Verzija 2
22.6.2015

MULTIVARIATNA ANALIZA: Zbirka rešenih primerov s komentarji
Denis Marinšek
1. natis
Ljubljana: Ekonomska fakulteta, 2015

ERRATUM

Stran 64

Odstavek, označen s prvo alinejo, se preoblikuje v:

- Čisto levo (Initial eigenvalues ❽) je prikazan rezultat, ki ga dobimo z metodo glavnih skupnih faktorjev (angl. *Principal component factoring*), ki privzame, da je tudi specifična varianca spremenljivke izražena kar preko skupnih faktorjev. Lastne vrednosti (angl. *Eigenvalues*) se seštejejo v vrednost 23 (podatki so standardizirani), s 23 faktorji pa zajamemo 100 % celotne variabilnosti začetnih spremenljivk. V tabeli komunalitet imamo stolpec, poimenovan Initial ❺. V njem so zapisani koeficienti determinacije med izbrano spremenljivko in vsemi ostalimi proučevanimi spremenljivkami. Primer: 39,8 % celotne variabilnosti spremenljivke »Standardni odklon me navdušuje« je pojasnjeno z variiranjem preostalih 22 spremenljivk.

Stran 99–100

1. $H_0: \mu_1 = \mu_2 = \mu_3 \qquad H_1: \text{vsaj en } \mu_j \text{ je različen od ostalih}$

2. $F = \left(\frac{1-\Lambda_i}{\Lambda_i}\right)\left(\frac{n-g}{g-1}\right) = \left(\frac{1-0{,}640}{0{,}640}\right)\left(\frac{40-3}{3-1}\right) = 10{,}399$

3. $m_1 = g-1 = 3-1 = 2 \qquad m_2 = n-g = 40-3 = 37$

$(p = 0{,}000) < (\alpha = 0{,}005) \rightarrow$ razlika je močno statistično značilna

4. **Na podlagi vzorčnih podatkov zavrnemo ničelno domnevo pri zanemarljivi stopnji značilnosti in sprejmemo sklep, da je vsaj ena aritmetična sredina različna od ostalih. To pomeni, da se število doseženih točk pri izpitu iz mikroekonomije statistično značilno razlikuje med letniki študija.**

Enako velja tudi za ostale tri predmete.

Nato sledi tabela z lastnimi vrednostmi (angl. *Eigenvalues* ❹). V konkretnem primeru imamo ocenjeni dve diskriminantni funkciji, saj imamo tri skupine (*g*) in štiri merjene spremenljivke (*p*): $min(3-1,\ 4) = 2$. Prva diskriminantna funkcija ima vedno najvišjo lastno vrednost oz. najvišjo varianco – v konkretnem primeru znaša 4,760 ❺. Prva diskriminantna funkcija zajame 78,3 % ❻ celotne variance merjenih spremenljivk in jo izračunamo kot 4,760/(4,760+1,317). V zadnjem stolpcu je navedena še vrednost koeficienta kanonične korelacije (0,909 ❼), ki predstavlja povezanost med vsemi štirimi merjenimi spremenljivkami na eni strani in letnikom študija na drugi strani. Če prvi koeficient kanonične korelacije kvadriramo dobimo mero, ki predstavlja delež variabilnosti med skupinami (med letniki študija), ki ga pojasnjuje linearna kombinacija štirih merjenih spremenljivk, izražena s pomočjo obeh diskriminantnih funkcij:

$$R^2 = 0{,}909^2 = 0{,}826$$

Z linearno kombinacijo štirih merjenih spremenljivk, izračunano v obliki obeh diskriminantnih funkcij, pojasnimo 82,6 % variabilnosti letnika študija.

Nato sledi tabela, poimenovana *Wilks' Lambda* ❽. V njej se nahajata preizkusa, ali sta koeficienta kanonične korelacije različna od vrednosti nič. V prvi vrstici se test nanaša na oba koeficienta kanonične korelacije hkrati (v SPSS-u označeno z »1 through 2«), v drugi vrstici pa se test nanaša samo na drugi koeficient kanonične korelacije. V drugem stolpcu sta podani vrednosti Wilksove lambde, ki se ju v konkretnem primeru izračuna po naslednji enačbi:

$$\prod_{l=1}^{n_\Lambda}(1 - \text{Koeficient kanonične korelacije}_l^2)$$

$$\text{Prva vrstica: } (1 - 0{,}909^2) \cdot (1 - 0{,}754^2) = 0{,}075$$ ❾

$$\text{Druga vrstica: } (1 - 0{,}754^2) = 0{,}432$$ ❿

S pomočjo Wilksove lambde (ta predstavlja delež nepojasnjene variance, saj je definirana kot 1−R^2) je izveden χ^2−preizkus, ki preizkuša, ali je koeficient kanonične korelacije diskriminantne funkcije enak nič. V konkretnem primeru znaša vrednost χ^2−preizkusa v prvi vrstici 91,996 ⓫ in je močno statistično značilna. Stopinje prostosti se izračuna po formuli $(p - r + 1)(g - r)$, kjer je *p* število merjenih spremenljivk, *r* analizirana diskriminantna funkcija, *g* pa število skupin. V prvi vrstici torej (4−1+1)(3−1)=8, v drugi vrstici pa (4−2+1)(3−2)=3.[1] V primeru, če hipoteze o enakosti koeficienta kanonične korelacije vrednosti nič v prvi vrstici ne bi mogli zavrniti, bi to pomenilo, da obe diskriminantni funkciji hkrati neuspešno diskriminirata med skupinami. Opazimo lahko, da tako kombinacija obeh diskriminantnih funkcij (prva vrstica) kot tudi druga diskriminantna funkcija sama (druga vrstica) uspešno razlikujeta med tremi skupinami pri zanemarljivi stopnji značilnosti (p < 0,000).

[1] V primeru binarne diskriminantne analize, kjer analiziramo samo dve skupini, se formula za stopinje prostosti poenostavi v $p(g-1)$.