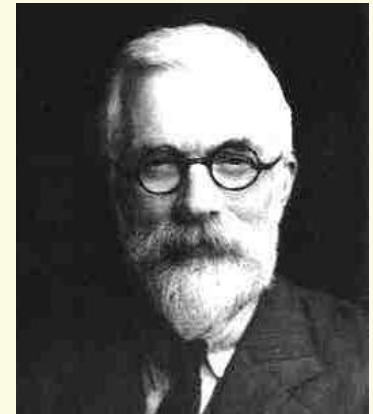


UNIVARIJATNA ANALIZA VARIJANCE

Univarijatna analiza varijance ili *ANOVA* je statistički postupak koji je osmislio Ronald A. Fisher, a kojim se utvrđuje statistička značajnost razlika između aritmetičkih sredina dvije ili više grupa entiteta u jednoj varijabli.



Ronald A. Fisher
(1890. – 1962.)

UNIVARIJATNA ANALIZA

VARIJANCE

Pri tome je moguće postaviti sljedeću alternativnu (H1), odnosno nultu (H0) hipotezu:

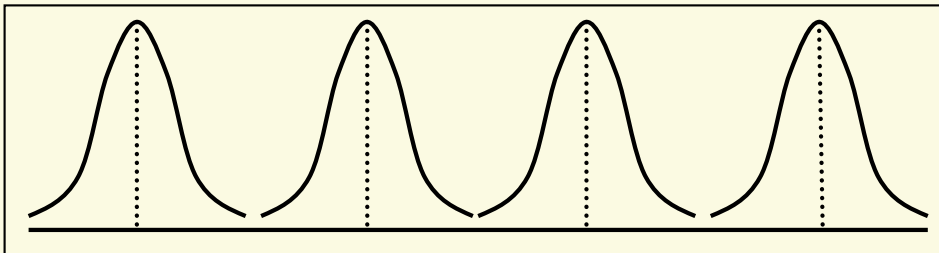
H1: $\bar{x}_1 \neq \bar{x}_2 \neq, \dots, \neq \bar{x}_k$ - Razlike između aritmetičkih sredina analiziranih grupa statistički su značajne uz pogrešku p .

H0: $\bar{x}_1 = \bar{x}_2 =, \dots, = \bar{x}_k$ - Uz pogrešku p ne možemo tvrditi da su razlike između aritmetičkih sredina analiziranih grupa statistički značajne.

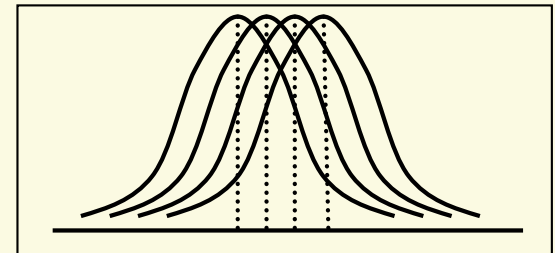
UNIVARIJATNA ANALIZA VARIJANCE

Ako je *varijabilitet između grupa* statistički značajno veći od *varijabiliteta unutar grupa* onda se grupe međusobno statistički značajno razlikuju.

Primjer 1



Primjer 2



(Prikaz aritmetičkih sredina četiriju grupa entiteta i odgovarajućih distribucija rezultata u slučaju kada je varijabilitet između grupa veći od varijabiliteta unutar grupa (*Primjer 1*) i u slučaju kada je varijabilitet unutar grupa veći od varijabiliteta između grupa (*Primjer 2*).

UNIVARIJATNA ANALIZA

VARIJANCE

Varijanca između grupa (σ^2_{ig}) je mjera varijabiliteta između grupa, a izračunava se formulom:

$$\sigma^2_{ig} = \frac{\sum_{g=1}^k n_g (\bar{x}_g - \bar{x}_t)^2}{k-1}$$

gdje je

➔ n_g - broj entiteta u grupi g gdje je $g=1,2,\dots,k$

➔ k - broj grupa

➔ \bar{x}_g - aritmetička sredina grupe g gdje je $g=1,2,\dots,k$

➔ \bar{x}_t - aritmetička sredina svih rezultata, odnosno zajednička aritmetička sredina

UNIVARIJATNA ANALIZA

VARIJANCE

Varijanca unutar grupa (σ^2_{ug}) je mjera varijabiliteta unutar grupa, a izračunava se formulom:

$$\sigma^2_{ug} = \frac{\sum_{g=1}^k \sum_{i=1}^{n_g} (x_{gi} - \bar{x}_g)^2}{n - k}$$

gdje je

- ➔ n_g - broj entiteta grupe g gdje je $g=1,2,\dots,k$
- ➔ k - broj grupa
- ➔ x_{gi} - rezultat entiteta i pripadnika grupe g gdje je $i=1,2,\dots,n_g$
- ➔ \bar{x}_g - aritmetička sredina grupe g gdje je $g=1,2,\dots,k$
- ➔ n - ukupan broj entiteta

UNIVARIJATNA ANALIZA VARIJANCE

Statistička značajnost razlika između grupa testira se putem *F-distribucije*. *F-vrijednost* se izračunava kao omjer varijance između grupa (σ^2_{ig}) i varijance unutar grupa (σ^2_{ug}).

$$F = \frac{\sigma^2_{ig}}{\sigma^2_{ug}}$$

Ako je izračunata F-vrijednost veća od *kritične F-vrijednosti* onda su razlike između grupa statistički značajne. Kritična F-vrijednost određuje se na temelju broja stupnjeva slobode $df_1=k-1$ i $df_2=n-k$ i pogreške statističkog zaključka p .

STATISTICA 7

Univarijatna analiza varijance

ANOVA se izvodi slijedom koraka: padajući izbornik *Statistics* → *Basic Statistics/Tables* → *Breakdown & one-way ANOVA*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti selektorsku varijablu (*Grouping variable*) i jednu ili više zavisnih varijabli (*Dependent variables*). Nakon izbora varijabli potrebno je odabrati opciju *ANOVA & tests* → *Analysis of Variance*.

Zadatak - U datoteci *Pejcic-318.sta* uz pogrešku $p=0,01$ testirajte da li se aritmetičke sredine učenika 1., 2., 3. i 4. razreda u varijabli *MPOL* statistički značajno razlikuju!

MULTIVARIJATNA ANALIZA VARIJANCE

Multivarijatna analiza varijance ili *MANOVA* je statistički postupak kojim se utvrđuje statistička značajnost razlika između centroida dviju ili više grupa ispitanika.

Pri tome je moguće postaviti sljedeću alternativnu (H1) odnosno nultu (H0) hipotezu:

H1: $m_1 \neq m_2 \neq \dots \neq m_k$ - Razlike između centroida analiziranih grupa statistički su značajne uz pogrešku p .

H0: $m_1 = m_2 = \dots = m_k$ - Uz pogrešku p ne možemo tvrditi da su razlike između centroida analiziranih grupa statistički značajne.

MULTIVARIJATNA ANALIZA

VARIJANCE

Centroid je vektor aritmetičkih sredina dviju ili više varijabli.

Primjer: Dvije grupe ispitanika (nogometaši i košarkaši) izmjerene su s tri testa motorike i dobiveni su sljedeći rezultati:

GRUPA	SDM	BML	T20M
N	220	570	3,4
	265	595	3,1
	240	615	3,1
\bar{x}	241,67	593,33	3,2

Centroid grupe nogometaša je

$$m_n = \begin{pmatrix} 241,67 \\ 593,33 \\ 3,2 \end{pmatrix}$$

GRUPA	SDM	BML	T20M
K	280	690	3,7
	310	720	3,5
	305	735	3,3
\bar{x}	298,33	715	3,5

Centroid grupe košarkaša je

$$m_k = \begin{pmatrix} 298,33 \\ 715 \\ 3,5 \end{pmatrix}$$

MULTIVARIJATNA ANALIZA VARIJANCE

Neka su u matricama B_g (gdje je $g=1,2,\dots,k$, a k broj grupa entiteta) podaci skupova $E_g = \{e_{gi} ; i = 1,\dots,n_g\}$ entiteta opisanih skupom $V = \{v_j ; j = 1,\dots,m\}$ varijabli, a u matrici B podaci svih entiteta odnosno skupa $E = \{e_i ; i = 1,\dots,n\}$ entiteta opisanih skupom $V = \{v_j ; j = 1,\dots,m\}$ varijabli.

MULTIVARIJATNA ANALIZA VARIJANCE

Operacijom

$$m = B^T \mathbf{1} n^{-1}$$

gdje je

→ $\mathbf{1}$ - vektor stupca s n jedinica

izračuna se zajednički centroid. Operacijom

$$D = B - \mathbf{1} m^T$$

izračuna se matrica odstupanja rezultata entiteta od zajedničkog centroida. Operacijom

$$T = D^T D$$

izračuna se *matrica suma kvadrata i krosprodukata za total.*

MULTIVARIJATNA ANALIZA VARIJANCE

Operacijom

$$\mathbf{m}_g = \mathbf{B}_g^T \mathbf{1}_g n_g^{-1}$$

izračuna se centroid grupe g . Operacijom

$$\mathbf{D}_g = \mathbf{B}_g - \mathbf{1}_g \mathbf{m}_g^T$$

izračuna se matrica odstupanja rezultata entiteta grupe g od centroida grupe g . *Matrica suma kvadrata i krosprodukata unutar grupa* (W) izračuna se operacijom

$$W = \sum_{g=1}^k W_g \quad \text{gdje je} \quad W_g = \mathbf{D}_g^T \mathbf{D}_g$$

MULTIVARIJATNA ANALIZA

VARIJANCE

Matrica suma kvadrata i krosprodukata između grupa može se izračunati operacijom

$$A = T - W$$

jer vrijedi da je

$$T = W + A$$

MULTIVARIJATNA ANALIZA

VARIJANCE

Statistička značajnost razlika između centroida grupa testira se putem *Wilksove lambda*.

$$\lambda_w = \frac{\det(W)}{\det(T)}$$

Wilksova lambda se kreće u intervalu od 0 do 1, a što je njena vrijednost manja to je veća vjerojatnost da je razlika između centroida grupa statistički značajna.

MULTIVARIJATNA ANALIZA

VARIJANCE

Pošto Wilksova lambda prati vrlo složenu distribuciju, u svrhu testiranja statističke značajnosti razlika može se izračunati i *aproksimativna F-vrijednost*.

$$F = \frac{1 - \lambda_w^{1/S}}{\lambda_w^{1/S}} \cdot \frac{df_2}{df_1} \quad \text{gdje je} \quad s = \sqrt{\frac{m^2(k-1)^2 - 4}{m^2 + (k-1)^2 - 5}}$$

$$df_1 = m(k-1)$$

$$df_2 = s \left((n-1) - \frac{m+k}{2} \right) - \left(\frac{m \cdot (k-1) - 2}{2} \right)$$

DISKRIMINACIJSKA ANALIZA

Diskriminacijskom analizom može se utvrditi statistička značajnost razlika između centroida dvije ili više grupa ispitanika te doprinos pojedinih varijabli razlikovanju među grupama.

Iz skupa od m manifestnih izračuna se $k-1$ *diskriminacijskih funkcija* ako je broj varijabli (m) veći ili jednak broju grupa (k), odnosno m diskriminacijskih funkcija ako je broj varijabli manji od broja grupa.

Diskriminacijske funkcije su linearne kombinacije manifestnih varijabli koje se izračunavaju uz uvjet da se centriodi grupa entiteta na njima što je moguće više razlikuju i uz uvjet da su međusobno potpuno linearno nezavisne.

DISKRIMINACIJSKA ANALIZA

Matricu diskriminacijskih funkcija (ψ) čine rezultati entiteta u $k-1$ diskriminacijskih funkcija, a izračunava se operacijom

$$\psi = D X$$

$$\begin{array}{c} \underbrace{\left[\begin{array}{cccc} \psi_{11} & \cdot & \cdot & \psi_{1k-1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \psi_{n1} & \cdot & \cdot & \psi_{nk-1} \end{array} \right]} \\ \cdot \\ \underbrace{\left[\begin{array}{cccc} d_{11} & \cdot & \cdot & d_{1m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ d_{n1} & \cdot & \cdot & d_{nm} \end{array} \right]} \\ \cdot \\ \underbrace{\left[\begin{array}{cccc} x_{11} & \cdot & \cdot & x_{1k-1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{m1} & \cdot & \cdot & x_{mk-1} \end{array} \right]} \end{array}$$

gdje je

→ X - matrica svojstvenih vektora matrice $W^{-1}A$

DISKRIMINACIJSKA ANALIZA

Operacijom

$$\Delta = \psi^T \psi n^{-1}$$

se izračuna dijagonalna matrica varijanci diskriminacijskih funkcija, a operacijom

$$\phi = \psi \Delta^{-1/2}$$

standardizirani rezultati entiteta na diskriminacijskim funkcijama, odnosno matrica standardiziranih diskriminacijskih funkcija.

DISKRIMINACIJSKA ANALIZA

Koeficijent kanoničke diskriminacije (R_{cj}) je mjera diskriminacijske moći odgovarajuće diskriminacijske funkcije, a predstavlja korelaciju diskriminacijske funkcije sa *selektorskom varijablom* odnosno varijablom koja određuje pripadnost entiteta grupama.

$$R_{cj} = \sqrt{\frac{\lambda_j}{1 + \lambda_j}}$$

Koeficijenti kanoničke diskriminacije se kreću u intervalu od 0 do 1. Što je koeficijent kanoničke diskriminacije veći, to znači da pripadajuća diskriminacijska funkcija bolje razlikuje grupe entiteta.

DISKRIMINACIJSKA ANALIZA

Statistička značajnost diskriminacijskih funkcija testira se redom od prve do posljednje. Utvrđi li se da neka diskriminacijska funkcija nije statistički značajna, tada niti jedna sljedeća diskriminacijska funkcija ne može biti statistički značajna.

Ako se želi zaključivati s uzorka na populaciju entiteta interpretiraju se samo one diskriminacijske funkcije za koje je utvrđeno da su statistički značajne.

Statistička značajnost razlika između centroida grupa entiteta u skupu manifestnih varijabli (MANOVA) istovjetna je statističkoj značajnosti prve diskriminacijske funkcije.

DISKRIMINACIJSKA ANALIZA

Testiranje statističke značajnosti diskriminacijskih funkcija vrši se putem χ^2 - distribucije s brojem stupnjeva slobode $df=(m-p) \times (k-p-1)$. χ^2 vrijednost se izračunava sljedećom formulom

$$\chi^2 = - \left(n - \frac{m+k}{2} - 1 \right) \log_e \left(\prod_{j=p+1}^q \frac{1}{1 + \lambda_j} \right)$$

gdje je

- ➔ n - broj entiteta
- ➔ m - broj varijabli
- ➔ k - broj grupa
- ➔ q - broj diskriminacijskih funkcija
- ➔ p - broj prethodno testiranih diskriminacijskih funkcija
- ➔ λ_j - svojstvene vrijednosti matrice $W^{-1}A$

DISKRIMINACIJSKA ANALIZA

Doprinos pojedinih manifestnih varijabli razlikovanju centroida grupa određuje se na temelju *matrice strukture diskriminacijskih funkcija* (F) koja se izračunava operacijom

$$F = Z^T \phi n^{-1}$$

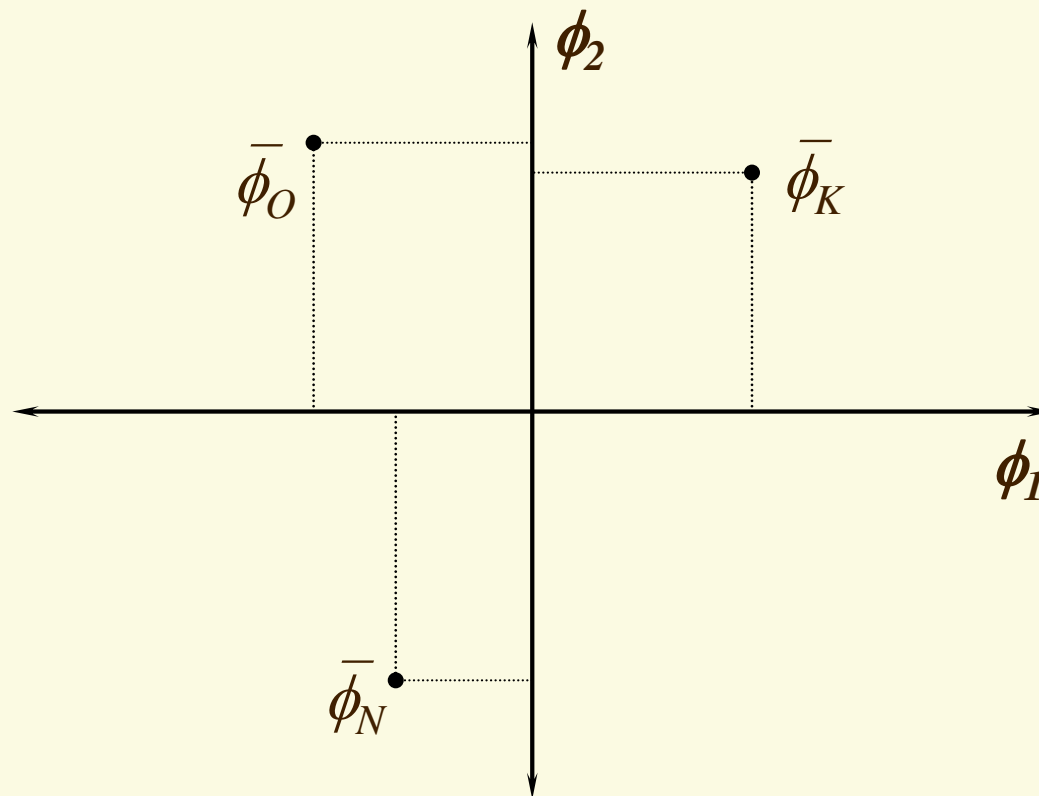
Matricu strukture diskriminacijskih funkcija čine korelacije manifestnih varijabli s diskriminacijskim funkcijama.

DISKRIMINACIJSKA ANALIZA

U interpretaciji razlika potrebno je utvrditi i *centroide grupa* na diskriminacijskim funkcijama kao i njihove međusobne udaljenosti. *Centroidi grupa* ($\bar{\phi}_g$) u prostoru diskriminacijskih funkcija izračunaju se operacijom

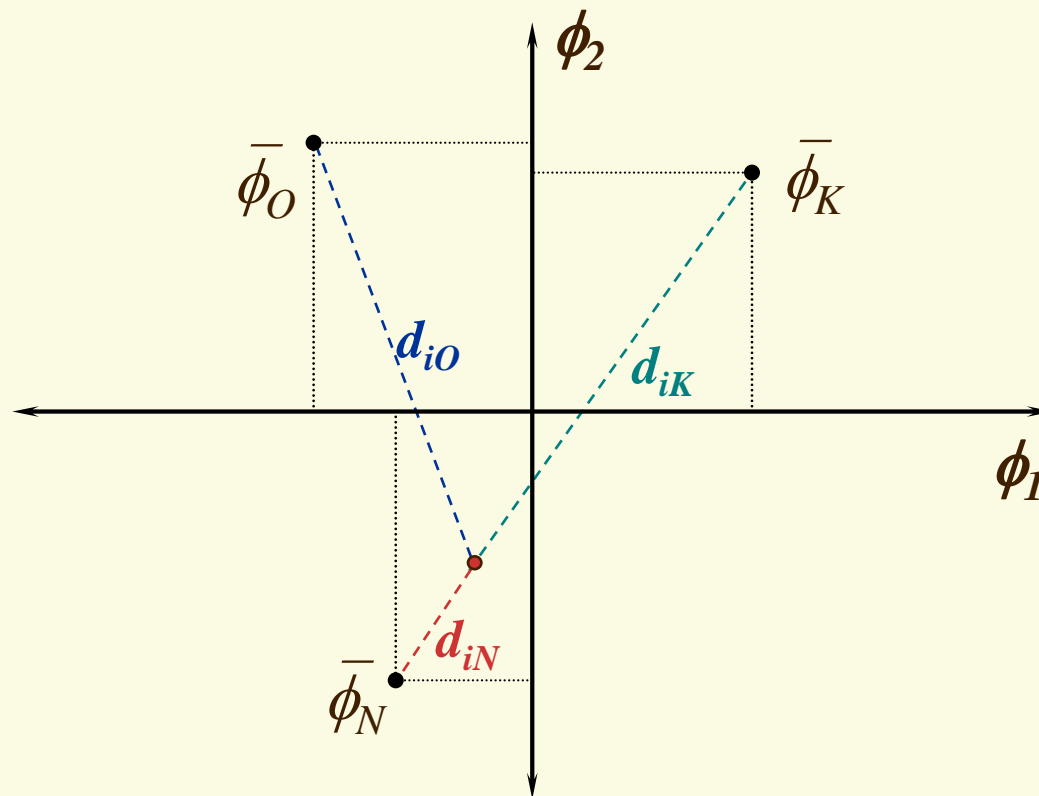
$$\bar{\phi}_g = \phi_g^T \mathbf{1}_g n_g^{-1}$$

DISKRIMINACIJSKA ANALIZA



(Prikaz koordinata centroida grupa K,N i O u koordinatnom sustavu dviju diskriminacijskih funkcija)

DISKRIMINACIJSKA ANALIZA



(Prikaz udaljenosti rezultata entiteta i od centroida grupa K, N i O u koordinatnom sustavu dviju diskriminacijskih funkcija)

DISKRIMINACIJSKA ANALIZA

Pripadnost nekog entiteta određenoj grupi može se procijeniti na temelju udaljenosti vektora njegovih rezultata na diskriminacijskim funkcijama od centroida svake pojedine grupe. Što je udaljenost od centroida grupe manja to je veća vjerojatnost da entitet pripada toj grupi.

STATISTICA 7

Diskriminacijska analiza

Diskriminacijska analiza provodi se slijedom koraka: padajući izbornik *Statistics* → *Multivariate Exploratory Techniques* → *Discriminant analysis*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti selektorsku varijablu (*Grouping variable*) i dvije ili više zavisnih varijabli (*Dependent variables*). Nakon odabira varijabli potrebno je odabrati opcije *Advanced* → *Perform canonical analysis* → *Advanced*.

STATISTICA 7

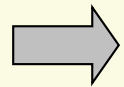
Diskriminacijska analiza

Nakon odabira prethodnih opcija moguće je izračunati: koeficijente kanoničke diskriminacije i testirati statističku značajnost diskriminacijskih funkcija (*Summary: Chi square tests of successive roots*), matricu strukture diskriminacijskih funkcija (*Factor structure*) i centroide grupa u prostoru diskriminacijskih funkcija (*Means of canonical variables*).

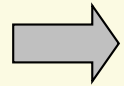
Zadatak - U datoteci *Pejcic-318.sta* utvrdite da li se učenici 1., 2., 3. i 4. razreda statistički značajno razlikuju prema građi tijela (varijable *ATV*, *ATT*, *AOP* i *ANN*)! Testirajte statističku značajnost razlika, izračunajte koeficijente kanoničke diskriminacije, matricu strukture i centroide grupa!

ANALIZA VARIJANCE, DISKRIMINACIJSKA ANALIZA

Literatura za pripremanje kolokvija



Dizdar, D. (2006). *Kvantitativne metode*. Zagreb: Kineziološki fakultet, str. 150-159, 245-257.



Mejovšek, M. (2003). *Uvod u metode znanstvenog istraživanja u društvenim i humanističkim znanostima*. Jastrebarsko: Naklada Slap, str. 224-242.