_i)	القطاعات (r_j)			
	r_1	r_2	r_3	r_4
t_1	y_{11}	٨.٦	8.9	٨.٦
t_2	٨.٧	y_{22}	٩.٠	٨.٧
t_3	٨.٥	٨.٤	٨.٦	٨.٥

وسنبدأ بالطريقة الكلاسيكية وهي تقدير القيم المفقودة بطريقة Yates باستخدام القانون (١)، وبعد ست دورات كانت القيمة المقدرة لـ y_{11} تساوي ٨.٦٥٤٢٨٥٧ ولـ y_{22} تساوي ٨.٦٧٤٢٨٥٧ بفرق يساوي ٠.٠٠٠٠٠٠٤ و٠.٠٠٠٠٠٠١- عن الدورة الخامسة على التوالي. ومن المعادلة (٢) وجد معاملاً تصحيح مجموع المربعات للمعاملات والقطاعات فكانا على التوالي: ٠.٠٢٢٢١٥ و٠.٠١٣٤٢٢، فأنشيء جدول تحليل التباين المصحح (الجدول ٦) للمعاملات والقطاعات.

الجدول (٦)

جدول تحليل التباين المصحح بطريقة Yates

Source	DF	Adj SS	Adj MS	F	P
treatments	2	0.129905	0.064952	15.50	0.013
blocks	3	0.123238	0.041079	9.80	0.026
Error	4	0.016762	0.004190		
Total	9				

ولتطبيق تحليل الانحدار بوجود القيمتين المفقودتين y_{11} وy_{22}، وضعت النتائج المتراكمة للتجربة والمتغيرات الوهمية ذات الرمز ألتأثيري في الجدول (٧).

الجدول (٧)

النتائج المتراكمة (y_i) والرموز التأثيرية

y_i	المعاملات		القطاعات		
	t_1	t_2	r_1	r_2	r_3
8.6	١	٠	٠	١	٠
8.9	١	٠	٠	٠	١
8.6	١	٠	-١	-١	-١
8.7	٠	١	١	٠	٠
9.0	٠	١	٠	٠	١
8.7	٠	١	-١	-١	-١
8.5	-١	-١	١	٠	٠
8.4	-١	-١	٠	١	٠
8.6	-١	-١	٠	٠	١
8.5	-١	-١	-١	-١	-١

وطبق تحليل الانحدار على محتوى الجدول (٩) في ثلاث مراحل هي: المعاملات والقطاعات سوية، والمعاملات منفردة، والقطاعات منفردة، ونتائج التحليل في الجدول (٨).

الجدول (٨)

تحليل الانحدار للمتغيرات الوهمية ذات الرمز التأثيري

تحليل الانحدار للمعاملات والقطاعات سوية: t_1 t_2 r_1 r_2 r_3

Source	DF	SS	MS	F	P
Regression	5	0.288238	0.057648	13.76	0.012
Residual Error	4	0.016762	0.004190		

Total	9	0.305000			
تحليل الانحدار للمعاملات: t_1 t_2					
Source	DF	SS	MS	F	P
Regression	2	0.16500	0.08250	4.13	0.066
Residual Error	7	0.14000	0.02000		
Total	9	0.30500			
تحليل الانحدار للقطاعات: r_1 r_2 r_3					
Source	DF	SS	MS	F	P
Regression	3	0.158333	0.05278	2.16	0.194
Residual Error	6	0.14667	0.02444		
Total	9	0.30500			

للتوصل إلى جدول تحليل التباين للتصميم، طبقت المساواة (١٦) فكانت:

$$SS(treatments)=SSR(t_1\ t_2\ r_1\ r_2\ r_3)-SSR(r_1\ r_2\ r_3)$$

$$=0.288238-0.158333=0.129905$$

$$SS(blocks)=SSR(t_1\ t_2\ r_1\ r_2\ r_3)-SSR(t_1\ t_2)=0.288238-0.165000=0.123238$$

إن هاتين القيمتين تتفقان تماماً مع قيم مجموع المربعات المصحح في الجدول (٦) الخاص بطريقة Yates، وبذلك تتطابق نتائج طريقة Yates والأسلوب المقترح لتحليل الانحدار.

ت. تصميم المربع اللاتيني: لهذا التصميم ثلاثة اتجاهات هي: المعاملات وهي تحت سيطرة الباحث، والاتجاهان الآخران تتحكم بهما الوحدات التجريبية وهما : الصفوف والأعمدة. ولبيان الأسلوب المقترح لمعالجة القيم المفقودة ومقارنتها مع طريقة Yates التقديرية، فقد اخترنا نتائج تجربة أشار إليها الراوي وخلف الله (٢٠٠٠، ١٥٤) وقد أقيمت لمقارنة كمية إنتاج أربعة أصناف من فول الصويا (المعاملات)، وقد استخدم لإجراء التجربة تصميم المربع اللاتيني (٤x٤)، ويوضح الجدول (٩) كمية الحاصل بالكيلوغرام/ قطعة تجريبية.

الجدول (٩)

نتائج تجربة أصناف فول الصويا طبقت في تصميم مربع لاتيني ٤x٤

الصفوف	الأعمدة			
	c_1	c_2	c_3	c_4
r_1	($t_٤$) ٥٠.٨	($t_٣$) ٥٠.٢	($t_١$) ٥٤.٦	($t_٢$) ٥٠.٤
r_2	($t_٢$) ٤٩.٧	($t_١$) ٥٣.٠	($t_٤$) ٥٣.٢	(t_3) 51.3
r_3	($t_٣$) ٥٠.٢	(t_4) ٥٢.٤	(t_2) ٥١.٣	($t_١$) ٥٥.٣
r_4	($t_١$) ٥٣.٣	($t_٢$) ٥٠.٨	(t_3) ٥١.٣	(t_4) ٥٤.٢

لنفرض أن الجدول (٩) احتوى على ثلاث قيم مفقودة هي: $y_{14(2)}$ و$y_{22(1)}$ و$y_{31(3)}$، وباستخدام القانون (٣) وجدت القيم التقديرية بطريقة Yates عند الدورة السابعة وهي: ٥١.٦ و٥٤.٢ و٤٩.٠ على التوالي، وبفرق يساوي: ٠.٠٠٠٠٠٢ و٠.٠٠٠٠٠٣ و٠.٠٠٠٠٠١ على

التـوالي عـن الـدورة السادسـة. بالاسـتعانة بالمعادلـة (٤) ووجـد أن: معامـل التصـحيح لمجمـوع مربعـات المعـاملات للقـيم المفقـودة هـي: ٩.٢٠١١ و٢.٦٦٧٨ و٤.٦٩٤٤ علـى التـوالي، أي بمجمـوع مقـداره: ١٦.٥٧٣٢ للمعـاملات وللصـفوف ٠.٥٣٧٦ وللأعمـدة ٤.٠٦٣٦. ويوضـح جدول (١٠) تحليل التباين المصحح لمجاميع المربعات ودرجات الحرية للخطأ.

الجدول (١٠)

تحليل التباين لتصميم المربع اللاتيني بطريقة Yates

Source	DF	Adj SS	Adj MS	F	P
treatments	3	22.0163	7.3388	91.7346	<0.01
rows	3	0.2123	0.0708	٠.٨٨٥٠	>0.05
columns	3	8.9268	2.9756	٣٣.٨١٣٦	<0.01
Error	3	0.2400	0.0800		
Total	12				

ولاتباع الأسلوب المقترح، وضعت القيم المتراكمة لنتائج التجربة (الجدول ٩) والرمز ألتأثيري في الجدول (١١).

الجدول (١١)

النتائج المتراكمة لتجربة فول الصويا (y_i) والرموز التأثيرية

	المعاملات			الصفوف			الأعمدة		
y_i	t_1	t_2	t_3	r_1	r_2	r_3	c_1	c_2	c_3
٥٤.٦	١	٠	٠	١	٠	٠	٠	٠	١
٥٣.٠	١	٠	٠	٠	١	٠	٠	١	٠
٥٥.٣	١	٠	٠	٠	٠	١	−١	−١	−١
٥٣.٣	١	٠	٠	−١	−١	−١	١	٠	٠
٥٠.٤	٠	١	٠	١	٠	٠	−١	−١	−١
٤٩.٧	٠	١	٠	٠	١	٠	١	٠	٠
٥١.٣	٠	١	٠	٠	٠	١	٠	٠	١
٥٠.٨	٠	١	٠	−١	−١	−١	٠	١	٠
٥٠.٢	٠	٠	١	١	٠	٠	٠	١	٠
٥١.٣	٠	٠	١	٠	١	٠	−١	−١	−١
٥٠.٢	٠	٠	١	٠	٠	١	١	٠	٠

٥١.٣	٠	٠	١	−١	−١	−١	٠	٠	١
٥٠.٨	−١	−١	−١	١	٠	٠	١	٠	٠
٥٣.٢	−١	−١	−١	٠	١	٠	٠	٠	١
٥٢.٤	−١	−١	−١	٠	٠	١	٠	١	٠
٥٤.٢	−١	−١	−١	−١	−١	−١	−١	−١	−١

وطبق تحليل الانحدار لمحتوى الجدول (٩) بعد حذف ما يعود إلى القيم المفقودة الثلاث ($y_{14(2)}$ و$y_{22(1)}$ و$y_{31(3)}$)، وطبق التحليل في أربع مراحل، ويشير الجدول (١٣) إلى نتائجها، استعداداً لإنشاء جدول تحليل التباين ذي مجموع مربعات ودرجات حرية مصححة للمعاملات والصفوف والأعمدة بعد الاستعانة بالمساواة (١٧)، والجدول (١٢) هو الجدول المنشود.

الجدول (١٢)

تحليل التباين لتصميم المربع اللاتيني بثلاث قيم مفقودة

Source	DF	Adj SS	Adj MS	F	P
treatments	3	20.0164	6.6721	83.40	0.002
rows	3	0.6055	0.2018	2.52	0.234
columns	3	7.0964	2.3655	29.57	0.010
Error	3	0.2400	0.0800		
Total	12				

الجدول (١٣)تحليل الانحدار في تصميم المربع اللاتيني٤x٤ بثلاث قيم مفقودة

تحليل الانحدار للمعاملات والصفوف والأعمدة سوية: t_1 t_2 t_3 r_1 r_2 r_3 c_1 c_2 c_3

Source	DF	SS	MS	F	P
Regression	9	37.9769	4.2197	52.75	0.004
Residual Error	3	0.2400	0.0800		
Total	12	38.2169			

تحليل الانحدار للمعاملات والصفوف سوية: t_1 t_2 t_3 r_1 r_2 r_3

Source	DF	SS	MS	F	P
Regression	6	30.881	5.147	4.21	0.052
Residual Error	6	7.336	1.223		
Total	12	38.217			

تحليل الانحدار للمعاملات والأعمدة سوية: t_1 t_2 t_3 c_1 c_2 c_3

Source	DF	SS	MS	F	P
Regression	6	37.3715	6.2286	44.20	0.000
Residual Error	6	0.8455	0.1409		
Total	12	38.2169			

تحليل الانحدار للصفوف والأعمدة سوية: r_1 r_2 r_3 c_1 c_2 c_3

Source	DF	SS	MS	F	P
Regression	6	17.961	2.993	0.89	0.556
Residual Error	6	20.256	3.376		
Total	12	38.217			

لم تتطابق مجموع مربعات تصميم المربع اللاتيني التي نتجت عن الأسلوب المقترح لتحليل الانحدار (الجدول ١٢) مع نتائج طريقة Yates (الجدول ١٠)، فالفرق المطلق بينهما فيما يخص المعاملات يساوي ٢.٠٠٠٢ .

ومن نتائج المحاكاة والأمثلة العديدة التي قمنا بدراستها لاحظنا بأن زيادة عدد القيم المفقودة يؤدي إلى زيادة الفرق بين قيم مجموع المربعات بين الطريقتين، فمثلاً لنفترض القيمتين: $y_{22(1)}$ و$y_{14(2)}$ مفقودتين، وأعطت طريقة Yates قيمتين تقديريتين لهما: ٥٣.٩ و٥١.٣ على التوالي، وكان معامل التصحيح لمجموع مربعات المعاملات يساوي ١٢.٤٨٤٤، فأصبح مجموع مربعات المعاملات المصحح يساوي ٢١.٩٥٨١ .في حين أعطى تحليل الانحدار مجموع مربعات مصحح للمعاملات مقداره ٢١.٦٤٨٠، ليكون الفرق المطلق بين الطريقتين يساوي ٠.٣١٠١ .كما لوحظ بأن نتائج تحليل التباين تتساوى في الطريقتين عند احتواء البيانات على مشاهدة مفقودة واحدة، فمثلا عند فقدان المشاهدة $y_{24(3)}$ تكون القيمة التقديرية لها تساوي ٥١.٤٣٣٣ ومعامل التصحيح مقداره ٢.٦٦٧٨ لنحصل على مجموع مربعات مصحح للمعاملات مقداره ٣٠.١٦٢٢ .

إن عدم التطابق بين طريقة Yates وأسلوب تحليل الانحدار المقترح عند تعدد القيم المفقودة في تصميم المربع اللاتيني لا يعني فشل الأسلوب المقترح في معالجة القيم المفقودة ولاسيما أَن لهما نفس مجموع مربعات الخطأ ودرجات الحرية في هذا التصميم، كما لوحظ تطابق النتائج بين الطريقتين في التصميم العشوائي الكامل وتصميم القطاعات العشوائية الكاملة وعند تعدد القيم المفقودة، كما ذكر هدبة (٢٠٠٥) عند مقارنته لخمس طرائق لمعالجة القيم المفقودة من بينها طريقة Yates، بأن الطرائق غالباً ما كانت تختلف في مجموع المربعات لمصادر التباين وبالتالي اختلفت قيمة الدلالة الإحصائية (α) اللازمة لرفض فرضية العدم، ومع هذا فإن الطرائق كافة نجحت في التغلب على القيم المفقودة ولكل منها ميزاتها، إذ امتاز أسلوبنا المقترح ببساطته وسرعته في معالجة القيم المفقودة من دون اللجوء إلى التقدير، ويمكنه إعطاء معلومات إضافية يستفاد منها في زيادة المعرفة حول الظاهرة المدروسة. إننا نوصي باستخدام أسلوب تحليل الانحدار المقترح لمعالجة القيم المفقودة في التصاميم الأساسية كلما توفر البرنامج ألحاسوبي Minitab بإصداراته الحاوية على المسارات المشار إليها.

المصادر العربية

١. 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.

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

٣. الراوي، خاشع محمود، ١٩٨٧، "المدخل إلى تحليل الانحدار"، مديرية دار الكتب للطباعة والنشر، جامعة الموصل.

٤. فهمي، فيدان محمود، ٢٠٠٤، "دراسة التداخل في حالة عدم التكرار في تجربة ذات عاملين"، رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.

٥. كاظم، أموري هادي ومسلم، باسم شلبية، ٢٠٠٢، "القياس الاقتصادي المتقدم النظرية والتطبيق"، مطبعة دنيا الأمل، بغداد.

٦. المشـهداني، محمـود حسـن والمشـهداني، كمـال علـوان، ٢٠٠٢، "تصـميم وتحليـل التجـارب"، الدار الجامعية للطباعة والنشر والترجمة، جامعة بغداد.
٧. هدبة، وكاع علي، ٢٠٠٥، "مقارنـة لبعض طرائق تقدير ومعالجـة القيم المفقودة في تصاميم التجارب الأساسية"، رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.

المصادر الأجنبية

8. 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.
9. Coons, I., 1957, "The analysis of covariance as a missing plot technique", Biometrica, 13,p 307-405.
10. Hunt, D. and Bell, M., 1990, "Semi-Iterative missing value estimation", J. of App. Stat., 39, p 389-395.
11. Hunt, D. and Triggs, C., 1989, " Iterative missing value estimation", J. of App. Stat., 38,p 293-300.
12. John, J. and Prescott, P., 1975,"Estimating missing values in experiments", J. of App. Stat., 24, p 190-192.
13. Kuehl, R., 2000, " Design of experiments, statistical principles of research design and analysis", Duxbury Press, N.Y. .
14. Little, R. J. and Rubin, D.B., 2002,"Statistical analysis with missing data", 2nd edition, John Wiley and Sons, N.Y. .
15. 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.
16. Yates, F., 1933, "The analysis of replicated experiments when the field results are incomplete", Emp. J. Exp. Agric., 1, p 129-142.
17. Yates, F. and Hale, R., 1939, "The analysis of Latin squares when two or more rows/columns or treatments are missing", Sup. J. of Roy. Stat. Soc., 6, p 67-79.

This document was created with Win2PDF available at http://www.win2pdf.com.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.