Statistika
(za biologe, 4.dio)

Bojan Basrak

rujan 2006

Testiranje statističkih hipoteza

Ponekad je na osnovu skupljenih podataka i vjerojatnosnog modela nužno donijeti binarnu odluku o populaciji odn. parametrima samog modela.

Primjer 1

i) Novi lijek je efikasniji od do sada korištenog?

ii) Podaci prikupljeni na uzorku životinja potvrdjuju, do sada prihvaćenu znanstvenu teoriju.

iii) Nakon ispitivanja javnog mišljenja, moramo odlučiti da li će novi proizvod biti uspješan na tržištu?

Ponekad za prikupljene podatke lako nadjemo vjerojatnosni model.

Primjer 2

i) Standardni lijek nakon jednomjesečne terapije uzrokuje poboljšanje u 60% pacijenata. U uzorku od 420 pacijenata novi lijek je doveo do poboljšanja u njih 343. Moramo odlučiti da li je novi lijek efikasniji.

ii) Nakon 800 bacanja novčića trebamo odlučiti da li je novčić nepristran.

iii) Od 80 predloženih članova porote u jednoj južnoj državi SAD, samo ih je 4 crna, iako je njihov udio u populaciji oko 50%. Da li je to u skladu s pretpostavkom da svi gradjani imaju jednaku vjerojatnosti biti predloženi za člana porote ili ne?

Ako podatke koje smo prikupili možemo opisati vjerojatnosnim modelom u kojem nam je nepoznat parametar θ, često je moguće i hipotezu koju želimo testirati izraziti preko nepoznatog parametra na sljedeći način:

$$H_0 : \theta \in \Theta_0.$$

Njoj suprotstavljenu hipotezu, slično možemo izraziti kao

$$H_A : \theta \in \Theta_A.$$

Ovdje Θ_0 i Θ_A predstavljaju dva disjunktna skupa u prostoru svih mogućih vrijednosti parametra θ.

U prethodnom primjeru prirodan model za podatke je Bernoullijeva razdioba (odn. binomna). Tada se hipoteze mogu izraziti u obliku

$$H_0 : p \in B_0 \subseteq [0, 1].$$

Njoj suprotstavljenu hipotezu, slično možemo izraziti kao

$$H_A : p \in B_A = [0, 1] \setminus B_0.$$

Dvije hipoteze

U postupku testiranja dvije hipoteze **nisu** ravnopravne. Prvu od njih zovemo **nul-hipoteza** i označavamo s H_0. Ona pretpostavlja tipično da se prikupljene numeričke vrijednosti daju objasniti slučajem (npr. novi lijek nije značajno bolji od standardnog, novčić je nepristran, kao i postupak izbora porote). Dakle ako mislimo da bi statistički test mogao ukazati na istinitost neke tvrdnje, nul-hipoteza H_0 predstavlja njenu negaciju.

Prije svega testom utvrdjujemo da li imamo dovoljno dokaza da bismo odbacili H_0, odn. da li su podaci sukladniji nekoj drugoj tvrdnji o parametrima. Tu drugu hipotezu zovemo **alternativna hipoteza**, u oznaci H_A.

Primjer 2 (nastavak) Nul odn. alternativnu hipotezu u prethodnom primjeru je prirodno postaviti na sljedeći način.

i) $H_0 : p \leq 0.6$ i $H_A : p > 0.6$.

ii) $H_0 : p = 0.5$ i $H_A : p \neq 0.6$.

iii) $H_0 : p \geq 0.5$ i $H_A : p < 0.5$.

Dvije vrste pogreške

Očito možemo napraviti dvije vrste pogreške u testiranju.

Pogreška 1. vrste

Odbacujemo H_0 kad je ona točna.

Pogreška 2. vrste

Ne odbacujemo H_0 kad je ona pogrešna.

Kako je uzorak slučajan, a ishod testiranja ovisi o njemu možemo se pitati koliko su vjerojatne ove pogreške uz unaprijed odredjenu proceduru. Stoga uvodimo dvije vjerojatnosti:

Vjerojatnost pogreške 1. vrste

$$\alpha = \alpha_\theta = P\left(H_0 \text{ odbacujemo } \middle| H_0 \text{ istinita}\right),$$

Vjerojatnost pogreške 2. vrste

$$\beta = \beta_\theta = P\left(H_0 \text{ ne odbacujemo } \middle| H_0 \text{ nije istinita}\right).$$

Primjetite da procedura koja nikad ne odbacuje H_0, ima $\alpha = 0$, takodjer ako uvijek odbacujemo H_0, možemo postići $\beta = 0$. Cilj je naravno naći proceduru testiranja tako da imamo obje vjerojatnosti male istovremeno.

Kod testiranja tipično postavljamo gornju ogradu na vjerojatnost pogreške 1. vrste α, a zatim tražimo test koji ima malu vjerojatnost pogreške 2. vrste.

Ovako postavljena ograda se i sama tipično označava s α i naziva **nivo značajnosti** (signifikantnosti) testa.

Testna statistika

Nakon što formuliramo hipoteze H_0 i H_A, te nivo značajnosti testa, moramo izabrati **testnu statistiku** T i odrediti **kritično područje** C_α.

Kritično područje ovisi o α, i odredjujemo ga (prije nego što smo vidjeli podatke) tako da vrijedi

$$P(T \in C_\alpha | H_0 \text{ je istina}) \leq \alpha.$$

Kako je α tipično mala vrijednosti, kritično područje zapravo izabiremo tako da je vjerojatnost da testna statistika upadne u njega mala pod pretpostavkom da vrijedi nulhipoteza.

Nakon prikupljanja podataka, iz uzorka izračunamo vrijednost **testne statistike** $T = t$, te donosimo odluku na sljedeći način

$$t \in C_\alpha, \quad \text{odbacujemo } H_0,$$

ili

$$t \notin C_\alpha, \quad \text{ne odbacujemo } H_0.$$

Za ovakav test se sada lako vidi da je vjerojatnost pogreške 1. vrste

$$\alpha_\theta = P\left(T \in C_\alpha | H_0 \text{ istinita}\right),$$

dok je vjerojatnost pogreške 2. vrste

$$\beta_\theta = P\left(T \notin C_\alpha | H_0 \text{ nije istinita}\right).$$

Uočite da smo kritično područje C_α izabrali tako da je

$$\alpha_\theta = P\left(T \in C_\alpha | H_0 \text{ istinita}\right) \leq \alpha.$$

Testiranje hipoteza o parametru p binomne razdiobe

U primjenama ove procedure tipično želimo testirati na osnovi uzorka da li je proporcija jedinki sa zadanim karakteristikama u nekoj populaciji u skladu s našim ili tudjim pretpostavkama.

Primjer 2 (nastavak, za iii dio)
Ako su članovi porote izabrani iz populacije na slučajan način, tj. tako da svi gradjani u toj populaciji s pravom glasa imaju jednaku vjerojatnost biti članovima porote, prirodno je pretpostaviti da je broj crnih članova porote S binomna sl. varijabla s parametrima 80 i p. Ovdje p predstavlja vjerojatnost da je pojedini sl. odabrani glasač crn. Kako nas brine mogućnost da su crni gradjani nedovoljno reprezentirani u poroti, možemo postaviti hipoteze na sljedeći način:

$$H_0 : p = 0.5 \text{ i } H_A : p < 0.5.$$

Dakle, želimo detektirati da li se vjerojatnost p razlikuje od pretpostavljene vrijednosti 0.5 i to tako da je strogo manja. Za ovakav tip alternativne hipoteze kažemo da je jednostrana. Pretpostavimo da postavimo nivo značajnosti testa $\alpha = 0.05$.

Sada moramo odabrati testu statistiku. Mi ćemo uzeti već poznatu vrijednost

$$Z = \frac{S - n/2}{\sqrt{n\frac{1}{2}\frac{1}{2}}},$$

jer znamo da je pod pretpostavkom da vrijedi nul-hipoteza Z približno $N(0,1)$ distribuirana.

Moramo odrediti i C_α. Uočite da ako vrijedi H_A očekivana vrijednost testne statistike bit će manja od 0, zato postavljamo

$$C_\alpha = (-\infty, -z_\alpha].$$

Jasno je da je pod $H_0 : p = \frac{1}{2}$

$$P(Z \in C_\alpha) = P(Z \leq -z_\alpha) \approx \alpha.$$

U našem primjeru je $\alpha = 0.05$, pa je $z_\alpha = 1.65$, a kako je $n = 80$ i $S = 4$, dobijemo

$$Z = \frac{4 - 20}{\sqrt{20}} = -8.04,$$

pa odbacujemo nulhipotezu.

Općenito, razmatramo 3 mogućnosti za alternativnu hipotezu o parametru p ako je želimo usporediti s nekom fiksnom vrijednošću p_0:

i) $H_0 : p = p_0$ i $H_A : p > p_0$.

ii) $H_0 : p = p_0$ i $H_A : p \neq p_0$.

iii) $H_0 : p = p_0$ i $H_A : p < p_0$.

Prva i treća od alternativnih hipoteza su **jednostrane**, dok je druga **dvostrana**. Za testnu statistiku možemo uzeti

$$Z = \frac{S - np_0}{\sqrt{np_0(1-p_0)}},$$

koja pod H_0 ima približno $N(0,1)$ razdiobu za velike uzorke.

Kritična područja su redom:

i) $C_\alpha = [z_\alpha, \infty)$.
ii) $C_\alpha = \mathbb{R} \setminus (-z_{\alpha/2}, z_{\alpha/2})$.
iii) $C_\alpha = (-\infty, -z_\alpha]$.

Ako dakle Z padne u neko od njih, odbacili bismo H_0 u korist odgovarajuće alternative H_A.

Test i p-vrijednost testa

Kod odlučivanja o ne/odbacivanju nul-hipoteze na osnovu zadane testne statistike za koju znamo konstruirati kritično područje C_α za svaki α, moguće je postupiti i na sljedeći način:

▷ za dane H_0 i H_A, te testnu statistiku T odredimo iz podataka vrijednost testne statistike $T = t$, a zatim

▷ nadjemo najmanji p tako da je $t \in C_p$. Ovakav p ovisi o uzorku i zove se **p-vrijednost** testa. Uočite, p-vrijednost je najmanji broj p takav da bismo na osnovu $T = t$ uz nivo značajnosti p odbacili nulhipotezu,

▷ nul-hipotezu odbacujemo na nivou značajnosti α ako je p-vrijednost manja od α.

Možemo reći da p-vrijednost zapravo kaže koliko je vjerojatno toliko ekstremno odstupanje testne statistike ako bi bilo istina da vrijedi H_0.

Primjer 2 (nastavak, za iii dio još jednom)

Naša testna statistike iznosi $Z = -8.04$, a p-vrijednost u ovom slučaju je

$$P(Z < -8.04) \approx \Phi(-8.04) = 4.5 \cdot 10^{-16}.$$

Mi bismo naravno odbacili nulhipotezu o nepristranosti izbora porote uz bilo koji tipično korišteni nivo značajnosti α. No p-vrijednost nam govori i koliko su malo vjerojatni podaci uz pretpostavku da H_0 zaista vrijedi. U ovom slučaju dobivena p-vrijednost je manja od vjerojatnosti da ćete bacajući novčić 50 puta zaredom dobiti sama pisma. Dakle, ekstremno mala.

Složena nul-hipoteza

Na posljednjim slajdovima smo pretpostavljali da je nul-hipoteza uvijek **jednostavna**, tj. oblika $H_0 : p = 0.5$, no ponekad bismo željeli testirati i složenu nulhipotezu, npr. $H_0 : p_0 \geq 0.5$. Općenito, pretpostavimo da želimo testirati:

$H_0 : p \leq p_0$ i $H_A : p > p_0$,

ili

$H_0 : p \geq p_0$ i $H_A : p < p_0$.

U oba slučaja, **test provodimo jednako** kao i da je $H_0 : p = p_0$, što intuitivno argumentiramo činjenicom da ako su podaci ekstremni u odn. na $p = p_0$ onda su još ekstremniji ako je p još manja odn. veća u odnosu na alternativnu hipotezu.

Testiranje hipoteza o parametru μ normalne razdiobe

Neka $X_1, \ldots, X_n$, $n \geq 1$, čine sl. uzorak iz normalne razdiobe s parametrima μ i σ^2, te neka su nam oba parametra nepoznata. Pretpostavimo da želimo testirati:

i) $H_0 : \mu = \mu_0$ i $H_A : \mu > \mu_0$.

ii) $H_0 : \mu = \mu_0$ i $H_A : \mu \neq \mu_0$.

iii) $H_0 : \mu = \mu_0$ i $H_A : \mu < \mu_0$.

Za testnu statistiku odabiremo

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{\hat{s}}.$$

Za T znamo da ima Studentovu t-razdiobu s $n-1$-im stupnjem slobode, ako vrijedi H_0 .

Stoga testove provodimo odredjujući kritično područje na sljedeći način

i) $C_\alpha = [t_\alpha, \infty)$.

ii) $C_\alpha = \mathbb{R} \setminus (-t_{\alpha/2}, t_{\alpha/2})$.

iii) $C_\alpha = (-\infty, -t_\alpha]$.

Ovdje je t_α kao i prije $(1-\alpha)$-kvantil t razdiobe s $n-1$ stupnjem slobode.

U slučaju jednostranih alternativnih hipoteza i) odn. iii), postupak testiranja bismo izveli jednako i da je H_0 složena hipoteza oblika i) $H_0 : \mu \leq \mu_0$ ili iii) $H_0 : \mu \geq \mu_0$.

Naime ako imamo dovoljno razloga za odbaciti $H_0 : \mu = \mu_0$ u korist $H_A : \mu > \mu_0$, tada je jasno da imamo još jače razloge za odbaciti $H_0 : \mu < \mu_0$ ili $H_A : \mu > \mu_0$.

Primjer (t-test o trajanju putovanja od kuće do Rooseveltovog trga) Na uzorku od 37 studenata, nadjena je aritmetička sredina trajanja putovanja $\bar{X} = 39.03$ min i standardna devijacija $\hat{s} = 24.14$ min. Pretpostavimo da su podaci normalno distribuirani (što je vrlo upitno ovdje, v. qq-plot iz prethodnog poglavlja), te da želimo testirati da li je očekivano trajanje putovanja 30 min, kako bi možda netko mogao tvrditi. Postavimo:

$H_0 : \mu = 30$ nasuprot $H_A : \mu \neq 30$, $\alpha = 0.01$, te odredimo

$$T = \sqrt{37}\frac{\bar{X} - 30}{\hat{s}} = -2.27$$

Kritično područje je

$$C_{0.01} = \mathbb{R} \setminus (-t_{0.005}, t_{0.005}) = (-2.71, 2.71)^c.$$

Kako T ne leži u njemu ne možemo odbaciti nul-hipotezu na ovom nivou značajnosti. Uvjerite se ipak da bismo je odbacili da smo postavili $\alpha = 0.05$.

Možemo izačunati i p vrijednosti u ovom slučaju kao 0.0289, pa je i iz nje jasno da za $\alpha = 0.01$ ne možemo odbaciti nul-hipotezu.

Testiranje hipoteza o očekivanju proizvoljne razdiobe na osnovu velikog uzorka

Neka $X_1, \ldots, X_n$, $n \geq 1$, čine sl. uzorak iz proizvoljne razdiobe s očekivanjem μ i varijancom $0 < \sigma^2 < \infty$, te neka su nam oba ova broja nepoznata. Pretpostavimo da želimo testirati:

i) $H_0 : \mu \leq \mu_0$ i $H_A : \mu > \mu_0$.

ii) $H_0 : \mu = \mu_0$ i $H_A : \mu \neq \mu_0$.

iii) $H_0 : \mu \geq \mu_0$ i $H_A : \mu < \mu_0$.

Za testnu statistiku ponovo odabiremo

$$T = \sqrt{n}\frac{\bar{X} - \mu}{\hat{s}}.$$

Ako je n dovoljno velik, iz centralnog graničnog teorema slijedi da T ima približno $N(0,1)$ razdiobu ako vrijedi $\mu = \mu_0$.
Kritična područja su sada:

i) $C_\alpha = [z_\alpha, \infty)$.
ii) $C_\alpha = \mathbb{R} \setminus (-z_{\alpha/2}, z_{\alpha/2})$.
iii) $C_\alpha = (-\infty, -z_\alpha]$.

Ako dakle Z padne u neko od njih, odbacili bismo H_0 u korist odgovarajuće alternative H_A.

Napomena Ovaj test možemo koristiti i za parametar μ normalne razdiobe uz uvjet da nam je poznata standardna devijacija σ i da njenu vrijednost uvrstimo u formulu za T umjesto procjene $\hat{s}$.

Snaga testa

Primjetite da za zadane hipoteze i nivo značajnosti možemo pronaći više testnih statistika T i kritičnih područja C_α. Pitanje kako ih odabrati što optimalnije s obzirom na vjerojatnosti pogreške 1. odn. 2. vrste.

Kao što smo istakli, za test koji nikad ne odbacuje H_0, vjerojatnost pogreške 1. vrste je 0. No takav test nije koristan naravno.

Primjer Neka je naš statistički test ugradjen u protupožarni alarm baziran na detektoru dima. Hipoteze su ovdje:

H_0 : nema požara, i H_A : u prostoriji je požar.

Ako nikad ne odbacujemo H_0 u ovom primjeru, ovakav alarm se nikad neće oglasiti pa ga nema smisla ni postavljati. Naš bi cilj mogao biti: uz malu vjerojatnost da će se alarm oglasiti u slučaju da požara nema, dobiti i što veću vjerojatnost da će se oglasiti u slučaju kad nastupi požar. Tu vjerojatnost možemo zapisati kao

$$P\left(\text{test odbacuje } H_0 | H_A \text{ je istinita}\right) = 1 - \beta.$$

Ova vrijednost se u statistici zove **snaga testa**.

U parametarskim vjerojatnosnim modelima kakve smo susretali do sada, H_0 odn. H_A se daju izraziti preko vrijednosti parametra. Ako je npr. u binomnom modelu $H_0 : p \leq p_0$, tada će gornja vjerojatnost ovisiti o pravoj vrijednosti parametra p, stoga definiramo **funkciju snage** testa

$$\gamma(p) = P\left(\text{test odbacuje } H_0 | p \text{ je prava vrijednost parametra}\right) = 1 - \beta(p).$$

Dakako, idealno bi bilo

$$\gamma(p) = 0 \text{ za } p \leq p_0 \text{ i } \gamma(p) = 1 \text{ za sve } p > p_0,$$

takvi idealni testovi u našim primjerima ne postoje. Stoga smo koristeći nivo značajnosti α, ograničili $\gamma(p) \leq \alpha$, za $p \leq p_0$, a razne testove s tim svojstvom usporedjujemo gledajući funkciju γ za $p > p_0$.

Primjer (snaga testa)

Pretpostavite da na osnovu uzorka duljine 100, od kojih je S broj jedinki s izvjesnim karakteristikama, želimo testirati sljedeće hipoteze o postotku takvih jedinki u cijeloj populaciji

$$H_0 : p \leq 0.5 \ \text{ nasuprot } \ H_A : p < 0.5.$$

Neka je zadano $\alpha = 0.05$. Koristimo uobičajenu testnu statistiku

$$Z = \frac{S - n/2}{\sqrt{n\frac{1}{2}\frac{1}{2}}}.$$

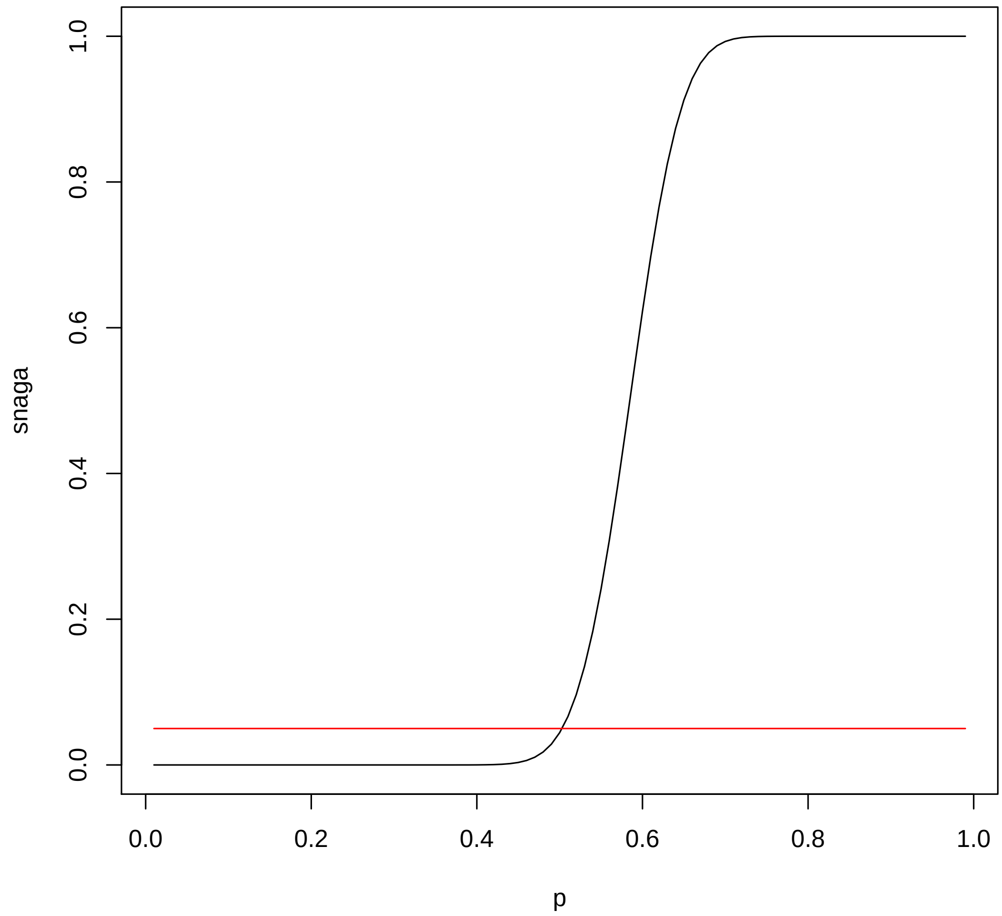
Kako smo objasnili, H_0 ćemo odbaciti za $Z \geq z_\alpha = 1.65$ u ovom slučaju. Dakle funkcija γ se može izračunati kao

$$\gamma(p) = P\left(Z \geq 1.65 | p \text{ je prava vrijednost parametra}\right).$$

Već znamo da je

$$\gamma(0.5) \approx \alpha = 0.05\,.$$

Snagu testa možemo vidjeti iz grafa funkcije γ. Što funkcija γ brže raste prema 1 za $p > 0.5$ test bismo smatrali snažnijim.



Funkcija snage $\gamma(p)$ za test $H_0 \leq \frac{1}{2}$.

Testiranje hipoteza o očekivanjima dvije normalne razdiobe

Uz jednake varijance

Neka $X_1, \ldots, X_n$, $n \geq 1$, čine sl. uzorak iz normalne razdiobe s parametrima μ_1 i σ^2, a $Y_1, \ldots, Y_m$, $m \geq 1$, neka čine sl. uzorak iz normalne razdiobe s parametrima μ_2 i σ^2 (nezavisan od X_i), te neka su nam sva 3 parametra nepoznata. Pretpostavimo da želimo testirati:

i) $H_0 : \mu_1 \leq \mu_2$ i $H_A : \mu_1 > \mu_2$.

ii) $H_0 : \mu_1 = \mu_2$ i $H_A : \mu_1 \neq \mu_2$.

iii) $H_0 : \mu_1 \geq \mu_2$ i $H_A : \mu_1 < \mu_2$.

Iz svakog uzorka posebno izračunamo njihove aritmetičke sredine $\bar{X}$ i $\bar{Y}$, te uzoračke varijance $\hat{s}_X^2$ odn. $\hat{s}_Y^2$. Za očekivati je da ćemo razliku izmedju μ_1 i μ_2 moći uočiti izmedju razlike $\bar{X}$ i $\bar{Y}$, no pitanje je kako?

Za testnu statistiku odabiremo

$$T = \frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y}}{\hat{s}},$$

gdje je

$$\hat{s}^2 = \frac{1}{n+m-2} \left((n-1)\hat{s}_X^2 + (m-1)\hat{s}_Y^2 \right).$$

Može se pokazati da, uz uvjet $\mu_1 = \mu_2$, statistika T ima Studentovu t razdiobu s $n+m-2$ stupnja slobode.

Kritično područje za svaki od tri testa odredjujemo na sljedeći način

i) $C_\alpha = [t_\alpha, \infty)$.
ii) $C_\alpha = \mathbb{R} \setminus (-t_{\alpha/2}, t_{\alpha/2})$.
iii) $C_\alpha = (-\infty, -t_\alpha]$.

Gdje je t_α $(1-\alpha)$-kvantil t razdiobe s $n+m-2$ stupnja slobode.

Na ovaj način možemo usporediti npr. rezultate kolokvija za one studente koji su posjetili predavanja i one koji to nisu činili. Pretpostavka da su podaci normalno distribuirani uz jednake varijance je katkad upitna u praksi. Ako ona nije zadovoljena testnu statistiku za 3 hipoteze o odnosu očekivanja dvije razdiobe možemo malo promijeniti, i ponovo kontruirati test o odnosu μ_1 i μ_2, uz uvjet da imamo veliki uzorak (v. Bhattacharyya i Johnson npr.).

Usporedba varijanci dva normalno distribuirana uzorka

Neka $X_1, \ldots, X_n$, $n \geq 1$, čine sl. uzorak iz normalne razdiobe s parametrima μ_1 i σ_1^2, a $Y_1, \ldots, Y_m$, $m \geq 1$, neka čine sl. uzorak iz normalne razdiobe s parametrima μ_2 i σ_2^2, te neka su nam sva 4 parametra nepoznata. Pretpostavimo da želimo testirati:

$H_0 : \sigma_1 = \sigma_2$ nasuprot $H_A : \sigma_1 \neq \sigma_2$.

ili što je potpuno ekvivalentno

$H_0 : \frac{\sigma_1}{\sigma_2} = 1$ nasuprot $H_A : \frac{\sigma_1}{\sigma_2} \neq 1$.

Izračunajmo uzoračke varijance $\hat{s}_X^2$ odn. $\hat{s}_Y^2$, a za testnu statistiku postavimo

$$F = \frac{\hat{s}_X^2}{\hat{s}_Y^2}.$$

Za ovu testnu statistiku je poznato da pod nul hipotezom H_0 ima takozvanu Fisherovu F razdiobu s $(n-1, m-1)$ stupnjeva slobode. Sada kritično podučje za nivo značajnosti α možemo konstruirati koristeći tablice za F razdiobu.

Usporedba očekivanja dva sparena normalno distribuirana uzorka

Primjer Pretpostavimo da želimo usporediti razliku izmedju krvnog tlaka za osobe koje uzimaju i one koje ne uzimaju odredjeni lijek. Mogli bismo postupiti kao i do sada kad smo željeli usporediti da li je očekivana vrijednost neke varijable u dvije populacije jednaka, npr. mogli bismo koristiti t-test.

Pretpostavite, medjutim, da je varijanca mjerenja σ^2 vrlo velika, tako da t-test može detektirati samo vrlo značajne razlike izmedju μ_1 i μ_2. Ovakav test ima malu snagu ako se μ_1 i μ_2 razlikuju za relativno malu vrijednost. Imate li bolju ideju?

Bolja ideja je da pokušamo eliminirati dio varijabilnosti izmedju podataka tako da isto mjerenje napravimo na jednoj osobi po dva puta, tj. jednom u periodu uzimanja lijeka i jednom u periodu kada ne uzima lijek.

Zbog ovog istog razloga, tj. smanjenja varijabilnosti, u medicini (ali i biologiji) istraživanja se često oslanjaju na podatke dobivene na parovima jedinki iz iste obitelji, ili čak na parovima jednojajčanih blizanaca.

Nakon što odlučimo kako i što spariti u ovakvom pokusu, poželjno je provesti i **randomizaciju**, tj. svaku jedinku iz para treba dodijeliti 1. ili 2. tretmanu u ovisnosti o ishodu nezavisnih bacanja novčića. Npr. ako bacimo pismo, prvo mjerimo tlak osobe uz uzimanje lijek, a zatim bez uzimanja lijeka, a ako padne glava napravimo obrnuto. Na ovaj način smanjujemo prostor za druge nekontrolirane izvore varijabilnosti.

Dakle ovdje na svakoj jedinki napravimo dva mjerenja. Tako da slučajni uzorak možemo označiti kao niz parova $(X_1, Y_1), \ldots, (X_n, Y_n)$. Ako želimo testirati razliku izmedju očekivanja pod jednim odn. drugim tremanom, za sve jedinke možemo promatrati razlike $D_i = X_i - Y_i,\ i = 1, \ldots, n$. Uočite da hipoteze o μ_1 i μ_2 možemo prevesti u hipoteze o očekivanju razlike, npr.

$$\mu_1 > \mu_2 \text{ je ekvivalentno } ED_i > 0.$$

Stoga sad D_i tretiramo kao uzorak iz jedne populacije, često je razumno pretpostaviti da su D_i normalno distribuirane s očekivanjem $d = \mu_1 - \mu_2$ i varijancom σ_d^2.

Pod gornjim uvjetima za hipoteze

i) $H_0 : \mu_1 \leq \mu_2$ i $H_A : \mu_1 > \mu_2$.

ii) $H_0 : \mu_1 = \mu_2$ i $H_A : \mu_1 \neq \mu_2$.

iii) $H_0 : \mu_1 \geq \mu_2$ i $H_A : \mu_1 < \mu_2$.

definiramo testnu statistiku

$$T = \sqrt{n}\frac{\bar{D} - \mu}{\hat{s}_d},$$

gdje je $\hat{s}_d^2$ uzoračka varijanca a $\bar{D}$ aritmetička sredina uzorka $D_1, \ldots, D_n$.

Za ovakav T znamo da ima Studentovu t-razdiobu s $n-1$-im stupnjem slobode, ako vrijedi $ED_i = 0$, tj. $\mu_1 = \mu_2$.
Stoga testove provodimo odredjujući kritično područje na sljedeći način

i) $C_\alpha = [t_\alpha, \infty)$.

ii) $C_\alpha = \mathbb{R} \setminus (-t_{\alpha/2}, t_{\alpha/2})$.

iii) $C_\alpha = (-\infty, -t_\alpha]$.

Usporedba parametra p za dvije binomne razdiobe

Na dva uzorka iz dvije različite populacije provodimo istraživanje kako bismo utvrdili postoji li razlika u postotku jedinki u jednoj odn. drugoj populaciji sa zadanim obilježjem.

Primjer Na uzorku od 40 studentica matematike 2. godine utvrdjeno je da je 25 njih zadovoljilo sve svoje obaveze iz prve godine, na uzorku od 40 studenata taj broj je nešto manji i iznosi 21. Možemo li tvrditi da postoji razlika izmedju uspješnosti muških odn. ženskih studenata matematike?

Podatke ovdje općenito reprezentiramo s dvama nezavisnim binomnim sl. varijablama $S_1 \sim B(n_1, p_1)$ i $S_2 \sim B(n_2, p_2)$, koje označavaju broj jedinki sa traženim krakteristikama u 1. odn. 2. uzorku. O odnosu p_1 i p_2 možemo postavite sljedeća tri para hipoteza

i) $H_0 : p_1 \leq p_2$ i $H_A : p_1 > p_2$.

ii) $H_0 : p_1 = p_2$ i $H_A : p_1 \neq p_2$.

iii) $H_0 : p_1 \geq p_2$ i $H_A : p_1 < p_2$.

Za testnu statistiku izabiremo

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

gdje su naravno

$$\hat{p}_1 = \frac{S_1}{n_1} \text{ i } \hat{p}_2 = \frac{S_2}{n_2},$$

a

$$\hat{p} = \frac{S_1 + S_2}{n_1 + n_2}.$$

Centralni granični teorem nam omogućuje da za velike uzorke koristimo sljedeća tri već dobro poznata kritična područja:

i) $C_\alpha = [z_\alpha, \infty)$.

ii) $C_\alpha = \mathbb{R} \setminus (-z_{\alpha/2}, z_{\alpha/2})$.

iii) $C_\alpha = (-\infty, -z_\alpha]$.

Testiranje hipoteze o nezavisnosti dviju Bernoullijevih sl. varijabli

Procedura koju ćemo opisati zasniva se na podacima koji su vrlo jednostavni. Svaku od n jedinki u uzorku klasificiramo u odn. na dva (kvalitativna) obilježja. Dakle, za svaku jedinku imamo dvije sl. varijable s vrijednostima 0 i 1, koje indiciraju ima li jedinka obilježje x i/ili obilježje y. Mogli bismo dakle reprezentirati podatke kao niz parova Bernoullijevih sl. varijabli $(X_i, Y_i),\ i = 1, \ldots, n$. Ono što želimo provjeriti je da li je prisutnost ovih obilježja kod pojedine sl. odabrane jedinke nezavisno. U jeziku vjerojatnosti to bi značilo da su nezavisne sl. varijable X_i i Y_i. Mogućih testova je i ovdje više.

Primjetite da bismo sakupljene podatke moglu prikazati u frekvencijskoj tablici oblika 2×2:

	0	1
0	n_{00}	n_{10}
1	n_{01}	n_{11}

gdje n_{kl} označava broj jedinki u uzorku (ili parova (X_i, Y_i)) za koje je $X_i = k$, a $Y_i = l$. Ovakve tablice se zovu i kontingencijske tablice [contingency tables], a mogu biti i reda $m \times n$, $n, m \geq 2$ općenito. Mi ćemo zbog jednostavnosti promatrati samo slučaj $m = n = 2$.

Primjer
Godine 1889. u istraživačkoj stanici u Rothamstedu (UK), za vrijeme popodnevne pauze za čaj, algolog dr. B.M. Bristol odbila je šalicu čaja uz napomenu kako pije čaj isključivo ako se prvo ulije miljeko u šalicu, a tek zatim čaj. Nazočan je bio i statističar R.A. Fisher, koji je ustvrdio da je to besmisleno, jer nitko ne može osjetiti razliku izmedju šalica čaja u kojima je mlijeko uliveno prije ili nakon čaja. Tada je dr. Bristol uzdahnula
"Oh, ali razlika stvarno postoji!"
Nakon čega je netko uzviknuo (kasnije će se ispostaviti, bio je to budući suprug gospodjice Bristol)
"Testirajmo je!"
Na sreću gospodin Fisher je brzo izašao s idejom o tome kako provesti ovaj test.

Postavimo u ovom primjeru $X_i = 1$ ako je u i-toj šalici uliveno prvo mlijeko, a 0 inače. Slično, nakon što dr. Bristol proba i-ti čaj ne znajući o kakvoj se šalici radi, postavimo $Y_i = 1$ ako je ustvrdila da je u i-tu šalicu uliveno prvo mlijeko, a 0 inače.

Ako su sl. varijable X_i i Y_i zaista nezavisne, trebalo bi vrijediti

$$P(X_i = k, Y_i = l) = P(X_i = k)P(Y_i = l)$$

za sve $k, l = 0, 1$.

Iako ne znamo pravu razdiobu sl. vektora (X_i, Y_i) prirodno je možemo aproksimirati na sljedeći način

	0	1
0	$\hat{p}_{00} = n_{00}/n$	$\hat{p}_{10} = n_{10}/n$
1	$\hat{p}_{01} = n_{01}/n$	$\hat{p}_{11} = n_{11}/n$

gdje je $\hat{p}_{kl}$ dakako procjena za $p_{kl} = P(X_i = k, Y_i = l)$.

Procjena za vjerojatnosti $P(X_i = k)$ i $P(Y_i = l)$ se sada može dobiti iz tablice, npr, $P(X_i = k)$ procjenjujemo kao

$$\hat{p}_k = \hat{p}_{k0} + \hat{p}_{k1},$$

a $P(Y_i = l)$ procjenjujemo kao

$$\hat{q}_l = \hat{p}_{0l} + \hat{p}_{1l}.$$

Za suprotstavljene hipoteze

$H_0: \; X_i$ i Y_i su nezavisne i $H_A: \; X_i$ i Y_i nisu nezavisne,

testna statistika sada je prirodno

$$C^2 = \sum_{k,l=0,1} \frac{(n_{kl} - n\hat{p}_k\hat{q}_l)^2}{n\hat{p}_k\hat{q}_l} = n \sum_{k,l=0,1} \frac{(\hat{p}_{kl} - \hat{p}_k\hat{q}_l)^2}{\hat{p}_k\hat{q}_l}.$$

Malo je teže argumentirati, no ponovo iz centralnog graničnog teorema slijedi da ova statistika ima χ^2 razdiobu s 1 stupnjem slobode.

Zbog toga je kritično područje ovog testa

$$C_\alpha = [\chi^2_\alpha, \infty)\,,$$

gdje je χ^2_α $(1-\alpha)$-kvantil χ^2 razdiobe s jednim stupnjem slobode.

Sažetak

Procedura testiranja statističkih hipoteza ima sljedeće korake:

- Formuliramo H_0 i H_A, te odredimo nivo značajnosti α.
- Odredimo testnu statistiku T i kritično područje C_α tako da

$$P(T \in C_\alpha | H_0) \leq \alpha.$$

- Izračunamo vrijednost testne statistike za naše podatke, npr. $T = t$.
- Odbacujemo H_0 ako $t \in C_\alpha$, inače je ne odbacujemo na nivou značajnosti α.

Posljednji korak ima i alternativni oblik ako možemo izračunati p-vrijednost dobivene testne statistike $T = t$, tada bismo odbacili H_0 ako je $p < \alpha$.