دراسة تأثير القيم الشاذة في بيانات تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة

المدرس رحيم جبار ظاهر
جامعة القادسية
كلية الإدارة والاقتصاد
قسم الإحصاء

ملخص البحث

من أهم المشاكل التي تظهر في البيانات هي وجود قيمه واحدة أو اكثر من القيم الشاذة ، هذه المشكلة تم دراستها بعمق في مسائل تحليل الانحدار ولكنها لم تلقى الاهتمام نفسه

في بيانات تصميم التجارب. في هذا البحث تم توضيح كيفية الكشف عن القيم الشاذه في بيانات تصميم خطط المعاينه المتزنة باستثناء الوحدات المجاورة وكيف يمكن لهذه القيم ان تؤثر على التحليل الاحصائي للتجربة ، وكذلك تم توضيح اسلوب المعالجة في حالة وجود قيم شاذة.

المقدمة

اقام الباحث سلسلة من التجارب في السنوات 1999 و 2000 [1], [2] لدراسة تاثير بعض مستويات سمادي النيتروجين والفسفور على بعض صفات نبات الذرة الصفراء مثل صفة

طول العرنوص ، طول النبات، حاصل الحبوب ، حاصل الوزن الجاف ، نسبة البروتين و نسبة الزيت مطبقة بتصميم خطط المعاينه المتزنة باستثناء الوحدات المجاورة وتوصل الباحث الى ان هنالك فروق معنوية مابين توافيق مستويات العوامل (المعالجات) وجميع الصفات المدروسة وعند تحليل مستوى الاستجابه وجد ان هنالك علاقة خطية مابين مستويات العوامل وجميع الصفات المدروسة وعند تكرار التجربة في عام 2001 وجد انه لايوجد أي فرق معنوي مابين توافيق مستويات العوامل على صفة وزن المادة الجافة وحللت التجربة في ذلك الوقت على انه لايوجد أي فرق معنوي بين المعالجات على صفة وزن المادة الجافة على الرغم من ان هذه النتيجة تتعارض مع خبرة الباحث السابقة في هذا المجال.

في عام 2002 قام كل من Bhar و Gupta بتطوير احصائية Cook [4] بحيث يمكن استخدامها في تصميم التجارب للكشف عن القيم الشاذة في تصميم القطاعات الكاملة العشوائية ، وقد استخدم الباحث هذه الاحصائية في تحليل بيانات التجربة (التي ظهرت فيها الفروق غير معنوية) بعد توظيف معالم تصميم خطط المعاينه المتزنة باستثناء الوحدات المجاورة في احصائية Cook المطورة فوجد ان هنالك قيمتان تشذان عن بقية القيم بحيث أثرت هذه القيم على كل التحليل الإحصائي للتجربة. ويهدف هذا البحث الى توضيح كيفية الكشف عن القيم الشاذة في تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة وكذلك الى كيفية معالجتها.

إحصائية Cook

في عام 1977 درس Cook [3] تأثير القيم الشاذة وقد اقترح صيغة تستخدم للكشف عن القيم الشاذة في النماذج الخطية ، حيث درس Cook النموذج الرياضي التالي :

$$Y = X\theta + e \qquad \text{.....(1)}$$

X تمثل مصفوفة الثوابت المعلومة [4].

θ تمثل متجه المعلمات المجهولة.

$$E(e) = 0$$
$$D(e) = \sigma^2 I_n \quad , \sigma^2 > 0$$

وقام بتحديد درجة تاثيرالقيمة الشاذة (i) من البيانات من خلال الصيغة التالية والتي سميت بإحصائية Cook :

$$D_i = \frac{(\hat{\theta} - \hat{\theta}_{(i)})'[D(\hat{\theta})]^{-1}(\hat{\theta} - \hat{\theta}_{(i)})}{rank\,[D(\hat{\theta})]} \qquad .....(2)$$

حيث ان:

$\hat{\theta}$ تمثل تقدير الى θ.

$\hat{\theta}_{(i)}$ تمثل تقدير المربعات الصغرى الى θ عند حذف القيمه الشاذة i .

هذه الإحصائية مفيدة جدا لبيان درجة التأثير لمجموعة جزئية من المعلمات وعلى الرغم من ذلك لم يتم استخدامها بشكل مباشر لاختبار الشواذ في بيانات تصميم التجارب ، ومن المعلوم ان مصفوفة التصميم X هي مصفوفة غير كاملة الرتبة وبالتالي فان التقدير الوحيد للمعالم هو غير ممكن ، وعلى كل حال في تصميم التجارب نحتاج الى مجموعة جزئية من المعلمات اكثر من حاجتنا الى كل المعلمات الخاصة بالتصميم فعلى سبيل المثال في تصميم القطاعات غير الكاملة نحن نهتم بتقدير تاثير المعالجات اكثر من حاجتنا الى تقدير تاثير بقية المعلمات مثل تاثير القطاعات او المتوسط العام ، وقد انتبه كل من Bhar و Gupta [**4**] الى هذه المسألة وقاموا بتطوير احصائية Cook بحيث يمكن تطبيقها على بعض التصاميم ومنها تصميم القطاعات الكاملة العشوائية حيث قام الباحثان باعادة كتابة النموذج (1) كما في الشكل (3) وذلك بافتراض ان هنالك t من المشاهدات مشكوك فيها على انها قيم شاذه.

$$Y = Z\gamma + e \qquad .....(3)$$

حيث ان:

$$Z = (X \quad U) \quad , \quad \gamma = (\theta' \quad \delta')'$$
$$U = (u_1, u_2, \ .... \ , u_t)$$
$$u_i = (0,0, \ ...., \ i(t-th),0,0,....,0)'$$
$$\delta = (\delta_1, \delta_2, \ ... \ , \delta_t)$$

وبالتالي قام الباحثان باعادة صياغة احصائية Cook كما في الصيغة (4) والتي تم تطبيقها على البيانات الخاصة بتصميم القطاعات الكاملة العشوائية في حالة كون القيمة i هي قيمه شاذة

$$D_i = \frac{\hat{\delta}_i u_i' S u_i \hat{\delta}_i}{(v-1)\hat{\sigma}^2} \qquad ....(4)$$

حيث ان :

$$S = \Phi \Delta' C_\tau \Delta \Phi$$
$$\Phi = I_n - D' K^{-1} D$$
$$K = DD'$$
$$C_\tau = \Delta \Phi \Delta'$$

D' هي مصفوفة تصميم $n \times b_{(0-1)}$ والتي تمثل تاثير القطاعات.

I_n مصفوفة الوحدة.

Δ' هي مصفوفة تصميم $n \times v_{(0-1)}$ والتي تمثل تاثير المعالجات

تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة

يعرف تصميم خطط المعاينه المتزنة باستثناء الوحدات المجاورة $BSEC(v, k, \lambda)$ على انه تصميم قطاعات غير كاملة متزنة جزئيا حيث يتم ترتيب (v) من المعالجات في (b) من القطاعات والتي حجم كل منها هو (k) بحيث تتحقق الشروط الاتية [5]:

1- كل معالجة تتكرر (r) من المرات ، ولاتوجد معالجة تظهر اكثر من مرة واحدة في أي قطاع من قطاعات التجربة.

2- أي زوج من المعالجات المجاورة لبعضها البعض لايظهر معا في أي قطاع من قطاعات التجربة.

3- أي زوج من المعالجات غير المجاورة لبعضها البعض يظهر في λ قطاع من قطاعات التجربة.

كما ان هذا التصميم يحقق العلاقات التالية [6]:

$(i) \quad bk = vr$

$(ii) \quad \lambda(v-3) = r(k-1)$

$(iii) \quad v \geq 3k$

$(iv) \quad b \geq v$

ويمكن كتابة النموذج الرياضي لتصميم خطط المعاينه المتزنه باستثناء الوحدات المجاورة بالشكل التالي :

$$Y = \mu 1_n + \Delta'\tau + D'\beta + e \qquad \text{.....(5)}$$

حيث ان μ يمثل المتوسط العام

τ يمثل متجه تاثير المعالجات ويتكون من v من المكونات.

β يمثل متجه تاثير القطاعات ويتكون من b من المكونات.

$\Delta 1 = r = (r_1, r_2, ..., r_v)$
$D1 = k = (k_1, k_2, ..., k_b)$

ويتم حساب مجاميع المربعات في جدول تحليل التباين كما هو موضح في الجدول الاتي :

جدول رقم (1)

جدول تحليل التباين (داخل القطاعات) لتصميم

$BSEC(v, k, \lambda)$

Source of variation	Degrees of freedom	Sum of square	Mean square
Treatments (adjusted)	$v-1$	$S_t^2 = \sum_{i=1}^{t} \hat{t}_i Q_i$	$\frac{S_t^2}{v-1}$
Blocks (unadjusted)	$b-1$	$S_b^2 = \frac{1}{k}\sum_{j=1}^{b-1} B_j^2 - \frac{G^2}{vb-k}$	$\frac{S_b^2}{b-1}$
Error	$vr-v-b-1$	S_e^2 (بالطرح)	$\frac{S_e^2}{vb-v-b-1}$
Total	$vr-1$	$\sum y_{ij}^2 - \frac{G^2}{vb-k}$	

حيث ان :

$\hat{t}_i$ تمثل تقدير تاثير المعالجة i.

Q_i تمثل المجموع المصحح للمعالجة i علما بان :

$$Q_i = T_i - \frac{V_i}{k} \quad .....(6)$$

T_i تمثل مجموع نتائج المعالجة i .

V_i تمثل مجموع مجاميع القطاعات التي تظهر فيها المعالجة i .

y_{ij} تمثل مجموع المعالجة i الواقعة في القطاع j .

B_j تمثل مجموع القطاع j .

G المجموع العام .

و قام الباحث بتوظيف معالم تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة في احصائية Cook المطورة وقد تم التوصل الى الصيغة التالية وذلك بافتراض ان القيمة التابعة للمعالجة i والواقعة في القطاع i تمثل قيمة شاذة

$$D_i = \frac{s_{ii}}{(v-1)\hat{\sigma}^2}\delta_i^2 \qquad ....(7)$$

حيث ان s_{ii} يمثل العنصر الاول في القطر الرئيسي للمصفوفة S .

الجانب التطبيقي:

التجربة:

البيانات التالية تمثل بيانات لتجربة حقلية اجريت وفق تصميم خطط المعاينه المتزنة باستثناء الوحدات المجاورة خلال الموسم الزراعي 2001 في مزرعة الباحث جواد عبد الكاظم

كمال[1] لدراسة تاثير بعض مستويات السماد النيتروجيني والفسفوري المضاف على حاصل الوزن الجاف لنبات الذرة الصفراء صنف بحوث 106، وقد اشتملت التجربة على ثلاث مستويات للسماد النيتروجيني هي(30 , 60 , 90) كغم N /هكتار وثلاث مستويات للسماد الفسفوري هي(30 ، 60 ، 90)كغم P/ هكتار و تم تقدير حاصل الوزن الجاف من وزن النبات من مساحة 1 م ^{2} والممثل لحاصل المادة الجافة الكلية فوق سطح التربة ثم حول طن / هكتار ويظهر في جدول (1) البيانات الخاصة بحاصل الوزن الجاف للقطعة التجريبية في مخطط التجربة ، كما تم اعطاء رقم لكل مشاهدة من المشاهدات.

جدول (2)

مخطط يوضح توزيع معالجات التجربة

ويحتوي على رمز وقيمة ورقم كل مشاهدة من المشاهدات .

Bocks	Treatments		
1	$n_{30}p_{90}$ 5.200 (1)	$n_{30}p_{30}$ **4.100** (10)	4.600 (19) $n_{90}p_{30}$
2	$n_{30}p_{60}$ 3.970 (2)	4.100 (11) $n_{60}p_{30}$	5.100 (20) $n_{90}p_{60}$
3	$n_{30}p_{90}$ 4.300 (3)	4.900 (12) $n_{60}p_{60}$	4.950 (21) $n_{90}p_{90}$
4	5.500 (4) $n_{60}p_{30}$	5.370 (13) $n_{60}p_{90}$	4.280 (22) $n_{30}p_{30}$
5	4.700 (5) $n_{60}p_{60}$	4.170 (14) $n_{90}p_{30}$	5.400 (23) $n_{30}p_{60}$
6	$n_{60}p_{90}$ **5.400** (6)	4.650 (15) $n_{90}p_{60}$	4.700 (24) $n_{30}p_{90}$

[1] الدكتور جواد عبد الكاظم كمال استاذ مساعد في قسم الجغرافية / كلية الآداب / جامعة القادسية.

7	4.900 (7) $n_{90}p_{30}$	4.480 (16) $n_{90}p_{90}$	5.200 (25) $n_{60}p_{30}$
8	5.840 (8) $n_{90}p_{60}$	5.100 (17) $n_{60}p_{60}$	4.700 (26) $n_{30}p_{30}$
9	5.700 (9) $n_{90}p_{90}$	4.480 (18) $n_{30}p_{60}$	5.390 (27) $n_{60}p_{90}$

تم إجراء التحليل الإحصائي لبيانات هذه التجربة وكما مبين في الجدول رقم (3)

جدول رقم (3)

يبين مكونات جدول تحليل التباين للتجربة

Source of variation	Degrees of freedom	Sum of square	Mean square	Computed F	Tabular F	
					5%	1%
Treatments (adjusted)	8	2.083	0.26	0.79	3.07	5.06
Blocks (unadjusted)	8	1.730	0.22	0.67		
Error	10	3.295	0.33			
Total	26	7.108				

يتضح من بيانات جدول تحليل التباين ان قيمة F المحسوبة هي اقل من قيمة F الجدولية عند المستويين 0.05

، 0.01 وهذا يعني عدم وجود فروق معنوية بين المعالجات على صفة وزن المادة الجافة.

بعد ذلك تم حساب احصائية Cook لكل مشاهدة من المشاهدات وكما موضح في الجدول (4).

جدول(4)

قيمة كوك لكل مشاهدة من المشاهدات

No. Of Observation	D_i
1	1.20
2	1.61
3	0.21
4	2.03
5	1.30
6	**3.70**
7	1.25
8	1.17
9	0.24
10	**4.32**
11	1.08
12	2.45
13	0.19
14	1.16
15	2.29
16	1.31
17	1.25
18	1.68
19	1.93
20	2.01
21	1.05
22	0.03
23	1.25
24	2.34
25	0.19
26	1.28

27	1.04

من الجدول في اعلاه نلاحظ ان قيمة Cook للمشاهدة التابعة للمعالجة $n_{60}p_{90}$ الواقعة في القطاع السادس وكذلك قيمة Cook للمشاهدة التابعة للمعالجة $n_{30}p_{30}$ الواقعة في القطاع الاول هي بارزة بالمقارنة مع قيمة Cook لبقية المشاهدات وبالتالي من الممكن ان تكون هذه القيم قد اثرت على التحليل الاحصائي لكل التجربة، وعند مقارنة جميع قيم Cook مع قيمة F الجدولية بمستوى معنوية 0.05 ودرجة حرية (10 , 8) والمساوية الى (3.07) وجد ان جميع قيم Cook هي اقل من (3.07) ماعدا القيمتان البارزتان . وبالتالي فان هاتان القيمتان يؤثران إحصائيا وهذا يزيد الشك لدينا في ان هذه القيم ربما قد اثرت على التحليل الاحصائي للتجربة.

وعند حذف القيمة التابعة للمعالجة $n_{30}p_{30}$ الواقعة في القطاع الاول والتعامل معها وكأنها قيمه مفقودة ومن ثم اجراء التحليل الاحصائي للتجربة وحساب قيمة F فكانت مساوية الى (4.11) وبمقارنتها مع قيمة F الجدولية بمستوى معنوية 0.05 ودرجة حرية (9 , 8) وجد ان الفروق بين المعالجات معنوية تحت مستوى احتمال 0.05 وكخطوة ثانية تم حذف القيمة التابعة للمعالجة $n_{60}p_{90}$ الواقعة في القطاع السادس ومن ثم اجراء التحليل الاحصائي وحساب قيمة F فكانت مساوية الى (3.88) وبمقارنتها مع قيمة F الجدولية وجد ان الفروق بين المعالجات معنوية تحت مستوى احتمال 0.05 .

كخطوة ثالثة تم حذف القيمتان معا وتم اجراء التحليل الاحصائي وحساب قيمة F فكانت مساوية الى (6.71) وبمقارنتها مع قيمة F الجدولية وجد ان الفروق بين المعالجات معنوية تحت مستوى احتمال 0.05وكذلك تحت مستوى احتمال 0.01 .

وللتاكد من ان قيمة المشاهدة التابعة للمعالجة $n_{30}p_{30}$الواقعة في القطاع الاول و قيمة المشاهدة التابعة للمعالجة $n_{60}p_{90}$ الواقعة في القطاع السادس هما القيمتان الوحيدتان اللتان تؤثران على التحليل الاحصائي للتجربة قام الباحث باجراء التحليل الاحصائي لبيانات التجربة عدة مرات وذلك بحذف قيمة واحدة في كل مرة من القيم المتبقية ومن ثم حساب قيمة F فوجد انه لاتوجد فروق معنوية في جميع الحالات ماعدا في حالة القيمتان السابقتان.

الاستنتاجات والتوصيات

في حالة الشك في ان بيانات تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة تحتوي على قيم شاذة يمكن استخدام الصيغة (7) للكشف عن القيم الشاذة ومن ثم تتم عملية معالجة هذه القيم. لذلك يوصي الباحث بضرورة استخدام هذه الصيغة في حالة وجود قيم شاذة في تصميم خطط المعاينة المتزنة باستثناء الوحدات المجاورة.

المصادر

[1] الحمزاوي ، رحيم جبار . كمال ، جواد عبد الكاظم " اثر السماد النيتروجيني والفسفوري على
طول العرنوص لنبات الذرة الصفراء" مجلة جامعة القادسية للعلوم الادارية والاقتصادية ، المجلد

6 العدد 2 لسنة 2004

[2]الحمزاوي ، رحيم جبار " تاثير بعض مستويات سمادي النيتروجين والفسفور على بعض صفات

حاصل الذره الصفراء (صنف بحوث 106) " مجلة جامعة القادسية للعلوم الادارية والاقتصادية

، المجلد 7 العدد 3 لسنة 2005.

[3] Cook, R.D. (1977). Detectionof influentialobservation in linear regression,
technometrics, 19 , 15-18

[4] Bhar, L and V.K.Gupta(2001). A Useful statistic for studying outliers in
experimental designs. Sankhya: the Indian Journal of Statistics, 63, 338-350

[5] Hedayat, A.S., Rao, C.R., And Stufken, J. (1988). A sampling plan excluding contiguous units. J. Stat. plann. Infer. 19, 159-170.

[6] See, K., Song, S. Y., and Stufken, J. (1997). on a class of partially balanced incomplete block designs with application . Commun. Statist.-Theory
meth., 26(1), 1-13