1. **MANOVA**

U analizi varijance (ANOVA) računamo razlikuju li se dvije ili više grupa (NV) s obzirom na jednu varijablu (ZV), dok se u MANOVI taj **nacrt** proširuje s više ZV. Tako imamo jednosmjernu MANOVU koja ima jednu kategorijalnu varijablu (NV) i 2 ili više kontinuiranih ZV, dvosmjernu MANOVU koja ima dvije NV i 2 ili više ZV itd. **Prednost** MANOVE u odnosu na ANOVU je što pomoću nje možemo proširiti nacrt na više glavnih efekata i interakciju.

**Primjer** nacrta MANOVA-e je kad želimo usporediti različite studije psihologije u Hrvatskoj s obzirom na znanje koje operacionaliziramo pomoću ocjena iz pet kolegija. Dakle, imamo jednu NV koja se sastoji od 4 grupe: Sveučilište u Zadru, Filozofski fakultet u Zagrebu, Hrvatski studiji i Sveučilište u Rijeci te pet ZV operacionaliziranih kroz ocjene iz 5 kolegija. Nacrt se može proširiti i s još jednom NV, npr. spol, tako da imamo dvosmjernu MANOVU s ćelijama 4X2 (1.NV ima 4 grupe, 2. NV ima 2 grupe) i s 5 ZV.

Zapravo bismo mogli pomoću 5 ANOVI izračunati postoje li razlike između skupina s obzirom na ocjenu iz pojedinog kolegija, ali povećanjem broja usporedbi povećava se i vjerojatnost α pogreške (da zaključimo da je razlika među grupama značajna kad razlika u stvarnosti ne postoji). Drugi način **računanja** razlike među grupama **putem ANOVE** je izračunati prosjek ocjena iz 5 kolegija i taj prosjek koristiti kao novu ZV te provesti ANOVU. Na tome principu počiva i MANOVA, samo što nova ZV nije samo prosjek ocjena iz 5 kolegija, već je optimalno ponderirana linearna kombinacija svih ZV. Pri tome će veći ponder imati ona ZV koja više razlikuje grupe. Npr. ako na sva 4 studija psihologije ocjene iz kolegija Percepcija kod većine studenata iznosi 5, ta varijabla neće dobro razlikovati grupe pa će dobiti i manji ponder.

Nova ZV koja nastane takvom optimalnom linearnom kombinacijom naziva se **diskriminativna funkcija** (DF). U MANOVi je središte našeg interesa prva DF zato što je ona formirana tako da maksimalno razlikuje grupe, odnosno da objašnjava maksimalni varijabilitet razlika među grupama.

U MANOVA-i se **prvo** **testira multivarijatna značajnost** za svaki od glavnih efekata i interakciju. Taj se test temelji na *karakterističnim korijenima (λ)* koji se bazira na omjeru objašnjene i neobjašnjene varijance (suma kvadrata između grupa/suma kvadrata unutar grupa). Značajnost karakterističnih korijena računa se jednim od 4 testa značajnosti s tim da je najčešće korišten *Wilksov lambda.* Jedino je on obrnuto proporcionalan λ, odnosno što je on manji, to je λ veća, a time je i razlika među grupama veća. Za sve ostale vrijedi što su njihove vrijednosti veće, to je veća i razlika među grupama, odnosno λ (*Bartlett-Pillai, Roy`s Largest Rc i Hotteling`s trace*).

Nakon testiranja multivarijatne značajnosti, analiza se svodi na **univarijatne testove** (ANOVE) za glavne efekte i interakciju, a na kraju se provode **post hoc analize** kako bi se utvrdilo među kojim točno grupama postoje razlike.

**Preduvjeti** koji trebaju biti osigurani da bi se MANOVA mogla provesti su:

1. *Varijable ne smiju biti visoko multikolinearne,*odnosno ne smije se 1 varijabla moći objasniti na temelju ostalih varijabli (npr. kad bi kao 4 varijable uzeli rezultate na 4 testa, a kao 5. varijablu zbroj rezultata na ta 4 testa – ta 5. varijabla ne daje nikakve nove informacije u odnosu na prethodne, odnosno ne objašnjava ništa novog varijabiliteta);
2. *Uni/multivarijatna normalnost distribucije ZV* mora biti zadovoljena;
3. *Broj sudionika po grupama* treba biti *ujednačen* s tim da u svakoj ćeliji mora biti najmanje 20 sudionika, a omjer najmanje i najveće grupe ne smije biti veći od 4;
4. *Jednakost varijanci* koja se provjerava analizom matrice kovarijanci, testira se Boxovim testom (ako je neznačajan, znači da su varijance podjednake i da je preduvjet zadovoljen: preosjetljiv, kod velikog broja sudionika lako bude značajan);
5. *Linearni odnosi*među varijablama**.**
6. **Diskriminativna analiza**

U MANOVI je NV kategorijalna, a ZV su kontinuirane, dok je u diskriminativnoj analizi obrnut slučaj, odnosno ZV postaje kategorijalna, a NV kontinuirana. Primjer **problema** u MANOVI je ispitati postojanje razlike među različitim studijima psihologije u RH s obzirom na njihovo znanje, odnosno s obzirom na njihove rezultate na 5 testova. U diskriminativnoj analizi provjeravamo možemo li na temelju rezultata na 5 testova znanja odrediti kojem studiju psihologije pripada pojedini sudionik (pripadnost grupi), odnosno razlikuju li dobro testovi znanja pripadnike različitih studija psihologije u RH. U diskriminativnoj analizi možemo imati samo jednu kategorijalnu, tj. zavisnu varijablu.

**Statistička logika** diskriminativne analize je formirati novu varijablu koja se naziva *diskriminativna funkcija*, a formira se na temelju optimalno ponderirane linearne kombinacije nezavisnih varijabli. Takvom kombinacijom osigurava se da ta diskriminativna funkcija maksimalno razlikuje grupe, odnosno da izlučuje maksimalni F omjer i ima najveći karakteristični korijen.*Karakteristični korijen λ* odražava omjer međugrupnog i unutargrupnog varijabiliteta te govori o tome koliko dobro pojedina diskriminativna funkcija objašnjava grupnu pripadnost.

Prva diskriminativna funkcija objašnjava maksimalni varijabilitet razlika među grupama, taj se varijabilitet uklanja iz matrice nezavisnih varijabli, a iz preostale, tzv. **rezidualne matrice** formira se druga diskriminativna funkcija tako da maksimalno objasni preostale razlike i tako sve dok se ne ukloni sav varijabilitet NV. Dakle, DF su ortogonalne, odnosno nezavisne te korelacija među njima iznosi 0 (kao i kod faktora u FA). Druga diskriminativna funkcija objašnjava ono što prva nije uspjela objasniti. Možemo izlučiti onoliko DF koliko ima grupa – 1 ili koliko ima NV.

Svaka DF je povezana s pojedinom NV, što su one bliže, to je njihova povezanost veća, kao što je i veći ponder s kojim ta NV sudjeluje u formiranju DF. Sudionici imaju projekcije na DF, odnosno može se odrediti njihov rezultat na pojedinoj DF i to se zovu **diskriminativni bodovi** (rezultati sudionika na diskriminativnoj funkciji). Također se može odrediti rezultat grupe na pojedinoj DF i to se naziva **centroid grupe** (prosjek grupe na diskriminativnoj funkciji).

Na temelju diskriminativnih bodova i centroida može se odrediti kojoj grupi pripada pojedini sudionik. Prvo se treba izračunati **cutt-point**, odnosno točku na temelju koje se određuje pripadnost grupi: sudionici čiji su rezultati iznad te točke (tog rezultata) pripadaju jednoj grupi, a sudionici čiji su rezultati ispod te vrijednosti pripadaju drugoj grupi. Kad imamo grupe podjednakih veličina, cutt-point se računa tako da se zbroj centroida dviju grupa podijeli s 2 **Zce**=(Za+Zb)/2. No, kad su grupe različitih veličina, potrebno je „pomaknuti“ centroid prema manjoj grupitako da se uvažavaju nejednake distribucije rezultata.To se postiže formulom: **Zcu**=(**Za\*Nb+Zb\*Na**)/(Na+Nb).

DF bolje razlikuje grupe što je njezin λ veći. **WIlksov λ** distribuira se po Chi kvadrat raspodjeli i testira se pripadajućim Chi kvadrat testom.

**Uvjeti provedbe** diskriminativne analize su:

1. Da nema *linearne zavisnosti NV*;
2. *Normalna univarijatna i multivarijatna distribucija NV*;
3. *Homogenost matrica kovarijanci.*
4. **Klaster analiza**
5. **Hijerarhijska klaster analiza**

Klaster analiza po svojoj je funkciji vrlo **slična faktorskoj analizi** zato što u oba slučaja analiziramo veći broj varijabli/sudionika i na temelju sličnosti među njima izdvajamo manji broj faktora/klastera. Međutim, u FA faktori se formiraju kao optimalno ponderirane linearne kombinacije varijabli s tim da se ponderi određuju na temelju korelacija među varijablama, dok u klaster analizi računamo bliskost/udaljenost svakog sudionika u odnosu na drugog preko promatranih varijabli. U FA najčešće imamo velik broj varijabli koji pokušavamo reducirati, dok u klaster analizi češće imamo veći broj sudionika (40-50, maksimalno 100) koje pokušavamo klasificirati u grupe tako da je klaster analiza vrlo slična Q tehnici FA gdje formiramo faktore iz sudionika.

Statistička logika hijerarhijske klaster analize je da se prvo izračuna bliskost ili udaljenosti svakog sudionika u odnosu na drugog preko promatranih varijabli. Oni slučajevi koji su najsličniji spajaju se u 1. klaster, drugi po sličnosti formiraju drugi klaster ili se pak prvome klasteru može pridružiti novi član ili čitavi klaster koji je u prethodnim koracima formiran. Taj proces spajanja odvija se sve dok se svi slučajevi ne spoje u 1 klaster, no istraživač sam odabire gdje će rezati, odnosno koliko će klastera odabrati. Mjesto rezanja može se odrediti prema grafičkom prikazu rezultata klaster analize – *dendogramu* tako da se reže tamo gdje su udaljenosti najveće/linije najduže ili pak prema koeficijentima prikazanima u tablici rasporeda aglomeracije (SPSS) – gdje je najveći skok u vrijednosti koeficijenata, tu se reže.

Za **određivanje koji će se članovi ili grupe spojiti,** možemo koristiti više **metoda** računanja**:**

1. **Međugrupna povezanost (between-group linkage)** – računa se kao prosječna udaljenost svakog člana jedne grupe od pojedinog člana druge grupe i kad je ta udaljenost minimalna, grupe (klasteri) se spajaju;
2. **Unutargrupna povezanost (within-group linkage)** – prvo se hipotetski formiraju novi klasteri pa se računa prosječna udaljenost među svim članovima potencijalnog novog klastera i kad je ona najmanja, ti se klasteri spajaju;
3. **Najbliži susjed** – udaljenost najbližeg člana grupe od novog člana ili grupe;
4. **Najudaljeniji susjed** – udaljenost najdaljeg člana grupe od novog člana ili grupe;
5. **Udaljenost centroida –** izračunaju se centroidi grupa po svim varijablama i računa se udaljenost centroida;
6. **Wardova metoda –** temelji se na ANOVI i veličini unutargrupnog varijabiliteta u odnosu na SStotal (varijabilitet svih sudionika u analizi). Ujedinjuju se oni sudionici ili klasteri koji nakon ujedinjenja rezultiraju minimalnim SSerror, odnosno oni koji maksimaliziraju r2.

Također razlikujemo više **metoda za određivanje udaljenosti među slučajevima**:

1. **Euklidske distance** – korijen iz sume kvadriranih razlika po varijablama;
2. **Kvadrirane euklidske distance** – suma kvadriranih razlika po varijablama;
3. Korelacije – prikladnije kad grupiramo varijable nego sudionike ili kad imamo velik broj slučajeva;
4. **Kosinus kuta** – kao korelacije samo što se izračunava preko kosinusa;
5. **Chebychev –** maksimalna apsolutna razlika među slučajevima po bilo kojoj varijabli.

Od iznimne je važnosti standardizirati početne rezultate ako su izraženi na različitim skalama mjerenja, u suprotnome će sve razlike biti pripisane varijablama s najvećim vrijednostima. Nema posebnih **preduvjeta** za provođenje hijerarhijske klaster analize, važno je samo da su varijable barem na intervalnoj skali te da su normalno distribuirane ukoliko provodimo standardizaciju rezultata u z-vrijednosti.

Hijerarhijska klaster analiza češće se koristi u ekonomiji, sociologiji i političkim znanostima, nego u psihologiji zato što su u psihologiji predmet interesa najčešće varijable, a kad analiziramo sličnosti među sudionicima češće ćemo koristiti diskriminativnu analizu budući da nas zanimaju i diskriminativne funkcije, odnosno po čemu su sudionici slični. Međutim, hijerarhijska klaster analiza jednostavna je metoda primjenjiva za zornu klasifikaciju manjeg broja sudionika ili slučajeva, primjerice za sortiranje brandova kave s obzirom na njihove sličnosti, analizu sličnosti među pripadnicima neke skupine (npr. sportski tim) itd.

1. **K-means clustering**

Hijerarhijska klaster analiza temelji se na matrici udaljenosti među slučajevima koju je prilično teško izračunati kad imamo puno entiteta i puno varijabli. U tom slučaju može se koristiti K-means clustering metoda s tim da je preduvjet za provođenje te analize da unaprijed odredimo željeni broj klastera. Najčešća je praksa da se prvo provede hijerarhijska klaster analiza na slučajno odabranom poduzorku i na temelju nje se odredi koliki bi broj klastera bio prihvatljiv.

**Statistička logika** ove metode je da u multivarijatnom prostoru varijabli računalo odabire onoliko poprilično različitih točaka/vrijednosti koliko smo unaprijed odredili da će biti klastera. Te vrijednosti postaju centri, odnosno aritmetičke sredine klastera. Potom se gleda koliko je svaki sudionik udaljen od pojedinog centra klastera te ga se smješta u onaj klaster od čijeg je centra najmanje udaljen. Zatim se računaju nove aritmetičke sredine svakog klastera na temelju rezultata trenutnih članova. Taj se proces ponavlja sve dok se aritmetičke sredine klastera ne prestanu značajno mijenjati, odnosno dok sudionici više ne prelaze iz klastera u klaster. Tad se klasteri smatraju homogenima, stabilnima i dobro se razlikuju.

**Preduvjeti** za provođenje K-means clustering metode su:

1. Prije analize varijable je potrebno *standardizirati*;
2. Ne smije biti *aberantnih rezultata* jer bi oni mogli biti uzeti za početne aritmetičke sredine;
3. Potrebno je *unaprijed odrediti broj klastera*.

Pri **interpretaciji obilježja klastera** koriste se posljednje izračunate aritmetičke sredine klastera na pojedinim varijablama. Razlike među klasterima mogu se testirati F-testom za svaku pojedinu varijablu.Također se može odrediti za svaki entitet kojem klasteru pripada i koliko je udaljen od aritmetičke sredine pripadajućeg klastera. Na temelju te udaljenosti može se prepoznati postoji li outlayer među rezultatima i ukloniti ga iz analize. Aritmetičke sredine klastera mogu se pohraniti i koristiti u daljnjim analizama i za postavljanje početnih sredina klastera.

**4. Kanonička analiza**

**Koristi se kad** imamo varijable izražene na intervalnoj ili ordinalnoj skali mjerenja te imamo 2 seta varijabli i računamo multivarijatnupovezanost među njima. Dok u regresijskoj analizi pokušavamo na temelju seta prediktora (NV) predvidjeti varijabilitet kriterija (ZV), u kanoničkoj analizi istovremeno pokušavamo predvidjeti varijabilitet u setu ZV na osnovi NV i onaj u setu NV na osnovi ZV, stoga u kanoničkoj analizi ne koristimo pojmove NV i ZV, već varijable lijevog i desnog seta.

**Cilj analize** je formirati kanoničke funkcije koje se sastoje od dvije kanoničke varijate i kanoničke korelacije. *Kanoničke varijate* predstavljaju optimalno ponderirane linearne kombinacije varijabli pojedinog seta s tim da su ponderi određeni tako da korespondentna kanonička varijata lijevog seta maksimalno korelira s varijatom desnog seta. Ta maksimalna korelacija između dvije varijate naziva se *kanonička korelacija*.

Nakon što se formira prvi par kanoničkih varijata, na isti se način formira drugi par kanoničkih varijata iz preostale rezidualne matrice lijevog i desnog seta varijabli i tako sve dok se ne iscrpi sav varijabilitet što znači da možemo izlučiti onoliko kanoničkih varijata u oba seta koliko ima varijabli u manjem setu. Zato će kanoničke varijate manjeg seta **objašnjavati sav varijabilitet** varijabli tog seta, dok kanoničke varijate većeg seta neće objašnjavati sav. Npr. ako imamo 4 varijable u lijevom setu i 3 varijable u desnom setu, moći ćemo formirati 3 para kanoničkih varijata. Tri varijate lijevog seta objasnit će ukupno npr. 71% varijance varijabli lijevog seta, dok će 3 kanoničke varijate desnog seta objasniti svu varijancu njegovih varijabli (100%).

Budući da se varijate formiraju iz rezidualnih matrica varijabli, kanoničke varijate istog seta međusobno su **ortogonalne**, odnosno nezavisne, tj. njihova je korelacija r=0 (kao i faktori u FA), dok je zbog optimalne linearne kombinacije varijabli pojedinog seta korelacija između korespondentnih varijata različitog seta maksimalna moguća.

Osim kanoničkih varijata, korelacija i funkcija, značajni su i:

1. **Kanonički korijeni** – kvadrirane kanoničke korelacije koje zapravo predstavljaju procjene zajedničkog varijabiliteta koji dijele parovi kanoničkih varijata;
2. **Kanonički ponderi** – predstavljaju betapondere s kojima varijable sudjeluju u formiranju kanoničkih varijata;
3. **Kanonička opterećenja (saturacije)** – povezanost varijabli jednog seta s pojedinom kanoničkom varijatom tog seta;
4. **Kanonička kros-opterećenja -** povezanost varijabli jednog seta s pojedinom kanoničkom varijatom suprotnog seta;
5. **Indeks redudancije** – količina varijance jedne kanoničke varijate koja se može objasniti drugom kanoničkom varijatom preko kanoničke funkcije.Zapravo predstavlja ukupan varijabilitet jednog seta varijabli koji se može objasniti na temelju drugog seta varijabli.

Za **interpretaciju** su nam bitne kanoničke funkcije koje su značajne, bilo testirajući značajnost karakterističnih korijena (npr. Wilks lambda), bilo testirajući značajnost bivarijatne korelacije parova kanoničkih varijata. Zanimljive su nam i one kanoničke funkcije koje imaju visoke indekse redundancije, odnosno koje objašnjavaju određenu količinu varijabiliteta u oba seta varijabli te one koje dijele interpretabilnu količinu zajedničke varijance, odnosno imaju visoke kvadrate kanoničkih korelacija.

Kanoničke funkcije interpretiraju se preko kanoničkih opterećenja, odnosno saturacija, kao što je slučaj i kod FA i DA. Mogu se interpretirati i preko kanoničkih kros-opterećenja jer se tu izravnije vidi povezanost jednog seta varijabli s drugim.

Preko kanoničke analize mogu se provesti i sve **ostale** **multivarijatne metode** koje poznajemo. Tako kvadrat kanoničke korelacije pri računanju ANOVE zapravo predstavlja F omjer, odnosno omjer međugrupnog i totalnog varijabiliteta, dok u regresijskoj analizi predstavlja R2, to jest koeficijent determinacije. FA se računa preko KA tako da se u oba seta stave manifesne varijable. U tom slučaju kanoničke varijate predstavljaju faktore, a indeksi redudancije zapravo predstavljaju karakteristične korijene faktora (korelacije između parova varijata su jednake 1 jer je zapravo riječ o istim varijatama). MANOVA se računa preko KA tako da se iz oba seta (ZV i NV) izluči po jedna kanonička varijata jer nas zanima samo 1 kanonička funkcija, dok se u diskriminativnoj analizi izlučuje onoliko varijata koliko ima varijabli u manjem setu, a diskriminativne funkcije su kanoničke varijate iz seta varijabli. Maksimaliziranjem kanoničkih korelacija maksimalizira se multivarijatno razlikovanje grupa i kod MANOVE i kod DA.

-