TESTIRANJE HIPOTEZA

Mirjana Kujundžić Tiljak i Davor Ivanković

Struke, koje svoje nove spoznaje pretežno izvode iz podataka, u pravilu polaze od uzorka ispitanika. Na uzorku se izvode mjerenja, s rezultatima tih mjerenja se računa, dobivaju se informacije u obliku aritmetičkih sredina, varijanci, proporcija i sličnog, a onda se dobiveno želi poopćiti na populaciju iz koje je uzorak uzet.
Znanstvena hipoteza predstavlja nagađanje, naslućivanje i pretpostavke koje motiviraju istraživanje. Iz znanstvene hipoteze, odnosno hipoteze istraživača (koja je u pravilu afirmativna) izvodi se statistička hipoteza.
Statistička hipoteza iskazuje se na način da može biti vrednovana statističko-analitičkim postupcima. Statistička hipoteza matematički je izraz koji predstavlja polaznu osnovu na kojoj se temelji kalkulacija statističkog testa.
Testiranje hipoteze je statistički postupak kojim se određuje da li i koliko pouzdano raspoloživi podaci podupiru postavljenu pretpostavku. Testiranje hipoteza, odnosno testiranje značajnosti u osnovi je postupak kvantifikacije impresija o specifičnoj hipotezi.

Slijed radnji u provjeravanju (testiranju) hipoteza:

- postavljanje *nul-* i *alternativne hipoteze*;
- izbor *razine značajnosti* (α);
- prikupljanje primjerenih podataka na odgovarajućem uzorku ispitanika;
- računanje *vrijednosti rezultata statističkog testa* specifičnog za *nul-hipotezu* (H_0);
- usporedba rezultata statističkog testa s vrijednostima iz poznate distribucije vjerojatnosti specifične za dati test;
- interpretacija rezultata statističkog testa u terminima vjerojatnosti *(P-vrijednost)*.

Nul-hipoteza, H_0 (engl. null hypothesis) pretpostavka je o izostanku efekta, tj. da ne postoji razlika među uzorcima u populaciji od interesa (npr. nema razlike u aritmetičkim sredinama). To je hipoteza koja se testira, *hipoteza da nema razlike* (engl. hypothesis of no difference). Postavlja se (najvećma) u svrhu odbacivanja. Odbacuje se ili prihvaća.
Primjer H_0: u muškaraca i žena u populaciji jednak je postotak pušača.

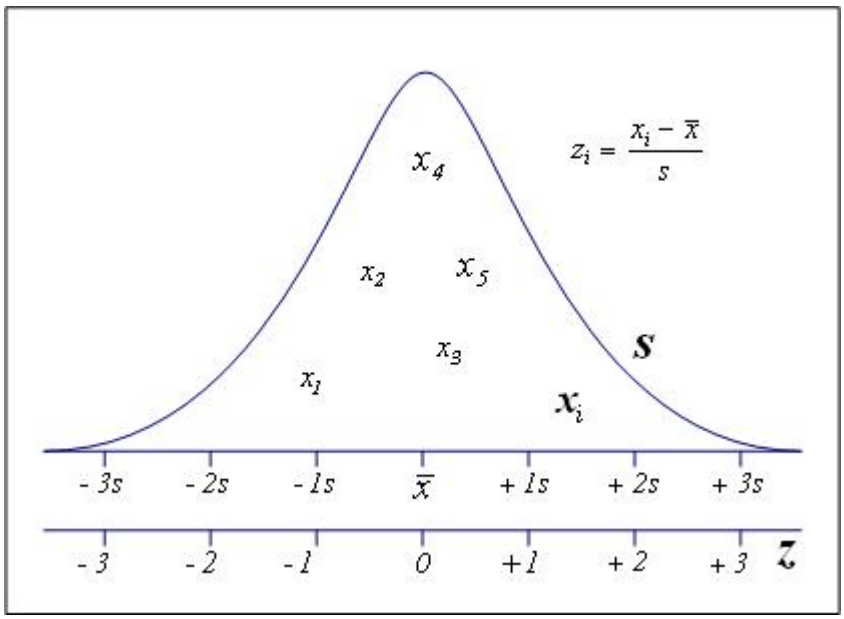
Alternativna hipoteza, H_1 (engl. alternative hypothesis) vrijedi ako nul-hipoteza nije istinita. Najčešće se direktno odnosi na teorijsku pretpostavku koja se želi istražiti, tj. često je alternativna hipoteza upravo hipoteza istraživača.
Primjer H_1: u muškaraca i žena u populaciji različit je postotak pušača

Kada se unaprijed ne može sa sigurnošću odrediti smjer neke razlike, ukoliko ona postoji, primjenjuje se *dvosmjerni test* (engl. two-tailed test). Ako na primjer, nije specificiran smjer razlike u postotku pušača, tj. da li je postotak pušača u muškaraca veći ili manji u odnosu na žene u populaciji primjenjuje se dvosmjerni test.

Jednosmjerni test (engl. one-tailed test) primjenjuje se kada je smjer efekta specificiran u alternativnoj hipotezi (H_1). Primjenjuje se znatno rjeđe; primjerice, u

istraživanju bolesti od koje svi neliječeni bolesnici umiru pa novi lijek ne može pogoršati situaciju.

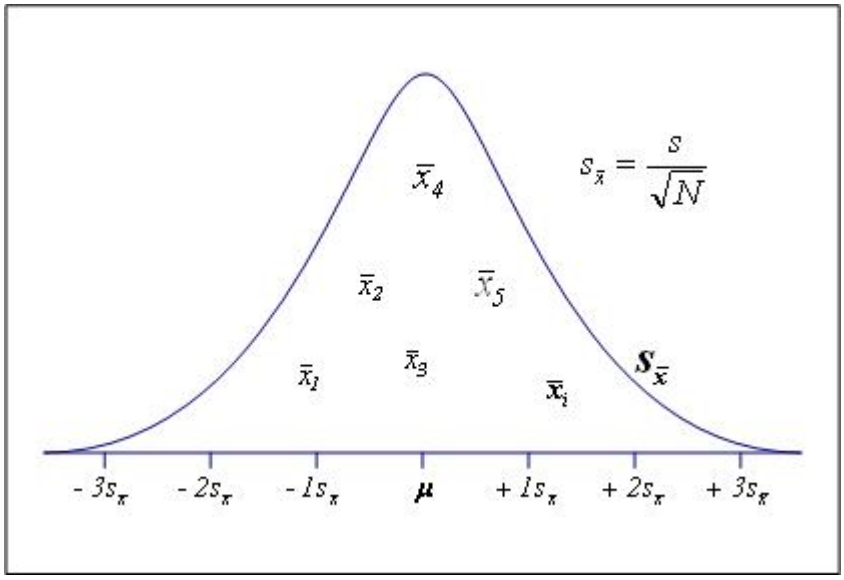
Test-statistika za određeni test određena je formulom. Uvrštavanjem podataka dobivenih mjerenjem na uzorku u takvu formulu dobiva se vrijednost test-statistike.
Pretpostavimo li neku varijablu koja se normalno raspodjeljuje, na reprezentativnom (velikom i slučajno odabranom) uzorku dobili bismo raspodjelu s procjenom parametara distribucije kao na slici 1.



Slika 1. Standardna normalna raspodjela

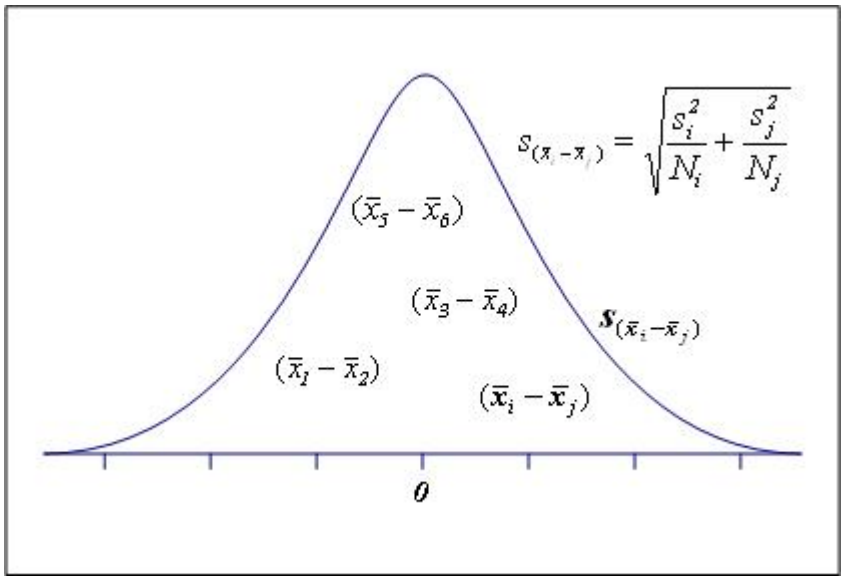
Moguće je jednostavnom algebrom odrediti položaj pojedinca u odnosu na aritmetičku sredinu uzorka izraženu u jedinicama varijabilnosti (z-vrijednost).

Kad bismo pak uzastopno iz iste populacije uzorkovali mnoštvo velikih slučajnih uzoraka bez ponavljanja a do iscrpljenja cijele populacije, aritmetičke bi se sredine tih uzoraka distribuirale normalno oko stvarne aritmetičke sredine populacije (μ) s mjerom varijabilnosti koja je standardna devijacije te konkretne normalne distribucije a naziva se, neslučajno, standardnom pogreškom. Radi se naime o pogrešci s kojom bismo procijenjivali pravu vrijednost aritmetičke sredine populacije temeljem slučajnog uzorka. To grafički prikazuje slika 2.



Slika 2. Distribucija aritmetičkih sredina uzoraka

Usporedimo li bilo koje dvije od tako dobivenih aritmetičkih sredina (a jasno je da su razlike među njima pripisive samo onome što smo s tom populacijom radili tj. slučaju) dolazimo do normalne distribucije koja očekivano ima aritmetičku sredinu jednaku nuli sa standardnom devijacijom koja se naziva standardnom pogreškom razlika aritmetičkih sredina, što prikazuje slika 3.



Slika 3. Distribucija razlika aritmetičkih sredina

To je zapravo ilustracija nul-hipoteze. Položaj pojedinca (ovdje razlike dviju aritmetičkih sredina) u odnosu na centar distribucije, izražen u jedinicama varijabilnosti za tu konkretnu distribuciju dade se prikazati kao z-vrijednost koja u tom jednostavnom slučaju istraživanja razlike između dvaju slučajnih, velikih i neovisnih uzoraka čini Studentov test (t-test) i izgleda ovako:

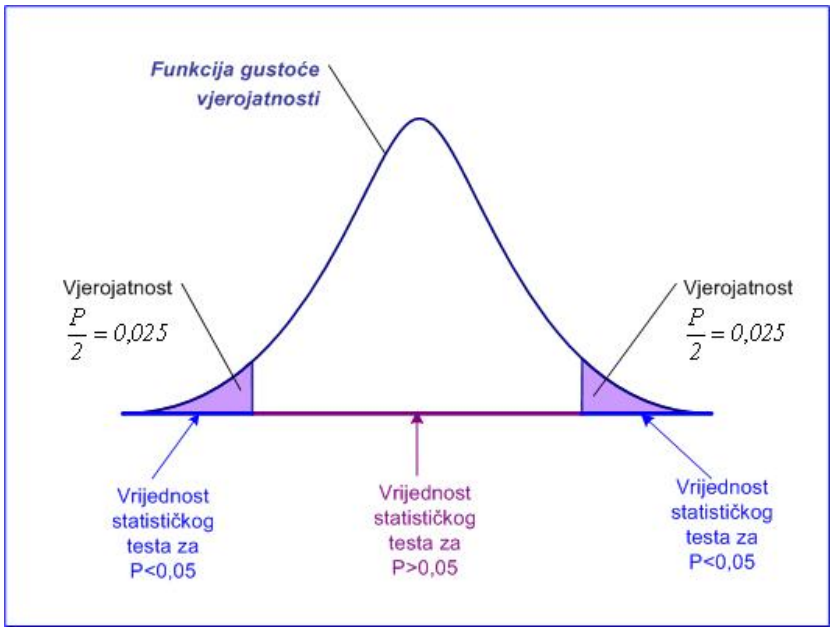
$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

Dobili smo vrijednost *slučajne varijable t*. Uz pretpostavku da distribucije podataka iz svakog pojedinog uzorka imaju normalnu distribuciju, slučajna varijabla izračunata prema prethodnoj formuli ima t-distribuciju. Kako je t-distribucija potpuno određena s parametrom *„broj stupnjeva slobode"*, preostaje nam da odredimo broj stupnjeva slobode za konkretni slučaj. Broj stupnjeva slobode, DF (engl. degrees of freedom), izračunava se po formuli:

$$DF = n_1 + n_2 - 2$$

Radi interpretacije rezultata testiranja treba pogledati u tablicu s graničnim vrijednostima za t-distribuciju.
Ako je izračunata vrijednost test-statistike t veća od granične vrijednosti t_g pročitane u tablici za određenu razinu značajnosti α, onda nul-hipotezu odbacujemo s vjerojatnošću $1-\alpha$. Naime, $t > t_g$ znači da je vjerojatnost da iz uzoraka računajući dobijemo veću vrijednost za t od graničnog t_g manja od α (manja od 0,05 ili 0.01 ili čak 0,001), što znači da razliku u aritmetičkim sredinama ne možemo pripisati slučaju. Drugim riječima, zaključak glasi: Uzorci A i B nisu uzeti iz iste populacije. Ta tvrdnja vrijedi s vjerojatnošću $1-\alpha$.

Valja, dakle, vidjeti da li se rezultat nalazi dovoljno blizu centru distribucije (0-hipotezi) što bi sugeriralo da hipotezu valja prihvatiti ili pak dovoljno daleko što bi nam omogućilo da hipotezu odbacimo, ne griješeći previše. Opisanu distribuciju prikazuje slika 4.

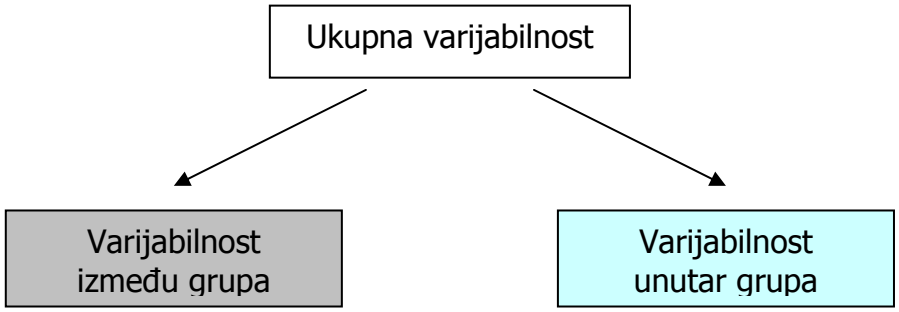


Slika 4. Distribucija vjerojatnosti statističkog testa s dvosmjernom vjerojatnošću, P =0,05

Pri izboru statističkih testova razlike (jednako tako i drugih statističkih testova) valja voditi računa o *tipu problema* koji želimo riješiti. Treba razmotriti *prirodu varijabli* (podataka) odnosno skale mjerenja, zatim imamo li jednu skupinu podataka, jedan uzorak, ili više njih, jesu li uzorci zavisni ili nezavisni. Ponavljanje mjerenja na istom uzorku rezultira zavisnim uzorcima (podataka). Primjerice želimo li ustanoviti da li određeni dodatak prehrani značajno povećava težinu eksperimentalnih životinja, izmjerit ćemo težinu životinja dva puta: prvi put na početku eksperimenta, prije prelaska na novu hranu i drugi put, nakon određenog vremena kroz koje su životinje dobivale ispitivani dodatak. Na taj način se dobiju dva uzorka podataka, težine na početku i težine na kraju eksperimenta. Ovdje se radi o *zavisnim uzorcima*. To će se odraziti i u formuli primjerenoj testiranju hipoteze o nepostojanju razlika. Ova će u nazivniku sadržavati mjeru varijabilnosti koja je primjereno umanjena za „srodnost" podataka.
Ako pak eksperiment planiramo na način da uzmemo dvije po težini podjednake skupine eksperimentalnih životinja, jednu skupinu ostavimo na standardnoj prehrani a drugoj u hranu dodamo ispitivani dodatak, onda će takav eksperiment rezultirati opet s dvije skupine podataka. Jednu skupinu podataka činit će težine skupine eksperimentalnih životinja na standardnoj prehrani a drugu težine životinja koje su jele hranu s ispitivanim dodatkom. No kako se radi o dvjema skupinama životinja, njihovi podaci neće biti zavisni, pa se ovdje radi o *nezavisnim uzorcima*.

ANALIZA VARIJANCE ZA NEZAVISNE UZORKE

T-test je primjenjiv isključivo kada imamo dva uzorka. Međutim često u istraživanjima imamo više uzoraka. Analiza varijance je test koji se primjenjuje u takvim slučajevima. Primjena analize varijance bit će moguća ako je mjerena varijabla normalno distribuirana i ako su varijance svih promatranih uzoraka jednake. Ideja analize varijance sastoji se u razdvajanju varijabilnosti mjerenog varijancom na dva dijela: varijabilnost među ispitanicima koji pripadaju različitim grupama odnosno uzorcima (engl. between-group variation) i varijabilnost među ispitanicima unutar svake pojedine grupe odnosno uzorka (engl. within-group variation). Ovaj drugi dio varijabilnosti često se naziva neobjašnjenom ili rezidualnom varijabilnošću.



Test se temelji na omjeru tih dviju varijabilnosti.

Nul-hipoteza koja se testira glasi:

$$H_0: \mu_1=\mu_2=\mu_{3=\ldots=}=\mu_k$$

ili aritmetičke sredine populacija iz kojih su uzeti uzorci jednake su

nasuprot alternativnoj hipotezi:

$$H_1: \mu_i \neq \mu_j$$

ili bar jedna se aritmetička sredina razlikuje od preostalih

U postupku primjene analize varijance prvo treba izračunati aritmetičke sredine za svaki od k uzoraka, zatim varijabilnost među ispitanicima između grupa (uzoraka) i varijabilnost među ispitanicima unutar grupa (uzoraka). Nakon toga treba načiniti test-statistiku F kao omjer varijabilnosti *između* grupa i varijabilnost *unutar* grupa, te broj stupnjeva slobode DF1=k-1 i DF2=n-k, pri čemu je k broj grupa (uzoraka) a n je ukupan broj ispitanika ($n=n_1+n_2+\ldots+n_k$, n_i je broj ispitanika u i-toj grupi odnosno uzorku).
Test-statistika F je slučajna varijabla koja ima F-distribuciju s DF1 i DF2 stupnjeva slobode.
Uz razinu značajnosti α koju istraživač sam izabire s obzirom na moguće posljedice (najčešće 0,05 ili 0,01) interpretacija rezultata je ista kao i u slučaju interpretacije t-testa. Drugim riječima nul-hipoteza – aritmetičke sredine svih grupa (uzoraka) su jednake – bit će prihvaćena ako je granični F veći od izračunatog. Ako je pak granični F manji od izračunatog, onda nul-hipotezu treba odbaciti. Drugim riječima, prihvatit će se hipoteza da se bar jedna aritmetička razlikuje od preostalih.
Postoji veliki izbor tzv. *post hoc* tj. testova koji se izvode nakon ANOVA-e kad ova dade statistički značajan rezultat. Nazivaju se još i *testovi višestruke usporedbe*. Svrha im je da otkriju koje razlike (između kojih od više uzoraka) su „zaslužne“ za ukupno statistički značajan rezultat. Ima više verzija ovakvih testova, a nazivaju se po autorima koji su ih opisali (Bonferroni, Student-Newman-Keuls - SNK, Tukey, Dunnett, Scheffé, Duncan itd.)

Razlikujemo parametrijske i neparametrijske statističke testove. *Parametrijski statistički testovi* su testovi koji, da bi se smjeli primijeniti u statističkom testiranju, postavljaju zahtjeve na distribuciju izvornih podataka. *Neparametrijski testovi* ne postavljaju takav tip zahtjeva, premda i oni pretpostavljaju određene uvjete pod kojima ih se smije primijeniti. Neparametrijske testove često zovu još i *testovima nezavisnim o distribucijama podataka* (engl. distribution free tests).
Neparametrijski testovi koriste se kod malih uzoraka, kada je nemoguće odrediti distribuciju podataka ili kada je primjenjena kategorijska skala mjerenja. Imaju manju snagu otkrivanja stvarnog efekta u odnosu na ekvivalentni parametrijski test.

Koji statistički test će se primijeniti u analizi prikupljenih podataka ovisi o *dizajnu studije*, *tipu varijable* koja se testira te o *raspodjeli* koju slijede istraživani podaci.

NEKI NAJČEŠĆE UPOTREBLJAVANI STATISTIČKI TESTOVI RAZLIKE

PARAMETRIJSKI TESTOVI	NEPRAMETRIJSKI TESTOVI	SVRHA TESTA
Studentov ili t-test za nezavisne uzorke	Mann-Whitney U-test Wald-Wolfowitz test	Usporedba dva nezavisna uzorka koji su uzeti iz iste populacije
t-test diferencija ili t-test za zavisne uzorke	Wilcoxon test sume rangova	Usporedba dva seta opažanja na istom uzorku
Jednosmjerna analiza varijance (F-test, ANOVA)	Kruskall-Wallis analiza varijance rangova (H-test) Medijan test	Usporedba više od dva nezavisna uzorka koji su uzeti iz iste populacije
Analiza varijance s ponavljanim mjerenjima	Freedmanova analiza varijance Kendall's W-test Cochran's Q-test	Usporedba više od dva seta opažanja na istom uzorku

Pri interpretaciji rezultata statističkih testova valja imati na umu da se *hipoteza ne dokazuje!* Rezultat statističkog testa ukazuje samo da li dostupni podaci podržavaju ili ne podržavaju hipotezu, tj. koliko bi hipoteza mogla biti prihvatljivom ili neprihvatljivom, dakako uz određenu vjerojatnost.
Na temelju podataka prikupljenih na uzorku ispitanika statističko-matematičkim postupkom izračunaju se *vrijednosti specifičnog statističkog testa*. Na temelju vrijednosti rezultata izračunatih statističkim testom donosi se odluka o odbacivanju ili prihvaćanju nul-hipoteze.
Dobivene vrijednosti statističkog testa reflektiraju veličinu dokaza *protiv* nul-hipoteze u ispitivanom uzorku. Dakle, što je veća apsolutna vrijednost statističkog testa (tj. bez obzira na njezin predznak + ili -), veći je i dokaz protiv nul-hipoteze, tj. manja je vjerojatnost da je nul-hipoteza istinita.

Sve vrijednosti statističkih testova slijede neku poznatu teorijsku distribuciju vjerojatnosti.. *Distribucija statističkog testa* distribucija je vjerojatnosti za vrijednosti konkretnog testa. Površina ispod krivulje predstavlja statističku vjerojatnost (P-vrijednost).

Sve moguće vrijednosti rezultata statističkog testa *točke su na horizontalnoj osi grafa* distribucije vjerojatnosti statističkog testa; na vertikalnoj osi je vjerojatnost. Razlikujemo dvije grupe vrijednosti: grupa vrijednosti u *području odbacivanja* H_0 te grupa vrijednosti u *području prihvaćanja* H_0. Prema tome u koje od ovih područja pada vrijednost statističkog testa prihvaća se ili odbacuje nul-hipotezu, odnosno odbacuje se ili prihvaća alternativnu hipotezu.

Odluka o tome koje vrijednosti spadaju u područje odbacivanja, a koje u područje prihvaćanja donosi se na temelju *razine značajnosti* (α) Razina značajnosti (α) određuje *površinu ispod distribucijske krivulje* vrijednosti statističkog testa koja je iznad vrijednosti na horizontalnoj osi u području odbacivanja H_0. *Razina značajnosti*

(α) predstavlja graničnu vjerojatnost uz koju još uvijek valja prihvatiti eventualno istinitu nul-hipotezu.

P-vrijednost, tj. vjerojatnost površina je ispod oba (ili u posebnim slučajevima jednog) kraja distribucije vjerojatnosti od izračunate vrijednosti statističkog testa. Većina računalne statističke programske podrške automatski izračuna dvosmjernu P-vrijednost.

Primjer: razina značajnosti $\alpha = 0,05$:
- ako je $P < 0,05$ odbacuje se nul-hipoteza, tj. rezultati su statistički značajni (signifikantni) na 5% razini značajnosti[1];
- ako je $P \geq 0,05$ prihvaća se nul-hipoteza; tj. rezultati nisu statistički značajni (signifikantni) na 5% razini značajnosti, odnosno nema dovoljno dokaza za odbacivanje nul-hipoteze.

Izbor veličine razine značajnosti statističkog testa (0,05; 0,01; 0,01; 0,001) je proizvoljan. Veličina razine značajnosti govori o tome u kojem postotku si istraživač dopušta načiniti grešku odbacivanja istinite nul-hipoteze. U situacijama kada su kliničke implikacije odbacivanja nul-hipoteze ozbiljne, valja zahtijevati snažnije dokaze za odbacivanje nul-hipoteze (razina značajnosti od 0,01 ili 0,001).

Navođenje rezultata uz određenu razinu značajnosti (npr. $P<0,05$ ili $P>0,05$) može navesti na krivi zaključak. Na primjer, ako je $P=0,04$ odbacuje se H_0, a ako je $P=0,06$ ne odbacuje se H_0. Postavlja se pitanje važnosti, odnosno značenja utvrđene razlike. Preporuka je, a danas i zahtjev većine uvaženih medicinskih časopisa, da se navede precizna P-vrijednost što je danas lako rješivo uz pomoć računala i primjenu dostupnih statističkih programskih podrški kao što su SAS, Statistica, SPSS, S-plus i druge, budući da one uz rezultat statističkog testa i ekspliciraju *egzaktnu P-vrijednost*.

Budući da se zaključivanje u statistici provodi na temelju informacija o uzorku, moguće je pogriješiti i donijeti krivu odluku.

Greška tipa I nastaje kada se odbaci istinita nul-hipotezu. Razina značajnosti statističkog testa α (alpha) predstavlja maksimalnu šansu, odnosno vjerojatnost da se načini greška tipa I. Veličinu razine značajnosti valja odrediti prije prikupljanja podataka.

Greška tipa II nastaje kada se ne odbaci neistinita nul-hipotezu te zaključi da nema efekta kada on stvarno postoji. Šansa da se načini greška tipa II naziva se β (beta).

---

[1]*Budući da je korijen riječi značajnost po nekima zapravo značaj ili karakter, statistička značajnost kao konvencionalna sintagma možda i nije najbolje odabrana. Možda bi bilo bolje govoriti o statistički znakovitim razlikama (ili pak sličnostima, povezanostima i sl.) jer znakovito znači da bi nađeni rezultat mogao upućivati na stvarnu razliku, sličnost, povezanost. Alternativni izbor pojma znatan vjerojatno ne bi bio primjeren. Razlike naime ne moraju biti velike da bi bile statističke upućujuće, tj. znakovite, ili kako je to uobičajeno reći, statistički značajne.*

Saga testa računa se kao (1- β) te predstavlja vjerojatnost odbacivanja neistinite nul-hipoteze. Snaga testa zapravo je šansa detektiranja, kao statistički značajnog, određenog realnog efekta liječenja. Uobičajeno se izražava u postotcima.

Na snagu statističkog testa utječu:

- veličina uzorka - snaga raste kako raste veličina uzorka;
- varijabilitet opažanja - snaga pada kako raste varijablitet opažanja;
- efekt od interesa (efekt razlike) - snaga je veća što je veći efekt;
- razina značajnosti (α) - snaga je veća što je veća razina značajnosti.

Uočljivo je da je parametar na koji je najlakše utjecati zapravo veličina uzorka. Stoga se tzv. *analiza snage (engl. power analysis)* koristi za izračunavanje potrebne veličine uzorka za istraživanje s visokom vjerojatnošću okrivanja stvarnog efekta zadane veličine

Snagu predloženog statističkog testa važno je poznavati već u stadiju planiranja istraživanja. Adekvatna snaga statističkog testa ukazuje da on ima "dobru" šansu otkrivanja klinički relevantnog efekta, ako on postoji. Snaga "dobrog" statističkog testa trebala bi biti barem 70-80%. Etički je neprihvatljivo, a također je i gubitak vremena i sredstava, provoditi klinička istraživanja koja imaju manju šansu otkrivanja realnog efekta liječenja.

.

Literatura:

1. *Ivanković D, i sur. Osnove statističke analize za medicinare. Zagreb: Medicinski fakultet Sveučilišta u Zagrebu, 1989.*
2. *Petrie A, Sabin C. Medical Statistics at a Glance (2nd Ed). Oxford: Blackwell Science Ltd, 2005.*
3. *Glantz. SA. Primer of Biostatistics (4th Ed). New York: McGraww-Hill: 1997.*
4. *Altman DG. Practical Statistics for Medical Research. London. Chapman & Hall, 1991.*
5. *Bland M. An Introduction to Medical Statistics (3rd Ed). Oxford: Oxford University Press, 2005.*
6. *Armitage P, Berry P. Statistical Methods in Medical Research. Oxford: Blackwell Science Ltd, 1994.*