YYÜ TAR BİL DERG
(YYU J AGR SCI)
2012, 22(3): 134–146

Geliş Tarihi(Received):16.04.2012
Kabul Tarihi (Accepted): 20.06.2012

Araştırma Makalesi/*Research Article (Original Paper)*

Grouping Lentil Genotypes by Cluster Methods Related to Linear Regression Model and Genotype × Environment Interaction Variance

Naser SABAGHNIA [1*], Rahmatollah KARIMIZADEH[2], Mohtasham MOHAMMADI [2]

[1]Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Maragheh, Maragheh, Iran.
[2]Dryland Agricultural Research Institute (DARI), Gachsaran, Iran
*e-mail: sabaghnia@maragheh.ac.ir; Tel: +98 (421) 2273070; Fax: +98 (421) 2273070

Abstract: Lentil (*Lens culinaris* Medik.) as a rich source of protein for human consumption is the fourth most important pulse crop in the world. Genotype × environment (GE) interaction is observed in lentil like the other crops performance trials and is an important issue for plant breeders. Eighteen lentil genotypes were evaluated under rainfed conditions using a randomized complete block design with 4 replications for 3 years between 2008 and 2010 and at 4 different locations. GE interaction was analyzed using clustering techniques. There was considerable variation for grain yield among both genotypes and environments based on combined ANOVA. Also, combined analysis of variance indicated that three way interaction (GYL) or GE interaction was highly significant. The high significance of GE interactions is indicating the studied genotypes exhibited both crossover and non-crossover types of GE interaction. According to dendrograms of regression method (methods 1) there were 2 different genotypic groups based on G (intercept) and GE (line slope) sources while there were none different genotypic groups based on GE (line slope) source. Also, the dendrograms of ANOVA methods indicated 12 different genotypic groups based on G and GE sources and 11 different genotypic groups based on GE sources. Considering both mean yield and stability performance, genotypes G1 (1418.7 kg ha^{-1}), G5 (1324.4 kg ha^{-1}), G14 (1401.9 kg ha^{-1}) and G15 (1307.3 kg ha^{-1}) were found to be the most favorable genotypes which could be recommended for national releases. Such an outcome could be regularly applied in the future to exploration lentil genotypes and other crops based on regression or ANOVA models in the other areas of the world.

Key words: Cluster analysis, GE interaction, Grain yield, *Lens culinaris* Medik.

Mercimek Genotiplerinin Doğrusal Regresyon Model ve Genotip x Çevre İnteraksiyon Varyansı ile İlgili Kümeleme Yöntemleri ile Gruplanması

Özet: İnsan tüketimi için zengin bir protein kaynağı olan mercimek (*Lens culinaris* Medik.), dünyanın dördüncü en önemli baklagil ürünüdür. Genotip x Çevre (GÇ) interaksiyonu, diğer ürün performans denemelerinde olduğu gibi mercimekte de gözlenen ve bitki ıslahçıları için önemli bir konudur. On sekiz mercimek genotipi, 2008 ve 2010 yılları arasında 3 yıl ve 4 lokasyonda tesadüf blokları deneme desenine göre 4 tekerrürlü olarak yağmurla beslenen koşullar altında değerlendirilmiştir. GÇ interaksiyonu, kümeleme teknikleri kullanılarak analiz edilmiştir. Kombine ANOVA'ya göre, verim konusunda hem genotipler arasında hem de çevre koşulları arasında önemli farklılıklar bulunmuştur. Ayrıca, kombine varyans analizi, üç yönlü interaksiyonun (GYL) veya GÇ interaksiyonun son derece önemli olduğunu göstermiştir. GÇ interaksiyonun çok önemli olması, çalışılan genotiplerin hem crossover hem de non-crossover GÇ interaksiyon tiplerini sergilemekte olduğunu göstermiştir. Regresyon yöntemi dendogramına göre (1. Yöntem), GE (eğim) kaynağına göre hiçbir farklı genotipik grup bulunmazken, G (sabit) ve GE (eğim) kaynaklarına göre 2 farklı grup bulunmaktadır. Ayrıca, ANOVA yöntemleri dendogramları, GE kaynağına göre 11 farklı genotipik grup gösterirken, G ve GE kaynaklarına göre 12 farklı grup göstermiştir. Hem verim hem de istikrar performansı dikkate alındığında, G1 (1418,7 kg/ha), G5 (1324,4 kg/ha), G14 (1401,9 kg/ha) ve G15 (1307,3 kg/ha) genotipleri ulusal sürüm için tavsiye edilebilecek en olumlu genotipler olarak bulunmuştur. Böyle bir sonuç, gelecekte dünyanın diğer alanlarında regresyon ve ANOVA modellerine bağlı olarak mercimek genotiplerinin ve diğer ürünlerin araştırılmasında düzenli bir şekilde uygulanabilecektir.

Anahtar kelimeler: GÇ interaksiyonu, Kümeleme analizi, *Lens culinaris* Medik, Tohum verimi

Introduction

Lentil (*Lens culinaris* L.) is the fourth most important pulse crop in the world after bean, pea, and chickpea which has important role in human consumption in some developing countries. Iran has the leading lentil breeding programs in recent years and the history of lentil production in Iran has been reviewed by Sabaghpour et al. (2004). Lentil seed is a rich source of protein for human consumption and its straw is a valued animal feed. It is adapted to less favorable environments, where it is predominantly grown in winter in regions where annual average rainfall ranges between 300 and 400 mm (Mohebodini et al. 2006). Cereals are more tolerant to drought, poor soils and salinity. The planting of legumes in rotation with cereals has been demonstrated to be beneficial in many arid and semi-arid areas (Dehghani et al. 2008). Although lentil-breeding programs have some priorities in common, the major objective of increasing the genetic potential of yield for most, if not for all, can be achieved via breeding for higher yield potential or eliminating hazards that reduce yield (Sabaghnia et al. 2008a).

Ensuring the stability of high yield performance under unfavorable environment conditions is the main problem facing breeders producing improved different crop genotypes (Annicchiarico 2002). A genotype is considered to be more adaptive or stable one if it has a high mean yield but a low degree of fluctuation in yielding ability when grown over diverse environments (Kang 1998, Gauch 2006). Most of the GE interactions are statistically non-additive, indicating that differences in yields among genotypes will depend on the environment (Yue et al. 1997). Consequently, selection procedures based on the mean yield of genotypes in a given test environment are less efficient (Hopkins et al. 1995). Lin et al (1986) compared 9 univariate stability methods to exploration of GE interactions in multi-environment trials and grouped them based on three stability types. Flores et al (1998) compared 22 univariate and multivariate methods to analyze GE interactions and classified them into three main groups: univariate parametric methods, univariate nonparametric methods and multivariate methods such as AMMI (additive main effects and multiplicative effects).

Cluster analysis as a nonparametric multivariate method classifies genotypes into categories and its final goal is to identify the actual groups. Several cluster analyses have been developed which some of them classify individuals for the similarity according to the one-way method (Callinski and Corsten 1985); and some others classify individuals for similarity of interactions based on the two-way method (Lin and Butler, 1990). Lin and Thompson (1975) used the deviation mean square from regression analysis of GE interaction (Finlay and Wilkinson 1963) as dissimilarity index for clustering. In contrast Lin (1982) used the GE interaction mean square as dissimilarity index for genotypes classification through a slight adjustment of distance coefficient. Lin and Butler (1990) introduced a new dissimilarity index according to regression analysis which benefits only genotype main effect. Also, Lin and Butler (1990) proposed another new dissimilarity index based on mean square of only GE interaction in contrast of dissimilarity index of Lin and Thompson (1975) which uses both effects of genotype and GE interaction in ANOVA procedure. The purpose of this study is to indicate the practical application of cluster analysis at study of GE interaction in lentil using four clustering methods.

Materials and Methods

Sixteen recently improved lentil genotypes with two registered cultivars (Gachsaran and Cabralia) were analyzed in completely randomized block design layout with 4 replications. Sowing was performed manually in rows 25 cm apart in 1 × 4 m plots consisting of 4 rows in the 2003-2004 growing seasons in four locations. The planted plot size was 4 m^2 and the harvested plot size was 1.75 m^2 (two 3.5 m rows at the center of each plot). Plots were not irrigated because the lentil is grown in rain-fed conditions. Control by hand weeding was carried out twice when the weed density was high, in the pre-flowering and post-flowering stages. The plots were fertilized with 20 kg N ha^{-1} and 80 kg P_2O_5 ha^{-1} at planting. The test location Gorgan in the north-east of Iran is characterized by semi-arid conditions with sandy loam soil. Kermanshah in the west of Iran is characterized by semi-arid conditions with clay loam soil. Gachsaran, in southern Iran, is relatively arid and has silt loam soil. Shirvan in the north-east of Iran is characterized by moderate conditions, relatively high rainfall and have clay loam soil. The test locations were selected as sample of lentil growing areas of Iran and to vary in latitude, rainfall, soil types, temperature and other agro-climatic factors. The properties and the location of the experimental environments are given in Table 1.

Table 1. Geographical properties and mean yield of the 18 lentil genotypes, studied in 4 locations

Code	Location	Altitude (meter)	Longitude Latitude	Soil Texture	Rainfall (mm)	Yield (kg ha^{-1})
1	Gorgan	45	55 12 E 37 16 N	Silty Clay Loam	367	767
2	Kermanshah	1351	47 19 E 34 20 N	Clay Loam	455	1923
4	Gachsaran	710	50 50 E 30 20 N	Silty Clay Loam	460	1747
5	Shirvan	1131	58 07 E 37 19 N	Loam	267	384

Cluster analysis procedures of Lin and Thompson (1975), Lin (1982) and Lin and Butler (1990), consists of methods depending on the linear regression model or the ANOVA model were used. Detailed illustration of clustering and computation of dissimilarity index are given in Lin and Butler (1990). Suppose *m* is the number of genotypes, *n* is the number of environments, *r* is the number of genotypes in a newly formed cluster and *rep* is the number of experiment replication. The first degrees of freedom (v_1) in method 1 is *2(r-1)*, and the second degrees of freedom (v_2) in method 1 is *(m-1)(n-1)*; the v_1 in method 2 is *(r-1)*, the v_2 in method 2 is *(m-1)(n-1)*; the v_1 in method 3 is *n(r-1)*, the v_2 in method 3 is *n(m-1)(rep-1)*; and the v_1 in method 4 is *(n-1)(r-1)*, the v_2 in method 4 is *(m-1)(n-1)*. The dissimilarity indices of methods 1 and 2 are the numerators of the test statistics for a common regression line and for parallelism, respectively. The dissimilarity indices of Methods 3 and 4 are mean squares of genotypes (G) and mean squares of genotypes plus GE interaction (G + GE), respectively. Detailed explanation and simple numerical examples of the procedure can be seen in Lin and Butler (1990). A statistical package as Sl16 (Lin et al. 1992) which is based on FORTRAN-77 program is used for all four methods of cluster analysis.

Results

Combined analysis of variance was conducted to determine the main effects of year, location, genotype, and their interactions on grain yield of lentil genotypes (Table 2). The main effects of year and location were not significant, but their interaction (YL) was highly significant ($P < 0.01$). The main effect of genotypes was significant, the genotype by year interaction was not significant, the genotype by location interaction was significant ($P > 0.05$) and three way interaction (GYL) or GE interaction was highly significant (Table 2). The high significance of GE interactions is indicating the studied genotypes exhibited both crossover and non-crossover types of GE interaction. Analyses of the quantitative traits like grain yield indicates important sources of genetic variation attributed to GE interaction. Complexity of grain yield is a result of diverse processes that occur during plant development. The GE interaction has been assumed to be a source of error that can be summed to zero by evaluating genotypes in large samples of environments representing the target of test environments (Cooper et al. 2002). The relative large contributions of GE interaction effects in grain yield of lentil which found in this study is similar to those found in other multi-environment trials' studies of lentil in rain-fed conditions (Mohebodini et al. 2006, Sabaghnia et al. 2008b).

Table 2. ANOVA analysis of lentil performance trial yield

Source	DF	MS	% of G+E+GE†
Year (Y)	2	8400774^{ns}	0.21
Location (L)	3	3962077^{ns}	0.15
Y × L	6	4579496**	0.34
R (Y × L)	36	38152	
Genotype (G)	17	320003**	0.07
G × Y	34	80769^{ns}	0.03
G × L	51	134137*	0.09
G × L × Y	102	84021**	0.11
Error	612	31713	

†E=Y+L+YL; GE=GY+GL+GLY

The linear regression model (Finlay and Wilkinson 1963) was calculated and for each genotype (Table 3). The pooled error estimate is 23641.32 which is the sum of deviation from linear regression of all genotypes and was used to performing F-test and stopping the clustering process. The clustering cycle, grouped genotypes, dissimilarity index of each clustering cycle, degrees of freedom in each step and related F-test statistic are given in Table 4. The F-test statistic was significant in the final cycle (17) where the dissimilarity index was 55331.91. In this step genotypes G6, G8, G10 and G17 were grouped with a cluster which containing other remained genotypes and so there was significant difference between these groups due to G and GE sources of linear regression model. The positions of the all studied genotypes and significant cutoff point in this method are seen in Fig. 1. According to this dendogram, there were two different genotypic groups. Lin and Thompson (1975) to improve the effectiveness of this cluster method indicate that most of the variation among genotypes is included in the between group component. The determination coefficient of linear regression model for 18 lentil genotypes ranged from 52.6 (G10) to 97.6 (G5) (Table 3). Genotypes with high coefficient of determination (R^2) values can be evaluated adequately via linear regression model and the genotype response to environments is predictable to considerable degree (Pinthus 1973, Crossa 1990). Regarding relatively moderate to high values of coefficient of determination for lentil genotypes, it can be concluded that using regression clustering method is useful to some extent for this dataset.

Table 3. Linear regression parameters and regression analysis of variance statistics

Genotype	Mean	Intercept	Slope	SS Total	SS Reg.†	SS Res.‡	R^2
1	1418.7	1418.8	1.27	1515242.3	1250305.0	26493.7	82.5
2	1365.6	1365.8	1.05	1018665.7	863545.9	15512.0	84.8
3	1287.3	1287.5	0.82	721705.0	525693.9	19601.0	72.8
4	1272.0	1272.1	1.17	1284266.9	1070198.4	21406.9	83.3
5	1324.4	1324.5	1.25	1263167.0	1227363.5	3580.2	97.2
6	1096.5	1096.6	0.85	700746.9	565836.4	13491.1	80.7
7	1304.1	1304.2	0.69	617923.7	375358.4	24256.6	60.7
8	1191.1	1191.2	0.73	539627.7	416570.8	12305.8	77.2
9	1329.5	1329.5	0.92	1189801.0	659870.4	52993.0	55.5
10	1187.9	1188.0	0.74	809204.0	425368.8	38383.6	52.6
11	1374.1	1374.4	0.77	602236.9	467730.8	13450.7	77.7
12	1334.7	1334.9	1.01	1011974.9	794669.0	21730.5	78.5
13	1292.1	1292.2	0.95	842195.7	705067.3	13712.8	83.7
14	1401.9	1402.1	1.25	1475832.9	1226959.8	24887.3	83.1
15	1307.3	1307.5	1.25	1394695.0	1210375.5	18431.8	86.8
16	1272.4	1272.5	1.11	1203809.0	956080.7	24772.8	79.4
17	1203.3	1203.5	1.11	1459865.0	959035.7	50082.8	65.7
18	1314.6	1314.8	1.05	930617.7	862452.3	6816.6	92.7

†Linear regression model sum of squares

‡ Residual sum of squares

Table 4. The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through regression model and based on line slope and intercept

Step	Grouped genotypes	Diss. index	v_1‡	v_2¶	F-test
1	8,10	39.55	2	170	0.00
2	4,16	803.24	2	170	0.03
3	1,14	860.90	2	170	0.04
4	5,15	883.64	2	170	0.04
5	12,18	1556.79	2	170	0.07
6	9 (12,18)	2413.94	4	170	0.10
7	3,13	3352.22	2	170	0.14
8	2 (9,12,18)	4275.51	6	170	0.18
9	(5,15) (4,16)	5980.07	6	170	0.25
10	7 (3,13)	6886.04	4	170	0.29
11	11 (2,9,12,18)	9215.73	8	170	0.39
12	(7, 3,13) (2,9,11,12,18)	13175.34	14	170	0.56
13	17 (8,10)	14919.57	4	170	0.63
14	6 (8,10,17)	19099.95	6	170	0.81
15	(4,5,15,16)(2,3,7,9,11,12,13,18)	23618.05	22	170	1.00
16	(1,14)(4,5,15,16,2,3,7,9,11,12,13,18)	31305.80	26	170	1.32
17	(6,8,10,17) (1,14,4,5,15,16,17,2,3,7,9,11,12,13,18)	55331.91	34	170	2.34**

Most genotypes with the highest line slope which indicated specific adaptability to favorable environments clustered as big group (Fig. 1). Also, genotypes G6, G8, G10 and G17 (small group) were stable and had specific adaptability to poor environments due to low line slopes. The genotypes clustering based on similarity of linear regression parameters (both intercept and slope parameters) indicated that there are considerable variation among lentil genotypes and there are different with each other in response to environmental changes. This can be due to different origin of these improved genotypes, different pedigree and different breeding procedure. Also, this is noticed that all variations could not be investigated due to R^2 values. The question of whether similarity must be used (line slope alone or on both intercept and slope parameters), depends on the degree of emphasis the plant breeder wishes to put on GE interaction. For clustering genotypes, the similarity of both linear regression parameters (intercept and slope) may be more proper but for clustering test locations, the similarity of slope (GE interaction) is often more suitable. Of course, Lin and Butler (1990) conclusion is correct in those situations which magnitude of genotype effect (or intercept) is greater than GE interaction (or line slope). Brandle and Brule-Bable (1991) indicated that cluster analysis based on regression analysis may be a suitable tool of selecting the most stable, high yielding and responsive genotypes in rapeseed.

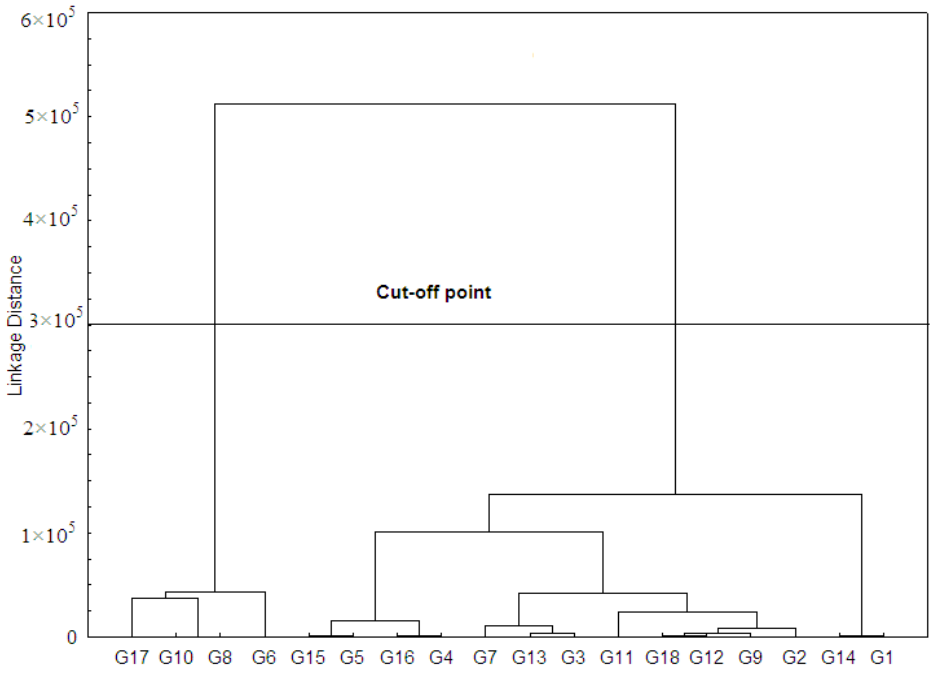

Fig. 1. Dendrogram of dissimilarity indices based on line slope and intercept of regression model for 18 genotypes of lentil which were evaluated across 12 environments.

Similar to method 1, joint linear regression (Finlay and Wilkinson 1963) was used to clustering. The properties of cutoff point determination in this clustering method and detailed information including clustering cycles, grouped genotypes, dissimilarity indices of cycles and F-test statistic are given in Table 5. The F-test statistic was significant in none of cycles and so there was not significant difference among lentil genotypes based on GE sources or lines slopes of linear regression model. The clustering cycles were summarized in dendogram of Fig. 2. Although, Lin and Butler (1990) to improve the effectiveness of Lin and Thompson (1975) cluster method proposed this clustering method based on only lines slopes or GE source but this procedure could not distinguish genotypic variations due to linear slopes. Similar to method 1, the validity of regression model could be proved by values of determination coefficient (Table 4) which were at moderate or high values. Karimizadeh et al. (2006) declared that there were good agreements between clustering methods which are based on similarity of both slopes and intercepts or only slopes in studying multi-environment trials of the different maize hybrids. In contrast we did not find any agreement in methods 1 and 2.

Table 5. The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through regression model and based on line slope

Step	Grouped genotypes	Diss. index	v_1‡	v_2¶	F-test
1	2,18	0.21	1	170	0.00
2	5,14	0.54	1	170	0.00
3	16,17	0.93	1	170	0.00
4	15 (5,14)	19.34	2	170	0.00
5	8,10	22.92	1	170	0.00
6	1 (5,14,15)	54.91	3	170	0.00
7	3,6	368.98	1	170	0.02
8	9,13	374.09	1	170	0.02
9	11 (8,10)	422.17	2	170	0.02
10	12 (2,18)	469.68	2	170	0.02
11	7 (8,10,11)	853.67	3	170	0.04
12	4 (16,17)	1044.15	2	170	0.04
13	(4,16,17)(2,12,18)	2561.07	5	170	0.11
14	(3,6)(7,8,10,11)	2750.07	5	170	0.12
15	(9,13)(2,4,12,16,17,18)	5550.39	7	170	0.23
16	(2,4,9,12,1316,17,18)(1,5,14,15)	11785.80	11	170	0.50
17	(3,6,7,8,10,11) (2,4,9,12,1316,17,18,1,5,14,15)	30610.35	17	170	1.29

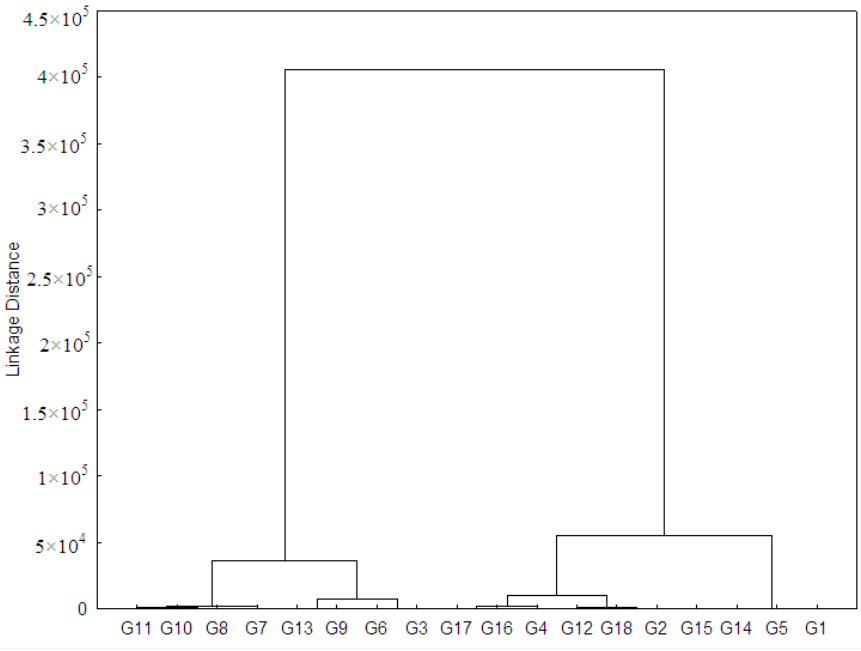

Fig. 2. Dendrogram of dissimilarity indices based on line slope of regression model for 18 genotypes of lentil which were evaluated across 12 environments.

The GE interaction in linear regression model was partitioned to heterogeneity (randomized variation) and residual components. In other word, GE interaction split to heterogeneity with df=17, MS=30610.59 and residual with df=170, MS=23641.32. The heterogeneity component as the randomized variation was not significant and indicated contribution of non-random effects was grater than random effects in GE interaction nature. Mandel (1961) indicated that if the lines slopes were identical for all studied genotypes, this heterogeneity component is distributed as χ^2 and is independent of environmental effects. Therefore, with considering high values of determination coefficient for linear regression, the model is appropriate and the GE interaction partitioning provides a method of testing for systematic GE interaction.

Although, like to other crops, the seed yield of each lentil genotype is a combined result of the of the G, E and GE interaction effects, only G and GE interaction are responsible to genotypes evaluation in multi-environment trials. Usually, E source describes most of the total seed yield variation, while G and GE interaction are usually small (Yan and Kang, 2003). Lin and Butler (1990) proposed a dissimilarity index using both G and GE interaction in terms of distance adjusted for the average effects of these sources in ANOVA. The numerical results of clustering process: clustering cycles, clustered genotypes, dissimilarity index of each step, degrees of freedom and F-test statistic are given in Table 6. According to the results of this table, F-test statistic was significant in cycle 7 where the dissimilarity index was 13279.33. In this step genotypes G11 and G12 were grouped with a cluster which containing genotypes G13 and G15 and so there was significant difference between these clusters based on G and GE sources of ANOVA model.

Table 6. The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through ANOVA model and based on genotype main effect (G) and GE interaction

Step	Grouped genotypes	Diss. index	v_1‡	v_2¶	F-test
1	5,18	5768.00	12	612	0.91
2	4 (5,18)	7887.17	24	612	1.24
3	13,15	10643.67	12	612	1.68
4	11,12	11475.00	12	612	1.81
5	3 (4,5,18)	11484.89	36	612	1.81
6	2,7	11696.33	12	612	1.84
7	(11,12)(13,15)	13279.33	36	612	2.09**
8	16 (3,4,5,18)	13741.17	48	612	2.17
9	1 (11,12,13,15)	14794.83	48	612	2.33
10	8 (3,4,5,16,18)	16420.27	60	612	2.59
11	14 (1,11,12,13,15)	16712.53	60	612	2.63
12	(2,7)(3,4,5,8,16,18)	18952.38	84	612	2.99
13	10 (2,3,4,5,7,8,16,18)	21252.83	96	612	3.35
14	(1,11,12,13,14,15)(2,3,4,5,7,8,10,16,18)	23583.05	156	612	3.72
15	9(1,11,12,13,14,15,2,3,4,5,7,8,10,16,18)	25456.89	180	612	4.01
16	17(1,11,12,13,14,15,2,3,4,5,7,8,9,10,16,18)	27331.00	192	612	4.31
17	6(1,11,12,13,14,15,2,3,4,5,7,8,9,10,16,17,18)	28923.45	204	612	4.56

For obtaining the dendrogram cutoff point, 20% of pooled error of combined ANOVA (Robert 1997) was used. According to this dendrogram of Fig. 3, there were 12 different genotypic groups consist on: genotypes G1, G6, G8, G9, G10, G14 and G17 as individual groups; G3 and G16; G2 and G7; G11 and G12; G13 and G15; and G4, G5 and G18 as the composite groups which had more than one genotype. This graphic presentation indicated high variation among studied lentil genotypes considering both genotypic main effects and GE interaction. Method 3 clustering procedure which benefits from G and GE interaction effects can be useful for identifying the most stable genotypes according to type I stability (Lin et al. 1986). Although, successful applications of Type I stability have been reported for small area tests (Francis and Kannenberg 1978) and some international experiments (Mohebodini et al. 2006, Dehghani et al. 2008, Karimizadeh et al. 2012), but the other stability types (Type 2 and Type 3) are very popular among plant breeders.

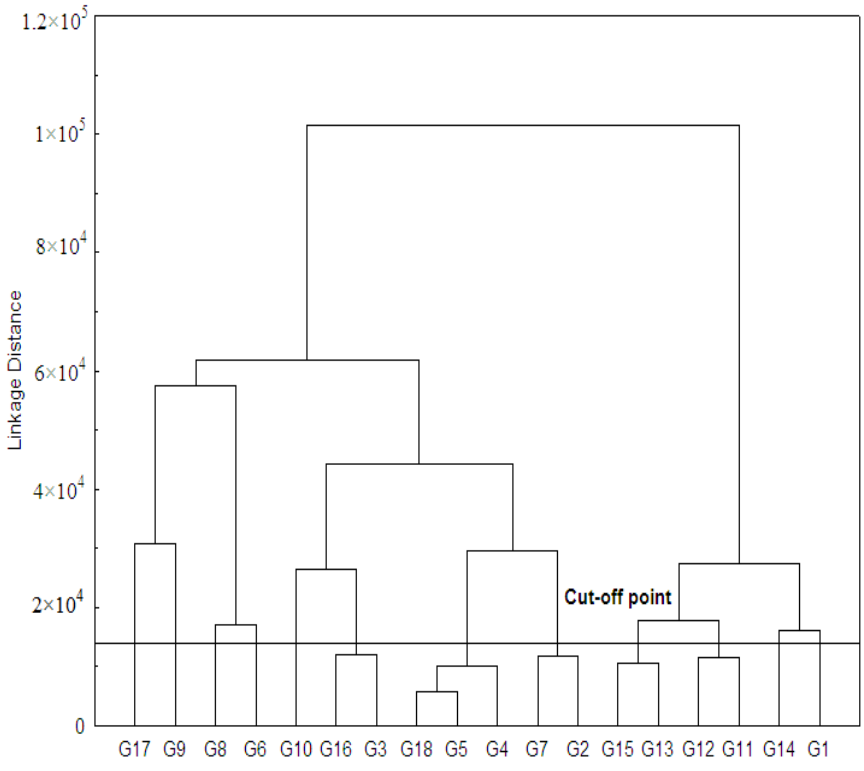

Fig. 3. Dendrogram of dissimilarity indices based on G and GE interaction of ANOVA model for 18 genotypes of lentil which were evaluated across 12 environments.

The dissimilarity index of method 4 is defined in terms of distance adjusted for the average effects of genotypes and it to be equivalent to within group MS of GE interaction in ANOVA (Lin 1982). The un-weighted pair-group clustering algorithm of Sokal and Michener's (1958) was used. The clustering process including clustering cycles, genotypes which grouped, dissimilarity index of each clustering cycle, degrees of freedom of F-test in each step and related F-test statistic are given in Table 7. In this investigation, we regarded that the optimized number of genotypes must still be very informative for GE interaction interpretation. According to Robert (1997) cutoff point was fixed 20% of pooled error in combined ANOVA and therefore GE interaction within clusters must thus be less than 20% of total variation. The F-test statistic was significant in cycle 9 where the dissimilarity index was 11890.42 and in this step genotypes G6 and G11 were grouped with a cluster which containing genotypes G8 and G13 (Table 7). Therefore, there was significant difference between these clusters based on GE interaction.

Table 7. The smallest dissimilarity index at each cluster step and the determination of the cutoff point in genotypes clustering through ANOVA model and based on GE interaction

Step	Grouped genotypes	Diss. index	v_1‡	v_2¶	F-test
1	5,18	6241.46	11	612	0.98
2	4 (5,18)	7755.27	33	612	1.22
3	6,11	7962.91	11	612	1.26
4	8,13	8745.82	11	612	1.38
5	1,15	9513.09	11	612	1.50
6	2,7	10685.45	11	612	1.65
7	3,16	10694.55	11	612	1.67
8	12 (1,15)	11679.09	22	612	1.84
9	(6,11)(8,13)	11890.42	33	612	1.87**
10	14 (3,16)	14252.47	22	612	2.25
11	10 (2,7)	15308.36	22	612	2.41
12	9,17	15776.55	11	612	2.49
13	(4,5,18)(3,14,16)	15849.64	55	612	2.50
14	(1,12,15)(6,8,11,13)	17795.27	66	612	2.81
15	(1,12,15,6,8,11,13)(4,5,18,3,14,16)	20509.76	132	612	3.23
16	(1,12,15,6, 8,11,13,4,5,18,3,14,16)(2,7,10)	22574.18	165	612	3.56
17	(9,17)(1,2,3,4,5,6,7,8,10,11,12,13,14,15,16,18)	24275.34	187	612	3.83

The visualization of this groping procedure (Fig. 4) indicated that there were nine different genotypic groups including; genotypes G9, G10, G14 and G17 as individual groups; G2 and G7; G3 and G16; G4, G5 and G18; G1, G15 and G12; G6, G8, G11 and G13 as the composite groups which had more than one genotype. Lin (1982) declared that the cluster analysis based on similarity of GE interaction is as an analytical tool for investigating multi-environment trials dataset, provides a logical base to compare the individuals within clusters by their average effect, and makes it possible to identify the structure of the GE interaction. The most prominent findings according to Fig. 4 are: genotypes 2 and 10 with the relatively high mean yield and low stability were groped as a same cluster while the other most stable genotypes or high yielding genotypes were cluster individually or mixed to each other. Similar to method 1, the genotypes clustering based on ANOVA and similarity of GE interaction showed huge variation among wheat genotypes and so 20 genotypes cluster into 11 groups. These mentioned results have many similarities to the cluster method involving both G and GE interaction effect based on ANOVA. Karimizadeh et al. (2006) reported similar results for clustering different maize hybrids through both methods 3 and 4, and declared that due to small proportion of G in comparison to GE interaction effect (about three times), many similarities are observed in allocation of these two clustering methods.

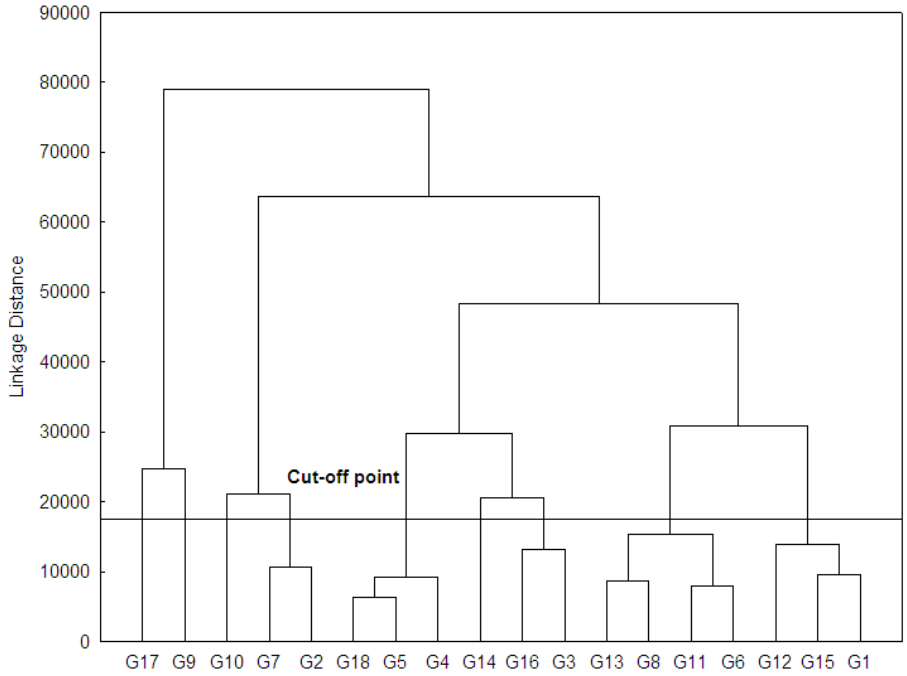

Fig. 4. Dendrogram of dissimilarity indices based on GE interaction of ANOVA model for 18 genotypes of lentil which were evaluated across 12 environments.

Discussion

Results of combined ANOVA indicated the significant effects of genotypes, environments and GE interaction but contributions of environments (70%) are not comparable with genotypes (7%) or GE interaction (23%). According to Yan et al. (2002) and Yan and Tinker (2005), environment explains 80% or higher of the total variation while only G and GE interaction are relevant to cultivar evaluation. The presence of GE interaction reduces the correlation between phenotype and genotype and selection progress. However, analysis of variance is uninformative in the explanation of GE interaction in analyzing lentil multi-environment trials dataset. Although the GE interaction is an important source of variation in any crop, geographic differentiation of landraces of lentil emphasizes the specific adaptation of this crop (Dehghani et al. 2008). We exploited the cluster analysis as the statistical method for evaluating lentil genotypes using the grain yield dataset in Iran. This paper demonstrated that different clustering procedures were very effective for studying the pattern of GE interaction and grouping lentil genotypes from multi-environment trials. As a general rule the effectiveness of linear regression model is insert verb here when 50% of the total sum of squares is accounted for by linear GE interaction (Crossa 1990, Hayward et al. 1993). However, the coefficients of determination for all studied genotypes were more than 50% and so method 1 and 2 could be regarded as the most favorable methods. Also, it seems that other strategies such as GE-based or G and GE-based procedures are useful for understanding and describing GE interactions. These procedures revealed that the GE interaction was an important source of lentil yield variation and its dendograms were effective enough for visualizing the response patterns of genotypes.

Joint linear regression model is used to explore GE interaction in multi-environment trials and it is major contribution was to quantify an environment effect using an environmental index (Finlay and Wilkinson, 1963). Lin and Thompson (1975) and Lin and Butler (1990) developed special types of cluster analyses to group genotypes for similarity of GE interaction plus G effect or only GE interaction via linear regression model. Brandle and Brule-Bable (1991), Lin and Lin (1994) and Karimizadeh et al. (2006) showed that this cluster analysis based on regression analysis has good ability for distinguish of similarities and dissimilarities. According to Lin et al. (1986), most of the linear regression models of multi-environment trials data analysis have Type II stability concept and a genotype is considered to be stable if its response to environment is parallel to the mean response of all genotypes in the trial and this type of stability beside Type III are very popular among plant breeders. Clustering of 18 lentil genotypes via regression

and based intercept and slope determined 2 distinct genotypic groups while clustering genotypes through regression and based slope revealed none distinct genotypic groups. Although, some authors (Karimizadeh et al. 2006) report that there are relatively similar results from the regression-based procedures, but we did not achieve similar results. This difference could be associated with the nature of the crop, environmental factors or diverse genetic background of lentil genotypes.

There are several ways to define the distance or similarity between two clusters which some of them have been proposed for classification of genotypes or environments (Abou-El-Fittouh et al. 1969, Mungomery et al. 1974, Lin 1982, Corsten and Denis 1990). Whatever method is selected, the question concerning the determination of cutoff point for the dendrograms is raised. All of the clustering methods enable breeders to explain the dataset into homogeneous subsets and to find out the GE interaction structure via the pattern grouping. The results of this investigation indicate that there are complex GE interactions in lentil genotypes in rain-fed conditions. The clustering results indicated that there are distinct genotypic groups for studied lentil genotypes from both G and GE interaction aspects. Finally, the results of this research indicate that cluster analysis may be a suitable tool of choosing stable genotypes with high yielding performance. Of further interest is the fact that the breeding genotypes in some groups was more stable with low mean yield or high yielding genotypes with low stability, indicating that the cluster analysis was successful in identifying variations among studied lentil genotypes. Such an outcome could be regularly applied in the future to delineate predictive, more rigorous recommendation strategies as well as to help define stability concepts for recommendations for durum wheat and other crops in the Middle East and other areas of the world. In conclusion, the following findings can be summarized from the present investigation: (1) genotypes G1 (1418.7 kg ha^{-1}), G5 (1324.4 kg ha^{-1}), G14 (1401.9 kg ha^{-1}) and G15 (1307.3 kg ha^{-1}) were found to be the most favorable genotype and they were thus recommended as good candidates for national release; (2) the clustering method 3 and 4 which are based on both G and GE interaction or only GE interaction were found to be useful in detecting the differences of the genotypes; and (3) the significant GE interaction and the changes in ranks of genotypes across environments suggest a breeding strategy of specifically adapted genotypes in homogeneously grouped environments.

Acknowledgements

Sincere gratitude goes to Iran's Research and Education Organization (AREO) and its Agricultural Research Stations for providing plant materials, experimental sites, and technical assistance.

References

Abou-El-Fittouh HA, Rawlings JO, Miller PA (1969). Classification of environments to control genotype by environment interactions with an application to cotton. Crop Sci. 9, 135–140.

Annicchiarico P (2002). Defining adaptation strategies and yield stability targets in breeding programmes. In: Kang, M.S. Ed. Quantitative Genetics, Genomics, and Plant Breeding, p. 365–383. Wallingford, UK, CABI.

Brandle JE, Brule-Bable AL (1991). An integrated approach to oilseed rape cultivar selection using phenotypic stability. Theor. Appl. Genet. 81, 679-684.

Calinski T, Corsten LCA (1985). Clustering means in ANOVA by simultaneously testing, Biometrics 41, 39–4

Cooper M, Chapman SC, Podlich DW, Hammer GL (2002). The GP problem: Quantifying gene-to-phenotype relationships. Silico Biol. 2, 151–164.

Corsten LCA, Denis JB (1990). Structuring interaction in two way tables by clustering. Biometrics 46, 207–215.

Crossa J (1990). Statistical analyses of multilocation trials. Adv. Agron. 44, 55–85.

Dehghani H, Sabaghpour SH, Sabaghnia N (2008). Genotype × environment interaction for grain yield of some lentil genotypes and relationship among univariate stability statistics. Spanish J. Agric. Res. 6, 385–394.

Finlay KW, Wilkinson GN (1963). The analysis of adaptation in a plant breeding programme. Aust. J. Agric. Res. 14, 742–754.

Flores F, Moreno MT, Cubero JI (1998) A comparison of univariate and multivariate methods to analyze environments. Field Crops Res. 56, 271–286.

Francis TR, Kannenberg LW (1978). Yield stability studies in short-season maize: I. A descriptive method for grouping genotypes. Can. J. Plant Sci. 58, 1029–1034.

Gauch, H. G. 2006. Statistical analysis of yield trials by AMMI and GGE. Crop Sci. 46, 1488–1500.
Hayward MD, Bosemark NO Romagosa I (1993). Plant Breeding: Principles and Prospects. pp.550. Chapman and Hall, London.
Hopkins AA, Vogel KP, Moore KJ, Johnson KD Carlson IT (1995). Genotype effects and genotype by environment interactions for traits of elite switchgrass populations. Crop Sci. 35, 125–132.
Kang MS (1998). Using genotype-by-environment interaction for crop cultivar development. Adv. Agron. 62, 199–252.
Karimizadeh R, Dehghani H, Dehghanpour Z (2006). Using Cluster Analysis for Stability of Maize Hybrids. J. Crop Produc. Process. 10, 337–348.
Karimizadeh R, Mohammadi M. Sabaghnia N. Shefazadeh MK Pouralhossini J (2012). Univariate stability analysis methods for determining genotype × environment interaction of durum wheat grain yield. African J. Biotech. 11, 2563–2573.
Lin CS (1982). Grouping genotypes by a cluster method directly related to genotype–environment interaction mean square. Theor. Appl. Genet. 62, 277–280.
Lin CS, Butler G (1990) Cluster analyses for analyzing two-way classification data. Agron. J. 82, 344–348.
Lin CY, Lin CS (1994) Investigation of genotype-environment interaction by cluster analysis in animal experiments. Can. J. Animal Sci. 74, 607–612.
Lin CS, Thompson B (1975). An empirical method of grouping genotypes based on a linear function of the genotype environment interaction. Heredity 34, 255–263.
Lin CS, Binns MR, Lefkovitch LP (1986). Stability analysis: where do we stand? Crop Sci. 26, 894–900.
Lin CS, Butler G, Hall I, Nault C (1992). Program for investigating genotype-environment interaction. Agron. J. 84, 121–124.
Mandel J (1961). Non-additivity in two-way analysis of variance. J. Am. Statist. Ass. 56, 878–888.
Mohebodini M, Dehghani H, Sabaghpour SH (2006). Stability of performance in lentil (*Lens culinaris* Medik) genotypes in Iran. Euphytica 149, 343–352.
Mungomery VE, Shorter R, Byth DE (1974). Genotype environment interactions and environmental adaptation. 1 Pattern analysis - application to soya bean populations. Aust. J. Agric. Res. 25, 59–72.
Pinthus JM (1973). Estimate of genotype value: a proposed method. Euphytica 22, 121–123.
Robert N (1997). Structuring genotype × environment interaction for quality traits in bread wheat, in two multi-location series of trials. Euphytica 97, 53–66.
Sabaghnia N, Dehghani H, Sabaghpour SH (2008a). Graphic analysis of genotype by environment interaction for lentil yield in Iran. Agron. J. 100, 760-764.
Sabaghnia N, Sabaghpour SH Dehghani H, (2008b). The use of an AMMI model and its parameters to analyse yield stability in multienvironment trials. J. Agric. Sci. 146, 571-581.
Sabaghpour SH, Safikhni M, Sarker A, Ghaffri A, Ketata H (2004). Present status and future prospects of lentil cultivation in Iran. Proc 5th European Conference on Grain Legumes, 7-11 June, Dijon, France.
Sokal R, Michener C (1958). A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. 38, 1409–1438.
Yan W, Hunt LA, Johnson P, Stewart G, Lu X (2002) On-farm strip trials vs. replicated performance trials for cultivar evaluation. Crop Sci 42, 385–392.
Yan W, Kang MS (2003). GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC Press. Boca Raton, FL.
Yan W, Tinker NA (2005). An integrated system of biplot analysis for displaying, interpreting, and exploring genotype by-environment interactions. Crop Sci. 45, 1004–1016.
Yue GL, Roozeboom KL, Schapaugh Jr. WT, Liang GH (1997) Evaluation of soybean cultivars using parametric and nonparametric stability estimates. Plant Breed. 116, 271–275.