المجلة العراقية للعلوم الاحصائية (14) 2008
ص ص [113–132]

العلاقة بين تحليل التباين وتحليل انحدار المتغيرات الوهمية

مروان عبد العزيز دبدوب* وكاع علي هدبة**

الملخص

عند تطبيق تحليل الانحدار للمتغيرات الوهمية ذات الرمز التأثيري أمكن الحصول على جدول تحليل التباين لثلاثة تصاميم ذوات النموذج الثابت وهي: العشوائي الكامل والقطاعات العشوائية الكاملة والمربع اللاتيني. يتميز الأسلوب المقترح ان الأسلوب التقليدي لتحليل التباين بأنه يعطى معلومات إضافية عن العلاقة بين المتغير المعتمد مع المتغيرات التوضيحية كتلك التي يعطيها تحليل الانحدار مثل معامل التحديد، وتعيين القيم الشاذة، كما يمكن لهذا الأسلوب من التغلب على القيم المفقودة دون اللجوء إلى تقديرها.

The Relationship between Analysis of Variance and the Regression Analysis of Dummy Variables

ABSTRACT

Regression analysis of dummy variables with effect code showed the same results of the analysis of variance table when they were applied on three fixed model designs. The designs were: completely randomized, randomized complete blocks and Latin square. This procedure gave an advantage upon the classic one, it yield additional information about the relation between the response and predictors as the regression analysis does, such as: coefficient of determination, identification of the outliers, also tackled the missing observations without estimating them.

---

* أستاذ مساعد/قسم الإحصاء والمعلوماتية/كلية علوم الحاسبات والرياضيات/جامعة الموصل.
** مدرس مساعد/قسم الإحصاء/كلية الإدارة والاقتصاد/جامعة كركوك.

تاريخ التسلم : 9 /2007/10 ـــــــــــــــ تاريخ القبول :5 /2008/3

المقدمة

يطبق تحليل الانحدار على متغيرات ذات بيانات كمية quantitative variables، وعلى متغيرات وصفية qualitative variables والتي يمكن تحويلها إلى إحـد أشكال المتغيرات المعبرة أو الوهمية indicated or dummy variable (دبدوب، 2006). والمتغير الوهمي هو ذلك المتغير الذي يأخذ أرقاما محددة من القيم لصفات مختلفة.

وأشار دبدوب (2006) إلى أهمية استخدام الرمز التأثيري effect code للمتغيرات الوهمية للتمييز بين صفات المتغير بدلاً من الأرقام المتسلسلة التي تفقده الكثير من صفاته التحليلية، ملغياً بذلك الأهمية الكمية لفئات المتغير ألوصفي أو النوعي، وقد عرف المتغير الذي له k من الصفات بـعدد k-1 من المتغيرات الوهمية حتى تكون مصفوفة المعلومات (**X'X**) ذات رتبة كاملة (full rank).

أشار (2001) Vuchkov and Boyadjieva وإسماعيل وآخرون (2003)، كغيرهم ممن اهتم بمجال تصميم التجارب وتحليلها، إلى النماذج الرياضية والطريقة التقليدية لإيجاد جدول تحليل التباين واختبار الفرضيات الخاصة بالتصاميم ومنها التصاميم الأساسية وهي: التصميم العشوائي الكامل وتصميم القطاعات العشوائية الكاملة وتصميم المربع اللاتيني.

قد يواجه الباحث من خلال استخدامه للطريقة التقليدية في تحليل التباين بعض العقبات منها: عدم إمكانية الحصول على معلومات إضافية كتلك التي يعطيها تحليل الانحدار، وجوب تقدير القيم المفقودة ويتبع ذلك تصحيح جدول تحليل التباين. ولتخطي العقبات نقدم أسلوبا تحليليا مقترحا يستخدم فيه تحليل الانحدار للمتغيرات الوهمية ذات الرمز التأثيري.

هدف البحث

يوضح تحليل التباين العلاقة بين متغير استجابة y_i ومتغيرات التوضيحية X_{ij} وذلك من خلال اختبار فرضية معنوية الفروقات بـين الأوسـاط الحسـابية لتلـك المتغيرات، والتي يعبر عنها بالمعاملات في مفهوم تصميم التجارب وتحليلهـا. إن

تحليل الانحدار يوضح تلك العلاقة بطريقة مختلفة وذلك من خلال نموذج يعنى بالمعالم واختبار الفرضيات الخاصة بها.

ضمن إطار العلاقة بين متغير الاستجابة والمتغيرات التوضيحية وضع الهدف الآتي:

الكشف عن العلاقة بين تحليل التباين للنموذج الثابت fixed model في ثلاثة تصاميم هي: العشوائي الكامل والقطاعات العشوائية الكاملة والمربع اللاتيني، وبين تحليل الانحدار للمتغيرات الوهمية ذات الرمز التأثيري لنتائج تجارب تلك التصاميم.

النموذج الخطي العام: General linear model

بين Agresti and Franklin (2007) أن قيم متغير الاستجابة تتأثر بعاملين: العامل الأول: هو حد تنبؤي مقداره: $(B_0+B_1X_{i1}+B_2X_{i2}+\cdots+B_mX_{im})$، والعامل الثاني: هو حد الخطأ العشوائي ε_i. كلا العاملين يكون نموذج الانحدار وصيغته هي:

$$y_i = B_0 + \sum_{j=1}^{m} B_j X_{ij} + \varepsilon_i \quad ; \quad i = 1,2,...,n \quad ; \quad j = 1,2,...,m \qquad ... \qquad (1)$$

y_i : القيمة i لمتغير الاستجابة.

$\beta_o,...,\beta_m$: ثوابت تدعى معالم الانحدار.

$X_{i1},...,X_{im}$: متغيرات توضيحية بعدد m كل منها يحتوي على عدد n من المشاهدات.

ε_i : الخطأ العشوائي الخاص بالوحدة التجريبية i.

أشار دبدوب والنعيمي (2006) الى أن طريقة المربعات الصغرى هي أفضل طريقة غير متحيزة لتقدير معالم نموذج الانحدار وتعطي مجموع مربعات للخطأ ($\varepsilon'\varepsilon$) أقل ما يمكن.

وتكون الصيغة باستخدام المصفوفات الاتية:

$\hat{\beta}=(X'X)^{-1}X'y$ ... (2)

في نموذج الانحدار (1) يكون مجموع المربعات الكلي هو:

$SStotal\ (SST) = SS(R(X_1,\ldots,X_{t-1})) + SSe$ ... (3)

حيث أن:

$SS(R(X_1,\ldots,X_m))$: مجموع المربعـات العائـد إلـى الانحـدار SS(due to regression)، والذي يمثل تأثير المعاملات في متغير الاستجابة، ويرمز إليه أيضاً بـ $SS(R_t)$.

SSe: مجموع مربعات الخطأ SS(error)، وهو مجموع مربعات انحراف القـيم المتوقعة عن مستوى الانحدار.

ولتحديد أهمية النموذج المقدر (1) في التنبؤ حول متغير الاسـتجابة تختبـر فرضية العدم الآتية:

$$Ho: B_1 = B_2 = \cdots = B_m = 0$$

ضد الفرضية البديلة (H_1):

$$H_1: B_1 \neq B_2 \neq \cdots \neq B_m \neq 0,\ (at\, least\, one\, of\ \beta_{j's} \neq 0)$$

يعتمد في الاختبار على المختبر الإحصائي F عند درجة حرية m وn-m-1، حيث أن:

$$F=\frac{MS(R(X_1\ldots X_m))}{MSe(X_1\ldots X_m)}=\frac{(\beta' X' y - n\bar{y}^2)/m}{(y' y - \beta' X' y)/(n-m-1)} \quad \ldots \quad (4)$$

هناك اختبار فرضية تتعلق بمجموعة جزئية من معالم الانحدار، الغايـة منـه معرفة أهمية وجود أو إضافة عدد $m-k$ من المتغيرات التوضيحية $(X_{k+1}\ldots.X_m)$ إلى نموذج الانحدار الذي يحتوي مسبقاً على عدد k من المتغيـرات التوضـيحية الأخرى. وهذا يعني اختبار الفرضية الآتية: (Berenson et al., 2006)

$$Ho: B_{k+1} = B_{k+2} = \cdots = B_m \mid \beta_1\ldots\beta_k = 0$$

$$H_1: B_{k+1} \neq B_{k+2} \neq \cdots \neq B_m \mid \beta_1\ldots\beta_k \neq 0,\ (at\, least\, one\, of\ \beta_{k+1},\ldots,\beta_m \neq 0)$$

والمختبر الإحصائي هو اختبار F بدرجة حرية $m-k$ ، $n-m-1$ ، حيث أن:

$$F = \frac{MS(R(X_{k+1}\, X_{k+2} \cdots X_m \mid X_1 \cdots X_k))}{MSe(X_1 X_2 \cdots X_m)} \qquad \ldots \qquad (5)$$

إن اختبار أهمية احتواء نموذج الانحدار على متغير ما يعني اختبار الفرضية الآتية:

$$H_0 : B_j \mid B_1\, B_2 \ldots B_{j-1}\, B_{j+1} \ldots B_m \;= 0$$
$$H_A : B_j \mid B_1 B_2 \ldots B_{j-1}\, B_{j+1} \ldots B_m \neq 0$$

وفي هذه الحالة يمكن اتباع اختبار F بدرجة حرية 1 و m، حيث أن:

$$F = \frac{MS(R(X_j \mid X_1\, X_2 \ldots X_{j-1}\, X_{j+1} \ldots X_m))}{MSe(X_1\, X_2 \;\ldots\ldots X_m)} \qquad \ldots \qquad (6)$$

التصميم العشوائي الكامل Completely randomized design

هو مخطط ذو اتجاه واحد يحتوي على t من المعاملات توزع عشوائياً على عدد tr من الوحدات التجريبية المتجانسة، حيث أن: r يمثل عدد تكرار كل معاملة في التجربة، وأن tr هو عدد n من قيم متغير الاستجابة نتيجة تطبيق المعاملات على الوحدات التجريبية المتجانسة.

تحول المتغيرات (المعاملات) في تحليل الانحدار إلى عدد t-1 من المتغيرات الوهمية ذات الرموز التأثيرية، كالآتي:

$$X_{ij} = \begin{cases} 1 & \text{إذا كانت المشاهدة } y_{ij} \text{ تعود للمعاملة } i \\ -1 & \text{إذا كانت المشاهدة } y_{ij} \text{ تعود للمعاملة } t \text{ (الأخيرة)} \\ 0 & \text{إذا كانت المشاهدة } y_{ij} \text{ تعود لغير ما ذكر} \end{cases}$$

وبذلك تنشأ مصفوفة المتغيرات الوهمية $X_{n\times((t-1)+1)}$ ، وعناصر المتجه $y_{n\times 1}$ عبارة عن قيم متغير الاستجابة، حيث أن:

$$y_{n\times 1} \begin{bmatrix} y_{11} \\ y_{12} \\ . \\ . \\ . \\ . \\ y_{uj} \end{bmatrix} \quad ; \quad X_{n\times t} = \begin{bmatrix} 1 & 1 & . & . & . & 0 \\ . & . & . & . & . & . \\ 1 & 1 & . & . & . & 0 \\ 1 & 0 & . & . & . & 1 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & 0 & . & . & . & 1 \\ 1 & -1 & . & . & . & -1 \\ . & . & . & . & . & . \\ 1 & -1 & . & . & . & -1 \end{bmatrix}$$

إن العلاقة بين تحليل التباين للتصميم وتحليل الانحدار للمتغيـرات الوهميـة، توضح في الفقرات الثلاث الآتية:

أ- النموذج الرياضي:

النموذج الرياضي لتحليل التباين للتصميم هو:

$$y_{ij} = u + \tau_j + \varepsilon_{ij} \qquad ... \qquad (7)$$

μ: الوسط الحسابي العام للتجربة.

τ_j: تأثير المعاملة j.

ε_{ij}: الخطأ التجريبي للوحدة التجريبية j التي طبقت عليها المعاملة i.

بعد إنشاء المتغيرات الوهمية يصبح نموذج الانحدار المشار إليه في المعادلـة (1) كالآتي:

$$y_i = \beta_0 + \underbrace{B_1 X_1 + B_2 X_2 + ... + B_{t-1} X_{t-1}}_{\text{تأثير المعاملات}} + \varepsilon_i \qquad ... \qquad (8)$$

ب- اختبار الفرضيات:

إن الفرضية التي يختبرها تحليل التباين للتصميم والعائدة إلى النموذج (7) هي:

$$H_o : \tau_1 = \tau_2 = ... = \tau_t = 0 \qquad ... \qquad (9)$$

$$H_1 : \tau_1 \neq \tau_2 \neq ... \neq \tau_t \neq 0 \ , \ (at\ least\ two\ of\ \tau_{i's}\ are\ not\ equal)$$

والفرضية التي يختبرها تحليل الانحدار والعائدة إلى النموذج (8) هي:

$$H_o : B_1 = B_2 = \ldots = B_{t-1} = 0 \qquad \ldots \qquad (10)$$
$$H_1 : B_1 \neq B_2 \neq \ldots \neq B_{t-1} \neq 0 \; , (at\ least\ one\ of\ \beta_{j's} \neq 0)$$

ت- مصادر التباين:

يلاحظ من النموذجين (7) و(8) تناظر الفرضيتين (9) و(10)، وعلية يمكن اختبار الفرضية (9) الخاصة بالتصميم باختبار الفرضية (10) الخاصة بتحليل انحدار المتغيرات الوهمية، ويبين الجدول (1) اختبار الفرضيتين (9) و(10).

الجدول 1: استخدام الترميز التأثيري في تحليل التباين للتصميم العشوائي الكامل.

S.O.V.	D.F.	S.S.	F
$R(X_1,...,X_{t-1})$	$t-1$	$SS(R(X_1,...,X_{t-1}))=SS(R_t)$	$\frac{MS(R_t)}{MSe}$
$error(X_1,...,X_{t-1})$	$n-t$	SSe	
$total$	$n-1$	SST	

تصميم القطاعات العشوائية الكاملة

Randomized complete blocks design

يحتوي هذا التصميم على اتجاهين هما: المعاملات بعدد t والقطاعات بعدد r. يقسم كل قطاع إلى عدد t من الوحدات التجريبية المتجانسة التي توزع عليها المعاملات عشوائياً. ونتيجة لتطبيق المعاملات على الوحدات التجريبية نحصل على tr من قيم متغير الاستجابة.

يتم إنشاء عدد t+r-2 من المتغيرات الوهمية، منها t-1 خاصة بالمعاملات و r-1 خاصة بالقطاعات، حيث أن: $i = 1, \ldots ,t \; ; \; j = 1, \ldots ,r$. وكالآتي:

$X_{1, \ldots ,t-1} =$
- 1 إذا كانت المشاهدة y_{ij} تعود للمعاملة t
- -1 إذا كانت المشاهدة y_{ij} تعود لآخر معاملة
- 0 غير ذلك

$X_{t, \ldots ,t+r-2} =$
- 1 إذا كانت المشاهدة y_{ij} تعود للقطاع r
- -1 إذا كانت المشاهدة y_{ij} تعود لآخر قطاع
- 0 غير ذلك

وبذلك تنشأ مصفوفة المتغيرات الوهمية $X_{n\times(t+r-2)}$ ، والمتجه $y_{n\times1}$ الـذي عناصـره هي عن قيم متغير الاستجابة، حيث أن:

$$X_{n\times t+r-2}=\begin{bmatrix}1 & 1 & . & . & 1 & 0\\ . & . & . & . & . & .\\ 1 & 1 & . & . & -1 & -1\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ 1 & 0 & . & . & 0 & 1\\ 1 & -1 & . & . & -1 & -1\end{bmatrix} \qquad ; \qquad y_{n\times1}=\begin{bmatrix}y_{11}\\ y_{12}\\ .\\ .\\ .\\ .\\ y_{uj}\end{bmatrix}$$

لنلاحظ العلاقة بين تحليل التباين لهذا التصميم وتحليل الانحدار للمتغيرات الوهمية في الفقرات الثلاث الآتية:

أ- النموذج الرياضي:

إن النموذج الرياضي لتصميم القطاعات العشوائية الكاملة هو:

$$y_{ij}=\mu+\rho_i+\tau_j+\varepsilon_{ij} \qquad \dots \qquad (12)$$

حيث أن:

$\hat{\mu}=\bar{y}_{..}$ الوسط الحسابي العام:

تأثير المعاملة i : $\tau_i=\hat{\mu}_i-\hat{\mu}=\bar{y}_{i.}-\bar{y}_{..}$

تأثير القطاعات j: $\hat{\rho}_j=\hat{\mu}_j-\hat{\mu}=y_{.j}-\bar{y}_{..}$

الخطأ التجريبي: $\hat{\varepsilon}_{ij}=y_{ij}-y_{i.}-y_{.j}+\bar{y}_{..}$

وبعد إنشاء المتغيرات الوهمية يصبح نموذج الانحدار المشار إليه في المعادلة (1) كالآتي:

$$y_i=B_o+\underbrace{B_1X_1+\dots+B_{t-1}X_{t-1}}_{\text{تأثيرات المعاملات}}+\underbrace{B_tX_t+\dots+B_{t+r-2}X_{t+r-2}}_{\text{تأثيرات القطاعات}}+\varepsilon_i \qquad \dots \qquad (13)$$

ب- اختبار الفرضيات:

إن جدول تحليل التباين الخاص بالتصميم يختبر فرضيتين هما:

الفرضية الأولى خاصة بالمعاملات وهي:

$$H_o : \tau_1 = ... = \tau_t \qquad ... \qquad (14)$$

$$H_1 : \tau_1 \neq ... \neq \tau_t \ , (at\ least\ two\ of\ them\ are\ not\ equal)$$

والفرضية الثانية خاصة بالقطاعات وهي:

$$H_o : \rho_1 = \rho_2 = ... = \rho_r \qquad ... \qquad (15)$$

$$H_1 : \rho_1 \neq \rho_2 \neq ... \neq \rho_r \ , (at\ least\ two\ of\ them\ are\ not\ equal)$$

والاختبار العام لتحليل انحدار المتغيرات الوهمية يعني اختبار الفرضية الآتية:

$$H_o : \beta_1 = ... = \beta_{t-1} = \beta_t = ... = \beta_{r+t-2} = 0 \qquad ... \qquad (16)$$

$$H_1 : \beta_1 \neq ... \neq \beta_{t-1} = \beta_t = ... = \beta_{r+t-2} \neq 0 \ , (at\ least\ one\ of\ \beta_{j's} \neq 0)$$

والفرضية (16) تناظر فرضية تحليل تباين التصميم الآتية:

$$H_o : \tau_1 = \tau_2 = ... = \tau_t = \rho_1 = ... = \rho_r \qquad ... \qquad (17)$$

$$H_1 : \tau_1 \neq \tau_2 \neq ... \neq \tau_t \neq \rho_1 = ... \neq \rho_r \ , (at\ least\ two\ of\ them\ are\ not\ equal)$$

من الفرضيتين (16) و(17)، يتبين أن كلاً من الفرضيتين (14) و(15) هما فرضية جزئية من الفرضية (16). ولاختبار الفرضيتين (14) و(15) باستخدام تحليل انحدار المتغيرات الوهمية، تختبر الفرضيتان الجزئيتان(18) و(19) الآتيتان:

$$H_o : \beta_1 = ... = \beta_{t-1} \mid \beta_t \ ... \ \beta_{r+t-2} = 0 \qquad ... \qquad (18)$$

$$H_1 : \beta_1 \neq ... \neq \beta_{t-1} \mid \beta_t \ ... \ \beta_{r+t-2} \neq 0 \ , (at\ least\ one\ of\ \beta_{j's} \neq 0)$$

$$H_o : \beta_t = ... = \beta_{r+t-2} \mid \beta_1 ... \beta_{t-1} = 0 \qquad ... \qquad (19)$$

$$H_1 : \beta_t \neq ... \neq \beta_{r+t-2} \mid \beta_1 ... \beta_{t-1} \neq 0 \ , (at\ least\ one\ of\ \beta_{j's} \neq 0)$$

ت-مصادر التباين:

من النموذج (13) يصبح مجموع المربعات العائد إلى الانحدار في المعادلة (3) كالآتي:

$$SS(R(X_1,...,X_{t+r-2}))=SS(R_t)+SS(R_r) \quad ... \quad (20)$$

وأن:

$$SS(R_t) = SS(R(X_1,...,X_{t-1})) \quad ... \quad (21)$$

$$SS(R_r) = SS(R(X_t,...,X_{t+r-2})) \quad ... \quad (22)$$

يتم اختبار الفرضيتين (14) و(15) الخاصتين بالتصميم عن طريق اختبار الفرضيتين (18) و(19) الخاصتين بتحليل انحدار المتغيرات الوهمية، كما هو موضح في الجدول (2).

الجدول 2: استخدام الترميز التأثيري في تحليل التباين لتصميم القطاعات العشوائية الكاملة.

S.O.V.	D.F.	S.S.	F
$R(X_1,...,X_{r+t-2})$	$r+t-2$	$SS(R(X_1,...,X_{r+t-2}))$	
$R(X_1,...,X_{t-1} \mid X_t,...,X_{r+t-2})$	$t-1$	$SS(R_t)$	$\frac{MS(R_t)}{MSe}$
$R(X_t,...,X_{r+t-2} \mid X_1,...,X_{t-1})$	$r-1$	$SS(R_b)$	$\frac{MS(R_b)}{MSe}$
$error(X_1,X_2,...,X_{r+t-2})$	$n-r-t+1$	SSe	
$total$	$n-1$	SST	

تصميم المربع اللاتيني Latin Square Design

هو عبارة عن تصميم قطاعين متعامدين أحدهما يمثل الصفوف rows (r_j) والآخر يمثل الأعمدة columns (c_k)، توزع المعاملات (t_i) عشوائياً على عدد rc من الوحدات التجريبية لتظهر كل معاملة مرة واحدة في كل صف ومرة في كل عمود، حيث إن: $i=1,2,\dots,t\ ;\ k=1,2,\dots,c\ ;\ j=1,2,\dots,r$ وأن: $r = c = t$.

ونتيجة لتطبيق المعاملات على الوحدات التجريبية نحصل على عدد $tr = n$ من قيم متغير الاستجابة.

ويتم إنشاء متغيرات وهمية ذات رموز تأثيرية بعدد t+r+c-3، منها t-1 تعود للمعاملات وr-1 تعود للصفوف وc-1 تعود للأعمدة، وكالآتي:

$X_j =$
- 1 إذا كانت المشاهدة $y_{jk(i)}$ تعود للمعاملة i
- -1 إذا كانت المشاهدة $y_{jk(i)}$ تعود للمعاملة الأخيرة
- 0 غيـــر ذلـــــك

$X_j =$
- 1 إذا كانت المشاهدة $y_{jk(i)}$ تعود للصف j
- -1 إذا كانت المشاهدة $y_{jk(i)}$ تعود للصف الأخير
- 0 غيـــر ذلـــــك

$X_j =$
- 1 إذا كانت المشاهدة $y_{jk(i)}$ تعود للعمود k
- -1 إذا كانت المشاهدة $y_{jk(i)}$ تعود للعمود الأخير
- 0 غيـــر ذلـــــك

وبذلك تنشأ مصفوفة المتغيرات الوهمية $X_{n\times(t+r+c-3)}$ ، والمتجه $y_{n\times 1}$ الذي عناصـره هي قيم متغير الاستجابة، حيث أن:

$$y_{rc\times 1} = \begin{bmatrix} y_{11(1)} \\ . \\ . \\ . \\ . \\ . \\ y_{rc(t)} \end{bmatrix} \quad ; \quad X_{rc\times(r+c+t-2)} = \begin{bmatrix} 1 & 1 & 0 & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . \\ 1 & 1 & 0 & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . \\ 1 & 0 & 1 & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & 0 & 0 & . & . & . & . & 1 \\ 1 & -1 & -1 & . & . & . & . & -1 \end{bmatrix}$$

إن العلاقة بين تحليل التباين وتحليل الانحدار للمتغيرات الوهمية تلاحظ فـي الآتي:

أ- النموذج الرياضي:

حيث إن النموذج الرياضي لتصميم المربع اللاتيني هو:

$$y_{jk(i)} = \mu + \rho_j + \gamma_k + \tau_{(i)} + \varepsilon_{jk(i)} \quad \ldots \quad (23)$$

وتأثيرات حدود النموذج هي:

الوسط الحسابي العام : $\hat{\mu} = \bar{y}_{..}$

تأثير المعاملة i: $\hat{\tau}_{(i)} = \hat{\mu}_i - \hat{\mu} = \bar{y}_{.(i)} - \bar{y}_{..}$

تأثير الصف j: $\hat{\rho}_j = \hat{\mu}_j - \hat{\mu} = y_{j.} - \bar{y}_{..}$

تأثير العمود k: $\hat{\gamma}_k = \hat{\mu}_k - \hat{\mu} = y_{.k} - y_{..}$

الخطأ التجريبي: $\hat{\varepsilon}_i = y_{jk(i)} - y_{j.} - y_{.k} - y_{.(i)} + 2\bar{y}_{..}$

ونموذج الانحدار لهذا التصميم باستخدام المتغيرات الوهمية هو:

$$y_i = \underbrace{B_0 + B_1x_1 + \ldots + B_{r-1}x_{r-1}}_{\text{تأثير المعاملات}} + \underbrace{B_r x_r + \ldots + B_{2r-2}x_{2r-2}}_{\text{تأثير الصفوف}} +$$

$$\underbrace{B_{2r-1}x_{2r-1} + \ldots + B_{3r-3}x_{3r-3}}_{\text{تأثير الأعمدة}} + \varepsilon_i$$

$$= B_0 + \sum_{j=1}^{r-1} B_j X_j + \sum_{j=r}^{2r-2} B_j X_j + \sum_{j=2r-1}^{3r-3} B_j X_j + \varepsilon_i \quad \ldots \quad (24)$$

ب-اختبار الفرضيات:

يختبر جدول تحليل التباين الخاص بتصميم المربع اللاتيني ثلاث فرضيات هي:

الفرضية الأولى خاصة بالمعاملات وهي:

$$H_o : \tau_1 = \ldots = \tau_t \quad \ldots \quad (25)$$

$$H_1 : \tau_1 \neq \ldots \neq \tau_t \ , (at\ least\ two\ of\ them\ are\ not\ equal)$$

والفرضية الثانية خاصة بالصفوف وهي:

$$H_o : \rho_1 = \rho_2 = \ldots. = \rho_r \quad \ldots \quad (26)$$

$$H_1 : \rho_1 \neq \rho_2 \neq \ldots \neq \rho_r \ , (at\ least\ two\ of\ them\ are\ not\ equal)$$

والفرضية الثالثة خاصة بالأعمدة وهي:

$$H_o:\gamma_1=\gamma_2=....=\gamma_c \quad ... \quad (27)$$

$$H_1:\gamma_1\neq\gamma_2\neq ...\neq\gamma_c,(at\ least\ two\ of\ them\ are\ not\ equal)$$

إن الفرضية العامة لتحليل الانحدار (due to regression) باستخدام المتغيـرات الوهمية هي:

$$H_o:\beta_1=\beta_2=...=\beta_{r-1}=\beta_r=...=\beta_{2r-2}=\beta_{2r-1}=..=\beta_{3r-3}=0 \quad ... \quad (28)$$

$$H_1:\beta_1\neq\beta_2\neq ...\neq\beta_{r-1}\neq\beta_r\neq ...\neq\beta_{2r-2}\neq\beta_{2r-1}\neq..\neq\beta_{3r-3}\neq 0\ (at\ least\ one\ of\ them\neq 0)$$

الفرضية (28) تعني اختبار فرضية تصميم المربع اللاتيني الآتية:

$$H_o:\tau_1=\tau_2=...=\tau_t=\rho_1=...=\rho_r=\gamma_1=...=\gamma_c \quad ... \quad (29)$$

$$H_1:\tau_1\neq\tau_2\neq ...\neq\tau_t\neq\rho_1=...\neq\rho_r\neq\gamma_1\neq ...\neq\gamma_c \quad (at\ least\ two\ of\ them\ are\ not\ equal)$$

يتبين من العلاقة بين الفرضيتين (28) و(29) أن الفرضيات (25) و(26) و(27) هي فرضيات جزئية من الفرضية (28)، وعليه يمكن اختبار هذه الفرضيات من خلال اختبار ثلاث فرضيات جزئية في تحليل الانحدار، وهي على التوالي:

$$H_o:\beta_1=...=\beta_{r-1}|\beta_r...\beta_{2r-2}\ \beta_{2r-1}...\beta_{3r-3}=0 \quad ... \quad (30)$$

$$H_1:\beta_1\neq ...\neq\beta_{r-1}|\beta_r...\beta_{2r-2}\ \beta_{2r-1}...\beta_{3r-3}\neq 0\ (at\ least\ one\ of\ them\neq 0)$$

$$H_o:\beta_r=...=\beta_{2r-2}|\beta_1...\beta_{r-1}\ \beta_{2r-1}...\beta_{3r-3}=0 \quad ... \quad (31)$$

$$H_1:\beta_r=...\neq\beta_{2r-2}|\beta_1...\beta_{r-1}\ \beta_{2r-1}...\beta_{3r-3}\neq 0\ (at\ least\ one\ of\ them\neq 0)$$

$$H_o:\beta_{2r-1}=...=\beta_{3r-3}|\beta_1...\beta_{r-1}\ \beta_r...\beta_{2r-2}=0 \quad ... \quad (32)$$

$$H_1:\beta_{2r-1}\neq ...\neq\beta_{3r-3}|\beta_1...\beta_{r-1}\ \beta_r...\beta_{2r-2}\neq 0\ (at\ least\ one\ of\ them\neq 0)$$

ت-مصادر التباين:

في هذا التصميم يصبح مجموع المربعات العائد إلى الانحدار في المعادلـة (3) كالآتي:

$$SS(R(X_1,...,X_{3r-3}))=SS(R_t)+SS(R_r)+SS(R_c) \quad ... \quad (33)$$

$$SS(R_t)=SS(R(X_1,...,X_{t-1}))=SS(R(X_1,...,X_{r-1})) \quad ... \quad (34)$$

$$SS(R_r)=SS(R(X_t,...,X_{r-1}))=SS(R(X_r,...,X_{2r-2})) \quad ... \quad (35)$$

$$SS(R_c)=SS(R(X_r,...,X_{c-1}))=SS(R(X_{2r-1},...,X_{3r-3})) \quad ... \quad (36)$$

يتم اختبار الفرضيات (25) و(26) و(27) الخاصة بالتصميم عن طريق اختبار الفرضيات (30) و(31) و(32) الخاصة بتحليل الانحدار الجزئي للمتغيرات الوهمية، كما هو مبين في الجدول (3).

الجدول 3: استخدام الترميز التأثيري في تحليل التباين لتصميم المربع اللاتيني.

S.O.V.	D.F.	S.S.	F
$R(X_1,...,X_{3r-3}) = due\ to\ regression$	$3r-3$	$SS(due\ to\ reg.)$	
$R_t(X_1,...,X_{r-1} \mid X_r,...,X_{3r-3})$	$t-1$	$SS(R_t)$	$\frac{MS(R_t)}{MSe}$
$R_r(X_r,...,X_{2r-2} \mid X_1...X_{r-1}, X_{2r-1}...X_{3r-3})$	$r-1$	$SS(R_r)$	$\frac{MS(R_r)}{MSe}$
$R_c(X_{2r-3},...,X_{3r-3} \mid X_1,...,X_{2r-2})$	$c-1$	$SS(R_c)$	$\frac{MS(R_c)}{MSe}$
$error(X_1, X_2,...,X_{3r-3})$	$n-3r+2$	SSe	
$total$	$n-1$	SST	

تطبيق:

أخذت البيانات من هدبة (2005). وهي نتيجة تجربة أجراها قسم البحوث الزراعية في موقع الرشيدية - محافظة نينوى، كان الهدف من التجربة إجراء مقارنات إحصائية بين متوسطات 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). طبقت التجربة في تصميم القطاعات العشوائية الكاملة بخمسة قطاعات، كل قطاع احتوى على خمس وحدات تجريبية متجانسة، في كل وحدة تجريبية أنشأ أربعة مروز طول كل مرز خمسة أمتار والمسافة بين كل مرزين متتاليين 75 سنتيمتراً، وزرع نبات واحد في الجورة بمسافة بين نبات وآخر 25 سنتيمتر. وقد سجلت نتائج التجربة وهي أوزان حاصل القطن الزهر بالكيلوغرام لكل وحدة تجريبية.

إن الهدف الرئيسي في دراسة هدبة (2005) هو المقارنة بين التصاميم الثلاثة الأساسية، وكان من الضروري لإجراء هذه المقارنات أن تحتوي هـذه التصـاميم على بيانات موحدة. عليه فقد أنشأ هدبة (2005) تصميمين هما: العشوائي الكامـل والمربع اللاتيني، مستعيناً في ذلك بنتائج التجربة التي طبقت بتصـميم القطاعـات العشوائية الكاملة، وقد اخذ بنظر الاعتبار ميزات كـل تصـميم. فكانـت بيانـات التصاميم كالآتي: العشوائي الكامل (الجدول 4) والقطاعـات العشـوائية الكاملـة (الجدول 6) والمربع اللاتيني (الجدول 8).

اسـتخدم لتحليل البيانات البرنامج الحاسوبي (Cody and Smith, 2006) SAS.

التصميم العشوائي الكامل:

الجدول 4: حاصل القطن الزهر بالكيلوغرام/وحدة تجريبية في تصميم عشـوائي كامل.

التكرار ← / المعاملات ↓	1	2	3	4	5	$Y_{i.}$
نازلي 503	4.97	5.03	4.89	4.85	4.84	24.58
1k347	4.58	4.49	4.49	4.35	4.33	22.24
طاقة 2	3.75	4.02	3.81	3.77	3.5	18.85
مرسومي	5.42	6.19	5.87	5.16	5.15	27.8
دلتا باية 5409	6.48	6.35	5.29	5.00	4.91	28.03

حللت البيانات المشار إليها في الجدول (4) بأسـلوبين: الأسلوب الأول: تحليل التباين الذي احتـوى على مصدرين للتباين همـا: المعامـلات treatments والخطـأ التجريبـي experimental error، والأسلوب الثاني: تحليل الانحدار باستخدام المتغيرات الوهمية ذات الرمز التأثيري والتي أشار إلى أسلوبها الجدول (1). وقد أعطى الأسلوبان لنتائج التحليلية نفسها المشار إليها في الجدول (5).

الجدول 5: نتائج تحليل التباين وتحليل الانحدار في التصميم العشوائي الكامل.

S.O.V.	D.F.	S.S.	M.S.	F
treatments = $due\,to\,regression$ = $R(X_1,...,X_{t-1})$	4	12.04	3.01	17.71
error = $error(X_1,...,X_{t-1})$	20	3.32	0.17	
total	24	15.35		

تصميم القطاعات العشوائية الكاملة:

الجدول 6: حاصل القطن الزهر بالكيلوغرام/وحدة تجريبية عند تطبيـق تصـميم القطاعات العشوائية الكاملة.

القطاعات ←	1	2	3	4	5	$Y_{i.}$
نازلي 503	4.97	5.03	4.89	4.85	4.84	24.58
1k347	4.58	4.49	4.49	4.35	4.33	22.24
طاقة 2	3.75	4.02	3.81	3.77	3.50	18.85
مرسومي	5.43	6.19	5.87	5.16	5.15	27.80
دلتابابية 5409	6.48	6.35	5.29	5.00	4.91	28.03
$Y_{.j}$	25.21	26.08	24.35	23.13	22.73	121.50

وقد طبق تحليلان على بيانات الجدول (6) هما: تحليل التباين الذي أشار إلـى ثلاثة مصادر للتباين، هي: المعاملات treatments والقطاعات blocks والخطـأ التجريبي experimental error، وتحليل انحدار المتغيرات الوهميـة باسـتخدام الرموز التأثيرية التي أشار إلى أسلوبها الجدول (2)، وقد أعطى التحليلان النتـائج نفسها (الجدول 7).

الجدول 7: نتائج تحليل التباين وتحليل الانحدار في تصميم القطاعات العشوائية الكاملة.

S.O.V.	D.F	S.S.	M.S.	F
$due\ to\ regression = R(X_1,...,X_{r+t-2})$	8	13.61	1.70	15.45
treatments = $R(X_1,...,X_{t-1} \mid X_t,...,X_{r+t-2})$	4	12.04	3.01	27.36
blocks = $R(X_t,...,X_{r+t-2} \mid X_1,...,X_{t-1})$	4	1.57	0.39	3.54
error = $error(X_1, X_2,...,X_{r+t-2})$	16	1.75	0.11	
total	24	15.36		

تصميم المربع اللاتيني:

الجدول 8: حاصل القطن الزهر بالكيلوغرام/وحدة تجريبية في تصميم مربع لاتيني.

	C1	C2	C3	C4	C5	$Y_{j.}$	$Y_{(i).}$
R1	4.97 t1	6.35 t5	5.87 t4	3.77 t3	4.33 t2	25.29	24.58
R2	4.58 t2	5.03 t1	5.29 t5	5.16 t4	3.50 t3	23.56	22.24
R3	3.75 t3	4.49 t2	4.89 t1	5.00 t5	5.15 t4	23.28	18.85
R4	5.43 t4	4.02 t3	4.49 t2	4.85 t1	4.91 t5	23.70	27.80
R5	6.48 t5	6.19 t4	3.81 t3	4.35 t2	4.84 t1	25.67	28.03
$Y_{.k}$	25.21	26.08	24.35	23.13	22.73	121.50	

وقد حللت بيانات الجدول (8) بأسلوبين: الأول: تحليل التباين للتصـميم ولـه أربعة مصادر للتباين، هي: المعاملات treatments والصفوف rows والأعمـدة columns والخطأ التجريبي experimental error، والأسلوب الثـاني: تحليـل الانحدار باستخدام المتغيرات الوهمية ذات الرمز التأثيري (الجـدول 3). وأعطـى الأسلوبان النتائج نفسها (الجدول 9).

الجدول 9: نتائج تحليل التباين وتحليل الانحدار في تصميم المربع اللاتيني.

S.O.V	D.F.	S.S.	M.S.	F
$due\,to\,regression = R(X_1,...,X_{3r-3})$	12	14.57	1.21	17.29
treatments = $R_t(X_1,...,X_{r-1} \mid X_r,...,X_{3r-3})$	4	12.04	3.01	43.00
rows = $R_r(X_r,...,X_{2r-2} \mid X_1...X_{r-1}, X_{2r-1}...X_{3r-3})$	4	0.96	0.24	3.43
columns = $R_c(X_{2r-3},...,X_{3r-3} \mid X_1,...,X_{2r-2})$	4	1.57	0.39	5.57
error = $error(X_1, X_2,...,X_{3r-3})$	12	0.79	0.07	
total	24	15.35		

أستخدم دبدوب (2006) المتغيرات الوهمية في التصاميم الأساسية للتغلب على القيم المفقودة دون اللجوء الى تقديرها، وقد بين أن أسلوبه المستخدم أعطى نفـس نتائج طريقة Yates في معالجة القيم المفقودة، فضلا عن حصوله مباشـرة علـى جدول تحليل التباين المصحح وفق عدد القيم المفقودة في بيانات التجربة المصممة. إن هذه النتائج تضيف فائدة إلى أسلوبنا المقترح من خلال معالجته القيم المفقودة.

للتأكد مما تم التوصل إليه فقد طبق الأسلوب المقترح لعدة مرات على نتـائج تجارب صممت حسب التصاميم: العشوائي الكامل والقطاعات العشـوائية الكاملـة والمربع اللاتيني.

الاستنتاجات

إن جدول تحليل التباين لكل من التصاميم: العشوائي الكامل والقطاعات العشوائية الكاملة والمربع اللاتيني قد تم التوصل إليها بتطبيق أسلوبين هما: الأسلوب الأول: إتباع أسلوب تحليل التباين التقليدي للتصميم المعني. والأسلوب الثاني (المقترح): هو تحليل انحدار المتغيرات الوهمية ذات الرمز التأثيري والذي أعطى نتائج الأسلوب السابق نفسها. إن لكل أسلوب ميزاته، فالأسلوب التقليدي لتحليل التباين يتميز بسهولة التطبيق للحصول على جدول تحليل التباين. في حين أن أسلوب تحليل انحدار المتغيرات الوهمية يتميز بإعطائه نتائج إضافية كتلك التي يعطيها تحليل الانحدار منها: قيمة معامل التحديد (R^2) والتي يمكن التوصل إليها من خلال المعلومات المتوافرة في الجداول (1 و2 و3)، وتعيين القيم الشاذة ومعالجتها كما أشار إلى ذلك دبدوب ويونس (2006)، كما يمكن التغلب على القيم المفقودة دون اللجوء إلى تقديرها، فضلاً عن الحصول مباشرة على جدول تحليل التباين المصحح لـ Yates نتيجة معالجة القيم المفقودة (دبدوب، 2006)، وغيرها من النتائج التي يمكن أن يستفيد منها الباحث في عمله البحثي.

المصادر العربية

1- إسماعيل، أكرم عثمان ومصطفى، عبد الرحيم عمر وعمر، قاسم عبد ا ، 2003، "تصميم التجارب وتحليلها"، الأمم المتحدة منظمة الأغذية والزراعة FAO، جمهورية العراق.

2- دبدوب، مروان عبد العزيز، 2006، "المتغيرات الوهمية بديلة عن تقدير القيم المفقودة في تجارب التصاميم الأساسية"، مجلة جامعة تكريت للعلوم الإدارية والاقتصادية، كلية الإدارة والاقتصاد-جامعة تكريت، المجلد 2، العدد 4، ص 167-184.

3- دبدوب، مروان عبد العزيز والنعيمي، أسوان محمد طيب، 2006، "طرائق مقترحة في انحدار الحرف"، المجلة العراقية للعلوم الإحصائية، كلية علوم الحاسبات والرياضيات-جامعة الموصل، المجلد 6، العدد 10، ص 85-106.

4- دبدوب، مروان عبد العزيز ويونس، فرح عبد الغني، 2006، " تـأثير القـيم الشاذة على نتائج تحليل الانحدار مع التطبيق على المواليد الخدج"، مجلة علوم الرافدين، كلية العلوم-جامعة الموصل، المجلد 17، العدد 1، ص 61-82.

5- هدبة، وكاع علي، 2005، " مقارنـة لبعض طرائق تقدير ومعالجـة القيـم المفقودة في تصاميم التجارب الأساسية"، رسـالة ماجستير (غير منشورة)، كلية عـلوم الحاسـبات والرياضيات، جامعـة الموصل، العراق.

المصادر الأجنبية

6- Agresti, A. and C. Franklin; 2007; "Statistics The Art and Science of Learning From Data", Pearson Prentice Hall, USA.

7- Bererson, M.; D. Levine and T. Krehbiel; 2006; "Basic Business Stat- istics–Concept and Applications", 10th edition, Pearson Prentice Hall, USA.

8- Cody, R. and J. Smith; 2006; "Applied Statistics And SAS Programm- ing Language", 5th edition, Pearson Prentice Hall, USA.

9- Vuchkov, I. N. and L. N. Boyadjieva; 2001; "Quality Improvement with Design of Experiments: A Response Surface Approach"; Kluwer Academic Publishers; London.