

## تحليل المكون الرئيسي لاختبار ثبات الأصناف

مقالة تعليمية مرجعية

علي فدعم المحمدي  
كلية الزراعة - جامعة الأنبارمدحت مجيد الساهوكي  
كلية الزراعة - جامعة بغداد

## المستخلص

تستخدم طريقة تحليل المكون الرئيسي على نطاق واسع لتقليل بعد البيانات ذات التركيب العنقودي المتعدد بتحويلها الى بيانات ذات بعد رئيسي أبسط، وكذلك تقليل من مجموع مربعات الخطأ للبيانات، وبالتالي سهوله اختبار الأصناف بدرجة كبيرة في مناطق بيئية متعددة. يعتمد GGE biplot للأصناف وترتيبها على كل من متوسط الأداء والثبات، وبدا فان استخدامه يبين ان بعض الأصناف لها تطبع ضعيف عند عدم وجود صنف متفوق الأداء في كل البيئات. يساعد تحليل المكون الرئيسي في الانتخاب على اساس الثبات والأداء للأصناف المزروعة تحت بيئات متعددة. على سبيل المثال، يمكن الانتخاب على اساس ثبات الحاصل في الحنطة او على اساس المادة الجافة في المحاصيل العلفية أو حاصل الشعير في القطن أو المقاومة لمرض معين لأصناف مختلفه لمحصول ما. كما ان هذه الطريقة تؤدي الى اكتشاف الانتخاب غير المباشر للحاصل يعتمد الانتخاب لصفات اخرى، اذ يفسر 81% من نمط GGE مقترحاً ان نمط التداخل الوراثي  $\times$  البيئي (GE) للحاصل قد يكتشف بشكل فعال بالانتخاب غير المباشر لصفات من غير الحاصل مثل وزن الحبة والتزهير المبكر والمقاومة للاضطجاع. تم استعراض تحليل الصفات التي تأثرت بتداخلات GE بطريقة GGE biplot باعتماد التغيرات بين البيئات وتغيرات الأصناف. قدرت متوسطات الأداء كمعدل لأداء الأصناف في السنوات او في المواقع، وتم ايضاح كيفية الاستنتاج من التحليل لتحديد الصنف الأمثل في الأداء والثبات عبر البيئات. ادرج في مقاله البرنامج الاحصائي الخاص بالتحليل (GGE biplot) ورسمت الأشكال التي تم الحصول عليها من التحليل ومناقشتها والاستنتاج منها.

The Iraqi Journal of Agricultural Science 39 (1): 102-115 (2008)

Sahookie &amp; Mehemdi

## PRINCIPAL COMPONENT ANALYSIS TO TEST STABILITY OF CULTIVARS ACROSS ENVIRONMENTS:- A TUTORIAL REVIEWED ARTICLE

M.M. ELSAHOOKIE

A. F. ALMEHEMDI

University of Baghdad /College of Agriculture University of ALanbar/ College of Agriculture

## Abstract

Principal component analysis is a widely-used method for dimensionality reduction of multivariate clustering data transformed into low dimensional data. It minimizes the sum of squared errors. The low dimensional structure of data can be effectively captured by PCA, transformed into new variates. Cultivar stability simply tested at multi-environment, GGE-Biplot depends on performance and stability means. Some cultivars could show low adaptation by PCA because there is no cultivar of a superior performance at all environments. Therefore, the PCA assists to select (based on stability and performance) a cultivar at multi-environments. PCA analysis could be used for selection depends on yield stability, forage crop biomass, cotton lint yield or certain disease resistance of different cultivars of a certain crop. Furthermore this method leads to discover indirect selection for other traits. The covariate effect pattern explained 81% of the GGE pattern in some trials, suggesting that the GE pattern for yield could be effectively explored by indirect selection for these traits and others as seed yield, early flowering and lodging resistance. The traits analyses that were affected by GE interaction using GGE-Biplot method were reviewed. The differences based on environment and cultivars variations. Performance means were estimated as cultivars performance at years or locations. It was explained how to conclude from the analysis to determine the optimal cultivar in performance and stability across several environments, to show specific or wide adaptation of genotypes. Along with this article, GGE-Biplot program was included for use.

## المقدمة

تحليل PCA فسر 46.2% و 15.8% من مجموع مربعات GGE بالتتابع ، فقد اشارت نقاط PC<sub>1</sub> اكثر من صفر ان بعض من اصناف حنطة الخبز لها المقدرة على التأقلم ، في حين ان نقاط PC<sub>1</sub> اقل من صفر ميزت الاصناف غير المتأقلمة او منخفضة الحاصل . اما نقاط PC<sub>2</sub> التي اقتربت من صفر فهي توضح الاصناف الثابتة اما النقاط العالية جدا من PC<sub>2</sub> فهي للاصناف غير الثابتة . او بهج Kaya و Taner (19) ان محاور اول مكونين رئيسيين كانت معنوية وساهمت بنسبة 78.64% من التداخل الوراثي × البيئي ، وان Biplot بين ان الاصناف ذات PCA<sub>1</sub> الاعلى و PCA<sub>2</sub> الاوطا اعطت اعلى حاصل (اصناف ثابتة) ، والاصناف ذات PCA<sub>1</sub> الاوطا و PC<sub>2</sub> الاعلى كان لها حاصل منخفض (اصناف غير ثابتة) . يفسر PCA معدل 88.8% من التباين الموجود في المجتمع ، فقد بين PC<sub>1</sub> 47.9% من التباين للصفات المظهرية المرتبطة بالحاصل و PC<sub>2</sub> 22.9% من صفات التطور و PC<sub>3</sub> 18.0% من الصفات المرتبطة بالخصوبة (29). يقلل PCA من مجموع مربعات الخطأ كما انه حساس لوجود الشواذ (Outliers) وهو معروف بانه تقنية احصائية معتمدة لحل مشاكل مهمة مثل سمات الاستنباط وتقدير الاشارة وغيرها (7 و 8) . كما يستخدم لدراسة مصادر التلوث الرئيسية في البيئة (3) وفي حماية شبكات الانترنت واجهزة الاتصالات (23) وفي دراسة علم البكتريولوجي (24) والدراسات الوراثية والتعبير الجيني (22، 39) ، وفي الدراسات العصبية لمعامل الذكاء Intelligence Quotient (16). كذلك فانه يؤدي الى تقليل واختزال الابعاد (Dimensionality) لبيانات متعددة التباين بتحويل البيانات ذات البعد العالي الى بيانات ذات بعد واطيء (14، 17، 30). عند زراعة اصناف في عدة بيئات فقد لا يتغير حاصل او اية صفة للتركيب المزروعة ، وهذا يعني ان بعضها يمتلك نوعا من الثبات ، يطلق عليه بالثبات Buffering أو Homeostasis أو Stability (2، 12). ان الثبات هو نتيجة لسلوك التركيب الوراثي تحت تأثير عوامل البيئات ، لذا استخدمت عدة طرائق لتقديره من بينها معادلة معامل الانحدار (10، 12، 15، 21، 26). ان الأساس الرئيسي لتقليل البعد هو ان PCA يلتقط الابعاد ذات التباينات الاكبر ، لذا كتبت هذه المقالة بهدف تقديم تعريف وتفسير بسيطين مع امثلة توضح استخدام هذه الطريقة لتطبيقها في تقدير وحساب ثبات الاصناف المزروعة تحت عدة بيئات ، وكذلك لبيانات اخرى مماثلة سواء كانت حول الكائنات الحية او في الصناعة او البيئة او في دراسات علم الاجتماع والاجرام والقانون وعلم الارض وغيرها.

زاد اهتمام باحثي تربية النبات في تداخلات الاصناف في البيئات وحتى الاحصائيين بشكل كبير بدراسة التأثير الرئيسي للسنف (G) و التداخل الوراثي × البيئي (GEI) Genotype x Environment Interaction ، خلال العقد الماضي لاجل تحليل بيانات تجارب الاصناف المزروعة في بيئات متعددة باستخدام (Biplot analysis) (25، 36). اوضحت عدة برامج حاسوبية كيفية تسهيل تطبيق هذه الطريقة باستخدام (Biplot analysis) (4، 6، 33، 35) . افترض Yan واخرون (34) و Yan (37) من بيانات اداء الحنطة الشتوية (GGE biplot ، Genotype, Environment x Environment Interaction) يتألف من مكونين رئيسيين هما PC<sub>1</sub> و PC<sub>2</sub> (Principal Component Analysis) لبيانات حاصل الحنطة في بيئات مختلفة يظهر GGE biplot و Covariate effect biplot ان المستويات البيئية تقع في مكوني البيئة الكبرى (Mega Environment) ، اذ ان نمط تأثير التباين المشترك يوضح 81% من نمط GGE ، مفترضا ان GE للحصول يمكن ان يكتشف بالانتخاب غير المباشر لهذه الصفة ، لاسيما ان حاصل محصول حبوبى مثلا يمكن برهنته بالانتخاب على اساس وزن الحبة الكبير و التزهر المبكر والمقاومة للاضطجاع عندما يكون ناميا في بيئات مختلفة (32). تكون بيانات الحاصل عادة المأخوذة من تجارب مناطقية او (Multiple Environmental Trials) MET كبيرة جدا فيكون من الصعب ادراك الخط العام لها بدون نوع من العرض بالرسم ، لذا فان طريقة biplot تعد حلا كبيرا لهذه المشكلة ، وقد اشير لها بتعبير GGE biplot وهي اداة مثالية لتحليل بياناتها (34، 35، 38). تقيس معاملات PCA 7.17% من التباين الكلي بين السلالات لمحصول ما (27) وبقيمة 20% من التباين بين اصناف محصول اخر مزروعة في مواقع مختلفة (28) . حللت بيانات اربع مجاميع اصناف لمحصول ما كان فيها تركيب منخفض الحاصل بشكل عام وغير ثابت . و تركيبان اخران منخفضا الحاصل وثابتان باعتدال و تركيب رابع كان عالي الحاصل وغير ثابت ، امكن باستخدام طريقة PCA الانتخاب للتأقلم الواسع او الخاص للمحصول قيد الاختبار (9، 11) . كما بين PCA الى ان ارتفاع النبات وطول النورة قد ساهما في الانفراج الوراثي بين سلالات الخرة البيضاء (13) والثبات العالي لصفات الورقة والساق اكثر من ثبات صفات النورة (1). فيما أوضح تحليل المكون الرئيسي ان للبروتين والزيت اختلافاً واضحاً بين عام و اخر خلال سني الاختبار ، اذ ان بعض الاصناف كانت ثابتة الصفة في البيئة فيما كانت الاخرى قد استجابت بشكل مختلف باختلاف البيئات (28، 29). وجد Kaya واخرون (20) ان

## المواد و الطرائق

طريقة حساب تحليل المكون الرئيسي

الخطوة الاولى : الحصول على البيانات

لتكن D او X مصفوفة لبيانات اصلية ذات بعدين (X)

و (Y) وهي ابسط حالة يمكن رسمها لايضاح تحليل PCA

(31) كما في بيانات جدول 1.

جدول 1. بيانات صفة معينة لعشرة اصناف مزرعة في بينتين X و Y .

	X	Y
	2.5	2.4
	0.5	0.7
	2.2	2.9
	1.9	2.2
	3.1	3.0
	2.3	2.7
	2	1.6
	1	1.1
	1.5	1.6
	1.1	0.9
Total	18.1	19.1
Mean	1.81	1.91

وبالتالي تنتج مجموعة بيانات بمتوسط قيمته صفرا ، لانه من متطلبات تحليل PCA ، وبذا تكون البيانات كما في جدول 2.

الخطوة الثانية: طرح المتوسط العام لأجل أن يعمل PCA بشكل صحيح ، يجب طرح المتوسط العام للبعدين (Y و X) وهما 1.81 و 1.91 من كل أبعاد البيانات. ان المتوسط المطروح هو المعدل تجاه كل بعد، لذا فان كل قيم X (لها  $\bar{X}$  مطروح) وكذلك لقيم Y (لها  $\bar{Y}$  مطروح) ،

جدول 2. البيانات المعدلة بعد طرح المعدل العام منها وقيم التباين.

X	Y	XX(X <sup>2</sup> )	YY(Y <sup>2</sup> )	XY
0.69	0.49	0.4761	0.2401	0.3381
-1.31	-1.21	1.7161	1.4641	1.5851
0.39	0.99	0.1521	0.9801	0.3861
0.09	0.29	0.0081	0.0841	0.0261
1.29	1.09	1.6641	1.1881	1.4061
0.49	0.79	0.2401	0.6241	0.3871
0.19	-0.31	0.0361	0.0961	-0.0589
-0.81	-0.81	0.6561	0.6561	0.6561
-0.31	-0.31	0.0961	0.0961	0.0961
-0.71	-1.01	0.5041	1.0201	0.7171
$\Sigma$		5.549	6.449	5.539
$\bar{X}$		0.5549	0.6449	0.5539
V		.616555556	0.716555556	0.615444444

لايضاح ذلك ، نأخذ مثلا قيمة  $X_1=2.5$  و  $Y_1=2.4$  ونطرح منهما متوسطيهما العام (جدول 1) .

Data Adjusted of  $X_1=2.5-1.81=0.69$

Data Adjusted of  $Y_1=2.4-1.91=0.49$

(Variance) ، أما في حالة ثلاثة ابعاد (X و Y و Z) ، فيمكن قياس التباين المشترك بين X و Y ، و X و Z ، و Y و Z . يؤدي قياس التباين المشترك بين X و X او Y و Y او Z و Z الى الحصول على تباين X و Y و Z ، وكما يلي مع العودة الى جدول 2.

a- Formula of variance

وبنفس الطريقة تستخرج بقية البيانات.

الخطوة الثالثة: حساب مصفوفة التباين المشترك

(Covariance Matrix)

لتطبيق PCA نحتاج الى معرفة مصفوفة التباين المشترك من البيانات الاصلية. يقاس التباين المشترك من بعدين او متغيرين (X و Y) . اذا تم احتساب التباين المشترك بين متغير واحد وذاته ، فسوف يتم الحصول على التباين

$$\text{Var}(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{(n-1)} \text{ or } = \frac{\sum_{i=1}^n X_i X_i \text{ of data adjust. } (X^2)}{(n-1)}$$

$$= \frac{\Sigma(2.5-1.81)(2.5-1.81)+\dots+(1.1-1.81)(1.1-1.81)}{9}$$

$$= \frac{\Sigma(0.69*0.69)+\dots+(-0.71*-0.71)}{9} = \frac{5.549}{9} = 0.616555556$$

$$\text{Var}(Y) = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})}{(n-1)} \text{ or } = \frac{\sum_{i=1}^n Y_i Y_i \text{ of data adjust. } (Y^2)}{(n-1)}$$

$$= \frac{\Sigma(2.4-1.91)(2.4-1.91)+\dots+(0.9-1.91)(0.9-1.91)}{9}$$

$$= \frac{\Sigma(0.49*0.49)+\dots+(-1.01*-1.01)}{9} = \frac{6.449}{9} = 0.716555556$$

b- Formula of covariance:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \text{ or } = \frac{\sum_{i=1}^n X_i Y_i \text{ of data adjust.}}{n-1}$$

لذا فان تعريف مصفوفة التباين المشترك لمجموعة بيانات بعد  $n$  تكون:

$$C^{n \times n} = (C_{i,j}, C_{j,i} = \text{Cov}(\text{Dim}_i, \text{Dim}_j))$$

اذ ان  $C^{n \times n}$  مصفوفة ذات  $n$  من الصفوف و  $n$  من الاعمدة و  $\text{Dim}_i = \text{Dim}_j$  تمثل بعد  $X^{\text{th}}$  و  $Y^{\text{th}}$  لاستخراج قيم  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$  و  $\text{Cov}(Y, Y)$  و  $\text{Cov}(X, X)$  تستخدم البيانات في جدول 2 بتطبيق المعادلة b نحصل على القيم، علما ان قيم  $\text{Cov}(Y, Y)$  و  $\text{Cov}(X, X)$  تم استخراجها طبقا للمعادلتين في a. اما قيم  $\text{Cov}(X, Y)$  فتستخرج بتطبيق المعادلة b مع العودة الى بيانات جدول 2.

$$\frac{\Sigma(2.5-1.81)(2.4-1.91)+\dots+(1.1-1.81)(0.9-1.91)}{10-1}$$

$$= \frac{\Sigma(0.69*0.49)+\dots+(-0.71*-1.01)}{9} = \frac{5.539}{9} = 0.615444444$$

وبذا طبقا للمثال تكون مصفوفة التباين المشترك وقيمتها الاتي:

$$C = \begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{pmatrix}$$

$$\text{Cov} = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

الخطوة الرابعة: حساب قيم Eigenvalues و Eigenvectors لمصفوفة التباين المشترك

$A =$  مصفوفة مربعة،  $\lambda =$  هي eigenvalue  
 لإيجاد قيم eigenvalue تطبق المعادلة  $(A - \lambda I) = 0$   
 def التي تتبع المعادلة (4)، والتي تكافئ المعادلة  $(A - \lambda I)V = 0$  لان قيم  $(\lambda)$  هي جذور لها.  
 على هذا الأساس يمكن إيجاد قيم eigenvalues لمصفوفة التباين المشترك كالآتي:

من السهولة استخراج هذه القيم لان مصفوفة التباين مصفوفة مربعة. تسمى الأرقام الموجودة في المصفوفة القطرية eigenvalues لمصفوفة التباين المشترك  $C$ ، فيما تسمى أعمدة المصفوفة  $P$  eigenvectors و قيم eigenvalue  $(\lambda)$  و eigenvector  $(v)$  لها صفة خاصة التي تخضع للمعادلة  $Av = \lambda V$

$$= \det \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

( Identity matrix ) وهي مصفوفة تشخيص

لاحظ الصف القطري لقيمة واحد محاط بالصفرين في مصفوفة مربعة، فعندما يضرب المتجه بهذه المصفوفة لنفس البعد فالناتج يكون هو المتجه نفسه  $(I v = v)$

بما ان  $(\lambda)$  عبارة عن المصفوفة القطرية لمصفوفة التباين المشترك فتكون كالآتي:

$$\det = \begin{pmatrix} .616555556 - \lambda & .615444444 \\ .615444444 & .716555556 - \lambda \end{pmatrix}$$

$$\det A = |A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$= (.616555556 - \lambda) (.716555556 - \lambda) - 0.378771863$$

باستخدام قانون الفرق بين المربعين وبتطبيق قوانين الجبر الخطي لإيجاد قيم  $(\lambda)$   
الحد الأول = الطرف الأول \* الأول

$$= \lambda * \lambda = \lambda^2$$

الحد الثاني = الطرف الثاني \* الثاني

$$= 0.616555556 * 0.716555556 = 0.441796309$$

الحد الأوسط = الطرف الأول \* الثاني ± الطرف الثاني \* الأول

$$= (.616555556 - \lambda) (.716555556 - \lambda) = -1.33311112\lambda$$

$$= \lambda^2 - 1.33311112\lambda + .441796309 - .378771863$$

بتبسيط المعادلة تكون:

$$= \lambda^2 - 1.33311112\lambda + 0.063024446 = 0$$

$$= (\lambda - 1.28402771) (\lambda - 0.0490833989) = 0$$

$$= \lambda - 1.28402771 \quad \rightarrow \lambda = 1.28402771 \text{ or}$$

$$\lambda - 0.0490833989 \quad \rightarrow \lambda = 0.0490833989$$

إذن تكون مصفوفة قيم eigenvalues الآتي:

$$\text{eigenvalue} = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

إيجاد eigenvectors:

تستخدم المعادلة  $(A - \lambda I) V = 0$  مرة ثانية لإيجاد V بما انه أصبحت لدينا قيم  $\lambda$  معلومة، إذن نطبق المعادلة لإيجاد قيم v:

$$= \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix} - 0.0490833989 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} .616555556 - 0.0490833989 & .615444444 \\ .615444444 & .716555556 - 0.0490833989 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} .567472158 & .615444444 \\ .615444444 & .667472158 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

1- يضرب كل عنصر أو يقسم بعدد. 2- يتم تغيير موقع صفين. 3- مضاعفة صف قد يضاف أو يطرح من آخر، وكما يلي:

تختزل المصفوفة طبقاً للصيغة (rref) reduced row echelon form، إذ إنها وسيلة لحل المعادلات، وتحصل فيها ثلاث عمليات:

$$= \begin{pmatrix} .567472158 & .615444444 \\ .615444444 & .667472158 \end{pmatrix} - 1.08453681 * \text{row}(1)$$

لمعرفة كيفية الحصول على الرقم 1.08453681 نقسم:

$$\frac{\text{Var}(X,Y)}{\text{Var}(X,X) - \lambda} = \frac{.615444444}{.567472158} = 1.08453681$$

ثم نعود إلى المصفوفة التالية:

$$.615444444 * -1.08453681 = -.667472154$$

$$.567472158 * -1.08453681 = -.615444444$$

$$= \begin{pmatrix} .567472158 & .615444444 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = 0$$

$$.567472158 X + .615444444 Y = 0$$

نفترض  $X = -.735178656$

$$.567472158(-.735178656) + .615444444 Y = 0$$

$$-.417193418 + .615444444 Y = 0$$

$$.417193418$$

$$Y = \frac{.417193418}{.615444444} = .677873401$$

$$.615444444$$

أما عندما تكون  $\lambda = 1.28402771$  تكون المصفوفة:

$$= \begin{pmatrix} -.667472154 & .615444444 \\ .615444444 & -.567472154 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

نضرب السطر الأول بالرقم 1.08453681 ونجمعه مع السطر الثاني:

$$= \begin{pmatrix} -.667472154 & .615444444 \\ .615444444 & -.567472154 \end{pmatrix} + 1.08453681 * \text{row}(1)$$

$$= \begin{pmatrix} -.667472154 & .615444444 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = 0$$

$$-.667472154X + .615444444Y = 0$$

$$-.667472154(-.67873399) + .615444444Y = 0$$

$$Y = -.735178656$$

إذن تكون مصفوفة قيم eigenvector الآتي:

$$\text{eigenvectors} = \begin{pmatrix} -.735178656 & -.67873399 \\ .67873399 & -.735178656 \end{pmatrix}$$

الخطوة الخامسة: اختيار المكونات وتكوين متجه السمة (Feature Vector).

هنا تكمن فكرة ضغط البيانات واختزال ابعادها، فلو نظرنا الى قيم eigenvalue و eigenvectors، سنلاحظ ان هذه القيم مختلفة تماماً في الحقيقة، فيتبين ان eigenvectors بقيم eigenvalues الأعلى هو المكون الرئيسي لمجموعة البيانات، ففي المثال السابق، كانت قيم eigenvectors ذات قيم eigenvalues الاعلى وهي التي تحدد منتصف البيانات، وهي العلاقة الأكثر معنوية بين ابعاد البيانات. عموماً عندما توجد قيم eigenvectors من مصفوفة التباين المشترك، فان الخطوة

التائية هي ترتيب قيم eigenvectors اعتماداً على قيم eigenvalues من الاعلى الى الادنى، اذ ترتب المكونات طبقاً للمعنوية. ان متجه السمة هو مجرد اسم تصوري لمصفوفة eigenvectors ويتم بناؤه بأخذ eigenvectors المرغوب حفظها من قائمة eigenvectors وتكوين مصفوفة منها في اعمدة: Feature Vector (eig<sub>1</sub> eig<sub>2</sub> eig<sub>3</sub> ..... eig<sub>n</sub>)  
اذ ان Feature Vector = متجه السمة، و eig<sub>1</sub> و eig<sub>2</sub> و eig<sub>3</sub> ..... eig<sub>n</sub> = قيم eigenvectors بحسب عددها مرتبة في اعمدة، وعليه تصبح مصفوفة متجه السمة الآتي:

المستبط من قيمة eigenvalue الادنى، والاقبل معنوية، فيصبح لدينا عمود مفرد كالآتي:

$$\begin{pmatrix} -.677873399 \\ -.735178656 \end{pmatrix}$$

حيث ان: Final Data هي البيانات النهائية او الجديدة و Row Feature Vectors وهو مصفوفة eigenvectors في اعمدة منقولة كي تكون في صفوف، لذا تكون المتجهات الأكثر معنوية في القيمة و Row Data Adjust. تمثل Mean adjusted Data - المنقولة، بتعبير اخر فقرات البيانات في كل عمود وبكل صف تمسك بعدد منفصل.

من المثال السابق، أصبح لدينا متجهان، لذا يمكن ان نختار كلا المتجهين لتكوين متجه السمة او ان نهمل المتجه الاقل، اي

الخطوة السادسة: استخراج مجموعة البيانات الجديدة

هي الخطوة النهائية لتحليل PCA وهي كذلك الاسهل. حالما يتم اختيار قيم (eigenvectors) التي يرغب بحفظها في البيانات وتكوين متجه السمة، يؤخذ منقول المتجه ويضرب بمجموعة البيانات الاصلية المنقولة كما يلي:

$$\text{Final Data} = \text{Row Feature Vector} * \text{Row Data Adjust.}$$

جدول 3. المكونات الرئيسية للمثال.

	PC <sub>1</sub>	PC <sub>2</sub>
X	-.677873399	-.735178656
Y	-.735178656	.677873399
S <sup>2</sup> =λ <sub>i</sub>	1.28402771	.0490933989
S <sup>2</sup> PC <sub>i</sub>	.963174089	.03682591

من هذا نرى ان المكون الرئيسي الأول يفسر وحده 96.3174089% من التباين الكلي كالآتي:

$$S^2 PC_1 = \frac{\lambda PC_1}{\sum \lambda_i} = \frac{1.28402771}{1.333121109} = .963174089$$

نضرب مجموعة البيانات الاصلية D او X في قيم eigenvectors يدور متجه البيانات. فمثلاً قيمة X = .69.

وان جذره المميز اكبر من واحد (λ<sub>1</sub>=1.28402771) وعليه فيمكن الاكتفاء بالمكون الرئيسي الأول PC<sub>1</sub> ولانه يمكن إيجاد قيم الهجن بحسب المكون الرئيسي الأول والثاني PC<sub>1</sub> و PC<sub>2</sub> كما في جدول 4.

$$.69 \begin{pmatrix} -.677873399 \\ -.735178656 \end{pmatrix} = \begin{pmatrix} -.759218206 \\ -.823400094 \end{pmatrix}$$

D.V<sub>1</sub> = PC<sub>1</sub>  
تكرر المعادلة بمتجهات اخرى من eigenvector طبقاً للمعادلة.

D.V<sub>i</sub> = PC<sub>i</sub> i = 2 to n  
البيانات الاصلية 200 حزمة (n = 200)، فهذا يعني ان 10 من PCs يمكن ان تحفظ، ذات قيم eigenvalues الاعلى ومجموعة البيانات هذه هي 200/10 = 20/1 من الحجم الاصيل.

$$X_n = X \cdot V_1 \text{ or } D_n = DE \text{ or } D.V_1 = PC_1$$

اذ ان PC<sub>1</sub> هو المكون الرئيسي الأول، فهو متجه ذو m من المكونات، وللحصول على مكونات رئيسية اخرى، يمكن ان

اذ ان z = عدد eigenvector و PC = المكون الرئيسي من 2, n, ... بالتتابع.

يمكن ان تجمع (تدمج) من بيانات PC لتكوين مكعب بيانات جديدة. اذا طبقت، فيمكن حفظ PCs القليلة الاولى حتى يمكن اختزال حجم البيانات. على سبيل المثال، اذ امتلكت مجموعة

$$PC_1 = V_{11}X_{ij} + V_{21}Y_{ij}$$

$$PC_2 = V_{12}X_{ij} + V_{22}Y_{ij}$$

جدول 4. قيم PC<sub>1</sub> و PC<sub>2</sub> للهجن العشرة المزروعة في بيئتين.

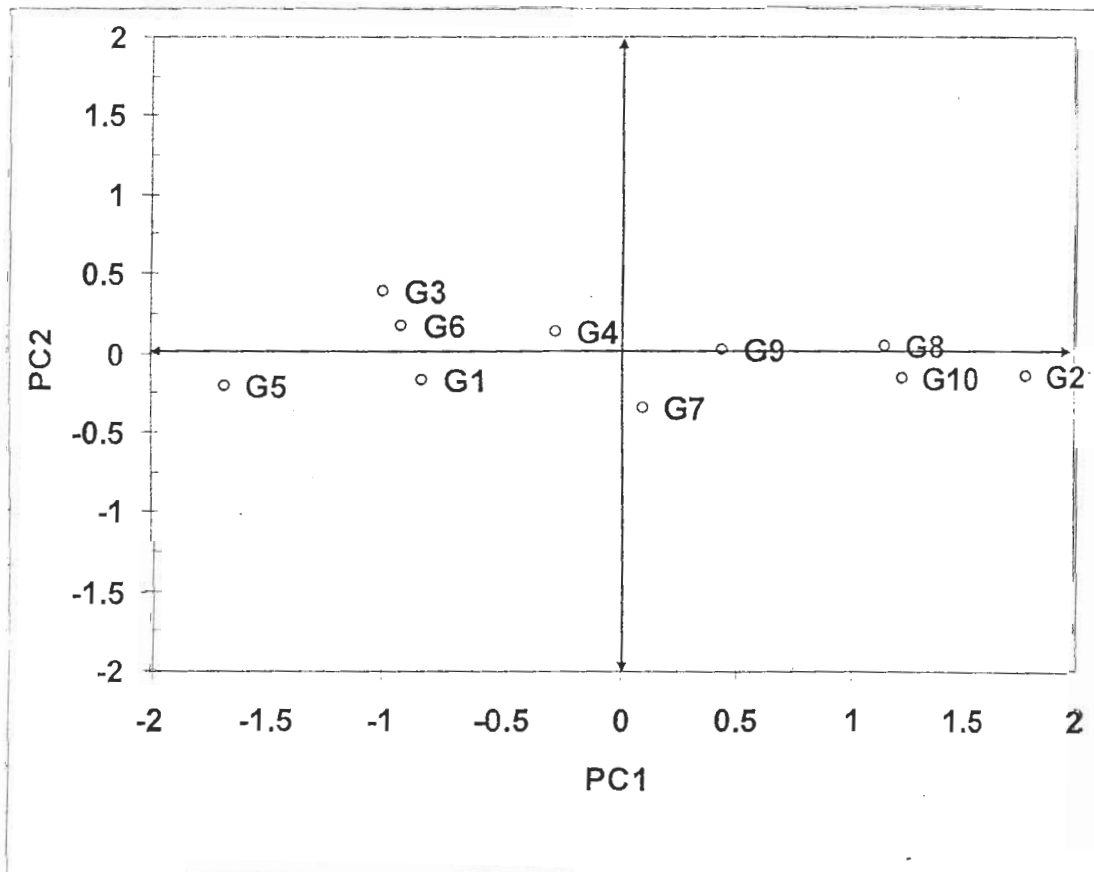
X	Y	PC <sub>1</sub>	PC <sub>2</sub>
.69	.49	-.827970186	-.175115306
-1.31	-1.21	1.777580326	.142857227
.39	.99	-.992197494	.38437499
.09	.29	-.274210415	.130417206
1.29	1.09	-1.675801419	-.209498462
.49	.79	-.912949103	.175282444
.19	-.31	.099109438	-.349824697
-.81	-.81	1.144572164	.046417258
-.31	-.31	.438046136	.017497853
-.71	-1.01	1.223820556	-.162675288

بتطبيق المعادلتين يمكن استخراج قيم الهجن كالآتي:

$$PC_1 \text{ for } G_1 = -.677873399 * .69 - .735178656 * .49 = -.827970186$$

$$PC_2 \text{ for } G_1 = -.735178656 * .69 + .677873399 * .49 = -.175115306$$

وهكذا تضرب بقية الأرقام لغاية الرقم 71 و -1.01 (شكل 1).



شكل 1. لوحة رسم البيانات الجديدة لعشرة هجن مزروعة في بيئتين.



الهجن  $G_3$  و  $G_4$  و  $G_6$  الواقعة في المربع A وكذلك الهجينان  $G_1$  و  $G_5$  الواقعان في المربع D غير ثابتة لأنها ذات قيم  $PC_1$  و  $PC_2$  أقل من صفر، إذ ان الهجن  $G_3$  و  $G_4$  و  $G_6$  ذات قيم  $PC_1$  أقل من صفر و  $G_1$  و  $G_5$  ذات قيم  $PC_2$  أقل من صفر.

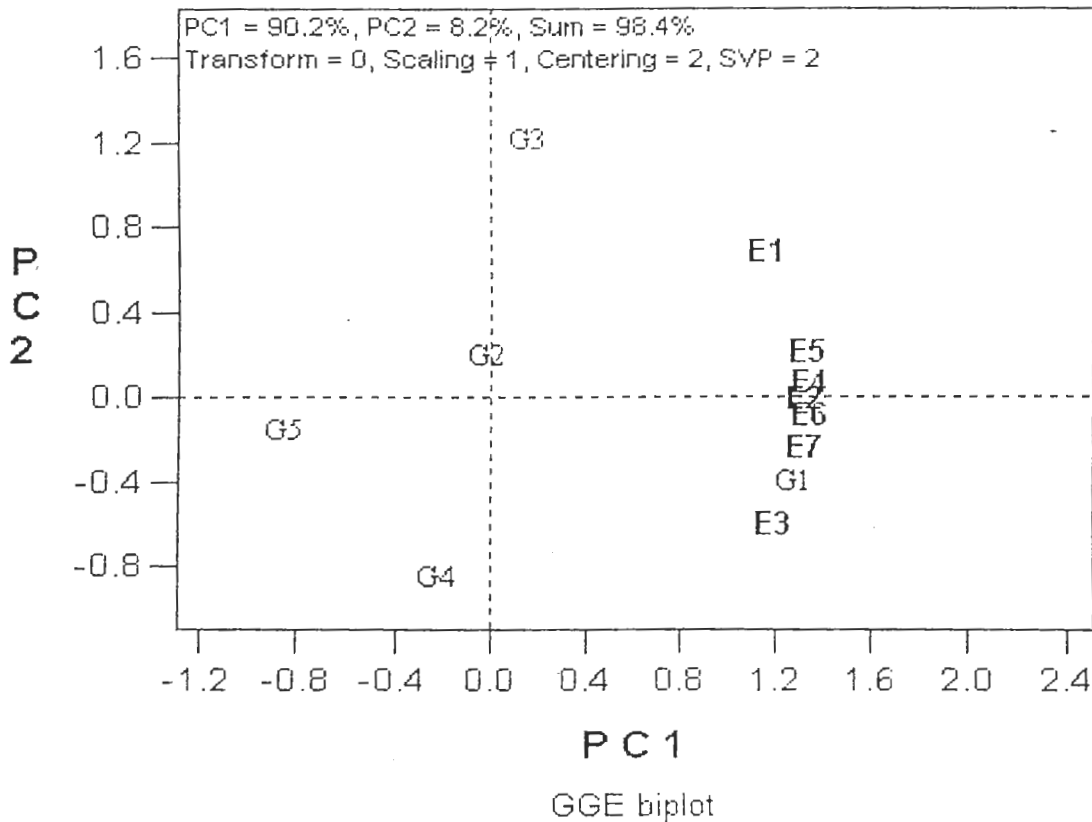
تشير لوحة البيانات ان الهجن التي تمتلك قيم  $PC_1$  أكبر من صفر ذات الحاصل الأعلى، أي التي تقع في المربعين B و C. أما الهجن ذات قيم  $PC_1$  أصغر من صفر فقد أعطت أقل حاصل، وبذا تكون الهجن  $G_9$  و  $G_8$  و  $G_2$  و  $G_{10}$  و  $G_7$  هي الهجن الأكثر ثباتاً والتي أعطت أعلى حاصل بالتتابع. تعد

## 2 - طريقة حساب معامل الانحدار

جدول 5. حاصل (طن/هـ) خمسة هجن من الذرة الصفراء زرعت في سبع بيئات (بيانات افتراضية).

$G_i$	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$\Sigma G_i$	$\bar{G}$
$G_1$	7.95	8.5	8.05	8.7	8.9	9.05	9.8	60.95	8.71
$G_2$	6	4.9	5.15	6.15	6.35	6.3	6.35	41.2	5.89
$G_3$	7.55	5.8	3.9	6.5	7	6.36	6.95	44.05	6.29
$G_4$	3.8	4.7	5.25	5.55	5.4	5.95	7	37.65	5.38
$G_5$	3.65	4	3.7	3.85	4.25	3.9	5.05	28.4	4.06
$\Sigma E_j$	28.95	27.6	26.05	30.75	31.9	31.55	35.15	212.25	
$\bar{E}$	5.79	5.52	5.21	6.15	6.38	6.31	7.03		6.06

تم رسم هذه البيانات والاستنتاج منها كما في شكل 2.



شكل 2. رسم GGE biplot للهجن الخمسة المزروعة في سبع بيئات.

للحاصل باستخدام النموذج المقتر من قبل Eberhart و Russell (10). هذا الذي يعتمد على الانحرافات عن المتوسط الاجمالي للحاصل في بيئات مختلفة كمؤشرات على الانتاج في البيئات، اذ يعطي مؤشرات استجابة للانحدار (قيم b) ومتوسطات من المربعات للانحرافات للخطا المجمع الناقص للانحدار (قيم S<sup>2</sup>d) كمؤشرات لاستجابة الانتاجية والثبات، بالتتابع، ثم يعرف أداء الصنف بالمعادلة:

المعامل الذي يقيس انحدار الطراز الوراثي <sup>th</sup> في بيئات مختلفة:  $\delta_{ij} = \text{انحراف عن انحدار الطراز الوراثي في البيئة } i^{\text{th}} = I_j \cdot i^{\text{th}}$  مؤشر البيئة محسوب كمتوسط لكل الطرز الوراثية في البيئة <sup>th</sup> اقل من المتوسط الاجمالي في كل البيئات

ثم تربع هذه القيم وتجمع للحصول على  $\sum d^2_{EG}$ ، ثم تقسم على درجات الحرية والتي هي مثل درجات الحرية في حالة الارتباط (n-2) للحصول على تغاير الارتداد  $S^2_{EG}$  كما في المعادلة:

$$S^2_{EG} = \sum d^2_{EG} / n - 2 \dots \dots \dots (1)$$

$$S_b = S_{EG} / \sqrt{\sum (E - \bar{E})^2} \dots \dots \dots (2)$$

$$S_d = \sqrt{\sum d^2 - (\sum d)^2 / n(n-1)} \dots \dots \dots (3)$$

يلاحظ من اللوحة أن الهجين G<sub>1</sub> الواقع في المربع الموجب بعد ثابتاً لأنه امتلك قيم PC<sub>1</sub> اكبر من الصفر يليه الهجين G<sub>2</sub> الواقع في المربع الموجب بالنسبة لقيم PC<sub>2</sub>، إذ امتلك قيم PC<sub>2</sub> اكبر من صفر. أما الهجين G<sub>3</sub> فامتلك قيم PC<sub>2</sub> بعيدة جداً عن نقطة الأصل، لذا بعد غير ثابت، إلا أنه امتلك قيم PC<sub>1</sub> موجبة اكبر من صفر، لذا فهو يفيد في الانتخاب لصفات خاصة. أما الهجينان G<sub>3</sub> و G<sub>4</sub> فهما غير ثابتين لأنهما واقعين في المربع السالب لكل من PC<sub>1</sub> و PC<sub>2</sub>. قدرت معالم الثبات

$$Y_{ij} = \mu_i + \beta_j I_j + \delta_{ij}$$

اذ ان  $Y_{ij}$  = متوسط حاصل الحبوب للطراز الوراثي <sup>th</sup> في البيئة <sup>th</sup>  $\mu_i$  = متوسط حاصل الطراز الوراثي <sup>th</sup>  $\beta_j \cdot I_j$

الحل:

لحساب قيم S<sup>2</sup>d تستخرج قيم G<sub>i</sub> في الجدول وهو (d) أي الفرق بين متوسط أداء الصنف ومتوسط البيئة (E) [d=G-E]

وبجذر قيمة (S<sup>2</sup><sub>EG</sub>) نحصل على الانحراف المعياري للارتداد S<sub>EG</sub> بالمعادلة:

كما يمكن استخراج الخطأ المعياري للفرق الذي يرمز له (Sd) بالمعادلة:

جدول 5. قيم d و d<sup>2</sup> و S<sup>2</sup><sub>d</sub> و S<sub>b</sub> و S<sub>d</sub> لحاصل الهجن المزروعة في البيئات.

	d <sub>G1</sub>	d <sup>2</sup> <sub>G1</sub>	d <sub>G2</sub>	d <sup>2</sup> <sub>G2</sub>	d <sub>G3</sub>	d <sup>2</sup> <sub>G3</sub>	d <sub>G4</sub>	d <sup>2</sup> <sub>G4</sub>	d <sub>G5</sub>	d <sup>2</sup> <sub>G5</sub>
	2.16	4.6656	.21	.0441	1.76	3.0976	-	3.9601	-2.14	4.5796
							1.99			
	2.92	8.5264	-.68	.4624	.22	.0484	-.88	.7744	-1.58	2.4964
	2.84	8.0656	-.06	.0036	-1.31	1.7161	.04	.0016	-1.51	2.2801
	2.55	6.5025	0	0	.35	.1225	-.6	.36	-2.3	5.29
	2.52	6.3504	-.03	.0009	.62	.3844	-.98	.9604	-2.13	4.5369
	2.74	7.5076	-.01	.0001	.05	.0025	-.36	.1296	-2.41	5.8081
	2.77	7.6729	-.68	.4624	-.08	.0064	-.03	.0009	-1.98	3.9204
Σ	18.1	49.291	-1.25	.96	1.61	5.3779	-	6.188	-	28.9115
	6						4.80		14.05	
S <sup>2</sup> <sub>dG</sub>		9.8582		.192		1.0756		1.2376		5.7823
S <sub>dG</sub>		3.1398		.4382		1.0371		1.1125		2.4046
S <sub>bG</sub>		2.1464		.2996		.7090		.7605		1.6439
S <sub>d</sub>		.2278		.1325		.3453		.2626		.1301

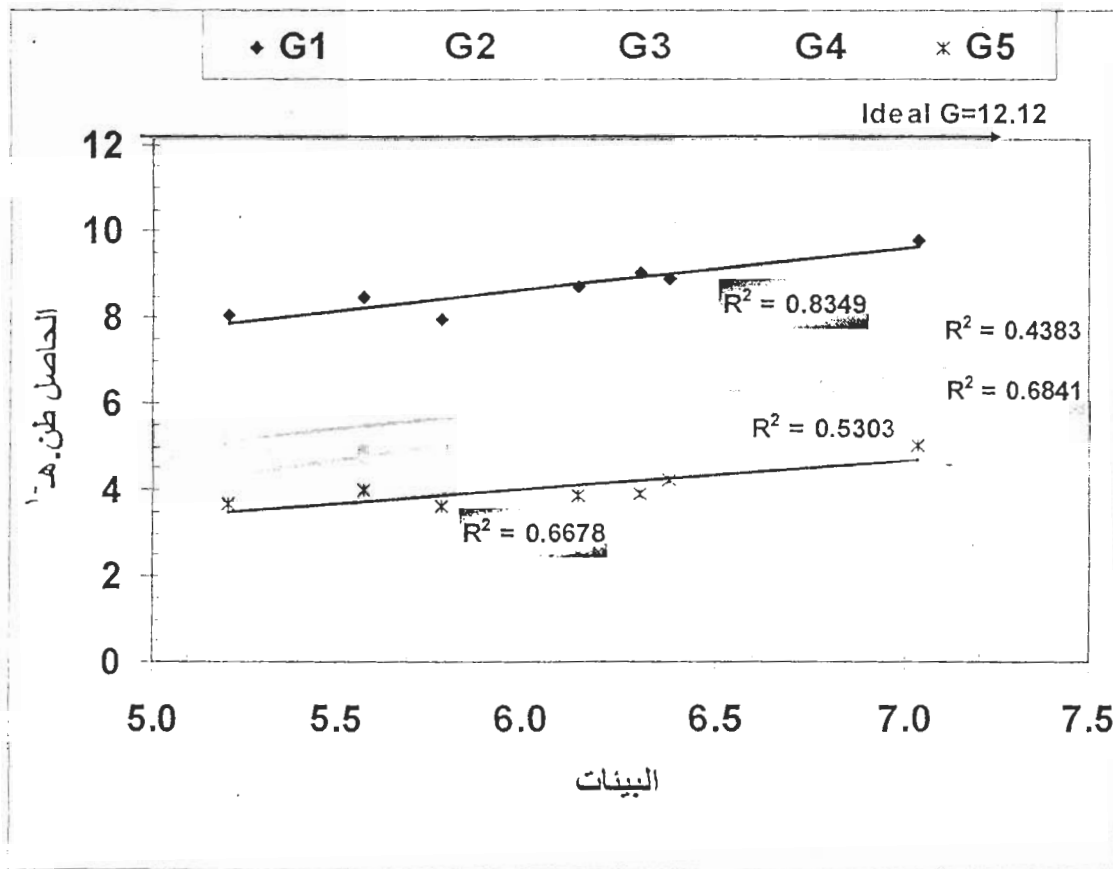
فمثلاً قيمة  $d_{G1} = 2.16 = G_{11} - E_{.1} = 7.95 - 5.79$ ، وبنفس الطريقة تستخرج بقية القيم. لحساب قيم  $S^2_{dG}$  تطبيق المعادلة (1) فمثلاً قيمة  $S^2_{dG1} = 9.8582$  أتت من تقسيم قيمة  $\sum d^2_{G1} = 49.291$  على  $2 - n = (2 - 7)$ ، وبنفس الطريقة لبقية الهجن. طبقت المعادلة (4) لحساب قيم b للهجن الخمسة:

$$b = \frac{\sum XY - \sum X \sum Y / n}{\sum X^2 - (\sum X)^2 / n} = \frac{\sum EG - \sum E \sum G / n}{\sum E^2 - (\sum E)^2 / n} \dots \dots \dots (4)$$

تم رسم البيانات كما في شكل 3. والاستنتاج المذكور تحته.

جدول 6. قيم معامل الانحدار  $b$  و  $R^2$  ومعادلة الخط المستقيم للهجن الخمسة المزروعة في سبع بيئات.

الهجن	$b$	$R^2$	معادلة الخط المستقيم
G1	.9671	.8349	$Y = .9671X + 2.8421$
G2	.8385	.6841	$Y = .8385X + .8011$
G3	1.3216	.4383	$Y = 1.3216X - 1.7219$
G4	1.2148	.5303	$Y = 1.2148X - 1.9883$
G5	.6580	.6678	$Y = .6580X + .0670$



شكل 3. معادلة الخط المستقيم لخمس هجن مزروعة في سبع بيئات.

فيو ثابت الا انه منخفض الحاصل. تشير متوسطات المربعات الناتجة عن الانحرافات ( $S^2_d$ ) إلى درجة الاعتماد التي توضع على الانحدار الخطي. ان ( $S^2_d$ ) تبدي استجابة غير خطية للهجن، وعندما تكون الانحرافات معنوية فان ثبات الهجن يحدد بتقدير ملحوظ لكل من  $\mu$  و  $\beta$  جدول (5). يعد الهجين ثابتاً عند قيمة ( $b$ ) مساوية لواحد (معدل الثبات)، اما اذا كانت قيمة ( $b$ ) اقل من واحد، فانه يشير الى درجة ثبات اعلى من المعدل، وقيمة ( $b$ ) اعلى من واحد تشير الى درجة ثبات اقل من المعدل مع قيم ( $S^2_d$ ) تقترب من الصفر وباستخدام هذه المعايير، فان الهجين  $G_1$  يعد

يلاحظ من الرسم ان معادلة الخط المستقيم للهجين الاول ( $R^2 = .8349$ ) تقترب من معادلة الخط المستقيم للهجين المثالي وبدا يعد هذا الهجين ثابتاً، كما انه اعطى أعلى معدل للحاصل (8.71 طن/هكتار). كذلك فان معادلة الخط المستقيم للهجين  $G_2$  اقتربت من معادلة الخط المستقيم للهجين المثالي بمتوسط حاصل 5.89 طن/هكتار. أما الهجينان  $G_3$  و  $G_4$  فقد اختلفت معادلتاهما عن معادلة الخط المستقيم للهجين المثالي بمتوسط حاصل 6.29 و 5.37 طن/هكتار اعتماداً على قيم ( $b$ )، أما الصنف  $G_5$  فكان موازياً للهجين المثالي بمتوسط حاصل 4.06 طن/هكتار

النبات و79.1 و81% للنسبة المنوية للشعر وحاصل الشعر في القطن بالتتابع. لإثبات النتائج في هذه المقالة ولفهم أفضل لتداخل GE لانجاز اهداف زراعية خاصة فان البحوث المستقبلية يمكن ان تعتمد استخدام PCA كاداة لتحليل ثبات الاصناف عبر البيانات. يطور GGE-Biplot العلاقات المتداخلة بين الاصناف التي تضمن ترتيبها اعتماداً على متوسط ادائها وثباتها، وكذلك العلاقات المتداخلة بين الاصناف والبيئة. ان دراسة الثبات والاداء مهمة جداً لاجل ادخال اصناف ذات حاصل وثبات عاليين سواء في بيئة الهدف او في عدة بيئات، عليه فان الاصناف التي تقترب من الصنف المثالي المقترض (Ideal cultivar) او توازيه تعد ثابتة وبهذا فهي تمتلك تطبعاً عاماً او واسعاً وبذا فان هذا التطبع هو الذي يمكن المحصول من الانتشار في عدة مناطق جغرافية متباينة، كما هو الحال في اداء الهجين G1 والذي اقترب من الصنف المثالي وكذلك الهجين G2 الذي كان موازياً للصنف المثالي وكما هو مؤكد تحليله ايضاً في لوحة GGE-Biplot شكل 2. فيما يلي وصف للبرنامج الاحصائي المستخدم لتحليل PCA كما اورده Yan (35).

البيانات التي حملتها على برنامج الاكسل سابقاً، فقطظهر لك البيانات.  
6. اضغط Biplot تظهر لديك شاشة GGE-Biplot.  
7. اختر Two-way data-matrix ثم ok تظهر لديك شاشة صغيرة GGE-Biplot، يسالك ما هو مؤشر فقدان خلايا من الملف فقط اضغط ok.  
8. تظهر لديك شاشة صغيرة Model selection فيه عدة ايعازات 0. Nosing  
1. Tester std Deviation  
2. Tester std error  
3. Tester LSD% اختر الامر رقم (1) ثم OK.  
9. تظهر شاشة الرسم النهائي GGE-Biplot للبيانات التي يمكنك مناقشتها والاستنتاج والتوصية منها.

ثابتاً، يليه  $G_2$  ثم  $G_3$  عند قيم معامل انحدار  $b = 0.9671$  و  $0.8385$  و  $0.658$ . بالتتابع اعطى الهجينان  $G_3$  و  $G_4$  قيم (b) اعلى من واحد، فيما اعطى  $G_3$  متوسط حاصل اعلى من المتوسط العام (6.29 ضن/هـ). هذا وقد ذكر الساهوكي (12) ان اعتماد قيمة b في اختبار ثبات الاصناف عبر البيانات لا تصف حالة الثبات بالمعنى المذكورة، وانما كانت صدفة ان هذا الاختبار الاحصائي يمكن تحويل مساره كي يفسر حركة البيانات بهذه الصورة. كذلك فقد اورد الساهوكي (12) ست عشرة معادلة قارن فيما بينها في دراسة موسعة، وشخص منها المعادلات  $(SD/\bar{X})$  او  $(1-SD/\bar{X})$  و  $s_i^2 = \sum (X_{ij} - \bar{X}.j)^2$  و  $b = \sum (X_{ij} - \bar{X}i) / \sum (X_{ij} - \bar{X}i) \bar{X}.j - \bar{X}..$  و  $s_i^2 = \sum (X_{ij} - \bar{X}.j)^2 / q - 1$ . بأنها الأفضل في تحديد ثبات الاصناف المزروعة عبر البيانات. عليه فان تحليل PCA هو اداة حديثة لتحديد ثبات واداء الاصناف المزروعة في عدة بيئات، وذلك لقدرة هذا التحليل في اختزال البيانات المشتتة والاستنتاج منها في ثبات عدة محاصيل، ففي صفة حاصل الحبوب في الحنطة الذي وجده Yan وآخرون (38) علل التحليل معن 78% و 79.5% و 92% من التباين لارتفاع

#### خطوات العمل على برنامج GGE-Biplot (version 5.2):

1. تحميل بيانات التحليل على الاكسل (excel(xls)).
2. تنصيب البرنامج تظهر (icon) التحليل بشكل لوحة رسم GGE-Biplot. يعمل البرنامج على windows 2001-2006 و winxp تجسده على [www.ggebiplot.com/biplot.htm](http://www.ggebiplot.com/biplot.htm).
3. ابدأ بالضغط عليها مرتين او من start/programefile/GGE-Biplot تظهر لدينا شاشة GGE-Biplot-start.
4. اضغط على start تظهر لدينا شاشة GGE-Biplot Data.
5. اضغط Data(\*) open تظهر لدينا شاشة بجدول متعددة، اختر جدولاً لعرضه من ايعاز select a table قريب من ايعاز open Data بهذه الحالة سوف تستدعي

#### المصادر

Wen Tai area in China. Agric., Ecosystems and Environ. J. 89: 241-252.

4- Blanche, S. B., G. O. Myers, J. Z. Zumba, D. Coldwell and J. Hayes. 2006. Stability comparisons between conventional and near-isogenic transgenic cotton cultivars. Cotton Sci. J. 10: 17-28.

5- Bregard, A., G. Belanger, R. Michaud and G.F. Tremblay. 2001. Biomass partitioning, forage nutritive value, and yield of contrasting genotype of Timothy. Crop Sci. 41: 1212-1219.

6- Campbell, B.T., and M.A. Jones. 2005. Assessment of genotype x

1- Aba, D.A., C.C. Nwasike, M. Yeye and A. A. Zaria. 2001. Studies on genetic variations in a sorghum variety irradiated with cobalt 60 ( $Co^{60}$ ). Afr. Crop. Sci. J. 9 (2): 377-384.

2- Allard, B.W., and A.D. Bradshaw. 1964. Implications of genotype x environment interaction in plant breeding. Crop Sci. 4: 503-508.

3- Bi, X., S. Chu, Q. Meng and X. Xu. 2002. Movement and retention of polychlorinated biphenyls in paddy field

- York, NY, USA. www.target.com/principal component analysis-Jolliffe-I-T/dp/.
- 18- Karamura, D., and B. MGenzi. 2004. On farm conservation of musa diversity in the great lakes region of East Africa. *Afr. Crop Sci. J.* 12(1): 75-83.
- 19- Kaya, Y., C. PaKa and S. Taner. 2002. Additive main effects and multiplicative interactions analysis of yield performances in bread wheat genotypes across environments. *Turk. J. Agric.* 26:275-279.
- 20- Kaya, Y., M. Akura and S. Taner. 2006. GGE-Biplot analysis of multi-environment yield trials in bread wheat. *Turk. J. Agric.* 30:325-337.
- 21- Kenga, R., S.O. Alabi and S.C. Gupta. 2003. Yield stability of sorghum hybrids and parental lines. *Afr. Crop Sci. J.* 11 (2) : 65-73.
- 22- Kirkpatrick, R., and K. Meyer. 2004. Direct estimation of genetic principal components: Simplified analysis of complex phenotypes. *Genetics* 168:2295-2306.
- 23- Labib, K., and V.R. Vermuri. 2004. An application of principal component analysis to the detection and visualization of computer network attacks. proceedings of SAR. <http://www.hds.utc.fr/sar04>.
- 24- Langenheder, S., E.S. Lindstrom and L.J. Tranvik. 2005. Weak coupling between community composition and functioning of aquatic bacteria. *Limnol. Oceanograph.* 50(3):957-967.
- 25- Mahajan, R.K., A.V. Rao and D.K. Aggarwal. 1981. Principal component analysis in some late maturing rice varieties. *Indian J. Agric. Sci.* 51 (1) : 9-12.
- 26- Nassiuma, D., and W. Wasike. 2002. Stability assessment of soybean varieties in Kenya. *Afr. Crop. Sci. J.* 10 (2) : 139-144.
- 27- Ogunbodede, B.A. 1997. Multivariate analysis of genetic diversity in Kenaf, *Hibiscus connabinus* L. *Afr. Crop Sci. J.* (2) : 127-133.
- 28- Peterson, D. M., D. M. Wesenberg, D. E. Burrup, and C. A. Erickson. 2005. Relationships among agronomic traits and environment interaction for yield and fiber quality in cotton performance trials. *Euphytica* 144 : 69-78.
- 7- Ding, C. and X. He. 2004. K-means clustering via principal component analysis. Proceedings of the 21th International conference on machine learning, Banff, Canada.
- 8- Ding, C., D. Zhou, X. He and H. Zha. 2006. R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization. Proceedings of the 23th International conference on machine learning, Pittsburg, PA.
- 9- Ebdon, J.S., and H. G. Gauch. 2002. Additive main effect and multiplication interaction analysis of national turf grass performance trials I. Interpretation of genotype x environment interaction. *Crop Sci.* 42 : 489-496.
- 10- Eberhart, S.A., and W.A. Russell. 1966. Stability parameters for comparing varieties. *Crop Sci.* 6 : 36-40.
- 11- Egesi, C.N., and R. Asiedu. 2002. Analysis of yam yields using the additive main effects and multiplicative interaction (AMMI) model. *Afr. Crop Sci.* 10 (3) : 195-201.
- 12- Elsahookie, M.M. 1996. Applications on stability analysis of genotypes. *Iraqi J. Agric. Sci.* 27 (2) : 11-20.
- 13- Ezeaku, I.E., S.C. Gupta and V.R. Prabhakar. 1999. Classification of sorghum germplasm accessions using multivariate methods. *Afr. Crop Sci. J.* 7 (1) : 97-108.
- 14- Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58(3):453-467.
- 15- Gebeyehu, S., and H. Assefa. 2003. Genotype x environment interaction and stability analysis of seed in navy bean genotypes. *Afr. Crop Sci.* 3. 11 (1) : 1-7.
- 16- Gray, J.R., and P.M. Thompson. 2004. Neurobiology of intelligence: Science and ethics. *Nat. Rev. Neurosci. J.*:471-482.
- 17- Jolliffe, I. T. 2002. *Principal Component Analysis*, Springer - Verlag, New

- 35- Yan , W . 2001 . GGE biplot – A windows application for graphical analysis of multienvironment trial data and other types of two-way data . Agron. J. 93 : 1111-1118.
- 36- Yan , W. and N. A. Tinker. 2005. Biplot analysis of multi- environment trial data : principles and applications . Presented at the Plant Canada 2005 Symposium, Contemporary Issues in Statistics . Data Management , Plant Biology and Agriculture Research.
- 37- Yan , W. 2002. Singular – value partitioning in biplot analysis of multienvironment trial data . Agron. J. 94 : 990-996.
- 38- Yan,W.,L.A.Hunt,Q.Sheng and Z,Szlavnic.2000.Cultivar evaluation and mega-environment investigation based on GGE-Biplot.Crop Sci.J.40:597-605.
- 39- Yeung,K.Y., and W.L.Ruzzo.2001.Principal component analysis for clustering gene expression data.Bioinformatics 17(9):763-774.
- grain composition in oat genotype grown in different environments . Crop Sci. 45 : 1249-1255 .
- 29- Pixley , K. V. and M. S. Bjarnason . 2002. Stability of grain yield , endosperm modification , and protein quality of hybrid and open – pollinated quality protein maize (QPM) cultivars . Crop Sci. 42 : 1882-1890.
- 30- Shlens , J . 2005. A tutorial on principal component analysis : Scientific papers .<http://www.sn1.salk.edu/shlens/>.
- 31- Smith , L. I. 2002. A tutorial on principal components analysis . Otago University , New Zealand.<http://www.csnet.otago.ac.nz/cosc433/>.
- 32- Yan , W.,and N. A. Tinker. 2005. An integrated biplot analysis system for displaying , interpreting , and exploring genotype x environment interaction . Crop Sci. 45 : 1004-1016 .
- 33- Yan , W.,and L.A. Hunt . 2002 . Biplot analysis of diallel data. Crop Sci. 42 : 21-30.
- 34- Yan , W. ,M. S. Kang , B. Ma , S. Woods , and P. L. Cornelius . 2007. GGE-Biplot vs. AMML analysis of genotype by environment data. Crop Sci. 47 : 641-653.