VII. STATISTIČKI TESTOVI

1. VRSTE POGREŠAKA. JAKOST TESTA

Označimo sa $f(x, \Theta)$ funkciju vjerojatnosti slučajne varijable x. Distribucija varijable x u potpunosti je određena ako je zadana funkcija f i vrijednost parametra Θ Pretpostavimo sada da za neku varijablu x poznajemo oblik funkcije f, ali da ne znamo vrijednost parametra Θ. Za vrijednost tog parametra Θ postavimo nul-hipotezu:

$$H_0 : \Theta = \Theta_0$$

i alternativnu hipotezu:

$$H_1 : \Theta = \Theta_1$$

Odluka koju od hipoteza treba prihvatiti kao istinitu donosi se na osnovi izvjesnog broja vrijednosti varijable x, drugim riječima, na temelju *uzorka*. Pokazat će se da postoji mogućnost prihvaćanja pogrešne hipoteze i da tome odgovara stanovita vjerojatnost; to je ujedno i najkarakterističnije obilježje statističkih odluka u teoriji testiranja. Pri svakom statističkom testiranju moguće su iste vrste pogrešaka, ali oblik funkcije vjerojatnosti tih pogrešaka varira od testa do testa.

Pretpostavimo da je:

$$x_1. x_2, \ldots, x_n$$

slučajni uzorak. Svakom takvom uzorku pridružimo točku $T(x_1, \ldots, x_n)$ n-dimenzionalnog prostora. Čitav prostor podijelimo zatim na dva disjunktna dijela A i B. Testiranje hipoteze H_0 prema alternativnoj hipotezi H_1 provedimo na slijedeći način: ako uzorak dade točku T iz dijela prostora A (kraće ćemo to označiti sa $T \in A$), prihvatimo hipotezu H_0 kao istinitu; na protiv, ako se desi $T \in B$, tj. točka T padne u dio prostora B, odbacimo hipotezu H_0 i prihvatimo H_1. Dio prostora A zovemo područjem prihvaćanja hipoteze H_0, a dio prostora B kritičnom domenom ili područjem signifikantnosti (sl. 61).

U mnogo slučajeva moguće je naći takvu slučajnu jednodimenzionalnu varijablu $t = t(x_1, x_2, \ldots, x_n)$ pomoću koje se izvodi testiranje hipoteze H_0. Prednost je toga da područje prihvaćanja, a stoga i područje signifikantnosti, postaju dijelovi pravca, a ne dijelovi n-dimenzionalnog prostora.

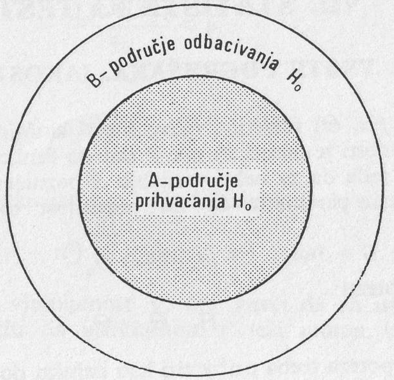
Pri opisanom načinu testiranja postoji mogućnost *dviju* vrsta pogrešaka. Pogreška *prve vrste* sastoji se od prihvaćanja hipoteze H_1 kada je stvarno istinita hipoteza H_0; vjerojatnost α takve pogreške jeste:

$$\alpha = P\{T \in B/H_0\} \tag{1}$$

jer nastaje kad točka T padne u dio B prostora, ali pri istinitosti hipoteze H_0. Kosa crtica i H_0 iza nje u vitičastoj zagradi izraza (1) ima ulogu da naglasi da je riječ o vjerojatnosti ako je istina hipoteza H_0. Mogli bismo također reći da pogreška prve vrste nastaje pri odbacivanju hipoteze H_0 kada je ona stvarno ispravna *(pogrešno odbacivanje H_0)*.

Pogreška *druge vrste* nastaje prihvaćanjem hipoteze H_0 kada je stvarno istinita hipoteza H_1 *(pogrešno prihvaćanje H_0)*. Vjerojatnost β takve pogreške pri opisanom načinu testiranja jeste:

$$\beta = P\{T \in A/H_1\} \tag{2}$$



Slika 61

Sve mogućnosti u pogledu istinitosti hipoteze H_0 i pravilnosti zaključivanja prikazane su u slijedećoj tablici:

Hipoteza H_0	Istinita	Neistinita
Odbacuje se	Pogreška 1. vrste	Pravilan zaključak
Prihvaća se	Pravilan zaključak	Pogreška 2. vrste

Osim vjerojatnosti α i β mogućih pogrešaka pri donošenju odluka, uvodi se još i pojam *jakosti* ili *moći* testa. Pod jakosti testa razumijeva se vjerojatnost odbacivanja hipoteze H_0 kada je ona zaista neispravna (vjerojatnost pravilnog odbacivanja H_0), dakle vjerojatnost:

$$p = P\{T \in B/H_1\} \tag{3}$$

Budući da je vjerojatnost da točka T padne u dio prostora A ili B jednaka 1, to je zbog (2) i (3):

$$\beta + p = 1 \tag{4}$$

odnosno:

$$p = 1 - \beta$$

Razabiremo odatle da je jakost testa p to veća, što je vjerojatnost pogreške druge vrste manja.

Vjerojatnost pogreške prve vrste α obično se unaprijed zadaje, u primjenama redovito se uzima 0,05 ili 0,01, a cijeli prostor nastoji se podijeliti na dijelove A i B tako da vjerojatnost β pogreške druge vrste bude po mogućnosti što manja.

Pretpostavimo da smo n-dimenzionalni prostor točaka T podijelili na dva načina: prvo na A i B i pri tom da vrijedi $P\{T \in B/H_0\} = \alpha$, a zatim na A_1 i B_1 i pri tom da je opet $P\{T \in B_1/H_0\} = \alpha$. Test koji se zasniva na prvoj podjeli ima istu vjerojatnost α pogreške prve vrste kao i test zasnovan na drugoj podjeli. Ako se desi da prvom testu odgovara manja vjerojatnost β pogreške druge vrste za svaku dopustivu vrijednost parametra Θ, kažemo da je test s domenama A i B *uniformno* jači od testa s domenama A_1 i B_1. U stvari to znači da ćemo pri upotrebi testa s domenama A i B činiti rjeđe pogrešku druge vrste negoli pri upotrebi testa s domenama A_1 i B_1. Ako je test s domenama A i B uniformno jači i od svakog drugog testa, kaže se da je on *uniformno najjači*. Na žalost, uniformno najjači testovi uvijek ne postoje, naime obično su neki od testova jači nad jednim područjem vrijednosti parametra Θ, a drugi opet nad drugim.

Primjer 1. Novčić X pokazuje grb uz vjerojatnost $p = \frac{1}{2}$. Novčić Y nije uravnotežen, pa pokazuje grb uz vjerojatnost $p = \frac{2}{3}$. Uzmimo jedan novčić i bacimo ga tri puta. Broj grbova u ta tri bacanja neka je slučajna varijabla t pomoću koje treba odlučiti da li je bacan novčić X ili novčić Y. Drugim riječima, t je slučajna varijabla kojom treba testirati hipotezu $H_0 : p = \frac{1}{2}$ prema alternativnoj hipotezi $H_1 : p = \frac{2}{3}$.

Naša varijabla t kojom obavljamo testiranje može poprimiti vrijednosti 0, 1, 2, 3, a distribuirana je po zakonu binomne razdiobe $B\{n = 3, p\}$. Stoga je:

$$P(t) = \binom{3}{t} p^t q^{3-t}$$

Ako je istinita hipoteza H_0 distribucija varijable t prikazana je ovom tablicom:

t	0	1	2	3
$P(t)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Pretpostavimo sada da je istinita hipoteza H_1. U tom slučaju distribuciju varijable t prikazuje tablica:

t	0	1	2	3
$P(t)$	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$

U oba slučaja vjerojatnosti $P(t)$ računane su po istoj formuli, jedino što je prvi puta uzeto $p = \frac{1}{2}$

a drugi puta opet $p = \frac{2}{3}$.

Razmotrimo testiranje hipoteze H_0 ako za područje B odbacivanja ove hipoteze uzmemo jednu jedinu vrijednost varijable t, i to $t = 0$. U tom slučaju područje A prihvaćanja hipoteze H_0 sastoji se od vrijednosti $t = 1, 2, 3$. Ukratko, ako u tri bacanja novčića bar jednom padne grb odlučimo se za izjavu da je bacan novčić X. Izračunajmo vjerojatnost pogrešaka prve i druge vrste pri ovakvom izboru domena A i B.

Po definiciji pogreške prve vrste slijedi da je njena vjerojatnost:

$$\alpha = P\{t \in B/H_0\} = P\{t = 0/H_0\} = \frac{1}{8} = 0{,}125$$

što zapravo znači da bismo u 12,5 % slučajeva izjavili da je bacan novčić Y kada je stvarno bacan novčić X (pogrešno bismo odbacili H_0). Vjerojatnost pogreške druge vrste jeste prema (2):

$$\beta = P\{t \in A/H_1\} = P\{t = 1, 2, 3/H_1\}$$

$$= \frac{6}{27} + \frac{12}{27} + \frac{8}{27} = \frac{26}{27} = 0{,}96$$

Dakle, prema našemu testu izjavili bismo u 96 % slučajeva da je bacan novčić X, kada je stvarno bacan novčić Y (pogrešno bismo prihvatili hipotezu H_0).

Područje B odbacivanja hipoteze H_0 proizvoljno smo odabrali, pa ne možemo biti sigurni da je taj test ujedno i najbolji. Da bismo pokazali da postoji i bolji test, uzmimo da domenu B' odbacivanja hipoteze H_0 predstavlja vrijednost $t = 3$; domena A' prihvaćanja hipoteze H_0 sastoji se sada od vrijednosti $t = 0, 1, 2$. Za test s ovim domenama odbacivanja i prihvaćanja hipoteze H_0 vjerojatnosti su pogrešaka prve i druge vrste:

$$\alpha' = P\{t \in B'/H_0\} = P\{t = 3/H_0\} = \frac{1}{8} = 0{,}125$$

$$\beta' = P\{t \in A'/H_1\} = P\{t = 0, 1, 2/H_1\}$$

$$= \frac{1}{27} + \frac{6}{27} + \frac{12}{27} = \frac{19}{27} = 0{,}70$$

Budući da je $\alpha = \alpha'$ i $\beta > \beta'$, bolji je test s domenom B' odbacivanja hipoteze H_0 nego li test s domenom B. Ukratko, manja je vjerojatnost pogreške druge vrste ako izjavimo da je bacan novčić Y ako je $t = 3$, negoli ako to izjavimo kada je $t = 0$.

Primjer 2. Varijabla t predstavlja proteklo vrijeme između uzastopnih reagiranja Geigerova brojača u izučavanju kozmičke radijacije. Poznato je da funkcija vjerojatnosti varijable t ima oblik:

$$f(t, c) = c\, e^{-ct},\ 0 < t < +\infty$$

gdje je c parametar koji zavisi o uvjetima pod kojima se vrši ispitivanje radijacije. Pretpostavimo da na temelju jednog jedinog mjerenja vremena koje je proteklo između dvaju uzastopnih reagiranja brojača treba testirati hipotezu $H_0 : c = 2$ prema alternativnoj hipotezi $H_1 : c = 1$. Kolike su vjerojatnosti pogrešaka prve i druge vrste ako je područje A prihvaćanja hipoteze H_0 interval $(0; 1)$?

Funkcija vjerojatnosti slučajne varijable t glasi:

$$f_0(t, 2) = 2e^{-2t}$$

ako je istinita hipoteza H_0, odnosno:

$$f_1(t, 1) = e^{-t}$$

ako je istinita hipoteza H_1. Budući da je područje B odbacivanja hipoteze H_0 interval $[1; +\infty)$, to je vjerojatnost pogreške prve vrste:

$$\alpha = P\{t \in B/H_0\} = P\{1 < t < \infty/H_0\}$$

$$= \int_1^\infty f_0(t, 2)\,dt = 2\int_1^\infty e^{-2t}\,dt = 0{,}135$$

odnosno vjerojatnost pogreške druge vrste:

$$\beta = P\{t \in A/H_1\} = P\{0 < t < 1/H_1\}$$

$$= \int_0^1 f_1(t, 1)\,dt = \int_0^1 e^{-t}\,dt = 0{,}632$$

2. NEYMAN-PEARSONOVA LEMA

Prije no što se upustimo u proučavanje pojedinih testova, moramo postaviti još jedno pitanje. Neka je zadana vjerojatnost α pogreške prve vrste; kako odrediti najbolju kritičnu domenu B (time bi ujedno bila određena i domena A prihvaćanja hipoteze H_0) u smislu da toj domeni pripada *najmanja* vjerojatnost β pogreške druge vrste?

Po pretpostavci su

$$x_1, x_2, \ldots, x_n$$

nezavisne slučajne varijable s istom funkcijom vjerojatnosti $f(x, \Theta)$. Stoga je funkcija vjerojatnosti njihove združene distribucije:

$$L(T) = f(x_1, \Theta) \cdot f(x_2, \Theta) \ldots f(x_n, \Theta)$$

Uvedemo li znak Π kao oznaku za produkt, možemo kraće pisati:

$$L(T) = \prod_{i=1}^{n} f(x_i, \Theta)$$

Za $\Theta = \Theta_0$ imamo:

$$L_0(T) = \prod_{i=1}^{n} f(x_i, \Theta_0)$$

Ako je $\Theta = \Theta_1$, onda je:

$$L_1(T) = \prod_{i=1}^{n} f(x_i, \Theta_1)$$

Stoga je vjerojatnost α pogreške prve vrste:

$$\alpha = P\{T \in B/H_0\} = \int_B L_0(T)\,dx \qquad (1)$$