

Uvod

Ova knjiga nema visok tehnički nivo jer je usmerena na onaj profil stručnjaka iz oblasti ekonomije koji će koristiti u praksi multivarijantne statističke tehnike. Statističke formule i modeli su svedeni na minimum. Da bi ekonomski stručnjak primenio određenu multivarijantnu analizu, dovoljno je da zna koje metode i tehnike je moguće primeniti na podacima koji su prikupljeni da bi se ostvario određeni istraživački cilj.

Poenta je da istraživač ne mora da razume i da izvodi računski deo multivarijantne statističke analize. Neophodno je da poznaje određenu multivarijantnu tehniku dovoljno dobro, da zna za šta se ona koristi, koji uslovi i pretpostavke moraju da budu zadovoljeni za njenu primenu i kako treba interpretirati dobijene rezultate analize.

Mnogi statistički metodi mogu da se izvedu uz pomoć običnih džepnih kalkulatora. Sa multivarijantnom statističkom analizom to nije slučaj. Za njeno izvođenje potreban je računar i neki od statističkih programa koji se nude na tržištu. U takvim uslovima, čak i neobučene i neiskusne osobe mogu da izvedu složene multivarijantne statističke tehnike, što je mač sa dve oštrice, jer iako se do rezultata dolazi relativno lako i brzo, postoji velika mogućnost da se pogreši.

Pojam multivarijantne statističke analize

Multivarijantna statistička analiza (MSA) je prisutna u nauci gotovo jedan vek. Međutim, njena primena u ekonomskim istraživanjima je počela tek kasnih 50-ih godina XX veka. Vremenom su aplikacije multivarijantne analize postajale sve češće s obzirom da su bile sve više cenjene kako od strane naučnika, tako i od strane poslovnih ljudi.

Pre pojave MSA, u većini istraživanja upotrebljavale su se analize koje su obrađivale najviše dve varijable (promenljive) istovremeno. Kao produkt takve analize najčešće su se javljale mere centralne tendencije (aritmetička sredina, modus, medijana...), mere varijacije (varijansa, standardna devijacija, kvartili...), intervali poverenja i testovi na osnovu normalnog rasporeda, t -rasporeda i slično. Najdalji domet u proučavanju odnosa dve pojave predstavljao je koeficijent korelacije.

MSA je pružila mnogo moćnije tehnike koji su omogućili istraživačima da otkriju obrasce ponašanja u međusobnom odnosu velikog broja promenljivih, obrasce koji bi inače bili sakriveni ili jedva приметni. Pored toga, većina tehnika je dovoljno precizna da se uz pomoć testa statističke značajnosti utvrdi da li je određena međuzavisnost zaista bitna ili je plod fluktuacije podataka u uzorku. Ove tehnike su znatno povećale količinu upotrebljivih informacija koje mogu da se izvuku

iz posmatrane statističke mase. „Dobra analogija je kontrast između slike u jednoj ili dve boje nasuprot iste priče u punom koloru.“³⁸

50-ih godina XX veka otežavajuća okolnost je bila što su se izračunavanja rezultata multivarijantne statističke analize vršila ručno ili uz pomoć džepnih digitrona. Takav način rada je oduzimao dane, a ponekad i nedelje dok bi se došlo do konačnog rezultata, a da pri tome nije izvršena nikakva računaska kontrola. Danas, uz pomoć računara visoke tehnologije i raspoloživog statističkog softvera, za rezultate multivarijantne statističke analize je potrebno svega nekoliko sekundi.

Varijabla ili promenljiva u kontekstu MSA je bilo koja pojava koja slobodno varira na takav način da se te varijacije mogu identifikovati i izmeriti. Kada se MSA primenjuje u ekonomskim istraživanjima, nailazimo na neisrpan skup pojava koje se mogu analizirati, kao, na primer, uvoz, izvoz, prodaja, troškovi, preferencije kupaca, društveni proizvod, nacionalni dohodak itd.

Tehnike multivarijantne statističke analize

Multivarijantna tehnika može da analizira međusobne odnose između više varijabli (više od dve), i to simultano prema odgovarajućem modelu na kojem se bazira ta tehnika. Većina tehnika identifikuje šablone (obrasce) podudarnosti ili odnosa između mnogo varijabli, ocenjuju relativnu važnost svake varijable i predviđa ili objašnjava moguće ishode.

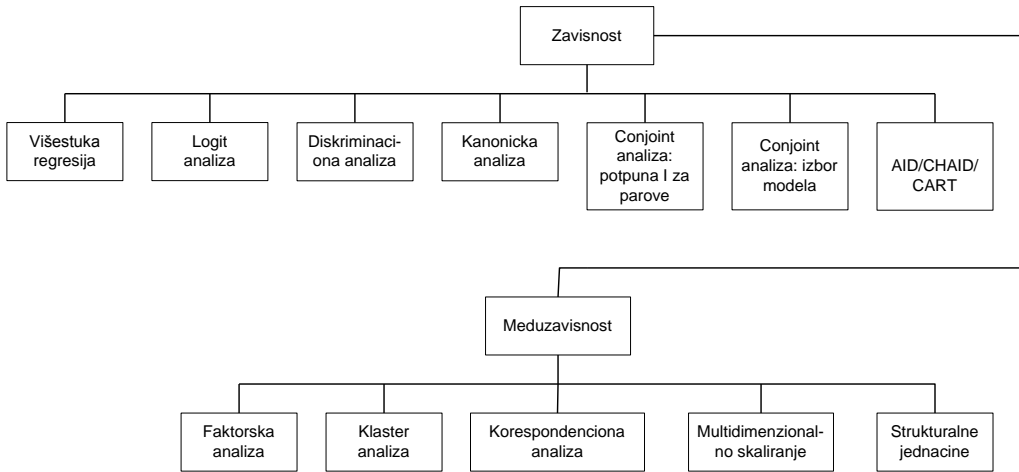
Sve multivarijantne statističke tehnike mogu da se podele u dve grupe:

- ↳ tehnike koje analiziraju zavisnost
- ↳ tehnike koje analiziraju međuzavisnost.

Kod tehnika koje analiziraju zavisnost cilj je da se jedna zavisna varijabla objasni i predvide njene varijacije na osnovu ostalih, nezavisnih varijabli. Na primer, kada je cilj da se predvidi kretanje prodaje automobila na nekom tržištu u zavisnosti od cene, boje, snage konkurencije, životnog standarda itd.

Kod tehnika koje analiziraju međuzavisnost, sve pojave u modelu se posmatraju na isti način, a cilj je da se utvrdi obrazac ponašanja varijabli, odnosno koje varijable su međusobno slične, a koje ne. Posebna korist ovih tehnika je kada se želi izvršiti segmentacija tržišta.

³⁸ Hanke, J.E., Wichern, D.W., Reitsch A.G. (2001), Business Forecasting, seventh edition, Prentice Hall, strana 4

Slika Pregled multivarijantnih statističkih tehnika³⁹

U jednom istraživanju često se javlja i potreba za kombinacijom tehnika iz dve navedene grupe. Na primer, uz pomoć klaster analize (tehnika međuzavisnosti) izvrši se segmentacija kupaca u grupe prema sličnim karakteristikama, a zatim se izvrši višestuka regresiona analiza (tehnika zavisnosti) da bi se identifikovale varijable koje imaju najveći uticaj na svaki segment pojedinačno.

Sve multivarijantne tehnike su razvijane na osnovu istraživanja stručnjaka iz različitih naučnih disciplina, kao što su matematika, statistika, psihologija, sociologija, ekonomija, poljoprivreda itd. Usled toga, multivarijantna statistička analiza je interdisciplinarna kako po svom poreklu tako i po svojoj primeni.

Tehnike koje su prezentovane u ovoj knjizi su one koje se trenutno najviše koriste u ekonomskim istraživanjima. Mnoge od ovih tehnika imaju strogu matematičku derivaciju i osnov i pripadaju „klasičnim“ statističkim modelima. Ostale tehnike uključuju različite aproksimacione metode i rešenja koja su se kroz praksu pokazala kao dovoljno tačna i upotrebljiva.

Korelacija ranga

Vrši se zamena originalnih vrednosti obeležja rangovima.

Ako je potrebno posmatrati korelaciju pojava čija se obeležja ne mogu meriti broјčano ili ako je potrebno brzo doći barem do približne vrednosti koeficijenta korelacije.

³⁹ Myers J. H., Mullet G. M., Managerial Applications of Multivariate Analysis in Marketing, American Marketing Association, Chicago, USA, 2003, strana 8

Najčešće se koriste Spearman-ov (za dve pojave) i Kendall-ov koeficijent korelacije ranga (za više od dve pojave).

Kendall-ov neparametarski koeficijent korelacije ranga se izračunava na osnovu sledeće formule:

$$r''_{12} = \frac{12}{m^2 \cdot n} \cdot \frac{n \sum_{i=1}^n S_i^2 - \left(\sum_{i=1}^n S_i \right)^2}{n^3 - n}$$

gde je:

m – broj numeričkih serija

n – broj podataka u svakoj seriji

$S_i, i = 1, 2, \dots, n$ – zbir rangova po redovima.

Ukoliko se pojave zajednički rangovi, koristi se sledeća formula:

$$r''_{12} = \frac{12}{m^2 \cdot n} \cdot \frac{n \sum_{i=1}^n S_i^2 - \left(\sum_{i=1}^n S_i \right)^2}{n^3 - n - \sum_{i=1}^k T}$$

Korektivni faktor se izračunava na osnovu sledeće formule:

$$\sum_{i=1}^k T = \frac{1}{12} \sum_{i=1}^k (3 - k)$$

gde je:

k – broj obeležja koja imaju zajedničke rangove.

Dobijeni koeficijent korelacije ne pokazuje jačinu i smer korelacione veze direktno između posmatranih pojava nego između njihovih rangova.

Primer:

Primer 271(strana 774) Spearman-ov koeficijent korelacije ranga

REG-002 Spearman-ov koeficijent korelacije ranga

REG-014 Spearman-ov koeficijent korelacije ranga, zajednički rangovi

Primer 272 (strana 776) Kendall-ov koeficijent korelacije ranga

MUL-002 Kendall-ov koeficijent korelacije ranga

Višestruka linearna regresija

"Ako se ispituje zavisnost jedne pojave od dve ili više nezavisnih pojava, onda se govori o višestrukoj ili multiploj regresiji."⁴⁰

Zadatak regresije: otkriti što više faktora.

Polazi se od pretpostavke: što više nezavisnih pojava u model, sve je manji uticaj latentne promenljive (slučajne greške) ε_i , $i = 1, 2, \dots, n$.

Voditi pažnju o selektivnom izboru promenljivih.

Višestruki linearni regresioni model za uzorak:

$$\hat{X}_{i1.23\dots m} = a_{1.23\dots m} + b_{12.34\dots m} X_{i2} + b_{13.24\dots m} X_{i3} + \dots + b_{1m.23\dots(m-1)} X_{im} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

gde je:

$a_{1.23\dots m}$ – slobodni član,

$\hat{X}_{i1.23\dots m}$, $i = 1, 2, \dots, n$ – pojedinačne vrednosti regresije,

$x_{i1}, x_{i2}, \dots, x_{ij}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ – vrednosti nezavisnih promenljivih,

$b_{12.34\dots m}, b_{13.24\dots m}, b_{1m.23\dots(m-1)}$ – regresioni koeficijenti,

ε_i , $i = 1, 2, \dots, n$ – latentna promenljiva (slučajna greška),

m – broj nezavisnih promenljivih,

n – broj jedinica u uzorku.

Važi uslov da broj jedinica u uzorku mora biti veći od broja nezavisnih promenljivih:

$$n > (m+1).$$

Parametri višestruke linearne regresije:

- ↖ ravan regresije,
- ↖ jednačina regresije,
- ↖ standardna greška.

Ravan regresije

⁴⁰ Stojković, M, "Statistika", Ekonomski fakultet u Subotici, 2001, strana 823.

Tri pojave u modelu → trodimenzionalna regresija, sa prikazom u obliku regresione ravni.

Više od tri pojave u modelu → hiperravan.

Jednačina trodimenzionalne regresije

$$\hat{x}_{i1.23} = a_{1.23} + b_{12,3} x_{i2} + b_{13,2} x_{i3}, i = 1, 2, \dots, n.$$

Promenljive X_1 , X_2 i X_3 mogu promeniti značenje.

Standardna greška

Predstavlja srednju meru odstupanja originalnih vrednosti obeležja zavisne promenljive od regresione ravni, mereno paralelno sa X_1 osom i izraženo u jedinici mere zavisno promenljive X_1 .

Interval poverenja

Testiranje

Višestruka linearna korelacija

Osnovni parametri:

- ↗ koeficijent višestruke linearne determinacije,
- ↗ koeficijent višestruke linearne nedeterminacije,
- ↗ koeficijent višestruke linearne korelacije,
- ↗ koeficijent višestruke linearne alienacije.

Parcijalna korelacija

Testiranje kod višestruke linearne korelacije

Izbor najbolje jednačine višestruke linearne regresije

Potrebno je doneti odluku koje promenljive uvrstiti u model, a koje ne.

Metodi za izbor:

- ↗ Metod izbora iz skupa svih jednačina višestruke linearne regresije.
- ↗ Metod eliminisanja od kraja.
- ↗ Metod uključivanja od početka.
- ↗ Poboljšani metod uključivanja od početka.

↪ Metod m-varijabilne redukcije.

Multikolinearnost

Kod višestruke linearne regresije i korelacije polazi se od pretpostavke da između nezavisnih promenljivih ne postoji nikakva veza.

Savršena multikolinearnost je retka, bliska multikolinearnost je vrlo česta.

Forme multikolinearnosti:

- ↪ otvorena – kada postoji linearna funkcionalna veza,
- ↪ prikrivena – postoji korelaciona veza.

Faze u posmatranju problema:

1. utvrđivanje postojanja multikolinearnosti,
2. traženje metoda za smanjivanje multikolinearnosti.

Problem multikolinearnosti do danas nije rešen u potpunosti.

Indikatori za otkrivanje multikolinearnosti:

- ↪ r_{23} ,
- ↪ M' .

Primer:

MUL-001 Višestruka regresija

Primer 278 (strana 840) – Višestruka linearna regresija i korelacija