Vjerojatnost i statistika

Testiranje statističkih hipoteza

16. siječnja 2024.

Testiranje statističkih hipoteza

- **statistička hipoteza** - hipoteza (slutnja) koja je formulirana u terminima distribucije slučajne varijable
- postoje dvije vrste hipoteza:
  - **nul-hipoteza** H_0
  - **alternativna hipoteza** H_1
- hipoteze treba postaviti tako da alternativna hipoteza odražava ono što želimo potvrditi

Testiranje statističkih hipoteza

- **statistički test** je pravilo na temelju kojeg donosimo odluku o odbacivanju ili neodbacivanju nul-hipoteze
- ukoliko na temelju statističkog testa **odbacimo** nul-hipotezu kažemo da prihvaćamo alternativnu hipotezu
- ukoliko na temelju statističkog testa **ne odbacimo** nul-hipotezu NE možemo reći da prihvaćamo nul-hipotezu

Primjer

Tužiteljstvo na temelju svojih indicija odlučuje podići optužnicu protiv jedne osobe. Pritom se drže načela "nitko nije kriv dok mu se ne dokaže krivnja". Njihove hipoteze su:

$$H_0 \ : \text{optuženi nije kriv}$$
$$H_1 \ : \text{optuženi je kriv}$$

Tužiteljstvo ima dvije mogućnosti:

- odbaciti H_0 (i prihvatiti H_1) $\Rightarrow$ postoji dovoljno dokaza za krivnju
- ne odbaciti H_0 $\Rightarrow$ ne postoji dovoljno dokaza za krivnju

U slučaju nepostojanja dovoljno dokaza za krivnju, bilo bi pogrešno reći da prihvaćamo H_0 jer, unatoč tome što tužiteljstvo nije uspjelo dokazati krivnju, to ne znači da osoba zaista nije kriva.

Pogreška I. i II. tipa

- **pogreška I. tipa**: odbaciti H_0 ako je ona istinita
- **pogreška II. tipa**: ne odbaciti H_0 ako je H_1 istinita

ISTINA	REZULTAT TESTA	
	ne odbaciti H_0	odbaciti H_0
H_0	dobra odluka	pogreška I. tipa
H_1	pogreška II. tipa	dobra odluka

- primjer pogreške I. tipa: proglasiti nevinu osobu krivom
- primjer pogreške II. tipa: osloboditi krivnje osobu koja je doista kriva

Razina značajnosti

- statistički testovi dizajniraju se tako da pogreška I. tipa bude što manja
- **razina značajnosti testa** α je maksimalna dozvoljena vjerojatnost pogreške I. tipa
- najčešće se uzima $\alpha = 0.01, 0.05$ ili 0.1
- **p-vrijednost** je *aproksimacija* maksimalne vjerojatnosti pogreške I. tipa na temelju podataka

Uspoređivanjem p-vrijednosti s razinom značajnosti α donosimo odluku:

- $p < \alpha \Rightarrow$ odbacujemo H_0 na razini značajnosti α i prihvaćamo H_1
- $p > \alpha \Rightarrow$ ne odbacujemo H_0 na razini značajnosti α

Testiranje hipoteza o očekivanju

Veliki uzorci: z-test

Nul-hipoteza:

$$H_0 : \mu = \mu_0$$

Test-statistika:

$$Z' = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

n - veličina uzorka

$\overline{X}_n$ - aritmetička sredina uzorka

σ - standardna devijacija

U uvjetima istinitosti H_0, Z' ima približno $\mathcal{N}(0, 1)$ distribuciju. Ako je σ nepoznata, možemo ju zamijeniti sa s_n.

Veliki uzorci: z-test

Realizacija test statistike:

$$\hat{z} = \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

Računanje p-vrijednosti ovisi o obliku alternativne hipoteze:

- $p = P(|Z'| \geq |\hat{z}|)$ ako je $H_1 : \mu \neq \mu_0$
- $p = P(Z' \geq \hat{z})$ ako je $H_1 : \mu > \mu_0$
- $p = P(Z' \leq \hat{z})$ ako je $H_1 : \mu < \mu_0$

Primjer - televizija.sta

Godine 1979. osnovna kabelska televizija u SAD-u u prosjeku je stajala \$ 7.37 mjesečno. Godine 1983. udruženje kabelskih televizija, koje broji više od 4000 kabelskih sustava, zaključilo je da je kabelska televizija poskupjela za samo 8% u odnosu na 1979. te da ne stoji statistički značajno više od \$8 mjesečno. Udruženje potrošača sumnja u te izjave pa ćemo ih provjeriti na temelju 33 podatka u bazi `televizija.sta`. U tu svrhu postavljamo sljedeće hipoteze:

$$
\begin{aligned}
H_0 &: \mu = \mu_0 = 8 \\
H_1 &: \mu > 8
\end{aligned}
$$

Primjer - televizija.sta

Da bismo izračunali vrijednost $\hat{z}$, trebamo $\overline{x}_n$ i s_n:

$$\overline{x}_n = 8.33, \quad s_n = 2.18$$

Računamo $\hat{z}$:

$$\hat{z} = \frac{\overline{x}_n - \mu_0}{s_n/\sqrt{n}} = \frac{8.33 - 8}{2.18/\sqrt{33}} = 0.87$$

Računamo p-vrijednost (vjerojatnosni kalkulator u Statistici):

$$p = P(Z \geq \hat{z}) = P(Z \geq 0.87) = 0.19$$

Primjer - televizija.sta

- zadajmo si razinu značajnosti $\alpha = 0.05$
- $p = 0.19 > \alpha$
- na nivou značajnosti $\alpha = 0.05$ ne odbacujemo nul-hipotezu, tj. na nivou značajnosti $\alpha = 0.05$ ne možemo tvrditi da je cijena kabelske televizije statistički značajno veća od 8 dolara
- u Statistici:

> Statistics $\rightarrow$ Basic Statistics $\rightarrow$ Difference tests: r,%, means $\rightarrow$ Difference between two means (normal distribution) - označiti Single mean 1 vs. population mean 2 $\rightarrow$ unijeti vrijednosti $\rightarrow$ Compute

Zadatak - lopta.sta

Jedan se poduzetnik bavi proizvodnjom loptica za golf. U suradnji s projektantima u poduzeću napravio je preinake na jednom dijelu stroja (ubrizgavalici). Cijeli je proces dizajniran tako da proizvodi loptice prosječne mase 0.25 unci. Kako bi istražio radi li nova ubrizgavalica zadovoljavajuće, odabire 40 loptica i bilježi njihove mase. Je li na nivou značajnosti $\alpha = 0.05$ očekivana masa loptice statistički značajno veća od 0.25 unci?

Rješenje: $H_0 : \mu = 0.25, \quad H_1 : \mu > 0.25, \quad p < \alpha$

Mali uzorci: t-test

Ukoliko podaci dolaze iz normalne distribucije i σ je nepoznato, za male uzorke možemo koristiti **t-test**.

Nul-hipoteza:

$$H_0 : \mu = \mu_0$$

Test-statistika:

$$T = \frac{\overline{X}_n - \mu_0}{s_n / \sqrt{n}}$$

n - veličina uzorka

$\overline{X}_n$ - aritmetička sredina uzorka

s_n - standardna devijacija

U uvjetima istinitosti H_0, test statistika T ima studentovu distribuciju s $(n-1)$ stupnjeva slobode.

Mali uzorci: t-test

Realizacija test statistike:

$$\hat{t} = \frac{\overline{x}_n - \mu_0}{s_n/\sqrt{n}}$$

Računanje p-vrijednosti ovisi o obliku alternativne hipoteze:

- $p = P(|T| \geq |\hat{t}|)$ ako je $H_1 : \mu \neq \mu_0$
- $p = P(T \geq \hat{t})$ ako je $H_1 : \mu > \mu_0$
- $p = P(T \leq \hat{t})$ ako je $H_1 : \mu < \mu_0$

Primjer - automobili.sta

Varijabla *potrosnja* iz baze podataka `automobili.sta` sadrži potrošnju goriva novog modela automobila pri brzini od 110 km/h na autocesti za 300 nezavisnih mjerenja. Ovu varijablu ima smisla modelirati normalnom slučajnom varijablom $\mathcal{N}(5.12, 0.97^2)$. Pretpostavimo da nas zanima samo prvih 25 mjerenja te da želimo provjeriti je li prosječna potrošnja u tih 25 mjerenja statistički značajno manja od 6 litara. U tu svrhu postavljamo sljedeće hipoteze:

$$H_0 \ : \ \mu = \mu_0 = 6,$$
$$H_1 \ : \ \mu < 6.$$

Primjer - automobili.sta

S obzirom da se radi o malom uzorku za kojeg možemo pretpostaviti da dolazi iz normalne distribucije, znamo da test-statistika T ima Studentovu distribuciju s 24 stupnja slobode.

Da bismo izračunali vrijednost $\hat{t}$ trebamo $\overline{x}_{25}$ i s_{25}:

$$\overline{x}_{25} = 5.48, \quad s_{25} = 0.85.$$

Računamo $\hat{t}$:

$$\hat{t} = \frac{\overline{x}_{25} - \mu_0}{s_{25}/\sqrt{25}} = \frac{5.48 - 6}{\frac{0.85}{\sqrt{25}}} = -3.06.$$

Računamo p-vrijednost (vjerojatnosni kalkulator u Statistici):

$$p = P(T \leq \hat{t}) = P(T \leq -3.06) = 0.0026.$$

Primjer - automobili.sta

- zadajmo si razinu značajnosti $\alpha = 0.05$
- $p = 0.0026 < \alpha$
- na nivou značajnosti $\alpha = 0.05$ odbacujemo nul-hipotezu i prihvaćamo alternativnu hipotezu, tj. na nivou značajnosti $\alpha = 0.05$ možemo tvrditi da je prosječna potrošnja ovog automobila (u prvih 25 mjerenja) manja od 6 litara
- u Statistici:

> Statistics $\rightarrow$ Basic Statistics $\rightarrow$ t-test single sample $\rightarrow$ Variables $\rightarrow$ Test all means against $\rightarrow$ unijeti vrijednost μ_0 $\rightarrow$ Summary

Testiranje hipoteza o vjerojatnosti događaja

Veliki uzorci - binomni test

Nul-hipoteza:

$$H_0 : p = p_0$$

Test-statistika:

$$Z' = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

n - veličina uzorka

$\hat{p}$ - relativna frekvencija uspjeha

U uvjetima istinitosti H_0 test statistika Z' ima približno $\mathcal{N}(0, 1)$ distribuciju.

Veliki uzorci - binomni test

Realizacija test statistike:

$$\hat{z} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Računanje p-vrijednosti ovisi o obliku alternativne hipoteze:

- $p = P(Z \geq \hat{z})$ ako je $H_1 : p > p_0$
- $p = P(Z \leq \hat{z})$ ako je $H_1 : p < p_0$

Uzorak je dovoljno velik za provođenje ovog statističkog testa ako interval

$$\left[p_0 - 3\sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + 3\sqrt{\frac{p_0(1-p_0)}{n}}\right]$$

ne sadrži ni 0 ni 1.

Primjer - vrtic.sta

U nekom poduzeću zaposleno je više od 3000 ljudi. Uprava poduzeća želi ponuditi pomoć svojim zaposlenicima oko organizacije čuvanja djece. Predložene su dvije opcije - otvaranje vrtića u sklopu poduzeća ili plaćanje dijela troškova čuvanja djece koje bi roditelji organizirali sami. Da bi se utvrdilo koja je od ovih dvaju mjera popularnija među zaposlenicima, odabran je uzorak od 60 roditelja s malom djecom koji su se izjasnili o tome koju opciju preferiraju. Njihovi odgovori označeni su na sljedeći način:

- 0 - radije bih novčanu pomoć za samostalno organiziranje čuvanja djece
- 1 - radije bih da se otvori vrtić u sklopu poduzeća.

Primjer - vrtic.sta

Pretpostavimo da uprava neće organizirati vrtić u sklopu poduzeća ako se pokaže da je proporcija roditelja koji podržavaju tu ideju statistički značajno manja od 0.75. Da bismo to provjerili, postavljamo sljedeće hipoteze:

$$\begin{aligned} H_0 &: p = p_0 = 0.75, \\ H_1 &: p < 0.75. \end{aligned}$$

Da bismo izračunali vrijednost $\hat{z}$, treba nam proporcija $\hat{p}$:

$$\hat{p} = \frac{38}{60} = 0.63$$

Primjer - vrtic.sta

Računamo $\hat{z}$:

$$\hat{z} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.63 - 0.75}{\sqrt{\frac{0.75 \cdot 0.25}{60}}} = -2.15$$

Računamo p-vrijednost (vjerojatnosni kalkulator u Statistici):

$$p = P(Z \leq \hat{z}) = P(Z \leq -2.15) = 0.016$$

- zadajmo si **razinu značajnosti** $\alpha = 0.05$
- $p = 0.016 < \alpha$
- na nivou značajnosti $\alpha = 0.05$ odbacujemo nul-hipotezu i prihvaćamo alternativnu hipotezu, tj. na nivou značajnosti $\alpha = 0.05$ uprava neće organizirati vrtić u sklopu poduzeća

Zadatak - perec.sta

Lanac pekarnica je odlučio prodavati nove perece u svojim pekarama, no prije toga žele ispitati tržište o novom proizvodu. O tome ovisi hoće li se nastaviti prodaja novih pereca ili ne. U bazi podataka `perec.sta` nalaze se podaci dobiveni iz uzorka od 50 kupaca, pri čemu su njihovi odgovori označeni na sljedeći način:

0 - pereci mi se ne sviđaju,

1 - pereci mi se sviđaju,

2 - neodlučan/a sam.

Lanac pekarnica odlučio je da neće prodavati nove perece ukoliko je proporcija onih kojima se ne sviđaju statistički značajno veća od 0.5. Koju odluku treba donijeti lanac pekarnica?

Rješenje: $H_0 : p = 0.5, \quad H_1 : p > 0.5, \quad p > \alpha$

Testiranje hipoteza o distribuciji općenito

χ^2-test

- χ^2-test koristimo kada želimo testirati ima li diskretna slučajna varijabla neku zadanu teorijsku distribuciju
- za procjenu stvarne distribucije koristimo empirijsku distribuciju temeljenu na podacima

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_k \\ p_1 & p_2 & \dots & p_k \end{pmatrix}$$

- hipoteze:

$$\begin{aligned} H_0 &: \text{distribucija iz koje dolaze podaci jednaka je teorijskoj} \\ H_1 &: \text{distribucija iz koje dolaze podaci razlikuje se od teorijske} \end{aligned}$$

χ^2-test

- u Statistici:

Statistics $\rightarrow$ Nonparametrics $\rightarrow$ Observed vs expected χ^2 $\rightarrow$ odabrati varijablu kojom modeliramo svojstvo čija nas distribucija zanima i varijablu kojom modeliramo teorijsku distribuciju $\rightarrow$ Summary

Primjer

Tržišni analitičar želi istražiti imaju li potrošači neke posebne sklonosti prema jednom od okusa sokova koji su se pojavili na tržištu. Na uzorku od 100 ljudi prikupio je preferencije prema ponuđenih pet okusa. Frekvencije zabilježene tim istraživanjem dane su u tablici

višnja	jabuka	naranča	limun	grejp
32	28	16	14	10

U svrhu određivanja postojanja sklonosti potrošača prema određenom okusu soka, definiramo teorijsku distribuciju

$$\begin{pmatrix} \text{višnja} & \text{jabuka} & \text{naranča} & \text{limun} & \text{grejp} \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}$$

Primjer

- za provođenje χ^2-testa u Statistici, potrebne su nam teorijske i empirijske frekvencije

okus	empirijska frekvencija	teorijska frekvencija
višnja	32	20
jabuka	28	20
naranča	16	20
limun	14	20
grejp	10	20

- hipoteze:

H_0 : distribucija iz koje dolaze podaci jednaka je teorijskoj (potrošači nemaju sklonosti prema određenom okusu)

H_1 : distribucija iz koje dolaze podaci razlikuje se od teorijske (potrošači imaju sklonosti prema određenom okusu)

Primjer

- zadajmo si razinu značajnosti $\alpha = 0.01$
- $p = 0.001234 < \alpha$
- na razini značajnosti $\alpha = 0.01$ odbacujemo nul-hipotezu i prihvaćamo alternativnu hipotezu, tj. na razini značajnosti $\alpha = 0.01$ možemo tvrditi da postoji statistički značajna sklonost potrošača prema određenom okusu soka

Zadatak - gradjevina.sta

Varijabla *napredovanje* baze `gradjevina.sta` sadrži ocjene kadrovskih službi 100 građevinskih poduzeća o tome u kolikoj mjeri uspješno obavljanje posla utječe na mogućnost napredovanja na bolje radno mjesto. Zabilježene ocjene su: 1 - uspješnost obavljanja posla uopće ne utječe na mogućnost napredovanja, . . . , 5 - napredovanje na bolje radno mjesto isključivo ovisi o uspješnosti u obavljanju posla. Pretpostavimo da je teorijska distribucija slučajne varijable kojom se modelira ta ocjena

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1/10 & 1/10 & 1/10 & 1/5 & 1/2 \end{pmatrix}.$$

Postoji li na razini značajnosti $\alpha = 0.01$ statistički značajno odstupanje empirijske distribucije te slučajne varijable od pretpostavljene teorijske distribucije?

Rješenje: $p < 0.0000001$

Normalna distribuiranost

- za testiranje normalnosti podataka koriste se dva statistička testa: **Lillieforsova inačica Kolmogorov-Smirnovljeva testa i Shapiro-Wilk W test**
- hipoteze:

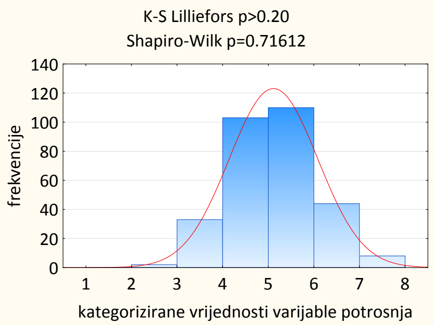
$$
\begin{aligned}
H_0 &: \text{varijabla ima normalnu distribuciju} \\
H_1 &: \text{varijabla nema normalnu distribuciju}
\end{aligned}
$$

- u Statistici

> Statistics → Basic Statistics → Descriptive Statistics → Normality → označiti polja: Kolmogorov-Smirnov & Lilliefors test for normality, Shapiro-Wilk W test → Histograms

Primjer - automobili.sta

U bazi podataka `automobili.sta` nalaze se rezultati mjerenja prosječne potrošnje novog tipa automobila pri brzini od 110 km/h na autocesti za 300 nezavisnih mjerenja. Ranije smo zaključili da ima smisla modelirati ovu varijablu kao normalnu slučajnu varijablu s očekivanjem $\overline{x}_{300} = 5.12$ i varijancom $s_{300}^2 = 0.97^2$. Sada možemo testirati hipotezu o normalnosti distribucije.



Primjer - automobili.sta

- zadajmo si razinu značajnosti $\alpha = 0.05$
- uočavamo da je za oba testa $p > \alpha$
- na razini značajnosti $\alpha = 0.05$ ne odbacujemo nul-hipotezu, tj. nemamo razloga sumnjati u to da varijabla ima normalnu distribuciju

Zadatak - MBA-studij.sta

Baza podataka `MBA-studij.sta` sadrži podatke o broju bodova na GMAT (Graduate Management Admission Test) testu za 100 studenata koji žele upisati neki studij. Možemo li na nivou značajnosti $\alpha = 0.05$ tvrditi da je slučajna varijabla kojom modeliramo broj bodova na tom testu normalno distribuirana?

Rješenje: Na nivou značajnosti $\alpha = 0.05$ ne odbacujemo nul-hipotezu da je varijabla normalno distribuirana. (KS test: $p > 2$, SW test: $p = 0.856$)