

FENOTIPSKA KARAKTERIZACIJA I MULTIVARIJACIONE ANALIZE

*V. Babić, M. Babić**

Izvod: Masovna upotreba PC računara i u agronomskim istraživanjima u novije vreme omogućila je primenu metoda multivarijacione analize koje omogućavaju sagledavanje simultane međizavisnosti između tri ili više nezavisno promenljivih. Ukoliko u svojoj bazi podataka imamo fenotipsku karakterizaciju selekcionog materijala po principima UPOV-og deskriptora, bilo da smo se odlučili za praćenje manjeg broja osobina ili za kompletan set osobina, jednostavnim skriningom željenih osobina uz primenu neke od metoda multivarijacionih analiza možemo za par minuta dobiti važne informacije.

Za anлізу je uzet set od 58 linija koje su fenotipski opisane u jednoj godini i jednom ponavljanju po principima UPOV-og deskriptora. Posmatrajući tri ocene: ugaod stabla i prvog lista iznad klipa, položaj lista i širina lista, cilj istraživanja je bio da se utvrdi da li ispitivane linije prave prirodne, homogene grupe na osnovu ove tri karakteristike.

Primenjena je hijerarhijska klaster analiza Ward-ov metod, a verodostojnost rezultata je testirana diskriminacionom analizom. Na osnovu unapred definisanih grupa klaster analizom, diskriminacionom analizom su formirane linearne kombinacije nezavisno promenljivih tako da je greška pogrešne klasifikacije minimalna. Prve tri diskriminacione funkcije obuhvatile su 100% varijanse to znači da je najmanja greška pogrešne klasifikacije onda kada je dati set podataka podeljen u 4 grupe.

Formiranje jedinstvenih baza podataka, koje postaju neiscrpan izvor korisnih informacija za stručnjake koji se bave planskim i usmerenim oplemenjivanjem, diktira potrebu iznalaženja efikasnog načina za korišćenje ogromnog obima informacija u čemu koršćena metodologija u datom radu može biti od izuzetnog značaja.

Ključne reči: baze podataka, fenotipska karakterizacija, diskriminaciona analiza

Uvod

Bavljenje selekcionim radom podrazumeva praćenje i ocenjivanje velikog broja svojstava linija i hibrida, kako u procesu selekcije, tako i kroz proces semenarstva i proizvodnje. UPOV-a konvencija radi zaštite prava oplemenjivača predviđa praćenje velikog broja osobina varijeteta (sorti, linija, hibrida). Kako se očekuje da će i naša zemlja postati potpisnica UPOV-e konvencije javila se potreba da se brojna zapažanja, koja se i inače redovno sprovode na našem elitnom materijalu, standardizuju u tom pravcu. Stoga je ustanovljen sistem gde se u momentu kada određeni hibrid ulazi u proces priznavanja u Sortnu

* Mr Vojka Babić, dr Milosav Babić, Institut za kukruz „Zemun Polje“, Zemun.

komisiju, kako sam hibrid, tako i njegove komponente opisuju po standardima testiranja koje je ustanovio UPOV (UPOV publication 2000). Pored seta od tridesetak komercijalnih hibrida i njihovih roditeljskih linija, od 2002. godine pa do danas, urađena je fenotipska karakterizacija i oko 230 hibrida i njihovih komponenti u trajanju od tri godine.

Različiti metodi mogu biti upotrebljeni za ocenu karakteristika u DUS testu u zavisnosti od tipa varijeteta i osobina koje se ocenjuju. Da bi jedno svojstvo moglo da se koristi za DUS testiranje mora da zadovolji određene uslove (TG/1/3). Mora da bude rezultat genotipa, da je u dovoljnoj meri konzistentno i ponovljivo u određenoj sredini, da se može jasno definisati i prepoznati, da ispoljava dovoljnu varijaciju između varijeteta da bi bilo moguće ustanoviti različite stepene ekspresije i time i različitost između varijeteta i da omogućava jasne zahteve u pogledu stabilnosti u smislu da obezbeđuje konzistentne i ponovljive rezultate posle svakog novog ciklusa umnožavanja (UPOV/DATA/BEI/04/8, TG/2/6).

U cilju pravilne upotrebe određenog svojstva u testiranju potrebno je razumeti različite načine njihovog ispoljavanja. Različiti su tipovi ekspresije svojstava i njihova primena u DUS testiranju. U osnovi postoje tri tipa osobina: kvalitativne, kvantitativne i pseudo kvalitativne. Različiti nivoi ekspresije svojstava su podeljeni u stadijume i označeni određenim brojem-ocenom. Testiranje takođe podrazumeva upotrebu odgovarajućih varijeteta određene biljne vrste kao repera za nivo ekspresije datog svojstva.

Kvalitativna svojstva su ona koja se ispoljavaju u diskontinuiranim stadijumima (pol biljke). Kod ovakvih osobina stadijumi su sami po sebi jasni i nezavisni. Po pravilu, ove karakteristike nisu pod uticajem sredine.

Kvantitativne karakteristike su one čija se ekspresija proteže kontinuirano od jednog do drugog ekstrema. Ekspresija se rangira u određeni broj stadijuma u zavisnosti od osobina i uz to ide određena ocena. Na primer visina biljke: 1-veoma nisko, 3-nisko, 5-srednje, 7-visoko, 9-veoma visoko. Stadijum ekspresije bi svakako trebao da ima puno značenje za DUS procenu mada Uputstva za testiranje ne preciziraju razliku potrebnu za različitost.

Postoje još i pseudo kvalitativna svojstva, to su ona koja se u krajnjoj meri ispoljavaju kontinuirano ali variraju od jednog do drugog oblika (oblik lista: ovalan- zašiljen, oblik klipa: koničan-cilindričan) i ne mogu da budu definisana sa dve krajnje pozicije. Tako su ocene klipa: koničan-1, konično –cilindričan-2 i cilindričan-3.

Međutim ovde neće biti reči o DUS testu samom po sebi već o tome kako se korisno mogu upotrebiti ocene pojedinih osobina genotipova za unapređenje procesa selekcije. Ocene fenotipskih karakteristika različitih genotipova su u osnovi jako jednostavne i do njih se dolazi vizuelnom ocenom bez nekog posebnog merenja bar kada se radi o kukuru. Za razliku od egzaktnih merenja, koja se sprovode u istraživačke svrhe, kao što su merenje dužine i širine lista ili merenje dužine i prečnika klipa, ovakve ocene ne zahtevaju obimna merenja na većem broju biljaka, većem broju ponavljanja pa i u trajanju od više sezona. Znači ne zahtevaju angažovanje veće količine radne snage već jedan stručnjak za određenu biljnu vrstu može da oceni veliki broj genotipova u toku jedne sezone. Naravno, fenotipska karakterizacija namenjena selekciji ne mora da sadrži sve preporučene osobine iz Uputstva za DUS testiranje već samo one koje su od interesa za odabir superiornog materijala. Stvaranje hibrida za uslove intenzivne poljoprivrede, gušće setve i intenzivnu primenu navodnjavanja i đubrenja, podrazumeva korišćenje početnog materijala koji se dobro ponaša u takvim, intenzivnim uslovima gajenja. Stvaranje nove početne populacije koja će se sastojati od genotipova, koji imaju uzan habitus, odnosno mali ugao stabla

i lista, uspravan i širok list podrazumeva izbor odgovarajućih linija iz ogromnog broja materijala kako iz banke gena tako i iz materijala koji se nalazi u tekućim komercijalnim programima oplemenjivanja. Predstoji dugotrajan i mukotrpan proces pronalaženja takvih linija. Sa povećanjem broja traženih parametara proces iznalaženja odgovarajućeg materijala se komplikuje. Postoji jedna količina materijala sa kojim se često radi i tu nema problema. Dobar oplemenjivač poznaje svoj materijal i moći će da odabere nekoliko linija. A šta sa brojnim materijalom koji nije u frekventnoj upotrebi?

Materijal i metod rada

Masovna upotreba PC računara i u agronomskim istraživanjima u novije vreme omogućila je jednostavnu primenu metoda multivarijacione analize koje omogućavaju sagledavanje simultane međuzavisnosti između tri ili više nezavisno promenljivih (Ruiz, 2001). Ukoliko se u bazi podataka nalazi fenotipska karakterizacija selekcionog materijala po principima UPOV-og deskriptora, bilo da je u pitanju praćenje manjeg broja osobina ili kompletnog seta osobina, jednostavnim skriningom željenih osobina uz primenu neke od metoda multivarijacionih analiza možemo za par minuta dobiti važne informacije. U ovom slučaju radilo se o izdvajanju materijala uskog habitusa i uspravnog i širokog lista.

Za analizu je uzet set od 58 linija koje su fenotipski opisane u jednoj godini i jednom ponavljanju po principima UPOV-og deskriptora. Posmatrajući tri ocene: ugao stabla i prvog lista iznad klipa, položaj lista i širina lista cilj je bio da se utvrdi da li ispitivane linije prave prirodne, homogene grupe na osnovu ove tri karakteristike. Ocene za ove osobine kod kukuruza kreću se od 1 do 9.

Primenjena je hijerarhijska klaster analiza Ward-ov metod, a kao mera sličnosti/različitosti upotrebljen je squared eucliden distance. Ukoliko se barata sa bazama podataka koje sadrže 1000 i više jedinica onda bi k-sredina klaster analiza ili dvostepena klaster analiza bile u stanju da efikasno savladaju takav set podataka.

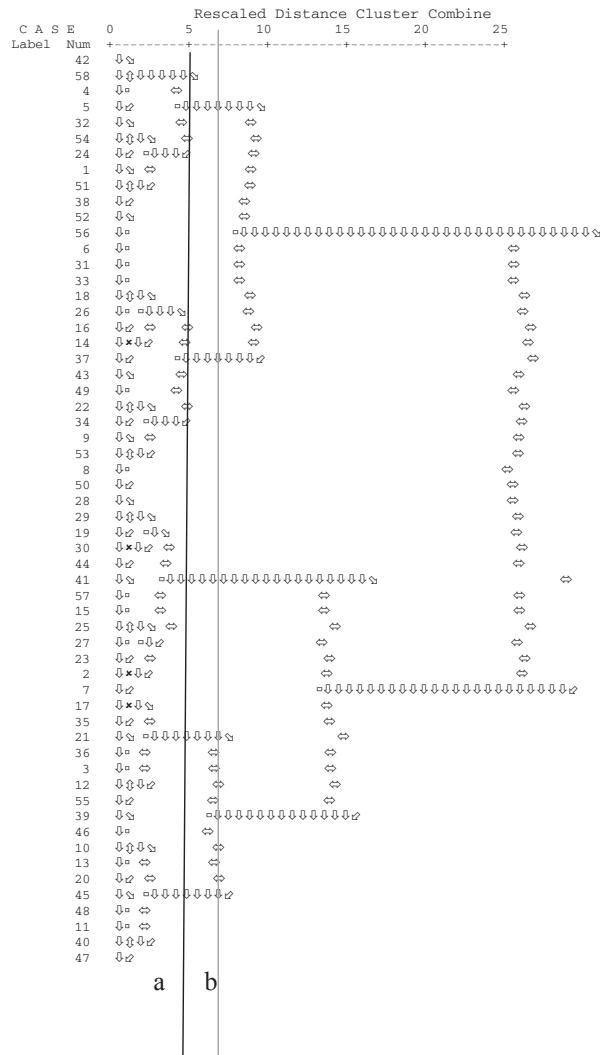
Adekvatnost primenjene klaster analize je testirana diskriminacionom analizom (Kovacic, 1998, Hastie et al., 1994). Osnovni cilj diskriminacione analize je formiranje linearnih kombinacija nezavisno promenljivih kojima će se diskriminacija između unapred definisanih grupa izvršiti tako da greška pogrešne klasifikacije bude minimizirana (Milligen and Cooper, 1985).

Rezultati istraživanja i diskusija

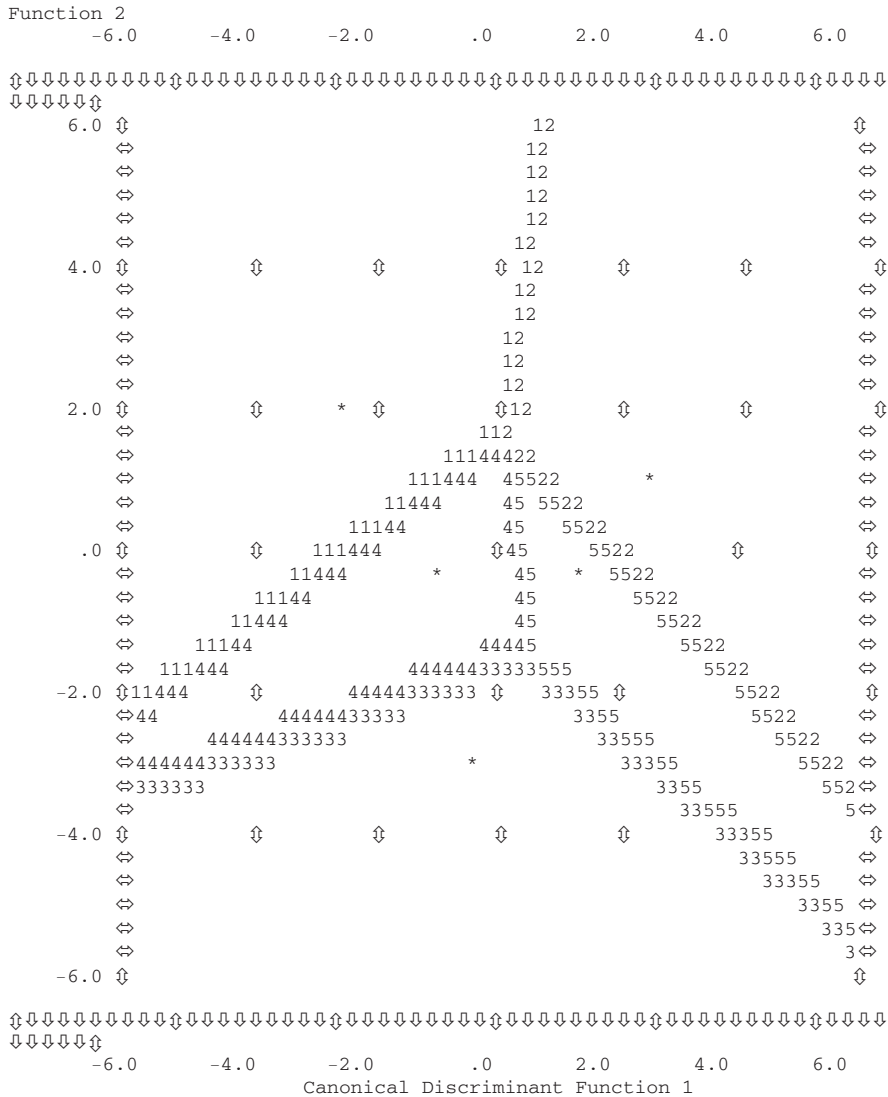
Na dendogramu slike 1 prikazani su rezultati klaster analize. Ispitivanih 58 linija uzimajući u obzir osobine: ugao stabla i prvog lista iznad klipa, položaj lista iznad klipa i širina lista u krajnjoj instanci se grupišu u dva klastera. U prvom klasteru nalaze se linije sa širokim i listom srednje širine veoma malog do malog ugla stabla i lista i veoma uspravnog do uspravnog lista. U drugom klasteru nalaze se linije sa uzanim listom ili sa listom srednje širine i uglom stabla i lista od veoma malog do srednjeg. Po logici stvari cilj je da se definišu klasteri tako da se u svakom klasteru nalazi veoma karakteristična, homogena grupa linija sa jedne strane, a da sa druge strane postoji dovoljna distanca između klastera za klasifikaciju.

U prvom pokušaju izvršeno je sečenje dendograma na odstojanju 5 i time definisano 5 klastera. Prvom klasteru pripada 10 jedinica i to su linije veoma malog ugla između stabla i lista sa veoma uspravnim do blago povijenim listom i veoma šorokom do širokom liskom. Drugom klasteru pripada 13 linija i to su linije sa srednjim uglom između stabla i lista srednje do jako povijenim listom i liskom srednje širine. U trećem klasteru nalazi se 7 linija koje imaju veoma mali do mali ugao stabla i lista blago povijen list i malu do veoma malu širinu liske. U četvrtom, najbrojnijem klasteru nalazi se 18 linija koje imaju srednju širinu liske, mali do veoma mali ugao stabla i lista i blago povijen list. Na kraju, u petom klasteru, nalazi se 10 linija koje imaju malu do srednju širinu lista, mali do veoma mali ugao stabla i lista i blago do srednje povijen list (slika 1 linija sečenja “a”).

Sl. 1. Dendrogram klaster analize za osobine ugao, položaj i širina lista
Dendrogram for a cluster analysis for the traits of leaf angle, position and width



Sl. 2. Teritorijalna mapa 5 klastera na osnovu 4 diskriminacione funkcije
Territorial map of 5 clusters based on 4 discriminant functions



Zatim je dobijena klaster analiza sa pet grupa testirana diskriminacionom analizom. Iako model prikazuje teritorijalnu mapu pet klastera tabela Eigenvalues pokazuje da su dovoljne tri kanonične diskriminacione funkcije da objasne varijabilnost datog seta podataka (sl. 2). Prva diskriminaciona funkcija obuhvata 69% varijanse, druga 23.4% i treća 7.6% (tab. 1).

Tab. 1. Eigenvrednosti diskriminacionih funkcija
Eigenvalues of discriminant functions

Funkcija	Eigenvrednost	% od varijanse	Kumulativno %	Kanonička korelacija
1	3.297 ^a	69.0	69.0	0.876
2	1.119 ^a	23.4	92.4	0.727
3	0.362 ^a	7.6	100.0	0.516

^a U analizi su upotrebljene prve tri kanoničke diskriminacione funkcije

Na osnovu unapred definisanih grupa diskriminacionom analizom su formirane linearne kombinacije nezavisno promenljivih tako da je greška pogrešne klasifikacije minimalna (Fonseca et al. 2004). Međutim kako prve tri diskriminacione funkcije obuhvataju 100% varijanse to znači da je najmanja greška pogrešne klasifikacije onda kada je dati set podataka podeljen u 4 grupe.

Prva promenljiva ili položaj prvog lista iznad klipa je u visio koj korelaciji sa prvom diskriminacionom funkcijom (0.81). Širina lista je u pozitivnoj korelaciji sa drugom diskriminacionom funkcijom, a ugao stabla i prvog lista iznad klipa u visokoj negativnoj korelaciji sa trećom diskriminacionom funkcijom (tab. 2).

Tab. 2. Struktura matrice prve tri diskriminacione funkcije u slučaju pretpostavke 5 grupa
Matrix structure of the first three discriminant functions in case of assumption of five groups

Osobina	Funkcija		
	1	2	3
Položaj lista	0.808*	0.183	0.560
Širina lista	-0.458	0.889*	-0.003
Ugao stabla i lista	0.624	0.338	-0.705*

* Najveća apsolutna korelacija između svake varijable i bilo koje diskriminacione funkcije

Može se zaključiti da inicijalna podela u pet grupa nije najadekvatnija. Sada se testira pretpostavka podele u 4 grupe. Seče se dendrogram na odstojanju oko šest (sl. 1, linija sečenja „b“). I prethodna analiza je ukazala na statističku značajnost samo prve tri diskriminacione funkcije. Isto se može zaključiti i na osnovu F testa diskriminacionih funkcija gde su p vrednosti za sve tri diskriminacione funkcije ispod 5% (tab. 3).

Tab. 3. F test značajnosti diskriminacionih funkcija
F test of significance for discriminant functions

Test funkcija	Wilksova Lambda	Chi-square	Stepeni slobode	Signifikantnost
1 prema 3	0.081	134.721	9	0.000
2 prema 3	0.346	56.721	4	0.000
3	0.734	16.543	1	0.000

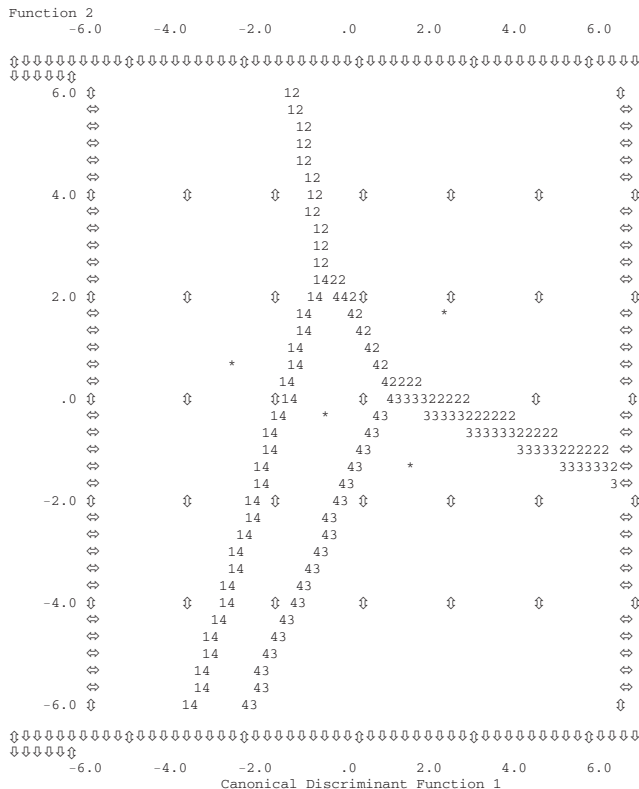
Naravno sve tri varijable ili nezavisno promenljive su značajne za diskriminaciju i model ih uzima u analizu. Minimalna vrednost parcijalnog F nezavisnih varijabli za ulazak u analizu ili bolje rečeno minimalna vrednost parcijalnog F potrebna za diskriminaciju iznosi 3.84. Sve tri promenljive imaju F vrednosti daleko iznad toga (tab. 4).

Tab. 4. F vrednosti nezavisno promenljivih u slučaju podele na 4 klastera
F values of independent variables in case of dividing into four clusters

Korak	tolerancija	F za povlačenje	Wilksova lambda
1 širina lista	1.00	35.696	
2 širina lista	0.98	35.361	0.388
ugao stabla i lista	0.98	28.118	0.335
3 širina lista	0.98	28.675	0.214
ugao stabla i lista	0.878	15.062	0.151
položaj lista	0.895	10.481	0.129

Centri klastera na teritorijalnoj mapi se sada pomeraju tako što se peti klaster pridružuje trećem klaseru (sl. 3).

Sl. 3. Teritorijalna mapa 4 klaster na osnovu 3 diskriminacione funkcije
Territorial map of 4 clusters based on 3 discriminant functions



Zaključak

U raznim oblastima istrživanja veoma često pokušavamo prirodu nekog objekta ili pojave da razumemo ili opišemo istovremenim merenjem većeg broja nezavisno promenljivih. Nažalost, takođe veoma često, nismo u stanju da, kompleksnu prirodu objekata ili pojava posmatranja, sagledamo u potpunosti. Sam prinos, koji predstavlja naš najčešći predmet interesa, po svojoj prirodi predstavlja višedimenzionu pojavu koja obuhvata više različitih karakteristika i na koju utiče veći broj faktora ili simultano međuzavisnih promenljivih. Masovna upotreba PC računara i u agronomskim istraživanjima u novije vreme omogućila je jednostavnu primenu metoda multivarijacione analize koje omogućavaju sagledavanje simultane međuzavisnosti između tri ili više nezavisno promenljivih. Od zadatog cilja istraživanja, problema koji treba razjasniti, kao i tipa podataka, zavisi izbor odgovarajuće metode (Amsalu, 1999). Takođe izbor metoda zavisi od toga da li su istraživanja orijentisana ka međuzavisnosti promenljivih ili im je osnovni zadatak ispitivanje međuzavisnosti objekata. U prvom slučaju, kada se ispituje međuzavisnost promenljivih posmatraju se kolone matrica podataka i tada su u osnovi ovih metoda kovarijacione i korelacione matrice. U drugom slučaju, u cilju poređenja dva objekta (hibrida, linija ili sredina), posmatramo redove matrica odnosno definišemo različite mere bliskosti. Osnov ovih metoda predstavljaju matrice odstojanja između objekata.

Formiranje jedinstvenih baza podataka koje postaju neiscrpan izvor korisnih informacija za stručnjake koji se bave planskim i usmerenim oplemenjivanjem, diktira potrebu iznalaženja efikasnog načina za korišćenje ogromnog obima informacija u čemu korišćenje metodologija u datom radu može biti od izuzetnog značaja.

Literatura

- 1.
2. *Amsalu A., Endashaw B. (1999):* Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea. *Genetic Resources and Crop Evolution* 46, 3.
3. *Fonsec, A. F. A. da, Sedyama T., Cruz C. D., Sakiyama N. S., Ferrão R. G., Ferrão M. A. G., Bragança S. M. (2004):* Discriminant analysis for the classification and clustering of robusta coffee genotypes. *Crop Breeding and Applied Biotechnology* 4, 3: 285-289.
4. *Hastie T. et al. (1994):* Flexible discriminant analysis by optimal scoring. *Journal of American statistical Association*, 89: 1255-1270.
5. *Kovacic Z. (1998):* Multivarijaciona analiza. Ekonomski fakultet-Beograd, pp 131.
6. *Kovacic Z. (1998):* Multivarijaciona analiza. Ekonomski fakultet-Beograd, pp 2-5.
7. *Milligan G.W. and Cooper M.C. (1985):* An Examination of Procedures for Determining the Number of Clusters in Data set. *Psychometrika*, 50c. 159-179.
8. *Ruiz J., de Galarreta I., Alvarez A. (2001):* Morphological classification of maize landraces from northern Spain. *Genetic Resources and Crop Evolution* 48, 4.
9. *UPOV, (1999):* Guidelines for the conduct of tests for distinctness, uniformity and stability for Maize . TG/2/6+Corr.
10. *UPOV, (2000):* What it is, what it does. UPOV Publication No.437(E)

11. *UPOV, (2002)*: General introduction to the examination of Distinctness, Uniformity and Stability and the development of Harmonized Descriptions of new Variete of Plants. TG/1/3
12. *UPOV, (2004)*: Designing the DUS tests. UPOV/DATA/BEI/04/4
13. *UPOV, (2004)*: Examining uniformity. UPOV/DATA/BEI/04/8

PHENOTYPIC CHARACTERISATION AND MULTIVARIATE ANALYSES

*V. Babić, M. Babić**

Summary

The mass utilisation of personal computers in agricultural studies in recent times has provided the application of multivariate analysis methods that facilitated recognition of simultaneous interdependence among three or more independent variables. If our data base encompasses records on the phenotypic characterisation of the breeding material according to principles of the UPOV descriptor, a simple screening of desired traits by the application of some of multivariate analysis methods, a very important information could be gained in just a couple of minutes regardless of monitoring of a few or a complete set of traits.

A set of 58 inbreeds, phenotypically described in one year and one replication according to UPOV descriptor, was analysed. According to the three estimates (the angle between the blade and stem on leaf just above upper ear, leaf attitude of blade and leaf width), the aim of the study was to determine whether the observed lines were actually natural, homogenous groups on the basis of these three properties.

The hierarchical cluster analysis, Ward's method, was applied, and credibility of results was tested by the discrimination analysis. Based on groups predefined by the cluster analysis, linear combinations of independent variables were formed by the discrimination analysis, hence the error of an incorrect classification was minimal. The first three discrimination functions encompassed 100% of a variance, meaning that the smallest error of the incorrect classification was if the given data set was divided into four groups.

The formation of the unique data bases, that become an inexhaustible source of information for experts working on planned and directed breeding, dictates a need to ascertain the efficient way for the utilisation of the enormous scope of information wherein the applied methodology in such a work can be of an exceptional importance.

Key words: data bases, phenotypic characterisation, discrimination analysis

* Vojka Babić, M.Sc., Milosav Babić, Ph. D., Maize Research Institute, Zemun Polje, Slobodana Bajićića 1, 11185 Belgrade, Serbia. Corresponding author (fax +381 11 3756 707; e-mail: vbabic@mrizp.co.yu)