

Diskriminaciona analiza

Multivarijaciona analiza

- ▶ **Multivarijaciona analiza** predstavlja niz multivarijacionih tehnika koje simultano analiziraju medjusobni odnos između velikog broja varijabli
- ▶ U samom nazivu ove analize uočavamo pojam „**multivarijacioni**“ koji nam ukazuje da je u pitanju veliki broj varijabli
- ▶ Osim ovog izraza može se u literaturi naići na termine „**multivarijantna**“ analiza, ali možda najispravniji termin bi bio „**multivarijabilna analiza**“ čime upravo naglašavamo veliki broj varijabli

Tehnike multivarijacione analize

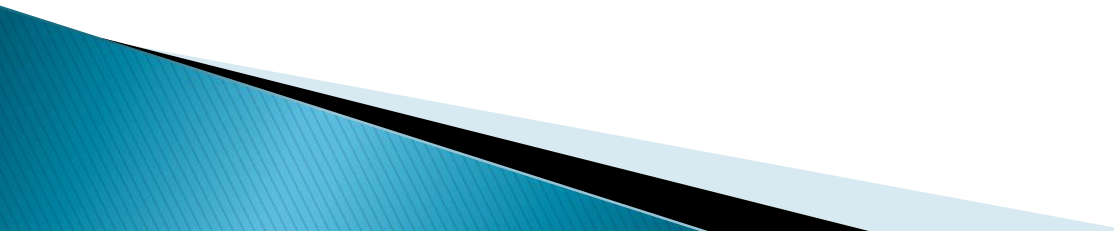
- ▶ Tehnike koje čine ovu analizu mogu da se podijele u dvije grupe:
- ▶ **tehnike zavisnosti**
- ▶ **tehnike međuzavisnosti**
- ▶ Kod **tehnika zavisnosti** možemo uočiti odnos između varijabli, gdje na jednoj strani imamo zavisne varijable na koje uticaj vrše nezavisne varijable
- ▶ U ovu grupu spadaju sledeće tehnike: **višestruka regresija i korelacija, logit analiza, diskriminaciona analiza, kanonička analiza, conjoint analiza**

Diskriminaciona analiza–uvod

- ▶ Od svih navedenih tehnika ističemo primjenu **diskriminacione analize**, koja je svojom pojavom omogućila identifikaciju bas one promjenljive koja najviše doprinosi razdvajanju grupa, a na osnovu vrijednosti skupa nezavisnih promjenljivih predviđa vjerovatnoću da će objekat pripasti jednoj od grupa
- ▶ Metod multivarijacione analize koji se bavi razdvajanjem različitih grupa i alokacijom opservacija u unapred definisane grupe nazivamo **diskriminaciona analiza**

- Prvi put ova tehnika javlja se negdje 1936.god.
- Međutim njena popularnost je rasla zajedno sa pojavom moćnih računara i razvojem tehnologije tokom 60-ih godina
- Npr. zamislite bankarskog službenika na šalteru za izdavanje kredita, koji želi da odredi da li da odobri molbu za kredit automobila
- Odluka treba da bude donesena tako što će se odrediti da li karakteristike (osobine) podnosioca molbe više liče osobama koje su u prošlosti vraćale uspješno kredit ili osobama koje to nijesu činile
- Informacije vezane za te dve grupe, dobijene iz arhive banke, uključuju različite faktore kao što su: starost, prihodi, bračno stanje i posjedovanje nekretnina
- Pr2. Na bazi podataka o depresiji pojedinaca, pokušavamo da predvidimo da li jedinka koja živi u zajednici ima veće ili manje šanse da bude depresivna

Sprovođenje diskriminacione analize

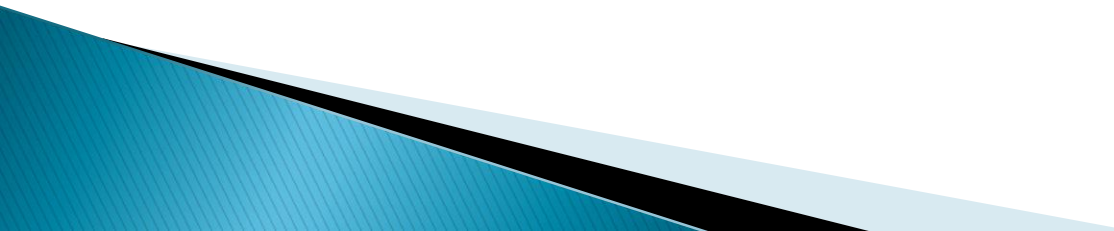
- ▶ Sprovođenje diskriminacione analize vrši se u nekoliko koraka:
 - ▶ **Formirati grupe**
 - ▶ **Ocjeniti diskriminacionu funkciju**
 - ▶ **Odrediti značajnost funkcije i varijabli**
 - ▶ **Interpretirati diskriminacionu funkciju**
 - ▶ **Izvesti klasifikaciju i validaciju**
- 

Ciljevi diskriminacione analize

- ▶ Postoje dva cilja diskriminacione analize:
- ▶ **Diskriminacija** (razdvajanje između grupa)
- ▶ **Klasifikacija** (alokacija opservacija)

- ▶ **Diskriminacija** – jedan od ciljeva diskriminacione analize kojim se utvrđuje da li postoji statistički značajna razlika u sredinama dvije ili više grupa i određuje se promjenljiva koja najviše doprinosi datoj razlici
- ▶ **Drugi cilj** se odnosi na utvrđivanje postupka za klasifikaciju opservacija na osnovu vrijednosti nekoliko promjenljivih u dvije ili više razdvojenih, unaprijed definisanih grupa
- ▶ U konkretnom istraživanju ova dva cilja se često međusobno preklapaju

Diskriminacioni skor

- ▶ Jedan od osnovnih pojmova kojim se služimo
 - ▶ **Diskriminacioni skor** predstavlja broj dobijen linearnom kombinacijom nezavisnih promjenljivih
 - ▶ Potom se taj skor transformiše u aposteriornu vjerovatnoću kojom provjeravamo pripadnost posmatranog objekta određenoj grupi
- 

Fišerov pristup


- ▶ **Ronald Aylmer Fisher** bio je poznati Engleski statističar, biolog, matematičar i genetičar
- ▶ Tokom 1936. godine **Fišer je prvi definisao preciznije cilj diskriminacione analize**, predstavivši postupak diskriminacije između dvije ili više grupa
- ▶ Dakle primjenom Fišerovog postupka za dvije grupe dobija se jedna linearna kombinacija nezavisnih promjenljivih
- ▶ Tom kombinacijom se postiže maksimizacija odnosa varijacija između i unutar grupa
- ▶ Međutim ukoliko se posmatra problem sa više grupa ne može se očekivati da će jedna diskriminaciona funkcija da razdvoji te grupe
- ▶ Time se upravo objašnjava **multivarijaciona diskriminaciona analiza**

Osnovne pretpostavke MDA

- ▶ Kako bi ovaj metod mogao da se primjeni neophodno je da su ispunjene i zadovoljene određene pretpostavke kako bi dobijeni rezultati bili vjerodostojni
- ▶ Osnovne pretpostavke MDA:
- ▶ između nezavisnih varijabli ne postoji multikolinearnost
- ▶ nezavisne varijable imaju multivarijacionu normalnu raspodjelu
- ▶ unaprijed su definisane grupe opservacija
- ▶ matrice varijansi i kovarijansi nezavisnih varijabli između grupa su jednake
- ▶ veza između varijabli je linearna
- ▶ veličine grupa se ne razlikuju mnogo

Analogija sa regresijom

- ▶ Multivarijaciona diskriminaciona analiza je vrlo slična višestrukoj linearnoj regresiji
- ▶ Tačnije preuzima isti zadatak prilikom predviđanja ishoda kao višestruka linearna regresija
- ▶ Osnovna razlika je u tome što kod multivarijacione diskriminacione analize zavisna varijabla je kategorijska, a kod višestruke regresije zavisna varijabla je numerička
- ▶ Cilj regresije jeste procjena parametara koji minimizuju sumu kvadrata reziduala

- ▶ Metod diskriminacione analize kao i metod multivarijacione regresione analize i multivarijacione analize varijanse spada u grupu metoda zavisnosti
 - ▶ Iako se ključno **razlikuju**, **karakteristično za sva tri metoda** jeste da formiraju predviđanje zavisne promenljive na osnovu skupa nezavisnih promenljivih
 - ▶ Ono što znamo jeste da je **kod regresione analize** linearna kombinacija formirana od mjerljivih promenljivih koje su **kvantitativne**
 - ▶ Takođe kreće se od pretpostavke da zavisna promenljiva ima normalnu raspodjelu i da je mjerljiva, a da nezavisne promenljive predstavljaju fiksne vrijednosti u ponovljenim uzorcima
- 

- ▶ Dok u **diskriminacionoj analizi**, već nam je poznato da se linearna kombinacija formira od mjerljivih promjenljivih, ali da je zavisna varijabla nemjerljiva, tj. ona je **kvalitativne prirode**.
- ▶ Zavisna promjenljiva je ovdje fiksna, dok nezavisne imaju normalnu raspodjelu
- ▶ U suštini metode su naizgled slične, ali im je drugačija osnova kojom pristupaju problemu
- ▶ **Slična situacija se dešava i prilikom poređenja analize varijanse i diskriminacione analize.** Kod diskriminacione analize zavisna promjenljiva je kvalitativna, a nezavisne promjenljive su kvantitativne, kod analize varijanse je skroz suprotna situacija

Faze diskriminacione analize

▶ 1.faza

- ▶ Ova faza se zasniva na samom nalaženju ciljeva diskriminacione analize, uglavnom se za ciljeve uzimaju sledeći statistički aspekti:
 - Postojanje statističkih značajnih razlika između određenih nezavisnih varijabli
 - Identifikacija nezavisnih varijabli koje doprinose klasifikaciji
 - Definisane procedure kojima se vrši klasifikacija objekata u konkretne grupe
- **2.faza–ovdje se sprovode 3 postupka**
 - Obično definišemo da su zavisne varijable deskriptivnog karaktera, a nezavisne varijable numeričkog karaktera
 - Veličina grupa mora da bude veća od broja nezavisnih varijabli, koje su uključene u analizu
 - Svrha same diskriminacione analize jeste podjela uzorka na grupe. Postoje dvije **grupe: uzorak za validaciju, uzorak za analiziranje**. Prva grupa služi za definisanje diskriminacione funkcije, dok se druga grupa koristi za testiranje validnosti definisane diskriminacione funkcije

▶ 3.faza

- Da bi diskriminaciona analiza mogla da se sprovede, vrši se testiranje osnovnih pretpostavki kao što su: normalna raspodjela nezavisnih varijabli, linearnost i odsustvo multikolinearnosti

• 4.faza

- U ovoj fazi utvrđujemo diskriminacionu funkciju na osnovu dva metoda kao što su: simultana (istovremena) procjena i postepena procjena

• 5. i 6. faza

- Ove faze se odnose na evaluaciju dobijenih rezultata

Problemi u primjeni diskriminacione analize

- ▶ Izuzetno je bitno naglasiti da postoje određeni problemi ukoliko dođe do napuštanja pretpostavki
- ▶ Kod svakog statističkog procesa neophodno je utvrditi da li su određeni parametri statistički značajni
- ▶ O ovome smo govorili na slajdu broj 10 gdje smo detaljno opisali koje pretpostavke moraju da se ispune da bi rezultati bili vjerodostojni
- ▶ Ukoliko ove pretpostavke ne bi bile ispunjene ovaj metod ne bi mogao da se primijeni

Rezime

- ▶ **Multvarijaciona analiza** predstavlja niz multivarijacionih tehnika koje simultano analiziraju međusobni odnos velikog broja varijabli
- ▶ Tehnike koje čini ovu analizu mogu da se podijele u dvije grupe:
 - ▶ **Tehnike zavisnosti**
 - ▶ **Tehnike međuzavisnosti**
- ▶ **Diskriminaciona analiza** je metod multivarijacione analize koji se bavi razdvajanjem različitih grupa i alokacijom opservacija u unaprijed definisane grupe
- ▶ Postoje dva cilja diskriminacione analize:
 - ▶ **Diskriminacija** (razdvajanje između grupa)
 - ▶ **Klasifikacija** (alokacija opservacija)
- ▶ **Diskriminacioni skor** predstavlja broj dobijen linearnom kombinacijom nezavisnih promjenljivih

- Sprovođenje analize se vrši u nekoliko koraka:
- **Formirati grupe**
- **Ocjeniti diskriminacionu analizu**
- **Odrediti značajnost funkcije i varijabli**
- **Interpretirati diskriminacionu funkciju**
- **Izvesti klasifikaciju i validaciju**
- **Fišer je prvi preciznije definisao cilj diskriminacione analize, predstavivši postupak diskriminacije između dvije ili više grupa**
- **Multivarijaciona diskriminaciona analiza je vrlo slična višestrukoj linearnoj regresiji, tačnije preuzima isti zadatak prilikom predviđanja ishoda**

Osnovna razlika je u tome što kod multivarijacione diskriminacione analize zavisna varijabla je kategorijska, a kod višestruke regresije zavisna varijabla je skalirana

Diskriminaciona analiza se sprovodi u 6 faza koje treba slijediti

Ukoliko ne pratimo ove faze i ne slijedimo pretpostavke, samim tim ni ovaj metod ne bi bio izvodljiv

Hvala na paznji!



▶ Boris Mišovic