

Prof. dr Mihajlo Mijanović
Boris Nikolić prof.
Univerzitet Crne Gore
Filozofski fakultet, Nikšić

OBJEKTIVNOST I DISKRIMINATIVNOST TESTOVA I MJERNIH INSTRUMENATA

OBJEKTIVNOST TESTA

Test je objektivan ako se njegova primjena i ocjenjivanje mogu tako standardizovati da dobiveni rezultati zavise isključivo od ispitanika, a ne od subjektivnog tumačenja ocjenjivača. Ako je test objektivan, različiti ispitivači ispitujući iste ispitanike doći će do jednakih (*istih*) rezultata¹

Kada za jedan test postoji čvrst i jasan kriteriji za razlikovanje tačnih od netačnih odgovora ili kada na nekom mjernom instrumentu (vaga, štoperica, metar, dinamometar, termometar i sl.), nema dvojbe o kojoj se vrijednosti i jedinici radi ma ko to očitavao, kažemo da je taj test ili mjerni instrument objektivan. Dakle rezultat ne zavisi od ocjenjivača već od mjernog instrumenta i preciznosti očitavanja. Kako se većina mjernih instrumenata digitalizirala, objektivnost se povećala odnosno greška kod očitavanja neke vrijednosti se minimizirala. Brojni testovi odnosno rezultati u testovima ne mjere se uz pomoć štoperice, vage, metra već su vrijednosti dobivaju nakon ocjene i procjene ispitivača. Primjer su brojne ankete i pitanja gdje se odgovori često svode na polovično tačne ili djelimično tačne. U takvim situacijama različiti ispitivači će se između sebe razlikovati u vrednovanju odgovora a time se narušava nivo objektivnosti. Neka pitanja u upitnicima i anketama su poluotvorenog ili otvorenog tipa pa se daje mogućnost dosta različite interpretacije tačnosti što uzrokuje različito vrednovanje odgovora od ispitivača do ispitivača. Problem

¹ Bujas, Z., Zagreb, 1959

objektivnosti je ozbiljan problem kod ocjenjivanja znanja u školama, fakultetima i sl. Ovaj problem je manje naglašen kada je riječ o prirodnim naukama, odnosno posebno je prisutan u ostalim naukama. Jedan sastav iz jezika dva i više nastavnika će različito ocijeniti, često se te ocjene bitno razlikuju od ocjenjivča do ocjenjivača. U postupku utvrđivanja objektivnosti ocjenjivači moraju biti potpuno neovisni ocjenjivači ne smiju biti u kontaktu, kako ocjenjivači ne bi uticali jedan na drugog. Dakle nema dogovora i bilo kakvog usaglašavanja.

Objektivnost se može statistički provjeriti na dva načina i to kao stepen slaganja između ocjenjivača i stepenom neslaganja između ocjenjivača. Step en slaganja između ocjenjivača utvrđuje se preko linearnog koeficijenta korelacije, ako ima više od dva ispitivača tada se izračunava korelacioni koeficijent u svim kombinacijama a nakon toga izračuna se prosječni korelacioni koeficijent. Prosječni korelacioni koeficijent poznat je kao *indeks objektivnosti testa*. Kako je poznato maksimalna korelacija iznosi 1 dakle indeks pouzdanosti 1 upućuje da se ispitivači ni u jednom slučaju ne razlikuju. Dovoljno je razlika u samo jednom slučaju pa će indeks pouzdanosti biti manji od 1. Udaljavanjem od 1 tj. približavanjem 0 upućuje na smanjenje objektivnosti. Statistički gledano objektivnost je dobra ako je koeficijent korelacije statistički značajan. Međutim empirija je nešto dugačija, pa se za objektivne testove smatraju oni čija je korelacija veća od 0.90 u nekim slučajevima čak veća od 0.98. Drugi pristup svodi se u biti na isto, dakle riječ je o utvrđivanju nivoa objektivnosti na osnovu utvrđivanja razlika između ocjenjivača. Razlike se utvrđuju preko metoda diskriminacije. Ako su samo dva ocjenjivača postupak se svodi na testiranje razlika aritmetičkih sredina t-test, ako ima više od dva ocjenjivača primjenjuje se metoda analize varijanse. Ako je razlika jednaka nuli tada je riječ o istim aritmetičkim sredinama odnosno može se reći da je objektivnost zadovoljavajuća i obratno. Statistički značajna razlika između ocjenjivača upućuje na zaključak da je objektivnost loša. Statistička značajnost razlika provjerava se preko t-testa u slučaju ako su dva ocjenjivača. Ako ima više od dva ocjenjivača statistička razlika utvrđuje se preko F-testa, odnosno pomoću multivarijantnih diskriminativnih postupaka.

Primjer

Izvršeno je ocjenjivanje nivoa tehnike kod 14 košarkaša. Ocjenjivači su bila trojica poznavalaca košarke. Prema utvrđenom dogovoru ocjene su bile u rasponu od 1 do 10. Rezultati su prikazani u sledećoj tabeli:

<i>košarkaši</i>	ocjenjivač (A)	ocjenjivač (B)	ocjenjivač (C)
1.	8	8	7
2.	7	8	8
3.	9	8	9
4.	10	10	10
5.	6	4	5
6.	7	7	7
7.	8	7	7
8.	4	4	3
9.	3	3	3
10.	2	1	2
11.	8	8	8
12.	7	6	6
13.	9	9	9
14.	10	8	9

- a) Izračunaj koeficijent objektivnosti preko linearne korelacije
b) Izračunaj koeficijent objektivnosti preko analize varijanse (*Anova*)

Matrica interkorelacija

<i>ocjenjivači</i>	A	B	C
A	1.00	.943	.968
B	1.00	.966	
C			1.00

a) U primjeru prosječna korelacija iznosi .959 što znači da se ocjenjivači slažu u mišljenjima, tako da su dobivene ocjene vrlo objektivne.

b) Statistička metoda jednofaktorske analize varijanse *Anova* upućuje da se ocjenjivači statistički značajno ne razlikuju u mišljenju, pa je zaključak isti, dakle radi se o značajnoj i visokoj objektivnosti. U primjeru vrijednosti jednofaktorske analize varijanse su: $F = .127$ $Sig. = .882^2$

Kako se vidi postupci utvrđivanja objektivnosti pomoću prosječnog korelacionog koeficijenta i metode jednofaktorske analize varijanse su ekvivalentni.

² U primjeru a) izračunat je koeficijent pravolinijske korelacije (Pirsonov koeficijent r)

DISKRIMINATIVNOST (*osjetljivost testa*)

Diskriminativnost testa podrazumijeva mogućnost mjernog instrumenta da razlikuje-diskriminira ispitanike.

Kada je riječ o fizičkim mjernim instrumentima kao što su: vaga, metar, dinamometar, terometar, kalorimetar, štoperica i sl. diskriminativnost-osjetljivost se kao problem u današnje vrijeme ne postavlja. U sportu brojni testovi odnosno rezultati u testovima izraženi su upravo u kilogramima, njutnima, stepenima, kalorijama itd., pa shodno već definisanim preciznim mjernim instrumentima sa digitalnim očitavanjem, diskriminativnost je teoretski i empiriski izuzetno visoka. Ako imamo elektronsko mjerenje vremena sa preciznošću na stote ili pak hiljadite dijelove sekunde, svi ispitanici mogu imati različit rezultat. Slično je sa preciznom vagom, metrom itd. Kako vidimo osjetljivost fizičkih mjernih instrumenata se kao problem ne postavlja. Problem osjetljivosti je prisutan kod tekstualnih testova, anketa, upitnika, zatim kod rješavanja raznih zadataka, ocjenjivanja neke sposobnosti, znanja u školi itd.

Za test koji riješe svi i test koji ne riješi niko kažemo da nema diskriminativnost. Ocjene od 5 do 10 imaju veću diskriminativnost od ocjena od 1 do 5. U prvom slučaju radi se o šestostepenoj skali a u drugom slučaju o petostepenoj. Ako neki zadatak iz matematike bodujemo od 1 do 30 te ako imamo 30 učenika, teoretski i empiriski može se desiti da nema dva ista slučaja ili pak da su svi isti. Ipak iz poznatih razloga ocjene se svode od 1 do 5 ili 5 do 10. Bez obzira koji način raspodjele tj. grupisanja bodova uradili desiti će se da učenici sa različitim brojem bodova svrstamo u istu grupu odnosno ocijenimo ih sa istom ocjenom. Optimalno grupisanje je u skladu sa normalnom (Gaussovom) raspodjelom. Normalna raspodjela podrazumijeva maksimalnu koncentraciju vrijednosti u aritmetičkoj sredini te simetrično opadanje vrijednosti sa lijeve i desne strane u odnosu na aritmetičku sredinu. Sa poznatim pravilima normalne funkcije vjerovatnoća da će na jednom testu biti najviše prosječnih, a podjednako i najmanje loših i odličnih. U praksi se dešava i nešto drugačije, tako da imamo slučajeve gdje većina dobije odlučne ocjene ili većina dobije loše ocjene. U jednom i drugom primjeru radi se o lošoj diskriminativnosti. Ako se desi da je najveći broj učenika dobio ocjenu 5 a ostale ocjene su podjednako zastupljene kažemo da je diskriminativnost odličnih loša i obratno ako je najveći broj učenika dobio ocjenu 1 a ostale ocjene su podjednako zastupljene kažemo da je diskriminativnost onih koji su dobili 1 loša. Testiranje empiriske

raspodjele u odnosu na teoretsku (*normalnu Gaussovu*) raspodjelu vrši se pomoću Kolmogorov-Smirnovljevog testa **K-S testa** ili Hi-kvadrat χ^2 testa.

Primjer

Na jednom prijemnom ispitu 50 kandidata postigli su sledeći broj bodova. Formirana je tabela distribucija frekvencija od sedam grupa, **K=7** gdje je veličina grupnog intervala 5

Hipoteza glasi: Dobivene vrijednosti su u skladu sa normalnom Gaussovom raspodjelom, odnosno nema statistički značajne razlike između empirijskih i teoretskih vrijednosti.

a) Testiraj normalnost distribucije pomoću Kolmogorov-Smirnovog **K-S testa**

b) Testiraj normalnost distribucije preko χ^2 – testa

Primjer a)

K	bod.	f	x	z	Y_z	F	K_f	K_F	K_{MC}	K_{FC}	D
1.	30-34	2	32	-2.02971	0.05186	2	2	2	0.04	0.04	0.00
2.	35-39	7	37	-1.36204	0.15822	5	9	7	0.18	0.14	0.04
3.	40-44	9	42	-0.69437	0.31443	11	18	18	0.36	0.36	0.00
4.	45-49	12	47	-0.02671	0.39886	13	30	31	0.60	0.62	0.02
5.	50-54	10	52	0.64096	0.32506	11	40	42	0.80	0.84	0.04
6.	55-59	9	57	1.30863	0.16915	6	49	48	0.98	0.96	0.02
7.	60-64	1	62	1.97629	0.05730	2	50	50	1.00	1.00	0.00
-	-	50	-	-	-	50	-	-	-	-	-

$x=47.20$, $s=7.5$, Y_z =visina ordinate za odgovarajuće standardizirane (z)

U primjeru je: $\frac{N}{\sigma} i = \frac{50}{7.5} 5 \approx 33.33$

očekivane frekvencije $F = Y_z \times \frac{N}{\sigma} i$

Vrijednost testa: $t_{(0.01)} = \frac{1.63}{\sqrt{N}} = \frac{1.63}{\sqrt{50}} = 0.2305$

$t_{(0.05)} = \frac{1.36}{\sqrt{N}} = \frac{1.36}{\sqrt{50}} = 0.1923$

$$D_{max} = 0.04 < Test_{(0.01)} = 0.2305 NS$$

Hipotezu o normalnosti distribucije smo potvrdili uz pomoć Kolmogorov-Smirnovog test **K-S testa**. Navedena distribucija je kako vidimo u skladu sa normalnom Gaussovom, gdje se empiriske vrijednosti-frekvencije statistički značajno ne razlikuju od teoretskih³.

Primjer b)

K	f	f'	F	F'	(f' - F)(f' - F)² / F'	
1.	2		2			
2.	7	9	5	7	2.00	0.57
3.	9	9	11	11	-2.00	0.36
4.	12	12	13	13	-1.00	0.08
5.	10	10	11	11	-1.00	0.09
6.	9	10	6	8	2.00	0.50
7.	1		2			
-	50	50	50	50	0.00	$\chi^2 = 1.60$

$$df = k-3, df = 5-3=2, gv_{(0.01)} = 5.99, c^2 = 1.60 < 5.99 N.S$$

Primjerom **b** smo takođe potvrdili da nema značajnih razlika između opaženih i očekivanih frekvencija⁴

K-S Test

$$D_{max} = .04 Test_{(0.01)} = .23 Test_{(0.05)} = .19 > .04 NS (distribucija je normalna)$$

χ^2 - Test

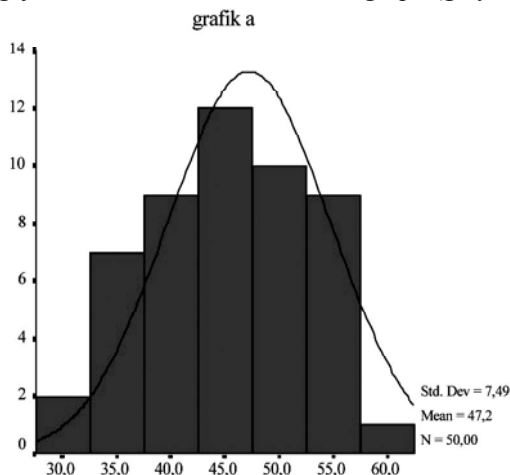
$$\chi^2 = 1.60 \chi^2_{(0.01)} = 5.99 > .04 NS (distribucija je normalna)$$

³ Maksimalno $D_{max} = .04$ je znatno niže od vrijednosti $test_{(0.01)} = .23$

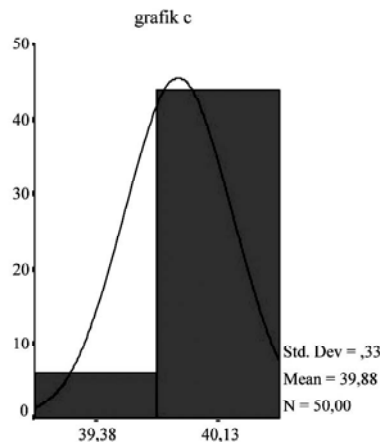
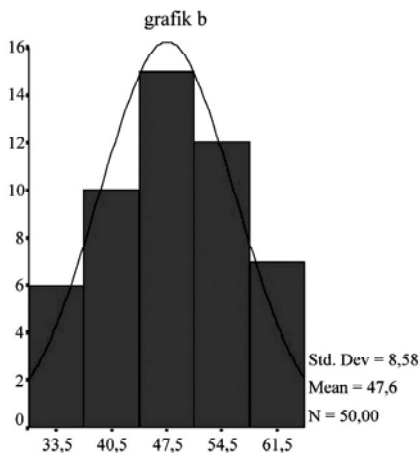
⁴ Granična vrijednost je znatno veća od χ^2_{-testa} tako da je hipoteza o statističkoj jednakosti između opaženih i očekivanih frekvencija potvrđena, čime je potvrđena hipoteza o normalnosti distribucije kao i u primjeru a). $\chi^2 = 1.60 < 5.99 N.S$

Diskriminativnost testa bila bi manja ako bi broj grupa bio pet a još manja ako kandidate svrstamo u dvije grupe npr. one koji su ostvarili do 39 bodova i nisu ispunili normu i one koji su imali više od 39 bodova i ispunili su normu. Navedeni primjeri prikazani su grafikonima (**a**, **b**, **c**)

Primjer gdje su studenti svrstani u sedam grupa (grafik - a)



Primjeri gdje su studenti svrstani u pet grupa (grafik - b) i dvije grupe (grafik - c)



U drugom primjeru (grafik b) vrijednost **K-S** testa je $D_{max} = .14$. Kako vidimo distribucija je i dalje u skladu sa normalnom, ali ostupanja su nešto veća u odnosu na prvi primjer gdje je $D_{max} = .04$

U trećem primjeru (*grafik c*) vrijednost **K-S** test je $D_{max} = .99$. S obzirom da je $D_{max} = .99$ znatno veće od $test_{(.01)} = .23$ donosimo pouzdan zaključak da distribucija u trećem primjeru bitno odstupa od normalne, te da ovakva raspodjela loše diskriminira kandidate, što je sasvim jasno i očigledno na prvi pogled.

Zaključak bi bio sledeći. Najbolja diskriminativnost je u prvom primjeru (*grafik a*), nešto lošija ali zadovoljavajuća je u drugom primjeru (*grafik b*). Loša diskriminativnost, koja se svodi na dihotomizaciju je primjer tri (*grafik c*).

LITERATURA

1. Bukvić, A. (1996). Načela izrade psiholoških testova, Zavod za udžbenike i nastavna sredstva, Beograd.
2. Fajgelj, S. (2003). Psihometrija, Metod teorija psihološkog mjerenja, Beograd.
3. Gilford, J.P. (1968). Osnove psihološke i pedagoške statistike, Savremena administracija, Beograd
4. Krković, A. (1966). Pouzdanost mjerenja psiholoških mjerenja, Društvo psihologa SRH, Zagreb.
5. Mijanović, M. (2002). Valjanost testiranja i mjerenja, Luča, Filozofski fakultet, Univerzitet u Podgorici.
6. Mijanović, M. (2004) Pouzdanost testiranja i mjerenja, Međunarodna naučna konferencija, Crnogorska sportska akademija, Bar.
7. Momirović, K. (1966). Valjanost psiholoških mjernih instrumenata, Društvo psihologa Hrvatske, SRH, Zagreb.
8. Momirović, K. (1988). Uvod u analizu nominalnih varijabli, Metodološke sveske, Jugoslovensko udruženje za sociologiju, Ljubljana.
9. Momirović, K. (1994). Jednostavan postupak za procjenu pouzdanosti, pod klasičnim modelom mjerenja, Institut za kriminološka i sociološka istraživanja, Beograd.
10. Momirović, K., Wolf, B., Popović, D. (1999). Uvod u teoriju mjerenja, Metrijske karakteristike kompozitnih mernih instrumenata, Fakultet za fizičku kulturu, Priština.
11. Rajtston, DŽ. R. (1966). Vrednovanje u savremenom obrazovanju, Savezni zavod za priručnike školskih i prosvetnih pitanja, Beograd.
12. Radžo, I., Wolf, B. (2002). Kvantitativne metode u sportu, Sarajevo
13. Wolf, B. (1988). Problem psihometrijskog vrednovanja S-R skala, Upitnici i skale za procenu ličnosti, Društvo psihologa i Filozofski fakultet, Novi Sad.
14. Yaffee, R.A. (2000). Common Correlation and Reliability Analysis with SPSS for Windows, New York University, <http://www.nyu.nyu.edu/acf/socsci/Docs/correlate.html>.