中華農業研究（Jour. Agric. Res. China）45(3)：218～229（1996）

作物區域試驗穩定性分析之缺區估計[1]

呂　秀　英[2]

摘要：育種程序中作物區域試驗施行之目的，在選出高產且穩定之品種。本省區域試驗一般採用的穩定性介量多以最受廣泛使用的直線迴歸分析法爲主。該法僅適用於每一基因型與環境組合都有資料之場合，因此一旦含有缺區，便無法正確估得環境指標，並進行穩定性分析；過去一般處理缺區的方式，係將含缺區的品系或環境全部刪除，使其呈均衡資料後，再進行穩定性分析，如此在缺區數多時勢必會造成品系或環境數減少，反而降低了評估結果的準確度。因此本研究擬在不摒除任何品系及環境數之情況下，利用傳統的缺區估計法使填入缺區估計值後，以均衡資料的型態再進行穩定性分析，其對穩定性分析評價結果的影響，將用來檢視缺區估計法在區域試驗資料的適用性。結果得知，填入缺區估計值後使呈均衡資料再進行標準穩定性分析，雖然簡單，但因爲會造成缺區品系之穩定性係數低估，故僅適用於穩定品系含少數一、二個缺區的場合，一旦缺區發生在不穩定品系時，容易使原來不穩定品系被誤認爲是穩定。

關鍵詞：缺區、穩定性、區域試驗、不均衡資料、迴歸分析。

目前本省水稻及各種雜糧作物之新育成品系，均規定在命名登記之前須於各農業改良及試驗場所至少進行兩年的區域試驗。所謂區域試驗，是由各場所依據育種程序中選出之優良品系，並加入當地優良推廣品種爲對照，於主要栽培季節分別舉行設計相同的品種比較試驗；而各地之栽培及管理方式則可按當地慣行辦法施行。育種程序中施行區域試驗之目的，在探討各品種於各特定地區之適應能力，並選出某一區域適應性較廣的品種。

品種之基因型及環境交感效應小者，其對環境的適應能力較大。適應性（adaptation）包括：穩定性（stability）、競爭性（competition）、忍耐性（tolerance）及可塑性（plasticity）[(1)]。育種程序中區域試驗施行之目的，大多著重於品種穩定性之探討。所謂穩定性，爲二種以上不同的品種在二種以上不同的環境下生活時，因環境的不同，使不同基因型對環境變異的反應產生相對間之差異。穩定性分析法的種類甚多，本省區域試驗一般採用的穩定性介量多以最受廣泛使用的直線迴歸分析法爲主。直線迴歸模式能有效地估計數量性狀對環境的適應能力，其所估得之迴歸係數並有容易解釋的優點，然而，由於其受限於某些統計前提，故僅適用於以下場合[(7,13,16,17,22,24,29,30)]：(1)基因型、環境及重複數很多；(2)基因型與環境交感效應大部份可用直線關係來解釋；(3)基因型之迴歸離差均方之間互爲均質；(4)基因型之機差變方均質（如一般變方分析所需之假設前提）。當這些任一條件不符合時，所求得之穩定性介量的估計值便不足以代表基因型對環境之適應能力。

1. 臺灣省農業試驗所　研究報告第1856號。
2. 本所農藝系副研究員。臺灣省　臺中縣　霧峰鄉。

作物區域試驗中常因人爲疏失而造成離值，此一般試驗者多將其拋棄不用，另外試驗期間受到天災人禍等因素，難免也會使部份品種在某些地區無法獲得資料，此等調查資料的漏失，即所謂的缺區（missing plot）現象。有關一般試驗資料如逢機完全區集設計、裂區設計或重複測量的缺區估計，已有理論上的根據[26,27]。而區域試驗資料實際上是一個含有基因型與環境交感效應的兩因子試驗，複因子試驗對於不均衡或不完整資料（unbalanced or incomplete data）的變方分析結果，雖可針對因子間之交感效應存在與否，利用常數配合法（fitting constant method）或均值加權平方法（weighted squares of means method）求得[4]，但由於區域試驗的探討重點及目的係在隨後的穩定性分析，即必須再對每一基因型以各環境所有重複之平均性狀觀測值隨環境指標（environmental index）作迴歸，而目前大部份的穩定性分析法僅能適用於每一基因型與環境組合都有資料之場合，尤其是直線迴歸分析模式，因此一旦含有缺區，便無法正確估得環境指標，並進行穩定性分析；至於每一基因型與環境組合都有資料，但重複不等，因爲仍然是以其重複平均值進行迴歸分析，故無影響，在此並不將之視爲區域試驗穩定性分析的缺區現象。過去一般處理區域試驗資料缺區的方式，係將含缺區的品系或環境全部刪除，使其呈完整資料後，再進行穩定性分析，如此在多量缺區的情況下勢必會犧牲不少辛苦調查收集之試驗資料，而且一旦欲刪除的品系爲重要品系或環境數原本即少的情形下，尤其可惜，同時也因爲品系或環境數減少，反而降低了評估結果的準確度。

因此，本研究爲使穩定性迴歸分析法能處理區域試驗的缺區資料，並在不摒除任何品系及環境之情況下，將利用傳統的缺區估計法，填入缺區估計值後以均衡資料的型態再進行穩定性分析。缺區估計法爲解決缺區現象之統計方法中最常見且簡單的技術，主要用來解決變方分析（analysis of variance）含有缺區的場合，本文將就此傳統方法對穩定性分析評價結果的影響，來檢視其在區域試驗資料的適用性。

統計方法

一、穩定性迴歸分析法

將基因型 i（i=1……m）在環境 j（j=1……n，爲地區與年度之組合）下的平均性狀觀測資料 y_{ij} 以下列數學模式表之：

$$y_{ij}=\mu+G_i+E_j+I_{ij}+e_{ij} \qquad [1]$$

式中各介量所代表之意義及其估值爲

$\hat{\mu}$＝總平均值＝$\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij}/mn=\bar{y}..$

$\hat{G}_i$＝基因型 i 的遺傳效應＝$\bar{y}_{i.}-\bar{y}..$，而 $\bar{y}_{i.}=\sum_{j=1}^{n}y_{ij}/n$

$\hat{E}_j$＝環境 j 的效應＝$\bar{y}_{.j}-\bar{y}..$，而 $\bar{y}_{.j}=\sum_{i=1}^{m}y_{ij}/m$

$\hat{I}_{ij}$＝基因型 i 與環境 j 之交感效應＝$y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}..$

e_{ij}＝試驗機差，而符合平均值爲0，變方爲 σ^2 之常態分布，e_{ij} 相互獨立。

並假定：

$\sum_{i=1}^{m}G_i=0$，$\sum_{j=1}^{n}E_j=0$，$\sum_{i=1}^{m}I_{ij}=\sum_{j=1}^{n}I_{ij}=0$，亦即假定 G_i，E_j 及 I_{ij} 均爲固定型效應[30]。

由於交感效應 I_{ij} 多呈環境指標 E_j 的直線函數，即

$$I_{ij}=\beta_i E_j+\theta_{ij} \qquad [2]$$

式中 β_i 爲 I_{ij} 隨 E_j 之迴歸係數，θ_{ij} 爲離迴歸機差。

故直線迴歸模式可寫成

$$y_{ij}=\mu+G_i+E_j+\beta_iE_j+\theta_{ij}+e_{ij} \quad [3]^{(21)}$$

$$=\mu+G_i+(1+\beta_i)E_j+\theta_{ij}+e_{ij} \quad [4]^{(12)}$$

$$=a_i+b_iE_j+\theta'_{ij} \quad [5]^{(10)}$$

〔5〕式中的 θ'_{ij} 爲離迴歸機差，截距估值 $\hat{a}_i=\bar{y}_{i.}$，b_i 爲各基因型平均性狀觀測值隨環境指標作迴歸所得的迴歸係數，一般皆以最小平方法（LS）來估計。農業育種學家多以 b_i 定義爲基因型 i 的穩定性係數（stability index）[10,12]：$b=1$ 爲合乎一般農業要求的穩定基因型，其表現在有利環境下可獲改善，但處於不利時也不致於太差；$0\leqq b<1$ 爲符合生物觀點之穩定性，該類基因型之表現在優勢中也許不十分傑出，但在劣勢裡可保持相當的忍耐力；$b>1$ 則爲不穩定或可塑的基因型，其對環境之變異相當敏感，環境之改善固然可獲致最大的增進，但環境變壞，其表現亦差。至於選取何者爲選拔準則，端視實際需要而決定。倘以 G×E 交感效應隨環境指標作迴歸所得的迴歸係數β而言[21]，則相當於各以 $\beta=-1$與$\beta=0$之間的值作爲準則，來選拔出符合生物或農業觀點的穩定品系。

二、缺區估計法

Allan and Wishart[6]於1930年首先提出缺區估計的概念，其係利用現有的觀測值，根據變方分析的模式，並利用最小平方法求出模式中各參數的估計值，然後用此估計值去預測缺失的觀測值，或是利用此估計值來取代不正常的觀測值。其後 Yates[31]將 Allan and Wishart 的方法推廣至二個以上的缺區值。往後的學者多依循 Yates 的方式，常用的教科書中也多採用此種觀念及方法，是解決變方分析缺區的一種傳統標準方法。其大致上均以逢機完全區集設計（RCBD）爲例，以符號 y 取代未知的缺區值，然後將機差平方和極小化，如此所解得之 y 值即爲所欲求的缺區值。

當缺失的資料僅有一、二時，如能進行缺區估計，並將此估計值填入，隨後以均衡資料用標準方法進行分析，可最爲簡單方便，此並適用於任何穩定性介量之場合。

由於本研究之主要目的是要估計基因型與環境組合的缺區值，故將區域試驗之平均觀測值資料視爲一個含有 m 個處理（品種）及 n 個區集（環境）的 RCBD 試驗。利用 Yates 所提之 RCBD 缺區估計法，填入估計值後，再以均衡資料用標準方法進行穩定性分析。

本研究之各項統計分析工作所需的計算程式，均係利用現今普遍使用的統計套裝軟體 SAS 編寫，在 IBM 相容個人電腦上建檔及執行。所編寫的穩定性迴歸分析之 SAS 程式，參見呂之報告[3]，而計算缺區估計值的 SAS 程式則引用彭[5]之方法。

試驗材料及方法

本研究利用呂及曾[2]之飼料玉米綜合新品系區域試驗資料爲分析材料，以檢視本研究所擬方法的適用性。試驗於1986年春作至1987年秋作進行，計二年四期作。參試品系包括 PC8、PC7、PCC、PUC 及對照品系（CK）之臺南選10號，分別於臺東縣東河鄉、臺東縣鹿野鄉及花蓮縣萬榮鄉共3個地區進行試驗。新品系 PC8 及 PC7 均爲自國際玉米小麥研究中心（International Maize and Wheat Improvement Center, CIMMYT）引進之抗倒伏的熱帶型半硬粒種，PCC 及 PUC 則爲本省朴子分場培育之熱帶型硬粒種。各試驗之田間設計採用 RCBD，四重複，按當地高級產量試驗標準，六行區，行長 5m，行株距75×25cm，小區面積22.5m²，田間作業管理採一般方式執行，調查公頃籽實產量（kg/ha）。

區域試驗所調查的產量結果及環境指標估值列於表 1。本研究之玉米區域試驗係以地區及年度之組合視爲不同環境，故合計 5 個品系 6 個環境進行穩定性分析。爲方便說明，並使一目瞭然，以下相關圖表及其說明將簡化符號，設品系代號各爲 V1～V5，而環境代號各爲 E1～E6。不均衡資料組則是利用人爲的方式，刪去該5×6兩向表資料中的一個或一個以上的單位（格）資料，使造成缺區現象。爲探討缺區發生在穩定或不穩定品種時，其對缺區估計分析所造成的可能影響，因此根據均衡資

料的穩定性分析結果，分別就表現出不同穩定度（穩定性係數 b<1、b=1、b>1）的品種，逢機刪去部份資料，以模擬成各種含有缺區的不均衡資料組。這些不均衡資料組均將利用缺區估計法，先填入估計值後，再進行穩定性分析，其穩定性評估結果並與原來未刪去數值之均衡資料組所得者作一比較印證，如偏誤在可接受範圍之內，即不影響穩定性評估結果，表示所建立之方法確能同時滿足統計理論及實際應用之要求。

表 1. 春作五個玉米品系於六個環境（兩年×三地區）下生長的四重複平均籽實產量資料（公斤/公頃）

Table 1. Mean kernel yield (kg/ha) of four replicates on five maize varieties grown in six environments (2 years×3 sites) under spring crop season

Variety	Environment[y]					
	E1	E2	E3	E4	E5	E6
V1 (PC8)	3289	2163	3692	3921	1589	1805
V2 (PC7)	3487	1660	3148	3897	1937	1404
V3 (PCC)	2651	1996	2375	2920	983	1158
V4 (PUC)	2467	2095	2575	3743	1846	1489
V5 (CK)[z]	2753	1761	3913	4179	1890	1678
Average	2929	1935	3141	3732	1649	1507
Envir. index	447	−547	659	1250	−833	−975

[z] CK：臺南選10號（Tainan sel. No. 10）
[y] E1：1986臺東縣東河鄉（Tongho, Taitung）
E2：1986臺東縣鹿野鄉（Louyai, Taitung）
E3：1986花蓮縣萬榮鄉（Wanlong, Hualien）
E4：1987臺東縣東河鄉（Tongho, Taitung）
E5：1987臺東縣鹿野鄉（Louyai, Taitung）
E6：1987花蓮縣萬榮鄉（Wanlong, Hualien）

結　果

一、原始均衡資料的穩定性分析

本研究飼料玉米綜合新品系區域試驗以原始均衡資料進行穩定性分析結果，如表 2 所示。由表 2 得知，各品系的分析結果均滿足直線迴歸法應用於穩定性分析的兩個重要統計前提，即各品系之迴歸係數（b）均顯著存在（直線關係成立），同時所有品系之迴歸離差均方（S^2d）不顯著且互爲均質，因此所求得之迴歸係數估值（b）足以解釋基因型對環境之適應能力。各參試品系之 b 值間的差異未達0.05之顯著水準，但如需進一步再互作比較，四個新品系的 b 值均相對地較臺南選10號之對照品系（CK）爲低，即表現可更爲穩定，其中 PC8 及 PC7 之 b 值最接近於1，爲符合農業觀念之穩定品種，其會隨著環境改善而表現增進，但也不致因爲環境不利而表現太差，PCC 及 PUC 之 b 值爲0.81，小於1，對環境的變異亦相當穩定，尤其在惡劣環境下可較 PC8 及 PC7 穩定，但因環境改善所能提高的產量卻不如 PC8 或 PC7 高，爲符合生物觀念的穩定性；對照品系（CK）的 b 值大於1，表示其穩定性相對地較其他四個新育成品系爲低。由於育種上欲選拔優良的品系，須兼顧穩定性及生產力，故一般均同時以穩定性介量與產量平均值爲選拔依據，如圖 1 所示，以產量平均值爲 X 軸，迴歸係數爲 Y 軸，觀察各品系點落在圖上區間的分布情形以進行相對間的比較：X 軸以產量總平均值（μ）爲界限，分爲高產區及低產區，Y軸以 b=1 爲界限，以最接近於 1 者爲最符合農業觀念之品系。因此，如以優於對照品系爲選拔之目標，而 PC8 應爲最理想的優良品系，其有較高的產量及穩定性。

表 2. 玉米品系春作籽實產量區域試驗以直線迴歸法進行穩定性分析之結果

Table 2. Stability analysis of kernel yield of maize varieties in a regional trial under spring crop season using linear regression method

Variety	Mean yield (a)	Stability index (b)	Deviation from regression (S^2d)
V1(PC8)	2743 a[z]	1.10**	41876
V2(PC7)	2589 ab	1.11**	96429
V3(PCC)	2014 c	0.81**	107299
V4(PUC)	2369 b	0.81**	92001
V5(CK)	2695 a	1.17**	136976
Average	2482	1.00	94917
LSD(0.05)	237	0.31	
Bartlett's χ^2			1.25[ns]

[z] Means with the same letter of a column are not significantly different at 5% level by LSD Test.
** : significant at 1% level under hypothesis b=0.
[ns] : nonsignificant at 5% level.

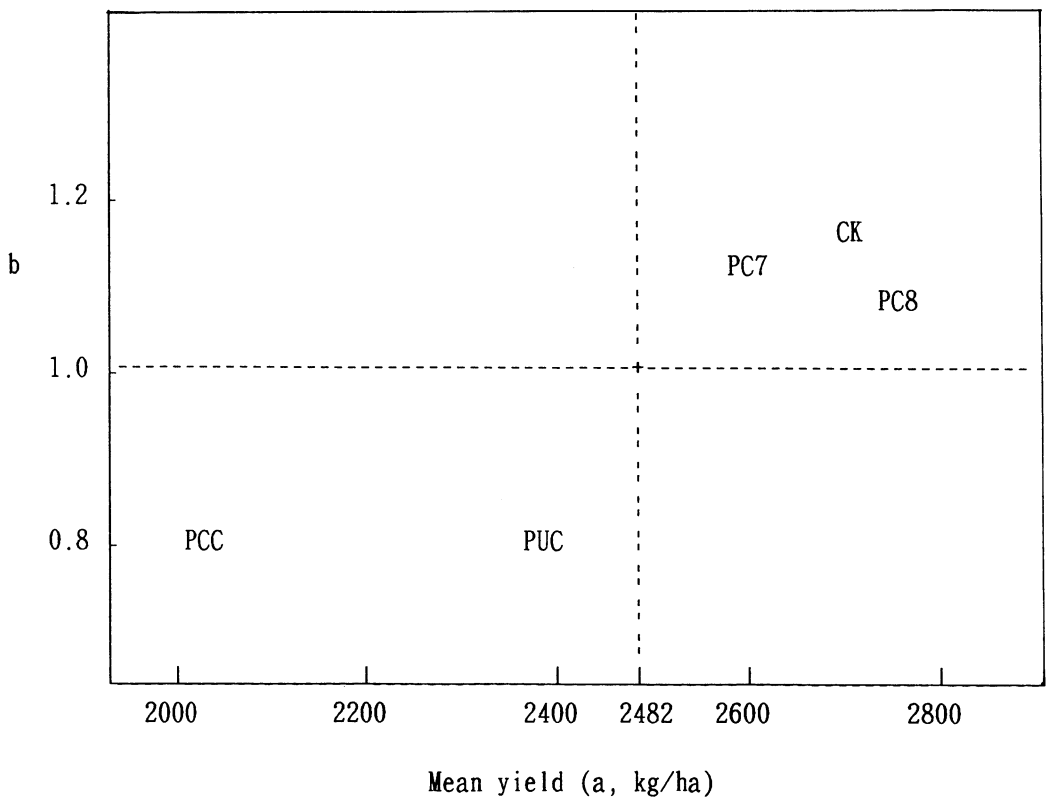


圖 1. 春作玉米品系籽實產量區域試驗品系之穩定性係數(b)與產量平均值(a)間的關係。

Fig. 1. Relationship between stability indices(b) and mean values(a)of kernel yield for the maize varieties in a regional trial under spring crop season.

二、缺區估計後的穩定性分析

由於 Yates 之缺區估計法僅適用於少數一、二個缺區的場合，因此本研究在此著重於僅有一個缺區的情況。缺區發生處，依上述玉米區域試驗均衡資料之穩定性分析結果，假設其可能是發生在符合生物穩定（b<1）、農業穩定（b=1）或不穩定（b>1）品種的某個環境上。選定這三種不同穩定度的玉米參試品系，分別爲V1（PC8）、V3（PCC）及V5（CK）。至於是在哪一個環境缺失資料，則於六個環境中逢機決定。因此所模擬成的單一缺區之三種規劃情形，如圖 2 所示。

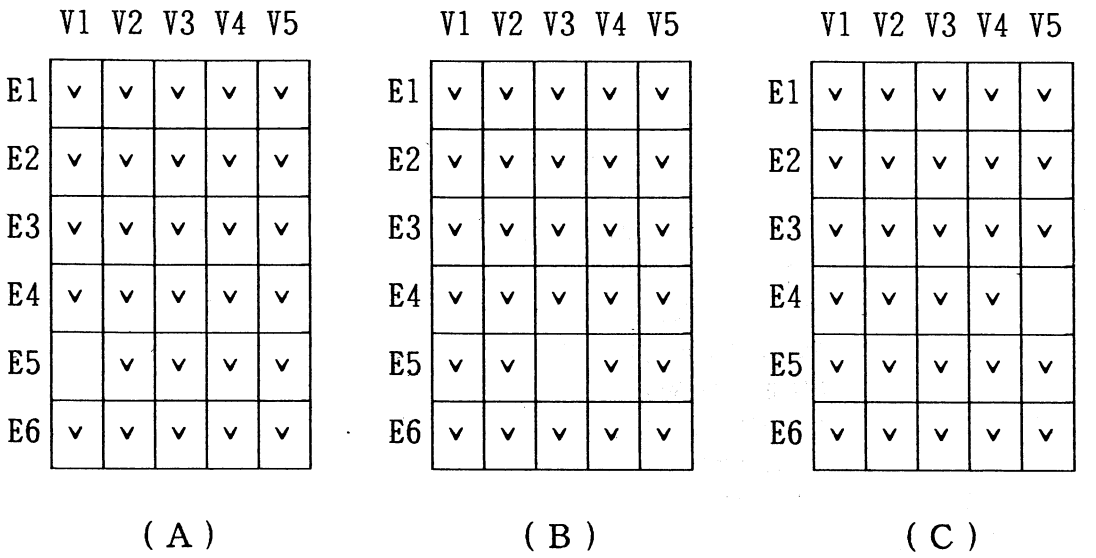


圖 2. 玉米區域試驗資料所模擬成的不同品系含單一缺區之三種規劃情形。(缺區各爲A：y_{15}, B：y_{35}, C：y_{54})

Fig. 2. Three schemes of simulating only one missing plot occurred on different varieties in a maize regional trial. (Missing plot occurred on A：y_{15}, B：y_{35}, C：y_{54})

利用 Yates 之缺區估計法，填入估計值後，所估算的產量平均值列於表 3，其與原始資料組比較，與原觀測平均值在數值上的差異甚小，且各品系之間的差異顯著性仍維持一致。由此可見，將環境因子暫時視爲區集效應，利用傳統之缺區估計法來預測區域試驗的缺區平均產量值，其效果可獲肯定。

表 3. 玉米區域試驗填入缺區估值後以完整資料所得的產量平均值（a, 公斤/公頃）

Table 3. Estimates of mean yield (a, kg/ha) for balanced data by imputating predicted value from missing-plot estimation in a maize regional trial

Variety	V1[z] ($\hat{y}_{15}$=2070)[y] (y_{15}=1589)		V3 ($\hat{y}_{35}$=1279) (y_{35}= 983)		V5 ($\hat{y}_{54}$=3829) (y_{54}=4179)	
	a	Diff.[x]	a	Diff.	a	Diff.
V1(PC8)	2823 a[w]	80	2743 a	0	2743 a	0
V2(PC7)	2589 ab	0	2589 ab	0	2589 ab	0
V3(PCC)	2014 c	0	2063 c	49	2014 c	0
V4(PUC)	2369 b	0	2369 b	0	2369 b	0
V5(CK)	2695 a	0	2695 a	0	2637 a	−58
Average	2498	16	2492	10	2470	−12

[z] genotype with missing-plot estimation.
[y] $\hat{y}_{ij}$=predicted value by missing-plot estimation on the i×j cell; y_{ij}=observed value on the i×j cell.
[x] Diff.=estimated difference between missing-plot estimation and original data.
[w] Means with the same letter of a column are not significantly different at 5% level by LSD Test.

缺區估計後所得的環境指標估值列於表 4 ，V1及V3於環境E5有缺區之場合，其環境變異範圍與原始均衡資料組所得者相同，而V5於環境E4有缺區之場合則數值略爲減少。接著，以填入估計值後呈均衡的資料，再利用標準方法進行穩定性分析，其迴歸離差均方列於表 5，各品系的迴歸離差均方之間皆互爲均質，表示符合穩定性直線迴歸模式的統計前提。所估得之迴歸係數（穩定性係數），如表 6 所示。當符合農業穩定的V1（PC8）品系在環境E5發生單一缺區時（即在此環境下缺失了V1的產量資料），所估得缺區平均預測產量值爲2070公斤／公頃，將此數值填入後，重新以完整的5×6兩向表進行穩定性分析，各品系所獲得的穩定性係數(b)與原始均衡資料組所得者作一比較，結果發現V1品系之 b 值由原先之1.10減少爲1.02，而其他品系之 b 值則都稍有增加，雖然各品系在 b 值上或有增減，但基本上對於各品系的穩定性評價結論大致無異：即PC8及PC7之 b 值最接近於 1 ，爲符合農業觀念之穩定品種；PCC及 PUC 之 b 值小於 1 ，爲符合生物觀念的穩定性；而對照品系（CK）的 b 值大於 1 ，爲最不穩定的品種。倘若缺區是發生在符合生物穩定的V3（PCC）品系時，所估得的缺區平均預測產量值爲1279公斤／公頃，將此數值填入後，重新進行穩定性分析，則只有V3品系的 b值變小，由原先之0.81減少爲0.76，至於其他品系的 b 值均略有增加，但整體而言，其對品種穩定性之評價與原始均衡資料組所得者，仍相當一致。倘若是不穩定品系V5（CK）在某環境下有缺區時，例如V5在環境E4沒有資料，填入缺區估計值（3829公斤／公頃）之後，再重新進行穩定性分析的結果，仍發現缺區品系之 b 值變小而其他品系之 b 值則略增的現象，但此 b 值上之改變卻會影響品種穩定與否的評價，亦即此時原先的不穩定品系V5（CK）之 b 值因爲大幅度變小，反而較其他品系更趨近於 1 ，造成相對地較其他品系表現爲穩定。綜合表 6 的結果得知，缺區估計法應用於穩定性迴歸分析的有效性，顯然與缺區發生在穩定或不穩定品系有關：缺區估計法會使得含缺區品系的穩定性係數明顯低估，此對原先被判斷爲穩定的品系之穩定度評價影響不大，但原先不穩定的品系卻很有可能因此而被誤認爲是穩定。

表 4. 玉米區域試驗資料填入缺區估值後以完整資料所得的環境指標估值（E）

Table 4. Estimates of environmental indices (E) for balanced data by imputating predicted value from missing-plot estimation in a maize regional trial

Envir.	V1[z] ($\hat{y}_{15}$=2070)[y] (y_{15}=1589)		V3 ($\hat{y}_{35}$=1279) (y_{35}= 983)		V5 ($\hat{y}_{54}$=3829) (y_{54}=4179)	
	E	Diff.[x]	E	Diff.	E	Diff.
E1	431	−16	437	−10	459	12
E2	−563	−16	−557	−10	−535	12
E3	643	−16	649	−10	671	12
E4	1234	−16	1240	−10	1192	−58
E5	−753	80	−783	50	−822	11
E6	−991	−16	−985	−10	−964	11
Max.	1234	−16	1240	−10	1192	−58
Min.	−991	−16	−985	−10	−964	11
Range	2225	0	2225	0	2156	−69

[z] genotype with missing-plot estimation.
[y] $\hat{y}_{ij}$=predicted value by missing-plot estimation on the i×j cell; y_{ij}=observed value on the i×j cell.
[x] Diff.=estimated difference between missing-plot estimation and original data.

表 5. 玉米區域試驗填入缺區估值後以完整資料所得的迴歸離差均方估值

Table 5. Estimates of deviation from regression for balanced data by imputating predicted value from missing-plot estimation in a maize regional trial

Variety	V1[z] ($\hat{y}_{15}$=2070)[y] (y_{15}=1589)	V3 ($\hat{y}_{35}$=1279) (y_{35}= 983)	V5 ($\hat{y}_{54}$=3829) (y_{54}=4179)
V1(PC8)	20044	50470	35074
V2(PC7)	83288	87967	93807
V3(PCC)	122405	73279	104556
V4(PUC)	87134	88793	102879
V5(CK)	129308	131788	143113
Average	88436	86459	119857
Bartlett's χ^2	3.11^{ns}	0.87^{ns}	5.77^{ns}

[z] genotype with missing-plot estimation.
[y] $\hat{y}_{ij}$=predicted value by missing-plot estimation on the i×j cell; y_{ij}=observed value on the i×j cell.
[ns] : nonsignificant at 5% level.

表 6. 玉米區域試驗填入缺區估值後以完整資料所得的穩定性係數估值 (b)

Table 6. Estimates of stability indices (b) for balanced data by imputating predicted value from mssing-data estimation in a maize regional trial

Variety	V1[z] ($\hat{y}_{15}$=2070)[y] (y_{15}=1589)		V3 ($\hat{y}_{35}$=1279) (y_{35}= 983)		V5 ($\hat{y}_{54}$=3829) (y_{54}=4179)	
	b	Diff.[x]	b	Diff.	b	Diff.
V1(PC8)	1.0185	−0.0768	1.1043	0.0090	1.1217	0.0264
V2(PC7)	1.1410	0.0269	1.1308	0.0167	1.1389	0.0248
V3(PCC)	0.8173	0.0062	0.7580	−0.0531	0.8301	0.0190
V4(PUC)	0.8251	0.0183	0.8182	0.0112	0.8173	0.0105
V5(CK)	1.1982	0.0255	1.1886	0.0159	1.0920	−0.0807

[z] genotype with missing-plot estimation.
[y] $\hat{y}_{ij}$=predicted value by missing-plot estimation on the i×j cell; y_{ij}=observed value on the i×j cell.
[x] Diff.=estimated difference between missing-plot estimation and original data.

討 論

導致缺區資料的機制已知或未知，是選擇適當的分析方法及結果解釋的主要因素。一旦機制在統計學家控制之下，例如導致缺區的機制純粹是樣本選擇之過程，且是以機率取樣選出單位，則可稱這種機制爲「可忽略的 (ignorable) 」。換言之，當缺區爲逢機缺失 (missing at random, MAR)

時，則謂缺區機制爲可忽略的；反之，若觀測值可被觀察到的機率與觀測值有關，則稱該缺區機制爲不可忽略，此時縮減資料分析是不允許的，因爲會導致偏誤。農業試驗的興趣，大多在於一依變數 Y（如作物產量）與一組因子（如品種、肥料、溫度、或其他環境影響因素等）之間的關係研究。當 Y 在 m-n 單位（假設爲 m 品種於 n 環境之試驗）上發生缺區，可能是由於無法發芽，或有記錄但遺失了，或因爲離值被拋棄不用。對此不均衡資料的標準分析法，一般是假設資料爲 MAR，即觀測值 y_{ij} 之缺區可能發生在任一個自變數上，即 y_{ij} 缺失之機率與 y_{ij} 值無關。在此情況下，可以使資料呈現近於均衡的方式來分析及敍述，以簡化計算[18]。由於區域試驗中發生缺區之機制，在統計上多屬於逢機性型態（符合 MAR），故就此特性以缺區估計來解決區域試驗缺區的問題，理論上應該可行。然而，由本研究結果發現，雖然傳統的缺區估計法在使用上最爲普遍且簡單，但因爲會造成缺區品系之穩定性係數低估，故僅適合於穩定品系含單一缺區的場合；倘缺區發生在不穩定品系，利用該法容易使原不穩定品系被誤認爲是穩定。

當區域試驗資料不符合 MAR 時，倘缺區之發生僅與自變數（環境）有關，傳統的條件推定程序[18,19]，只在環境指標能以實際的外在氣候因子來取代時，才能有效地藉由外在環境變數下來推定缺區，此固然有很多優點[7,11,15,23]，但在很多場合，環境的生產水準不可能單純地只由一些有限的環境因子所決定[8]；倘缺區之發生同時與隨變數及自變數有關，即缺區機制爲不可忽略，則爲最難分析的一種，其涉及的問題相當複雜[18,28]。

另外，針對以填入估計值進行缺區矯正的程序，少數學者如 Patterson[20]及 Silvey[25]則引用常數配合法來解決缺區的問題。常數配合法爲 Yates[32]所提倡，用以分析處理組合因子間無交感效應時的複因子試驗不均衡資料。Freeman[14]也提出類似的方法，並進而藉由無交感的複因子模式求得缺區估計值，填入估計值後再以均衡資料進行標準的穩定性分析。對此，Digby[9]認爲，該法會使穩定性係數產生偏誤而趨近於 1，故不適用於試驗中含有對環境反應較敏感的品種（即不穩定品種）。本研究中以缺區估計法應用於穩定性分析的結果，亦發現不穩定品種在缺區估計後有被誤認爲是穩定的缺失，但是缺區估計後的迴歸係數未必都會趨近於 1，例如第二種缺區估計情形（缺區發生在符合生物穩定的 V3 品系之 E5 環境）下，沒有缺區時的迴歸係數爲0.81，缺區估計後之迴歸係數則爲 0.76。雖然 Yates 的方法也是將〔1〕及〔2〕式中的交感效應 I_{ij} 視爲不存在，照理〔2〕式之迴歸係數 $\beta_i=0$，即 $b_i=1$，直覺上似乎 b_i 應趨近於 1，事實上則不然；這是因爲 V3 品系在 E5 環境下的表現值原爲一個極端的最小值（$y_{35}=983$），在缺區情況下經缺區估計結果其值增大爲1279，反而比 E6 環境下的表現值1158爲好（表 1），然而該缺區估計值之改變對環境指標的估計影響並不大，E5 環境指標由原來的－833（表 1）成爲－783（表 4），此意味著在相同環境變異下 V3 的表現會由原來的甚差變爲有所改善，換言之在較差之 E5 環境下 V3 原有的極端劣勢不見了，因此 V3 品系之 b_i 反而更遠離 1。至於其他兩種缺區情形，由於缺區的值原來都不是極端值，因此不致發生這種 b 偏離 1 的情況，而都趨近於 1。由此可知，Yates 缺區估計法對區域試驗穩定性係數的影響，不能單由無交感的複因子模式去推定，即直覺認定 $b_i=1$ 是缺區估值的特性。綜言之，無論是常數配合法或缺區估計法，以無交感之複因子模式欲探討實際上交感顯著存在的區域試驗資料，是不適當的。除非所有參試品種對環境變異的反應差異都不大，但此與實際情形完全不符。

誌　謝

本研究承蒙行政院國家科學委員會補助經費（NSC83-0409-B055-014），特此誌謝。

引用文獻

1. 呂秀英。1988。穩定性分析。科學農業 36：333-339。
2. 呂秀英、曾清田。1993。非介量穩定性分析法之研究 II.農業觀念之穩定性選拔。中華農學會報 164：1-16。
3. 呂秀英。1994。穩定性迴歸分析之 SAS 程式。中華農業研究 43：283-292。
4. 沈明來。1993。試驗設計學。220-236頁。九州。臺北。
5. 彭雲明。1994。以 Allan 與 Wishart 的觀點看缺區估計值的計算。中華農藝 4：199-208。
6. Allan, F. E. and J. Wishart. 1930. A method of estimating the yield of a missing plot in field experimental work. J. Agric. Sci. 20：399-406.
7. Becker, H. C. and J. Léon. 1988. Stability analysis in plant breeding. Plant Breed. 101:1-23.
8. Beckett, J. L. 1982. Variety × environment interactions in sugar beet variety trials. J. Agric. Sci., Camb. 98：425-435.
9. Digby, P. G. 1979. Modified joint regression analysis for incomplete variety × enviornment data. J. Agric. Sci., Camb. 93：81-86.
10. Eberhart, S. A. and W. A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6：36-40.
11. Eberhart, S. A., L. H. Penny and M. N. Harrison. 1973. Genotype by environment interactions in maize in Eastern Africa. East Afric. Agric. For. J. 39：61-71.
12. Finlay, K. W. and G. N. Wilkinson. 1963. The analysis of adaptation in a plant breeding programme. Austral. J. Agric. Res. 14：742-754.
13. Freeman, G. H. and J. M. Perkins. 1971. Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. Heredity 27：15-23.
14. Freeman, G. H. 1975. Analysis of interactions in incomplete two-way tables. Appl. Statis. 24：46-55.
15. Hadjichristodoulou, A. and A. Kari. 1984. Karpasia, a high yielding new durum wheat variety with improved grain quality characteristics. Techn. Bull. Agric. Res. Inst. 57：1-7.
16. Hardwick, R. C. and J. T. Wood. 1972. Regression methods for studying genotype-environment interactions. Heredity 28：209-222.
17. Lin, C. S., M. R. Binns and L. P. Lefkovitch. 1986. Stability analysis: Where do we stand? Crop Sci. 26：894-900.
18. Little, R. J. A. and D. B. Rubin. 1987. Statistical Analysis with Missing Data. John Wiley & Sons, Inc. New York. 278p.
19. Little, R. J. A. 1992. Regression with missing X's: A review. J. Amer. Statis. Assoc. 87：1227-1237.
20. Patterson, H. D. 1978. Routine least-squares estimation of variety means in incomplete tables. J. Natn. Inst. Agric. Bot. 14：401-412.
21. Perkins, J. M. and J. L. Jinks. 1968. Environmental and genotype-environmental components of variability: III. Multiple lines and crosses. Heredity 23：339-356.
22. Pritts, M. and J. Luby. 1990. Stability indices for horticultural crops. HortScience 25：740-745.
23. Rubin, D. B. 1976. Comparing regression when some predictor values are missing. Technometrics 18: 201-205.
24. Skroppa, T. 1984. A critical evaluation of methods available to estimate the genotype × environment interaction. Studia Forestalia Suecica 166：3-14.

25. Silvey, V. 1978. Methods of analysing NIBA variety trial data over many sites and several seasons. J. Natn. Inst. Agric. Bot. 14 : 385-400.
26. Steel, R. G. D. and J. H. Torrie. 1980. Priciples and Procedures of Statistics. 2nd ed. 藝軒. 台北. 633p.
27. Walters, D. E. and J. G. Rowell. 1984. A missing value technique for the analysis of data with repeated measurements on each experimental unit. J. Agric. Sci., Camb. 102 : 595-599.
28. Wang, R., Sedransk, J. and J. H. Jinn. 1992. Secondary data analysis when there are missing observations. J. Amer. Statis. Assoc. 87 : 952-961.
29. Westcott, B. 1986. Some methods of analysing genotype-environment interaction. Heredity 56 : 243-253.
30. Wright, A. J. 1976. Bias in the estimation of regression coefficients in the analysis of genotype-environmental interaction. Heredity 37 : 299-303.
31. Yates, F. 1933. The analysis of replicated experiments when the field results are incomplete. Emp. J. Exp. Agric. 1 : 129-142.
32. Yates, F. 1934. The analysis of multiple classifications with unequal numbers in the different classes. J. Amer. Statis. Assoc. 29 : 52-66.

Estimation of Missing Plot for Stability Analysis in Crop Regional Trials[1]

Hsiu-Ying Lu[2]

Summary

Crop regional trials were conducted for the purpose of selecting varieties with high yields and stability. Linear regression analysis is the most common and useful method to evaluate the adaptation of a quantitative character to the change of environments. However, a major practical limiation to date, stability analysis have been a requirement of no missing data, that is, data for every genotype and environment combination or treatment have been necessary. The precision of stability analysis can be unduly influenced by removing all the data of the genotypes or environments which contain missing plot, because of the reduction in number of genotypes and environments. In this study, a balanced dataset was obtained by substituting the missing values with the estimated ones, then one used the standard stability analyses for this assumed balanced data. The effect of missing-plot estimation on stability analysis was investigated to test its applicability in regional trial data. Results showed that the method was easy to use and involved a stable genotype with just one or two missing observations. An unstable genotype might be wrongly regarded as stable because that missing-plot estimation yielded a below-estimation of stability index when missing data occurred in an unstable genotype.

Key words : Missing plot, Stability, Regional trial, Unbalanced data, Regression analysis.

1. Contribution No.1856 from Taiwan Agricultural Research Institute. This study was supported by the National Science Council, Grant NO.83-0409-B055-014.
2. Statistician, Department of Agronomy, TARI, Wufeng, Taichung, Taiwan, ROC.