BUGI

western Balkans Urban aGriculture Initiative

Co-funded by the Erasmus+ Programme of the European Union

Eksperimentalna statistika

Senada Kalabušić
Zenan Šabanac
Pakeza Drkenda
Univerzitet u Sarajevu

Project number: 586304-EPP-1-2017-1-BA-EPPKA2-CBHE-JP This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein
2019.

Metod uzorka

Zašto posmatramo uzorak?

i) Osnovni skup čine sva moguća posmatranja obilježja X.
ii) U praktičnim situacijama često nije moguće vršiti ispitivanja na cijelom osnovnom skupu(npr.osnovni skup može imati beskonačno mnogo elemenata ili ako ih ima konačno mnogo, ispitivanje cijelog skupa može zahtijevati velike materijalne troškove).

Zbog navedenog, ispitivanje se vrši na dijelu osnovnog skupa koji se naziva *uzorački skup* ili jednostavno *uzorak*.
Broj jedinica osnovnog skupa ili uzorka naziva se *obim* ili *veličina*.
Broj jedinica osnovnog skupa označavat ćemo sa N, a broj jedinica uzorka sa n.

Metod uzorka

Neka su $x_1, x_2, \ldots, x_n$ rezultati nekog posmatranja obilježja X, koji čine uzorak. Metod, koji se sastoji u tome da se na osnovu osobina uzetog uzorka, zaključuje o numeričkim karakteristikama i zakonu raspodjele obilježa X, naziva se metod *uzorka*.
Da bi rezulatati dobiveni na osnovu uzorka bili objektivni, uzorak mora biti *reprezentativan*. Znači, da u najvećoj mjeri oslikava karateristike osnovnog skupa.
Metode dobivanja reprezentatitvog uzorka zasnivaju se na pretpostavci da svaka jedinica(element) osnovnog skupa ima istu mogućnost(vjerovatnoću) da bude jedinica u uzorku. Jedinice reprezentativnog uzorka trebaju se birati na slučajan način. U praksi je nekad ovaj zahtjev teško ispuniti(priroda ispitivane pojave ili zahtjevni materijalni troškovi).

Metod uzorka

Imamo sljedeće procedure izbora:

i) prost slučajan izbor(sa vraćanjem i bez vraćanja). Procedura formiranja uzorka: Tablice slučajnih brojeva

ii) kontrolirani izbor. Procedure formiranja uzorka:mehanički izbor, višestepeni izbor i stratifikovani izbor.

Raspodjela(distribucija) parametra skupa

Parametri raspodjele osnovnog skupa nisu slučajne veličine(njihova tačna vrijednost mogla bi se dobiti ispitivanjem svih jedinica skupa ili ako bi bila poznata funkcija raspodjele osnovnog skupa). Parametri uzorka izračunavaju se nakon što se formira uzorak.

Metod uzorka

Ovi parametri su slučajne veličine, jer su vrijednosti koje formiraju slučajan uzorak realizacije slučajne veličine X. Svaki parametar uzorka ima svoju raspodjelu:

i) teorija tačnih raspodjela,
ii) teorija asimptotskih raspodjela

Metod uzorka

Tačne raspodjele parametra iz osnovnog skupa sa normalnom raspodjelom

Aritmetička sredina Broj uzoraka veličine n koji se mogu dobiti iz osnovnog skupa veličine N :

i) sa vraćanjem $k = N^n$,

ii) bez vraćanja $k = \frac{N!}{n!(N-n)!} = \frac{N(N-1)(N-2)\cdots(N-n+1)}{1 \cdot 2 \cdots n}$.

Metod uzorka

Elementi uzorka $x_1, x_2, \ldots x_n$ se posmatraju kao realizacije nezavisnih slučajnih veličina $X_1, X_2, \ldots, X_n$ sa istom raspodjelom, zato je aritmetička sredina uzorka $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ jedna realizacija slučajne veličine

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Za raspodjelu aritmetičkih sredina uzoraka mogu se izračunati njeni pokazatelji

Metod uzorka

(a) Srednja vrijednost osnovnog skupa je poznata

i) Aritmetička sredina raspodjele sredina uzoraka $\overline{\overline{X}} = \frac{\sum_{i=1}^{n} \bar{x}_i}{k}$, gdje su $\bar{x}_i$ aritmetičke sredine uzoraka, a k broj uzoraka. Primijetimo da ako je μ aritmetička sredina osnovnog skupa, onda je $\overline{\overline{X}} = \mu$.

ii) Varijansa raspodjele sredina uzoraka: $s^2 = \frac{\sum_{i=1}^{n} (\bar{x}_i - \mu)^2}{n}$. Sa s ćemo označavati standardnu devijaciju uzorka, a sa σ standardnu devijaciju osnovnog skupa.

(b) Srednja vrijednost osnovnog skupa nije poznata

Metod uzorka

Ako sredja vrijednost osnovnog skupa nije poznata, varijansa uzorka je $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$.
Ako je poznata varijasna osnovnog skupa, varijansa raspodjele aritmetičkih sredina uzorka u slučaju prostih slučajnih uzoraka bez vraćanja(ponavljanja) je $s^2 = \frac{\sigma}{n}\frac{N-n}{N-1}$.
Ako su uzorci sa vraćanjem(ponavljanjem), varijansa uzorka je $s^2 = \frac{\sigma^2}{n}$.
Standardna devijacija raspodjele aritmetičkih sredina uzorka naziva se *standardna geška aritmetičke sredine* i jednaka je

Metod uzorka

$$s_{\bar{X}} = \sqrt{\frac{\sum\limits_{i=1}^{n}(\bar{x}_i - \mu)^2}{k}}.$$

Ako su poznate varijansa ili standardana devijacija osnovnog skupa, standardna greška aritmetičke sredine u slučaju uzorka bez vraćanja je $s_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$, odnosno s vraćanjem je $s_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. Izraz $\frac{N-n}{N-1}$ naziva se *korektivni faktor*. Vrijednost ovog faktora je uvijek manja od 1,zato ovaj faktor utječe na smanjenje greške pa su i zaključci precizniji. Ako je veličina uzorka relativno veća u odnosu na osnovni skup, onda se vrijednost korektivnog faktora smanjuje i postaje nula kada je $n = N$. U ovom slučaju je i standardna greška jednaka nuli.

Metod uzorka

Ako se vrijednost korektivnog faktora približava jedinici, tj. ako se veličina uzorka smanjuje i blizu je jedan ($n = 1$), onda je utjecaj korektivnog faktora na smanjenje greške zanemarljiv. Korektivni faktor se zanemaruje u slučaju beskonačnog osnovnog sklupa. U slučaju konačnog osnovnog skupa i biranja uzorka bez vraćanja obično se korektivni faktor zanemaruje ako je $\frac{n}{N} \leq 0,04$.

Ocjena nepoznatog parametra

Statističko zaključivanje je postupak pomoću kojeg se donose zaključci o vrijednostima nepoznatog parametra osnovnog skupa, a na osnovu informacija koje dobijemo iz uzorka. Osnovni vidovi statistučkog zaključivanja su:

i) statističko ocjenjivanje
ii) testiranje hipoteza

Kriterij za izbor između ova dva vida je sljedeći: Polazimo od osnovnog skupa i izvlačimo slučajni uzorak veličine n. Ako imamo neku informaciju o parametaru, onda testiramo hipotezu. Pretpostavljenu vrijednost parametra odbacujemo ili ne odbacujemo sa unaprijed definiranim rizikom. Ako nemamo informaciju o parametru, onda primijenjujemo ocjenjivanje parametra. Ocjenjenu vrijednost parametra izražavamo brojem ili intervalom sa određenom pouzdanošću.

Ocjena nepoznatog parametra

Ocjena nepoznatog parametra osnovnog skupa primijenjuje se:

i) kad je nepoznata veličina osnovnog skupa(broj N nije poznat)

ii) kad se ne mogu utvrditi sve vrijednosti obilježja osnovnog skupa

iii) kad je osnovni skup beskonačan.

Postupak je sljedeći: Iz osnovnog skupa sa funcijom raspodjele $F(x, \theta)$, gdje je θ nepoznati parametar, uzima se uzorak veličine n. Na osnovu dobivenih posmatranja $x_1, x_2, \ldots, x_n$ izračunava se približa vrijednost nepoznatog parametra. Ova vrijednost se zove *vrijednost ocjene nepoznatog parametra* i označava se sa $\hat{\theta}$. Za dobivanje ocjene nepoznatog parametra posmatra se funkcija $\hat{\theta} = u(X_1, X_2, \ldots, X_n)$, koja se naziva *uzoračka funkcija*(ocjenitelj) ili *statistika*.

Ocjena nepoznatog parametra

Cilj ocjene nepoznatog parametra θ svodi se na nalaženje takvih uzorčkih funkcija koje se mogu koristiti kao dobra ocjena nepoznatog parametra θ.

- Postoji mogućnost da se donese pogrešan zaključak
- U praksi se obično uzima jedan uzorak veličine n. Na osnovu tog uzorka ocjenjuje se nepoznati parametar osnovnog skupa.
- S obzirom da je uzorak slučajan i konačan, onda su sve uzoračke funkcije slučajne.
- Ocjena $\hat{\theta}$ je slučajna veličina, a njena vrijednost, izračunata na osnovu uzorka, $\hat{\theta} = u(x_1, x_2, \ldots, x_n)$ je jedna realizacija iz skupa mogućih realizacija.

Ocjena nepoznatog parametra

Ocjene se dijele na:

i) tačkaste

ii) intervalne

Tačkasta ocjena je određena jednim brojem $\hat{\theta}$, dok je intervalna ocjena određena sa dva broja $\hat{\theta}_1$ i $\hat{\theta}_2$ koji su krajnje tačke intervala koji pokriva ocjenjivani parametar θ.

Ocjena nepoznatog parametra

Osobine dobre ocjene su:

i) nepristrasnost(ocjena je nepristrasna kada je njeno matematičko očekivanje jednako ocjenjivanom parametru)

ii) saglasnost(konzistentnost)(ocjena je saglasna ako pri povećanju veličine uzorka ocjena u vjerovatnoći teži ocjenjivanom parametru)

iii) efikasnost(između svih mogućih nepristrasnih ocjena izračunatih na osnovu uzorka, efikasnom se smatra ona koja ima najmanju varijansu)

Ocjena nepoznatog parametra

Izračunavanje standardne greške aritmetičke sredine Ako je poznata varijansa(standardna devijacija) osnovnog skupa, onda se standardna greška aritmetičke sredine može izračunati na sljedeći način: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.
Ako je uzorak uzet iz velikog ili beskonačnog osnovnog skupa, onda je $\sigma\bar{X} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$.
S obzirom da je varijansa(standardna devijacija) osnovnog skupa uglavnom nepoznata, onda se koriste varijansa(standardnan devijacija) uzorka.
Ocjenjena standardna greška aritmetičke sredine za prost slučajan uzorka bez ponavljanja je $s_{\bar{X}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$, odnosno s ponavaljanjem $s_{\bar{X}} = \frac{S}{\sqrt{n}}$.

Ocjena nepoznatog parametra

Standardna greška aritmetičke sredine može se izračunati direktno iz podataka uzorka: $s_{\bar{X}} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n(n-1)} \frac{N-n}{n}}$, odnoso radna formula je $s_{\bar{X}} = \sqrt{\frac{\sum(x_i)^2 - \frac{(\sum x_i)^2}{n}}{n(n-1)} \frac{N-n}{n}}$.

Kod raspodjele frekvencija za ocjenu standardne greške aritmetičke sredine koriste se formule:

$$s_{\bar{X}} = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{n(n-1)} \frac{N-n}{n}},$$

odnosno

$$s_{\bar{X}} = \sqrt{\frac{\sum f_i(x_i)^2 - \frac{(\sum f_i x_i)^2}{n}}{n(n-1)} \frac{N-n}{n}}.$$

Interval povjerenja za ocjenu nepoznate sredine osnovnog skupa

Interval povjerenja nepoznatog parametra osnovnog skupa je interval u kome se sa određenom vjerovatnoćom nalazi parametar osnovnog skupa.
U praksi interval povjerenja se utvrđuje na bazi 95% ili 99%, što povlači da je mogućnost greške 5%, odnosno 1%.
Mogućnost greške naziva se *prag značajnosti*, označva se sa $\alpha = 0,05$ ili $\alpha = 0,01$.
Interval povjerenja za srednju vrijednost osnovnog skupa pri poznatoj varijansi ima sljedeći oblik

$$\bar{X} - Z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}}$$

Interval povjerenja za ocjenu nepoznate sredine osnovnog skupa

i pokriva srednju vrijendost osnovnog skupa sa zadatom vjerovatnoćom $1-\alpha$ i tačnost ocjene srednje vrijednosti osnovnog skupa $\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. Dužina intervala je $\Delta = 2 \cdot Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. Povećanjem veličine uzorka, pri istoj vjerovatnoći, povećava se i tačnost ocjene (ε je obrnuto proporcionalan veličini uzorka).

Interval povjerenja za ocjenu nepoznate sredine osnovnog skupa

Interval povjerenja za srednju vrijednost osnovnog skupa pri nepoznatoj varijansi ima sljedeći oblik

$$\bar{X} - t_{(n-1,\frac{\alpha}{2})} \cdot s_{\bar{X}} \leq \mu \leq \bar{X} + t_{(n-1,\frac{\alpha}{2})} \cdot s_{\bar{X}}$$

gdje se $t_{(n-1,\frac{\alpha}{2})}$ određuje iz uvjeta
$P(-t_{(n-1,\frac{\alpha}{2})} < t < t_{(n-1,\frac{\alpha}{2})}) = 1-\alpha$.
U slučaju velikog uzorka $t-$raspodjela se može aproksimirati standardizovanom normalnom raspodjelom, pa je interval povjerenja za $(1-\alpha)100\%$.

Primjeri

1. Nova sorta pšenice posijana je na 5 oglednih parcela. Nakon žetve dobiveni su sljedeći prinosi(u t/ha):

$$6 \quad 6,5 \quad 6,9 \quad 7,0 \quad 7,8$$

Ocijeniti prosječan prinos pšenice sa pozdanošću 99% (rizik je 1%). Kolika je tačnost ocjene?

Rješenje Primijetimo da nije poznata varijansa(standardna devijacija)osnovnog skupa. Na osnovu informacije iz uzorka, imamo

$$\bar{x} = \frac{\sum x_i}{n} = \frac{34,2}{5} = 6,84$$

$$s_{\bar{X}} = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n(n-1)}} = \sqrt{\frac{235,7 - 5 \cdot 46,78}{5 \cdot 4}} = 0,2976.$$

Primjeri

$$1-\alpha = 0,99 \Rightarrow \frac{\alpha}{2} = 0,005 \Rightarrow t_{(4;0,005)} = 4.604.$$

Sada je interval povjerenja:

$$\begin{aligned}
\bar{X} - t_{(n-1,\frac{\alpha}{2})} \cdot s_{\bar{X}} \leq & \quad \mu \quad \leq \bar{X} + t_{(n-1,\frac{\alpha}{2})} \cdot s_{\bar{X}} \\
6,84 - 4.604 \cdot 0,2976 \leq & \quad \mu \quad \leq 6,84 + 4.604 \cdot 0,2976 \\
5,4698 < & \quad \mu \quad < 8,2101.
\end{aligned}$$

Sa sigurnošću od 99% može se tvrditi da će prosječan prinos nove sorte pšenice biti u intervalu $[5,4698; 8,2101]$.

Primjeri

Tačnost dobivene ocjene je

$$\varepsilon = t_{(4;0,005)} \cdot s_{\bar{X}} = 4,604 \cdot 0,2976 = 1,3701$$

Primjeri

2. Iz osnovnog skupa normalne raspodjele izvučen je uzorak veličine $n = 16$, u cilju ocjene aritmetičke sredine skupa sa sigurnošću 95%. Neke je poznata standardna devijacija skupa $\sigma = 8$, a vrijednost aritmetičke sredine u uzorku $\bar{x} = 64$. Odrediti interval povjerenja za aritmetičku sredinu osnovnog skupa.
Rješenje Interval povjerenja sa 95% sigurnošću za aritmetičku sredinu osnovnog skupa pri poznatoj varijansi(standardnoj devijaciji) glasi:

$$[\bar{X} - 1,96\sigma_{\bar{X}}, \ \bar{X} + 1,96\sigma_{\bar{X}}].$$

Primjeri

Uvrštavanjem u formulu u izraz, dobivamo

$$64 - 1,96 \cdot \frac{8}{4} \leq \mu \leq 64 + 1,96 \cdot \frac{8}{4}$$

$$60,08 \leq \mu \leq 67,92.$$

Dakle, sa 95% sigurnošću tvrdimo da je aritmetička sredina osnovnog skupa jednaka jednoj od vrijednosti iz intervala $[60,08; 67,92]$.

Primjeri

3. Zasijano je 200 ha neke poljoprivredne kulture. Da bi se ocijenio srednji prinos te kulture posmatrano je 12 ha. Dobiveni rezultati su:

Prinos(t)	3-4	4-5	5-6	6-7
Površina(ha)	3	4	2	3

Ocijeniti srednji prinos poljoprivrede kulture za cijelo polje s pouzdanošću 99%. Kolika je tačnost procjene i kolika je procjena ukupnog prinosa?

Primjeri

Rješenje $N = 200$ i $n = 12$, $\frac{n}{N} = \frac{12}{200} = 0,06$ što je veće od 0,04. Zato se varijansa ocjenjuje pomoću formule

$$s_{\bar{x}} = \sqrt{\frac{\sum f_i x_i^2 - n\bar{x}^2}{n(n-1)} \cdot \frac{N-n}{N}}$$

$$= \sqrt{\frac{305 - 292.84}{12 \cdot 11} \cdot \frac{200-12}{200}} = 0,294$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = 4,92.$$

Kako je $1 - \alpha = 0,99 \Rightarrow \frac{\alpha}{2} = 0,005 \Rightarrow t_{(\frac{\alpha}{2},11)} = 3,106$

Primjeri

Interval povjerenja je

$$\bar{x} - t_{(\frac{\alpha}{2},n-1)} \leq \mu \leq \bar{x} + t_{(\frac{\alpha}{2},n-1)}$$

$$4,92 - 3,106 \cdot 0,294 \leq \mu \leq 4,92 + 3,106 \cdot 0,294$$

$$4,006 \leq \mu \leq 5,833$$

Dakle, sa sigurnošću 99% može se reći da će prosječan prinos biti u intervalu $[4,006; 5,833]$. Tačnost dobivene ocjene je $\varepsilon = 3,106 \cdot 0,294 = 0,913$. Procjena ukupnog prinosa je

$$N(\bar{x} - t_{(\frac{\alpha}{2},n-1)}) \leq N\mu \leq N(\bar{x} + t_{(\frac{\alpha}{2},n-1)})$$

Primjeri

tj.

$$801,02 \leq N\mu \leq 1166,1$$

Testiranje statističkih hipoteza

Statistička hipoteza je precizna pretpostavka o karakteristici osnovnog skupa koja se može provjeriti.
Statistički metod koji na osnovu uzorka provjerava prihvatljivost hipoteze zove se statističko testiranje, a sama procedura statističkim testom.
Statistička hipoteza može biti parametarska i neparametarska. Kod parametarskih hipoteza, pretpostavka se odnosi na vrijednost parametra funkcije raspodjele. Kod neparametarskih hipoteza, pretpostavka se odnosi na tip raspodjele osnovnog skupa.
Postavljena hipoteza se obično zove nulta i označava se sa H_0.
Svakoj nultoj hipotezi suprostavljamo alternativnu hipotezu koju označavamo sa H_1.

Testiranje statističkih hipoteza

Konkretna nulta hipoteza obično se postavlja u vidu "nema promjene", "nema razlike", "nema ujecaja", tj. u vidu status quo. Alternativna je suprotna nultoj, postavlja se u vidu "ima promjene", "ima razlike".
Nulta hipoteza može biti prosta i složena. Nulta hipoteza je prosta ako se njom tvrdi da je parametar jednak tačno jednoj unaprijed datoj vrijednosti. Ako nulta hipoteza obuhvata više od jedne vrijednosti onda se kaže da je složena. U postavci nulte hipoteze uvijek mora biti sadržan znak jednakosti ($" = "$, $\leq$, $\geq$).
Alternativna je uvijek složena, jer sadrži sve vrijednosti koje nisu obuhvaćene nultom hipotezom.

Testiranje statističkih hipoteza

Svi statistički testovu mogu se klasificirati po tri osnova:
1. Prema prirodi problema koje rješavamo imamo testove o parametrima skupa(aritm.sredina, varijansa itd.), o obliku raspodjele, o slučajnosti uzorka itd.
2. Prema broju uzoraka na kojima zasnivamo testiranje imam testove zasnovane na jednom uzorku, na dva uzorka i tri i više uzoraka.
3. Prema vrsti i jačini preduvjeta na kojima se zasnivaju imamo parametarske i neparametarske testove.
Parametarski testovi imaju zajednički preduvjet da osnovni skup ima normalnu raspodjelu. Neparametarski ne zahtijevaju ovaj preduvjet i zasnivaju se na puno blažim uvjetima.

Testiranje statističkih hipoteza

Testiranje statističkih hipoteza sastoji se iz sljedećih koraka:
1. Formulacija nulte i alternativne hipoteze i zadavanje nivoa značajnosto α,
2. Uzima se slučjani uzorak i bira optimalni test,
3. Provjeravaju se preduvjeti za odabrani test. Ako nisu ispunjeni bira se novi test, a ako jesu procedura se dalje nastavlja,
4. Izračunava se tzv.statistika testa(iz uzorka). Određuje $p-$ vrijednost alternativno određuje se tzv.kritična vrijednost iz tablica odgovarajuće raspodjele(određuje se kritična oblast K),
5. Donosi se zaključak o odbacivanju ili neodbacivanju nulte hipoteze na osnovu $p-$vrijednosti alternativno kritične oblasti,

Testiranje statističkih hipoteza

6. Ako dobivena statistika testa iz uzorka pripada K, nulta hipoteza se odbacuje sa nivoom značajnosti α. Ako izračunata statistka testa ne pripada K nult ahipoteza se ne odbacuje. Ako se desi da je izračunata statistika testa na granici kritične oblasti, onda se ne donosi zaključak, nego se rade daljna ispitivanja radi donošenja zaključka.

Alternativnu hipotezu kojom moguća odstupanja stvarne od hipotetičke vrijednosti pratimo u oba smjera nazovamo *dvosmjernom ili dvostranom*. Shodno tome odgovarajući test je dvosmjerni (dvostrani). Ako pratimo odstupanje samo u jednom smjeru, onda imao *jednosmjerne ili jednostrane* testove.

Testiranje statističkih hipoteza

Kako odabrati između dvosmjernog i jednosmjernog testa?
Ako testirana vrijednost ne smije odstupiti u bilo kom smjeru od standarda (hipotetičke vrijednosti)ili ako unaprijed ništa ne znamo o potencijalnom odsupanju parametra od njegove hipotetičke vrijednosti, onda koristimo dvosmjerne testove. (npr.SPSS nema mogućnost jednosmjernog testiranja, pa se o tome treba voditi računa).

Greške pri testiranju

1. Ako pri testiranju odbacimo istinitu nultu hipotezu, onda činimo **grešku prve vrste**.
2. Ako pri testiranju ne odbacimo netačnu nultu hipotezu, onda činimo **grešku druge vrste**.

Testiranje statističkih hipoteza

Nivo značajnosti testa(naziva se i **rizik greške prve vrste**) je vjerovatnoća da ćemo odbaciti istinitu nultu hipotezu. Označava se sa α. Obično se α bira da je 0.05 ili 0.01.
Vjerovatnoća da nećemo odbaciti netačnu nultu hipotezu naziva se **rizikom greške druge vrste** i označava se sa β.
$\alpha + \beta \neq 1$.
rizik greške **nije isto što i** greška!
Jačina(snaga) testa je vjerovatnoća da se odbaci netačna nulta hipoteza.
Najbolji je test onaj koji za odabrani nivo značajnosti α ima najveću jačinu(snagu).

Testiranje parametarskih hipoteza

Testiranje aritmetičke sredine

(i) Poređenje aritmetičke sredine uzorka sa aritmetičkom sredinom osnovnog skupa(Z-test ako je poznata standardna devijacija osnovnog skupa. $t-$test ako nije poznata standardna devijacija osnovnog skupa.

(ii) Poređenje dvije aritmetičke sredine dva nezavisna uzorka(Z-test ako su poznate standardne devijacije osnovnih skupova. t-test ako nisu poznate standardne devijacije osnovnih skupova) .

(iii) Poređenje tri i više sredina iz tri i više uzoraka (ANOVA-Analiza varijanse)

Testiranje parametarskih hipoteza

Statistika testa koju ćemo koristiti ima najčešće sljedeći oblik

$$\text{Statistika testa} = \frac{\text{Ocjena parametra-Hipotetička vrijednost parametra}}{\text{Standardna greška ocjene}}.$$

Mala vrijednost statistike testa ne daje nam osnov da odbacimo nultu hipotezu. Međutim, njene velike vrijednosti sugeriraju da nulta hipoteza nije tačna.
Koje vrijednosti ćemo smatrati malim a koje velikim? Postoje dva načina za donesemo zaključak o odbacivanju ili neodbacivanju nulte hipoteze:Neyman-Paersonov način(preko tablica raspodjele) i Fischerov način zasnovan na tzv.$p-$ vrijednosti.

Testiranje parametarskih hipoteza

p-**vrijednost** je vjerovatnoća da statistika testa uzme vrijednost jednaku ili još ekstremniju od vrijednosti koja se upravo realizirala u uzorku, pod uvjetom da je nulta hipoteza tačna. Što je manja p-vrijednsot jači su dokazi protiv nulte hipoteze.
Ako je $p-$ vrijendost manja od odabranog nivoa znčajnosti α, nulta hipoteza se odbacuje. U suprotnom nemamo dovoljno dokaza da odbacimo nultu hipotezu.
Ako je $p-$vrijednost manja od nivoa zančajnosti α, kažemo da je rezultat statistički značajan na nivou α. (Nedostatak: ako se obrađuju podaci bez korištenja računara, onda se može javiti veliki problem, jer se tačna p vrijednost može odrediti samo kod $Z-$testa na osnovu tablice).

Testiranje parametarskih hipoteza

1. Testiranje aritmetičke sredine osnovnog skupa ako je poznata standardna devijacija osnovnog skupa i osnovni skup ima normalnu raspodjelu. Mogućnosti za nultu i alternativnu hipotezu:

$$\begin{aligned} H_0 &: \mu = \mu_0 \quad H_1 : \mu \neq \mu_0 \quad \text{dvostrani test} \\ H_0 &: \mu \leq \mu_0 \quad H_1 : \mu > \mu_0 \quad \text{jednostrani test} \\ H_0 &: \mu \geq \mu_0 \quad H_1 : \mu < \mu_0 \quad \text{jednostrani test} \end{aligned} \tag{1}$$

$Z-$statistika glasi

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}, \; \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

Donošenje zaključka Zaključak o odbacivanju ili neodbacivanju nulte hipoteze donosi se na osnovu tablica normalne raspodjele. Najčešće se za nivo značajnosti α uzima 0,05 ili 0,01. Za $\alpha = 0,05$ kritična vrijednost za Z je 1,96, a za $\alpha = 0,01$ kritična vrijednost Z je 2,58. Ako je apsolutna vrijednost izračuntaog Z iz formule veća ili jednaka od navedenih vrijednosti iz tablica, onda odbacujemo nultu hipotezu. Ako je apsolutna vrijednost izračunatog Z iz formule manja od navedenih tabličnih vrijednosti, onda nemamo osnova da odbacimo nultu hipotezu. Važno je napomenuti da ovo ne znači da je nulta hipoteza tačna, nego samo da dokazi protiv nulte hipoteze nisu dovoljno jaki. Formulacija da se prihvata nulta hipoteza znači da rezultati iz uzorka podržavaju nultu hipotezu i da se ona ne može odbaciti.

Ako se odlučuje na osnovu $p-$vrijednosti, onda ako je p vrijednost manja od nivoa značajnosti α onda odbacujemo nultu hipotezu. Generalno,p vrijednost kod svakog Z testa izračunavamo:

(i) $H_1 : \mu > \mu_0 \quad p = P(Z \geq z)$.

(ii) $H_1 : \mu < \mu_0 \quad p = P(Z \leq z)$.

(iii) $H_1 : \mu \neq \mu_0 \quad p = 2P(Z \geq |z|)$.

Ako nije poznata standardna devijacija osnovnog skupa i osnovni skup ima normalnu raspodjelu, onda koristimo $t-$statistiku, odnosno $t-$test

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}, \; s_{\bar{X}} = \frac{s}{\sqrt{n}}.$$

Pravilo odlučivanja je pomoću kritičnih oblasti (ili $p-$vrijednosti ako se koristi računar). U tablicama za $t-$raspodjelu odredi se tablična vrijednost(kritična vrijednost) za odabrani nivo značajnosti i broj stepeni slobode. Ako je izračunata iz formule vrijednost t manja od tabliňe onda se nulta hipoteza ne odbacuje. Nasuprot tome, ako je veća ili jednaka onda se nulta hipoteza odbacuje.

Stepen slobode Broj stepeni slobode je jednak broju slobodnih opservacija u uzorku, koji se dobiva kada se od ukupnog broja jedinica u uzorku oduzme broj ograničenja koja se nameću ovim vrijednostima. Za svako uvedeno ograničenje gubi se po jedan stepen slobode.

Testiranje hipoteze zasnovano na dva uzorka

Moguće nulta i alternativna hipoteza

$$\begin{aligned} H_0: &\quad \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2 \,\text{dvosmjerna alternativna} \\ H_0: &\quad \mu_1 \leq \mu_2 \quad H_1: \mu_1 > \mu_2 \,\text{jednosmjerna alternativna} \\ H_0: &\quad \mu_1 \geq \mu_2 \quad H_1: \mu_1 < \mu_2 \,\text{jednosmjerna alternativna.} \end{aligned}$$

$Z-$test koristimo ako oba osnovna skupa imaju normalnu raspodjelu i ako su poznate standardne devijacije oba osnovna skupa.

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{(\bar{X}_1 - \bar{X}_2)}}, \quad \sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

$n_1$broj jedinica prvog uzorka,a n_2 broj jedinica drugog uzorka. (Zaključuje se na ranije opisani način).
$t-$test koristimo ako nisu poznate standardne devijacije, ali pretpostavljamo da su jednake, tj. da je uvjet homogenosti varijansi zadovoljen.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{(\bar{X}_1 - \bar{X}_2)}}, \quad s_{(\bar{X}_1 - \bar{X}_2)} = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

pod uvjetom da je H_0istinita ima Studentovu raspodjelu sa $n_1 + n_2 - 2$ stepena slobode. s_1 je standardna devijacija prvog uzorka,a n_2 je standardna devijacija drugog uzorka. Zaključak se donosi na opisani način.

Testiranje hipoteze zasnovano na tri ili više uzoraka(Analiza varijanse(ANOVA))Ako bismo koristili $t-$test i testirali po dva uzorka povećali bismo grešku prve vrste. ANOVA ukupan varijabilitet razdvaja na dva, tj. varijabilitet koji nastaje zbog utjecaja primijenjenih tretmana i slučajan varijabilitet. Ovdje se pojavljuju više od jedne alternativne hipoteze.

Posmatra se utjecaj više različitih faktora na posmatrano obilježje. Različite vrijednosti faktora nazivaju se nivoi. Npr.na prinos neke kulture utječu faktori:đubrivo, sorta, navodnjavanje i sl. dok su npr.različite vrste đubriva nivoi faktora đubrivo.
U zavisnosti od broja ispitivanih faktora razlikujemo jednofaktorsku, dvofaktorsku, višefaktorsku ANOVU.
Jednofaktorska ANOVA Polazimo od jednog faktora koji ima k različitih nivoa. Svakom od ovih k nivoa odgovara jedan osnovni skup sa normalnom raspodjelom $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots k$.

Testiranje parametarskih hipoteza

Parametri μ i σ_i^2 k osnovnih skupova su nepoznati, ali se pretpostavlja da su varijanse međusobno jednake, tj.$\sigma_i = \sigma_j$, $i \neq j$, $i, j = 1, \ldots, k$. Cilj testiranja je da se provjeri da li postoji statistički značajna razlika između aritmetičkih sredina datih skupova, tj. $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$, a alternativna H_1 je da postoji bar jedan osnovni skup sa različitom sredinom. Ako se H_0 ne odbaci, onda se k osnovnih skupova posmatraju kao jedan osnovni skup sa sredinom $\mu = \frac{1}{k}\sum_{i=1}^{k} \mu_i$. Ako se H_0 odbaci, onda se nekom od metoda grupnog poređenja izdvajaju skupovi sa različitim sredinama, tj. izdvajaju se nivoi faktora koji imaju najviši utjecaj na posmatrano obilježje.

Testiranje parametarskih hipoteza

Uzorci za različite nivoe faktora

Nivoi faktora							
Ponavljanja	1	2	$\cdots$	i	$\cdots$	k	
	x_{11}	x_{12}	$\cdots$	x_{i1}	$\cdots$	x_{k1}	
	x_{12}	x_{22}	$\cdots$	x_{i2}	$\cdots$	x_{k2}	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
	x_{1n_1}	x_{2n_2}	$\cdots$	x_{in_i}	$\cdots$	x_{kn_k}	
Obim	n_1	n_2	$\cdots$	n_i	$\cdots$	n_k	N
Suma	G_1	G_2	$\cdots$	G_i	$\cdots$	G_k	G
Sredina	$\bar{x}_i$	$\bar{x}_2$	$\cdots$	$\bar{x}_i$	$\cdots$	$\bar{x}_k$	$\bar{x}$

Testiranje parametarskih hipoteza

Sad se izračunavaju:

$$
\begin{aligned}
\bar{x}_i &= \frac{1}{n}\sum_{j=1}^{n_i} x_{ij} = \frac{G_i}{n_i} \\
\bar{x} &= \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} = \frac{G}{N} \\
\hat{\alpha}_i &= \bar{x}_i - \bar{x},\ i = 1, \ldots, k \\
G_i &= \sum_{j=1}^{n_i} x_{ij},\ G = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij},\ N = \sum_{i=1}^{n} n_i .
\end{aligned}
$$

Testiranje parametarskih hipoteza

Ukupna varijansa sume kvadrata razbija se na faktorsku sumu kvadrata i sumu kvadrata greške, tj. $SS_U = SS_F + SS_G$, gdje je

$$SS_U = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2,$$

iskazuje odstupanje posmatranja od generalne uzoračke sredine, a karakterizira je variranje u objedinjenom osnovnom supu od generalne sredine

$$SS_F = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2,$$

iskazuje odstupanje uzoračkih sredina od generalne uzoračke sredine, a karakterizira je međugrupno variranje

Testiranje parametarskih hipoteza

$$SS_G = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2,$$

iskazuje unutargrupno variranje.
Često se koriste tzv.radne formule:

$$\begin{aligned} SS_U &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 - C, \; C = \frac{G^2}{N} \\ SS_F &= \sum_{i=1}^{k} \frac{G_i^2}{n_i} - C \\ SS_G &= SS_U - SS_F. \end{aligned}$$

Testiranje parametarskih hipoteza

Ocjene ukupne, faktorske i varijanse greške:

$$
\begin{aligned}
MS_U &= \frac{SS_U}{N-1} \\
MS_F &= \frac{SS_F}{k-1} \\
MS_G &= \frac{SS_G}{N-k}.
\end{aligned}
$$

Za provjeru nulte hipoteze koristi se statistika

$$F = \frac{MS_F}{MS_G}.$$

Testiranje parametarskih hipoteza

Ova statistik aima F raspodjelu sa $\nu_1 = k - 1$ $\nu_2 = N - k$ stepeni slobode. Iz tablica se određuje vrijednost $F_{(\alpha,\nu_1,\nu_2)}$. Zatim se ova vrijednost upoređuje sa izračunatom statistikom F. Ako je $F < F_{(\alpha,\nu_1,\nu_2)}$, onda nema osnova za odbacivanje nulte hipoteze. Ako je $F > F_{(\alpha,\nu_1,\nu_2)}$, onda se nulta hipoteza odbacuje i smatra se da faktor utječe na posmatrano obilježje. Obično se rezultati izračunavanja prikazuju u sljedećoj tabeli

Testiranje parametarskih hipoteza

Izvor varijacije	Sume kvadrata	Stepeni slobode	Sredine kvadrata	$F-$vrijednost
Faktor	SS_F	$k-1$	$MS_F = \frac{SS_F}{k-1}$	$F = \frac{MS_F}{MS_G}$
Greška	SS_G	$N-k$	$MS_G = \frac{SS_G}{N-k}$	
Ukupno	SS_U	$N-1$		

Ako se nulta hipoteza odbaci znači da postoji razlika između sredina osnovnih skupova. Sada je potrebno izdvojiti skupove sa različitim sredinama. Postoji veliki broj testova za grupno poređenje sredina: $t-$ test, Dankanov test, test najmanje značajne razlike, itd. Broj poređenja je $\frac{k(k-1)}{2}$.

Testiranje parametarskih hipoteza-testovi grupnog poređenja

$t-$**test** Nulta hipoteza je $H_0 : \mu_i = \mu_j$, a alternativna $H_1 : \mu_i \neq \mu_j$, $i < j$, $1 \leq i \leq k$, $1 \leq j \leq k$. Izrav cunava se $t-$statistika $t = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{MS_G\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$, $i, j = 1, 2, \ldots, k$; $i \neq j$. Ako je $n_i = n_j$, onda je $t = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{\frac{2MS_G}{n}}}$. Nulta hipoteza se testira s istim nivoom značajnosti α. Izračunato t upoređuje se sa $t_{(\frac{\alpha}{2}, N-k)}$. Ako je $|t| \leq t_{(\frac{\alpha}{2}, N-k)}$, onda nema osnova za odbacivanje nulte hipoteze. U protivnom nulta hipoteza se odbacuje u korist alternativne.

Testiranje parametarskih hipoteza-testovi grupnog poređenja

Dankanov test Nulta i alternativna hipoteza se postavljaju kao i u prethodnom testu poređenja $H_0 : \mu_i = \mu_j$, $H_1 : \mu_i \neq \mu_j$, $i < j$, $1 \leq i \leq k$, $1 \leq j \leq k$. Za primjenu ovog testa treba biti jednak broj ponavljanja kod svakog ispitivanog tretmana. Za pragove značajnosti $\alpha = 0,05$ i $\alpha = 0,01$ formiraju se dvije tabele oblika

Interval	2 3 4 . . . k
Kritična vrijednost	
Najmanje značajni interval	

Testiranje parametarskih hipoteza-testovi grupnog poređenja

U prvom redu tabele upisuju se mogući intervali na osnovu broja posmatranih tretmana. Zatim se očitaju kritične vrijednosti iz tablica za višetruki test intervala za date pragove značajnosti i stepene slobode greške iz tabele ANOVE i to za svaki interval od 2,3 ,4,..., k i upisuju se u drugi red. Očitane i upisane kritične vrijednosti množe se sa izračunatom ocjenom standardne greške arit.sredine, a proizvod predstavlja vrijednost najmanjeg značajnog intervala i upisuje se u treći red. Sa najmanje značajnim intervalom porede se razlike aritm. sredina.

Testiranje parametarskih hipoteza-testovi grupnog poređenja

Test najmanje značajne razlike-NZR test Nulta i alternativna hipoteza se postavljaju kao i u prethodna dva testa. Ovdje se prvo izračunaju najmanje značajne razlike

$$NZR_{\alpha} = t_{(N-k,\alpha)} \cdot S_{(\bar{x}_1-\bar{x}_2)}.$$

Zatim se formira pomoćna tabela. U prvoj koloni se upisuju aritmetičke sredine po veličini od najveće do najmanje. U sljedeće kolone se unose razlike aritmetičkih sredina koje su pozitivne uvijek. Ove razlike se porede sa izračunatim najmanje značajnim razlikama.

Neparametarski testovi se mogu koristiti kad nije poznat tip funkcije raspodjele osnovnog skupa. Jedn od dobro poznatih neparmetarskih testove je χ^2-test.

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_i^*)^2}{f_i^*}$$

f_i^* su teorijske(očekivane) frekvencije. Test se može primijeniti samo ako su $f_i^* \geq 5$. Za odabrani nivo značajnosti i broj stepeni slobode $\nu = k - m - 1$(k je broj intervala, m broj nepoznatih parametara funkcije raspodjele) odredi se kritična vrijednost iz tablica i upoređuje sa izračunatom vrijednosti.

Ako je izračunato χ^2 manje od tabličnog onda nema osnova za odbacivanje nulte hipoteze, u protivnom odbacujemo nultu hipotezu.

Postoje i drugi neparametarski testovi, kao što su: Test znaka(Sign test), Wilcoxonov test, McNemarov test, Marginal homogeneity test itd.

Primjeri parametarskih testiranja

1. Za ispitivanje utjecaja dva načina prihranjivanja na prinos jedne sorte kruške, na 50 stabala je primijenjen prvi način, a na 40 drugi način prihranjivanja. Na osnovu podataka dobivenih iz uzorka izračunate su sredine $\bar{x}_1 = 53,5kg/stablu$ i $\bar{x}_2 = 57,7kg/stablu$, i varijanse $\sigma_1^2 = 21,4$ i $\sigma_2^2 = 23,8$. Testirati da li postoji značajna razlika izmedju ova dva načina prihranjivanja na nivou značajnosti $\alpha = 0,05$.
2. Ista vrsta jabuka uzgaja se u dva različita područja neke države. Označimo ta područja sa A i B. Na slučajan način odabrano je 9 stabala iz područja A i izmjeren je njihov prinos po stablu u kg: $31, 26, 28, 36, 37, 25, 24, 32; 25$ a prinos po stablu na slučajno odabranih 11 stabala iz područja B iznosi:

$$38, 36, 29, 25, 30, 38, 26, 28, 22, 24, 28.$$

Ako je poznato da je prinos normalna slučajna promjenljiva, na nivou značajnosti $\alpha = 0,05$ testirati hipotezu da jabuke u području B daju veći prinos. Možemo li uz isti nivo značajnosti zaključiti da se prinosi jabuka u području A i B razlikuju? (Zadatak uraditi u SPSS.)

Ispituje se utjecaj radnog staža na produktivnost pri proizvodnji nekog proizvoda. Rezultati posmatranja dati su u sljedećoj tabeli

Efikasnost	do 10 god.	10-15 god.	15-25 god.
Broj	135	176	155
proizvoda	156	196	160
za jednu	165	204	149
smjenu		180	171
			140

Pod pretpostavkom da produktivnost rada ima normalnu raspodjelu, sa istom varijansom za sva tri skupa , analizom varijanse provjeriti hipotezu da radni staž ne utječe na produktivnost. Hipotezu testirati na nivou značajnosti $\alpha = 0,05$.(SPSS)

Primjeri neparametarskih testiranja

U izabranom uzorku od 100 domaćinstava koji su anketirani o potrošnji biljnog ulja, 40 je izjavilo da upotrebljava biljno ulje, a 60 da koristi masnoću životinjskog porijekla. Da li se može smatratai da je odnos broja domaćinstava koji upotrebljavaju ove dvije vrste masnoća $1:1$?

Regresija i korelacija

Cilj je da se uspostavi veza između očekivane vrijednosti zavisno promjenjive na osnovu date nezavisno promjenjive. Formira se jednadžba regresije, zatim se izračunava standardna greška regresije.

Cilj korelacione analize je posmatranje jačine veze između ove dvije promjenjive. Izračunavaju se **koeficijent korelacije koeficijent determinacije**.

Analizira se tzv.**dijagram rasturanja** koji se dobije kad se na x osu nanesu vrijednosti nezavisno promjenjive a na y osu vrijednosti zavisno promjenjive.

Regresija i korelacija

Posmatrajmo neke dvije promjenljive X i Y. Ako su ove promjenljive detrminističke, između njih postoji funkcionalna veza $Y = f(X)$. Znači da svakoj vrijednosti promjenljive X iz oblasti definiranosti odgovara jedna i samo jedna vrijednost Y
Izabrani tip funkcije zove se **regresiona funkcija** ili **regresija**. U slučaju dvije promjenljive koristi se još i termin **prosta regresija** ili **parna regresija**.
Prosta linearna regresija Ispituje se zavisnost dva obilježja X i Y. I osnovnog skupa uzima se određeni broj parova (X_i, Y_i), tj. njihovih realizacija (x_i, y_i).
Jednadžba regresione prave je

$$y = a + bx.$$

Parametri a i b određuju se metodom najmanjih kvadrata.

Regresija i korelacija

Sistem jednadžbi iz kojih se određuju a i b su

$$\begin{aligned}\sum_{i=1}^{n} y_i &= an + b\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i &= a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2.\end{aligned}$$

a je ocjena prosječnog početnog nivoa zavsino promjenljive(jer za $x_i = 0$ je $y_i = a$.); b je ocjena prosječne promjene zavisno promjenljive na jedinicu promjene nezavisno promjenljive(jer je b koeficijent pravca prave).

Regresija i korelacija

Standardna greška regresije

$$s_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}.$$

Ova greška predstavlja mjeru odstupanja empirijskih podataka od jednadžbe regresije. Što je manja standardna greška onda linearni regresioni model bolje opisuje zavisnost posmatranih obilježja.

Regresija i korelacija

Koeficijent korelacije

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

$$|r_{XY}| \leq 1.$$

Koeficijent korelacije određuje jačinu korelacione veze posmatranih promjenljivih. Što je vrijednost r_{XY} bliža $+1$ ili -1 to je zavisnost jača, što je vrijednost bliža nuli zavisnost je slabija. Ako je $r_{XY} > 0$ onda su obje promjneljive istog smjera, tj. obje promjenljive istovremno ili rastu ili opadaju. Ako je $r_{XY} < 0$ onda rast jedne izaziva opadanje druge promjenljive i obrnuto.

Regresija i korelacija

Koeficijent determinacije

$$\hat{d_{XY}}^2 = \frac{SS_R}{SS_U}.$$

Predstavlja proporcionalni dio ukupnog varijabiliteta zavisno promjenljive objašnjen utjecajem nezavnisno promjenljive. Njegova dopuna do jedinice zove se **koeficijent nedeterminacije** i predstavlja proporcionalni dio ukupnog variranja zavisno promjenljive koji nije objašnjen utjecajem nezavinso promjenljive:

$$1 - \hat{d_{XY}}^2 = \frac{SS_G}{SS_U}.$$

Primjeri

Na osnovu podataka o broju kokoški u kokošinjcu i tjelesnoj masi kokoški, ocijeniti jednadžbu regresije i izračunati standradnu grešku(SPSS).

Vel.kokošinjca x	5	6	7	9	11	13
Masa u kg. y	1	0.5	0.7	0.4	0.9	1.1

Planiranje eksperimenta. Latinski kvadrat. Dizajn

Latinski kvadrat(pravougaonik)reda n je matrica (šema) nad skupom $S_n = \{1, 2, \ldots, n\}$, koja ima n (v)vrsta i n kolona, gdje svaka vrsta i svaka kolona sadrži različite elemente iz S_n. Elementi skupa S_n ne moraju biti brojevi nego općenito nekih n različitih elemenata.
Latinski kvadrati se u planiranju eksperimenta koriste za smanjenje greške u eksperimentu.

Primjeri

1. Iskoristiti latinski kvadrat L reda 3 za planiranje sljedećeg eksperimenta: 3 sorte pšenice tretiramo sa dva tipa đubriva i zemljište za sjetvu se priprema na 3 različita načina? Koja kombinacija sorte, đubriva i obrade zemljišta ce dati najbolje prinose.
2. Na prinos pšenice utječu tri varijacije: dubina oranja, količina vještačkog đubriva i sorta pšenice. Svaka od varijacija ima 4 različite vrijednosti. Dubine oranja su:20, 25, 30, 35 cm. Količine đubriva su u kg/hektaru i iznose:220, 250, 280, 300. Sorte pšenice označimo sa 1,2,3,4. Iskoristiti latinski kvadrat za optimalno planiranje eksperimenta.

Latinski kvadrati

Dva latinska kvadrata reda n su međusobno **ortogonalni** ako se na odgovarajućim mjestima iz oba kvadrata nalaze svi različiti uređeni parovi iz $S_n \times S_n$.

Ako postoje ortogonalni latinski kvadrati onda se potreban broj eksperimenata koje treba izvršiti može smanjiti.

Savršenija struktura od ortogonalnih kvadrata je **dizajn**. Da bi dizajn postojao potrebno je i dovoljno da postiji $n-1$ ortogonalnih kvadrata reda n.

Dizajn ili **blok šema** tipa $t-(v,k,\lambda)$ je skup $\mathcal{B}=\{B_1,B_2,\ldots,B_d\}$, $k-$točlanih podskupova, tzv.**blokova** skupa $S_v=\{1,2,\ldots,v\}$, $t<k<v$, pri čemu se svaki $t-$člani podskup iz S_v nalazi u tačno λ blokova.

Primjer

Cilj eksperimenta je da se testiraju 6 vrsta đubriva. Koristi se 18 parcela, 3 parcele na svakoj od 6 farmi u različitim dijelovima države. Postupak je da se đubrivo primjeni na parcele, posadi se biljka i mjeri rezultat. Kako rasporediti đubrivo po parcelama?

Bibliografija

1.R.Mead, R.N.Curnow, A.M.Hasted, Statistical methods in agriculture and experimental biology, Second edition, Springer, 1993.
2. Snežana Matić-Mekić, Primenjena matematika za biološke smerove Poljoprivrednog fakulteta, Univerzitet u Novom Sadu 2015.
3. D.C. Montgomery, Design and Analysis of Experiment, John Wiley & Sons, 2001.
4. B.Mutevelić, E.Nikolić Đori, Statistika, Univerzitet u Novom Sadu, Poljoprivredni fakultet, 2018.
5. J.Stanković, N.R.Ralević I.Ljubanović-Ralević, Statistika sa primjenom u poljoprivredi, Mladost Biro, Beograd, 2012.
6. I. Šošić, Zbirka zadataka iz statistike, Mikrorad i Ekonomski fakultet, Zagreb, 1998.