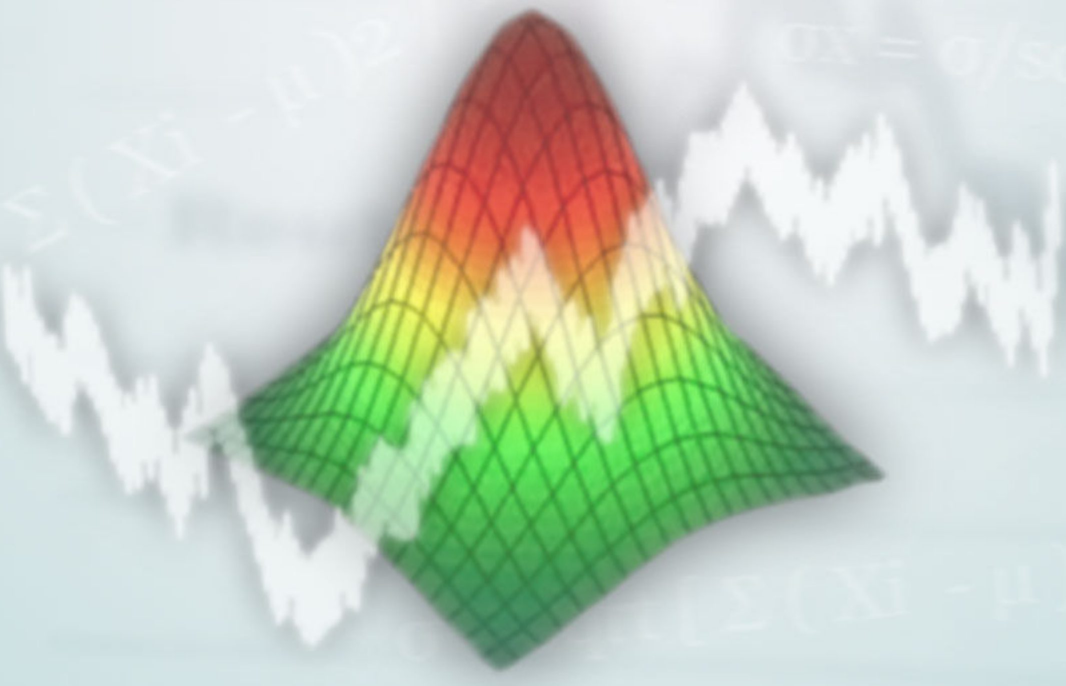
Maja Biljan-August Snježana Pivac Ana Štambuk

STATISTIČKA ANALIZA U EKONOMIJI



Rijeka, 2009.

UDŽBENICI SVEUČILIŠTA U RIJECI

MANUALIA UNIVERSITATIS STUDIORUM FLUMINENSIS

Prof. dr. sc. Maja Biljan-August
Prof. dr. sc. Snježana Pivac
Doc. dr. sc. Ana Štambuk

STATISTIČKA ANALIZA U EKONOMIJI

Izdavač:
Ekonomski fakultet Sveučilišta u Rijeci

Recenzentice:
Prof. dr. sc. Jasna Horvat

Doc. dr. sc. Alemka Šegota

Doc. dr. sc. Tea Baldigara

Lektorica:
Kerol Musul-Perić, prof.

Autor naslovnice:
Luka Mičetić, dipl. oec.

Pri izradi naslovnice korišteni su materijali objavljeni na www.sxc.hu.

Objavljivanje ovog sveučilišnog udžbenika odobrilo je Povjerenstvo za izdavačku djelatnost Sveučilišta u Rijeci Odlukom – klasa: 602-09/09-01/28, ur. broj: 2170-57-05-09-3 od 25. rujna 2009.

Objavljeno na URL: http://www.efri.hr/prikaz.asp?txt_id=6327
i http://oliver.efri.hr/~statist/biljan-pivac-stambuk-statisticka.pdf.

ISBN: 978-953-6148-85-1

Rijeka, rujan, 2009.

Prof. dr. sc. Maja Biljan-August
Prof. dr. sc. Snježana Pivac
Doc. dr. sc. Ana Štambuk

STATISTIČKA ANALIZA U EKONOMIJI

EKONOMSKI FAKULTET U RIJECI
RIJEKA, 2009.

MAJA BILJAN-AUGUST
SNJEŽANA PIVAC
ANA ŠTAMBUK

ISBN: 978-953-6148-85-1

PREDGOVOR

Za ovaj udžbenik može se reći da je nastavak i nadogradnja udžbenika «Upotreba statistike u ekonomiji» istih autorica. Stoga se objašnjenja svih osnovnih statističkih pojmova, koji nisu detaljno prikazani u ovom, mogu pronaći u navedenom udžbeniku.

«Statistička analiza u ekonomiji» namijenjena je prvenstveno studentima Ekonomskog fakulteta u Rijeci i u cijelosti prati nastavni plan i program kolegija «Statistička analiza». Itekako je mogu koristiti i svi drugi zainteresirani korisnici koji u svom praktičnom, stručnom i znanstvenom radu, baveći se društvenim istraživanjima, primjenjuju statističke metode i tehnike.

Kao nastavak rada i u ovoj knjizi su obuhvaćene teorijske osnove i objašnjenja za svako obrađeno i uključeno područje statistike. Kroz rješavanje konkretnih primjera daju se objašnjenja dobivenih rezultata i njihovo kritičko vrednovanje.

Kao posebnu vrijednost ovog udžbenika potrebno je istaknuti da su u privitku i ovdje dane upute za primjenu programskog paketa *Statistica* u statističkoj analizi, što pruža mnoštvo mogućnosti za provođenje statičkih metoda i tehnika na konkretnim analizama. Na taj način udžbenik zajedno s već spomenutim «Upotreba statistike u ekonomiji» čini cjelinu koja čitatelju/korisniku omogućuje da uz upoznavanje odabrane statističke metodologije savlada rad i rješavanje statističkih problema na računalu. Naime, upotrebom statističkih paketa, počevši već od pripremne faze statističkog istraživanja, znatno se skraćuje i pojednostavljuje vrijeme potrebno za primjenu statističkih tehnika. Na taj se način statistički postupci približavaju mnogim korisnicima, a u nastavu se uvode suvremene metode rada koristeći statističke pakete.

Svaki primjer prezentiran u poglavljima i u privitku knjige sadrži i rješenja u svrhu kontrole valjanosti usvojenoga gradiva. Na taj način se studentima ne ostavlja zagonetnim ni način pismene provjere znanja te ih se osposobljava da samostalno statistički analiziraju određene pojave na stručno zadovoljavajući način.

Studenti koji nastave obrazovanje na diplomskim i doktorskim studijima bit će pripremljeni za stručni i znanstveni rad u svojim istraživanjima, gdje se ekonomski problemi statistički rješavaju upotrebom računala. Dio studenata koji će se nakon diplomiranja neposredno uključiti u poslovnu praksu, također će imati koristi, zbog mogućnosti upotrebe usvojenog znanja statističke teorijske i programske potpore pri poslovnim i ekonomskim analizama.

Potrebno je napomenuti da je program *Statistica* kompatibilan s *Microsoft Excelom*, pa se već postojeći podaci iz jednog mogu jednostavno kopirati u drugi program. To je važna činjenica, s obzirom da je poznato da je *Microsoft Excel* jedan od najraširenijih programa za obradu velikih baza podataka te je lako dostupan većini korisnika.

U radu su korištene oznake i simboli, koji su preuzeti iz standardne statističke literature. Ako se, pak, u literaturi koriste različite oznake, upotrijebljena je češće spominjana verzija.

Na kraju istaknimo, ovaj udžbenik je nastao kao rezultat višegodišnjeg iskustva autora kod primjene statističkih metoda u izradi brojnih znanstvenih i stručnih radova, studija i analiza te kao rezultat predavačkog iskustva pri prenošenju znanja iz područja statistike na mnoge generacije studenata Ekonomskog fakulteta.

Za suradnju i korisne sugestije zahvaljujemo recenzenticama prof. dr. sc. Jasni Horvat, doc. dr. sc. Alemki Šegoti i doc. dr. sc. Tei Baldigari.

Rijeka, Split, rujan 2009.

Autorice

SADRŽAJ

1. TEMELJNI POJMOVI VJEROJATNOSTI

1.1. Kombinatorika i vjerojatnost

1.1.1. Kombinatorika

U statističkoj literaturi se smatra se da je **kombinatorika važan dio matematike** kojeg je dobro poznavati pri upoznavanju statističke teorije i primjene.

Kao jedan od osnovnih problema kombinatorike može se navesti problem sa 2 prazna mjesta. Neka su A_i elementi s kojima se može popuniti prvo prazno mjesto, i neka su B_j elementi s kojima se može popuniti drugo prazno mjesto:

$$A_1, A_2, ..., A_{n_1}; \qquad B_1, B_2, ..., B_{n_2}. \tag{1.1.1}$$

Iz (1.1.1) može se uočiti da elemenata A_i ima n_1, a elemenata B_j ima n_2.

Postavlja se pitanje na koliko se različitih načina mogu popuniti navedena dva prazna mjesta, vodeći računa o tome da na 1. mjesto treba staviti jedan od elemenata A_i, a na 2. mjesto jedan od elemenata B_j. Svako popunjavanje 1. mjesta može doći sa svakim popunjavanjem 2. mjesta, te se može zaključiti da se oba mjesta mogu popuniti na $(n_1 \cdot n_2)$ načina. To se može prikazati kao na slici 1.1. jednom dvostrukom tablicom:

Slika 1.1.

A_i \ B_j	B_1	B_2	...	B_{n_2}
A_1	A_1B_1	A_1B_2	...	$A_1B_{n_2}$
A_2	A_2B_1	A_2B_2	...	$A_2B_{n_2}$
...	...	...	...	...
A_{n_1}	$A_{n_1}B_1$	$A_{n_1}B_2$	...	$A_{n_1}B_{n_2}$

Slika 1.1. pokazuje kombinaciju od n_1 redaka i n_2 stupaca, a na sjecištu svakog retka i svakog stupca je uređeni par elemenata (A_i, B_j), a takvih parova ima $(n_1 \cdot n_2)$, te svaki od njih predstavlja popunjavanje 1. i 2. praznog mjesta.

Ako su zadana 3 prazna mjesta: 1., 2. i 3., te za popunjavanje 1. stoji na raspolaganju n_1 elemenata, za popunjavanje 2. stoji na raspolaganju n_2 elemenata, a za popunjavanje 3. stoji na raspolaganju n_3 elemenata, broj svih mogućih kombinacija popunjavanja svih triju mjesta je jednak $(n_1 \cdot n_2 \cdot n_3)$. To se može pojasniti činjenicom da se prva dva mjesta mogu popuniti na $(n_1 \cdot n_2)$ načina, te da svako takvo popunjavanje može doći i u kombinaciji s popunjavanjem 3. mjesta.

Poopćenje ovog problema vodi ka **osnovnoj lemi kombinatorike**:

Ako postoji k praznih mjesta, i pri tome se 1. može popuniti sa n_1 različitih elemenata, 2. se može popuniti sa n_2 različitih elemenata, ..., $k-to$ mjesto sa n_k različitih elemenata, tada se svih k mjesta može popuniti na $(n_1 \cdot n_2 \cdot ... \cdot n_k)$ različitih načina.

1.1.1.1. Permutacije bez ponavljanja

Ako postoji n različitih elemenata $A_1, A_2, ..., A_n$, njih je moguće poredati u jedan niz na više različitih načina. Svaki takav niz, odnosno poredak naziva se permutacija.

Dvije permutacije se međusobno razlikuju samo po redoslijedu elemenata.

Da bi se izračunao broj **permutacija bez ponavljanja** (jer su svi članovi niza različiti) može se zamisliti da su sva mjesta u nizu prazna. Svako moguće popunjavanje tih mjesta daje po jednu permutaciju.

Prvo mjesto u nizu je moguće popuniti na n različitih načina, jer na to mjesto može doći svaki od n elemenata. Kada je prvo mjesto popunjeno, ostaje $(n-1)$ element, pa se drugo mjesto može popuniti na $(n-1)$ način. Kada su prva dva mjesta popunjena treće se može popuniti na $(n-2)$ načina, i tako dalje. Prema osnovnoj lemi kombinatorike može se zaključiti da je broj mogućih popunjavanja svih n mjesta jednak produktu:

$$n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 3 \cdot 2 \cdot 1 \tag{1.1.2}$$

Produkt prirodnih brojeva od 1 do n ima oznaku $n!$ i čita se "n faktorijela". **Broj permutacija bez ponavljanja od n različitih elemenata je:**

$$P^{(n)} = n!, \tag{1.1.3}$$

gdje u oznaci $P^{(n)}$ predstavlja gornji indeks.[1]

Primjer 1.1.1.

Na koliko se načina može osigurati 6 velikih potrošača električnom energijom, ako svaka od 6 centrala može snabdijevati samo jednog potrošača, a svaki se potrošač može snabdijevati električnom energijom iz samo jedne centrale?

$n = 6$

$P^{(6)} = 6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720.$

Pod gornjim uvjetima 6 velikih potrošača, koristeći usluge 6 raspoloživih centrala, može se snabdjeti električnom energijom na 720 različitih načina.

1.1.1.2. Permutacije s ponavljanjem

Ako postoji n različitih elemenata, a između njih je r_1 jednakih, r_2 jednakih,..., r_k jednakih, naravno uz uvjet da je:

$$r_1 + r_1 + ... + r_k = n, \tag{1.1.4}$$

svaki mogući poredak tih n elemenata u nizu predstavlja po jednu permutaciju s ponavljanjem.

Ako su svi elementi u nizu različiti broj njihovih permutacija (bez ponavljanja) je $n!$. Ako je u nizu r_1 međusobno jednakih elemenata, skup svih permutacija dijeli se na grupe od po $r_1!$ međusobno jednakih permutacija, jer su u svakoj grupi one permutacije, koje na istim mjestima sadrže jednake elemente, te se njihovim permutiranjem ne dobivaju nove permutacije. Dakle, broj permutacija n elemenata s r_1 međusobno jednakih elemenata, odgovara broju tih grupa:

$$\frac{n!}{r_1!}. \tag{1.1.5}$$

[1] Prema dogovoru vrijedi da je: $0! = 1$.

Ako su kod preostalih $(n-r_1)$ elemenata njih r_2 međusobno jednaki, ponovo se skup permutacija dijeli na grupe od po $r_2!$ međusobno jednakih permutacija. Sada je broj permutacija jednak:

$$\frac{n!}{r_1! \cdot r_2!}. \qquad (1.1.6)$$

Poopćavanjem, **skupa od n različitih elemenata, gdje je r_1 jednakih, r_2 jednakih,..., r_k jednakih elemenata, vrijedi da je broj permutacija s ponavljanjem:**

$$P^{(n)}_{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! \cdot r_2! \cdot \ldots \cdot r_k!}. \qquad (1.1.7)$$

Primjer 1.1.2.

Koliko se različitih brojeva većih od 40.000 može napisati od znamenki: 3, 7, 7, 8, 9.

Iz zadanog skupa znamenki može se napisati ukupno 60 peteroznamenkastih brojeva:

$n = 5; r_1 = 2$

$$P^{(5)}_2 = \frac{5!}{2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 60 .$$

Svi brojevi koji počinju s 3 manji su od 40.000 pa ih treba isključiti, a njih ima:

$$P^{(4)}_2 = \frac{4!}{2!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 12 .$$

Stoga se od zadanog skupa znamenki može napisati 48 brojeva većih od 40.000:

$$P^{(5)}_2 - P^{(4)}_2 = 60 - 12 = 48 .$$

1.1.1.3. Varijacije bez ponavljanja

Ako postoji skup od n različitih elemenata: $A_1, A_2, ..., A_n$, svaki podskup od r elemenata:

$$A_1^{'}, A_2^{'}, ... A_r^{'}, \tag{1.1.8}$$

gdje je svaki $A_i^{'}$ neki A_j iz početnog skupa, uz uvjet da se elementi A_j ne ponavljaju u jednom podskupu, zove se varijacija n-tog reda i r-tog razreda bez ponavljanja.

Vrijedi da je: $r \leq n$. Ako je $r = n$, tada varijacije postaju permutacije bez ponavljanja.

Formiranje jedne varijacije može se opisati na način da se iz početnog skupa od n elemenata na slučajan način odabere r članova i složi ih se u niz. Da bi se izračunao broj varijacija bez ponavljanja može se zamisliti r praznih mjesta u nizu. Prvo prazno mjesto moguće je popuniti na n različitih načina. Nakon toga, drugo mjesto je moguće popuniti na $(n-1)$ način. Svako popunjavanje prvog mjesta može se dogoditi sa svakim popunjavanjem drugog mjesta, pa se može zaključiti da se prva dva mjesta u nizu mogu popuniti na $[n \cdot (n-1)]$ način. Prema tome, r mjesta se može popuniti na sljedećih načina:

$$n \cdot (n-1) \cdot (n-2) \cdot ... \cdot [n-(r-1)]. \tag{1.1.9}$$

Broj popunjavanja r mjesta bez ponavljanja iz skupa od n različitih elemenata je **broj varijacija n-tog reda i r-tog razreda bez ponavljanja**:

$$V_r^{(n)} = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot (n-r+1). \tag{1.1.10}$$

Ako se izraz (1.1.10) pomnoži i podijeli s $(n-r)!$ **vrijedi da je:**

$$V_r^{(n)} = \frac{n!}{(n-r)!}. \tag{1.1.11}$$

Primjer 1.1.3.

Koliko se troznamenkastih prirodnih bojeva može sastaviti od elemenata skupa: 1, 2, 3, 4, ako se znamenke u brojevima ne ponavljaju?

n = 4; r = 3

$$V_3^{(4)} = \frac{4!}{(4-3)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{1} = 24 \cdot$$

Iz zadanog skupa moguće je sastaviti 24 troznamenkasta prirodna broja.

1.1.1.4. Varijacije s ponavljanjem

Ako postoji n različitih elemenata iz kojeg se uzimaju podskupovi od po r elemenata, te ako se u svakom podskupu isti elementi mogu pojavljivati dva ili više puta, dobivaju se varijacije n-tog reda i r-tog razreda s ponavljanjem.

S obzirom da je dozvoljeno ponavljanje istih elemenata u tzv. podskupovima, može biti $r > n$.

Birajući prvi element podskupa od n zadanih elemenata, jasno je da ga se može izabrati na n različitih načina. Za popunjavanje drugog mjesta na raspolaganju je opet n elemenata, jer se za drugo mjesto u nizu može upotrijebiti i onaj već izabrani element za prvo mjesto. Prema tome prva dva mjesta u nizu se mogu popuniti na $(n \cdot n) = n^2$ različitih načina. Svako slijedeće mjesto u nizu veličine r se može popuniti na n načina.

Stoga vrijedi da je **broj varijacija s ponavljanjem n-tog reda i r-tog razreda:**

$$\overline{V}_r^{(n)} = n^r \,. \tag{1.1.12}$$

Primjer 1.1.4.

Proizvodi se označavaju oznakama od 5 slova. Koliko se takvih oznaka može sastaviti iz skupa od 25 slova?

n = 25; r = 5

$$\overline{V}_5^{(25)} = 25^5 = 9765625 \,.$$

Može se sastaviti 9.765.625 različitih oznaka.

1.1.1.5. Kombinacije bez ponavljanja

Ako postoji niz od n različitih elemenata iz kojeg se formiraju nizovi od po r elemenata, te pri tome nije bitan raspored samih elemenata u nizu, na taj se način dobivaju kombinacije n-tog reda i r-tog razreda bez ponavljanja. To znači sa se dvije kombinacije međusobno razlikuju samo ako nemaju iste elemente.

Broj kombinacija bez ponavljanja mogu se izračunati pomoću broja varijacija bez ponavljanja na slijedeći način:

Sve varijacije bez ponavljanja n-tog reda i r-tog razreda mogu se podijeliti u skupine koje sadrže samo one varijacije koje se ne razlikuju po elementima već samo po njihovom rasporedu. Svaka takva skupina ima $r!$ varijacija. Permutirajući elemente jedne varijacije, dobiju se nove varijacije, koje su s aspekta kombinacija ustvari sve jednake.

Stoga, **kombinacije n-tog reda i r-tog razreda bez ponavljanja su:**

$$K_r^{(n)} = \frac{V_r^{(n)}}{r!} = \frac{n!}{r! \cdot (n-r)!} . \qquad (1.1.13)$$

Izraz s desne strane jednadžbe (1.1.13) još se označava s:

$$\binom{n}{r}, \qquad (1.1.14)$$

što se čita: "n povrh r".

Ako se u izrazu (1.1.13) i brojnik i nazivnik pokrate s $(n-r)!$ vrijedi da je:

$$K_r^{(n)} = \binom{n}{r} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1)}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot (r-1) \cdot r} . \qquad (1.1.15)$$

Vrijedi da je:

$$\binom{n}{r} = \binom{n}{n-r} . \qquad (1.1.16)$$

Primjer 1.1.5.

Skup se sastoji od 10 različitih elemenata. Na koliko se načina može izabrati uzorak od 3 elementa?

n = 10; r = 3.

Kada se govori o uzorku misli se na njegov sadržaj pa nije bitan raspored samih elemenata. Sloga je svaki uzorak jedna kombinacija.

$$K_3^{(10)} = \binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot (7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} = 120 \cdot$$

Uzorak se može izabrati na 120 različitih načina.

1.1.1.6. Binomni poučak

Zadatak je pronaći red potencija po a i po b koji tvore razvoj potencije binoma (tzv. **binomni poučak**):

$$(a+b)^n = (a+b) \cdot (a+b) \cdot \ldots \cdot (a+b), \qquad (1.1.17)$$

gdje se faktor $(a+b)$ javlja n puta. U izrazu (1.1.17) a i b su proizvoljni brojevi, a n je prirodni broj.

Ovaj problem moguće riješiti pomoću metoda kombinatorike. Ako se izmnože sve zagrade iz (1.1.17), a ima ih n, jasno je da će se potencija a^n javiti 1 put. Ako se iz jedne zagrade član b izmnoži sa članom a iz preostale $(n-1)$ zagrade za rezultat se dobije $a^{n-1}b$. Zagradu, odnosno faktor, iz koje se uzima b može se izabrati na $\binom{n}{1}$ način, pa se u (1.1.17) javlja član:

$$\binom{n}{1} a^{n-1} b \, . \qquad (1.1.18)$$

Ako se iz dvije zagrade član b izmnoži sa članom a iz preostalih $(n-2)$ zagrade za rezultat se dobije $a^{n-2}b^2$. Dvije zagrade, odnosno faktora, iz kojih se uzima b može se izabrati na $\binom{n}{2}$ načina, pa se u (1.1.17) javlja i član:

$$\binom{n}{2} a^{n-2} b^2 . \tag{1.1.19}$$

Na isti način može se zaključiti da se u (1.1.17) član $a^{n-x} b^x$ može izabrati na $\binom{n}{x}$ načina, pa vrijedi da je:

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \ldots + \binom{n}{x} a^{n-x} b^x + \ldots + \binom{n}{n-1} a\, b^{n-1} + b^n \tag{1.1.20}$$

Ako se uvaži da je: $\binom{n}{0} = 1$ i $\binom{n}{n} = 1$ **binomni poučak se u skraćenom obliku može pisati**:

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x . \tag{1.1.21}$$

1.1.2. Vjerojatnost

Slučajni događaj je takav događaj koji se može, ali ne mora realizirati, tj. realizira se uz određenu vjerojatnost.

Vjerojatnost *realizacije* **slučajnog događaja "A"** jednaka je omjeru broja povoljnih ishoda i svih mogućih ishoda:

$$P(A) = \frac{m(A)}{n}, \quad 0 \le P \le 1 , \tag{1.1.22}$$

gdje je:

$m(A)$ - broj svih povoljnih ishoda za događaj A,

n - broj svih mogućih ishoda.

Za siguran događaj: $\Rightarrow P = 1$ (obrat ne vrijedi).

Za nemoguć događaj: $\Rightarrow P = 0$ (obrat ne vrijedi).

Vjerojatnost da se **slučajni događaj "A"** *ne realizira* jednaka je omjeru broja nepovoljnih ishoda i svih mogućih ishoda:

$$P(\overline{A}) = \frac{n - m(A)}{n} = 1 - P(A), \quad 0 \leq P \leq 1 , \qquad (1.1.23)$$

gdje je:

$n - m(A)$ - broj svih nepovoljnih ishoda za događaj A,

n - broj svih mogućih ishoda.

Vrijedi da je:

$$P(A) + P(\overline{A}) = 1 . \qquad (1.1.24)$$

Za ovakav slučaj kada je unaprijed poznat broj svih povoljnih/nepovoljnih i ukupnih ishoda izračunata vjerojatnost se naziva **vjerojatnost "a priori".**

Ako vjerojatnost realizacije slučajnog događaja A nije poznata unaprijed, može se izračunati tzv. **vjerojatnost "a posteriori"**:

$$P(A) = p\lim \frac{m(A)}{n} , \qquad (1.1.25)$$

gdje je:

$p\lim$ -"granična vrijednost po vjerojatnosti" (da se razlikuje od limesa u linearnoj algebri).

Primjer 1.1.6.

U paketu od 18 proizvoda 5 je oštećeno. Kolika je vjerojatnost da od 3 istodobno izvučena proizvoda 2 budu ispravna?

U paketu ima 13 (*18-5=13*) ispravnih proizvoda pa vjerojatnost da 1 slučajno izvučeni proizvod bude ispravan iznosi:

$$P(A) = \frac{m(A)}{n} = \frac{13}{18} = 0{,}72 .$$

Vjerojatnost da je slučajno izvučeni proizvod oštećen protivna je vjerojatnost i iznosi 0,28:

$$P(\overline{A}) = 1 - P(A) = 1 - 0{,}72 = 0{,}28 .$$

Primjer 1.1.7.

U skladište je stigla pošiljka od 200 proizvoda. Kontrola je utvrdila da je 10% proizvoda oštećeno. Kolika je vjerojatnost da u slučajno izabranom uzorku od 3 proizvoda budu upravo 2 ispravna?

a) U pošiljci od 200 proizvoda ima 10% oštećenih, a to je 20 proizvoda, stoga je 180 proizvoda ispravno. Povoljni ishodi su svi izabrani uzorci od 3 proizvoda koji sadrže 2 ispravna pa se broj povoljnih ishoda može izračunati kao kombinacije bez ponavljanja:

$$\binom{180}{2}\binom{20}{1} = \frac{180!}{2!(180-2)!} \cdot \frac{20!}{1!(20-1)!} = 16110 \cdot 20 = 322200 \ .$$

Broj svih mogućih ishoda također se može izračunati kao kombinacije bez ponavljanja:

$$\binom{200}{3} = \frac{200!}{3!(200-3)!} = 1313400 \ \cdot$$

Stoga je vjerojatnost izbora takvog uzorka jednaka:

$$P(A) = \frac{\binom{180}{2}\binom{20}{1}}{\binom{200}{3}} = \frac{322200}{1313400} = 0{,}24 \, .$$

b) Zadatak se može riješiti i pomoću binomne jednadžbe.

Vjerojatnost da 1 slučajno izabrani proizvod bude oštećen iznosi 0,1 ili 10%:

$$P(a) = \frac{20}{200} = 0{,}1 \cdot$$

Vjerojatnost da je 1 slučajno izabrani proizvod bude ispravan je 0,9:

$$P(b) = \frac{180}{200} = 0{,}9 \ \cdot$$

Vjerojatnost da u uzorku od 3 (n=3) proizvoda budu točno 2 ispravna (x=2) iznosi:

$$\binom{n}{x} a^{n-x} b^{x} = \binom{3}{2} 0{,}1^{3-2} \cdot 0{,}9^{2} = \frac{3!}{2!(3-2)!} \cdot 0{,}1 \cdot 0{,}9^{2} = 3 \cdot 0{,}1 \cdot 0{,}81 = 0{,}24 \, .$$

Oba načina izračuna daju isti rezultat i pokazuju da je vjerojatnost izbora uzorka od 3 proizvoda tako da sadrži 2 ispravna iznosi 0,24 pa izbor takvog uzorka nije vjerojatan.

1.1.2.1. Teorem o zbrajanju vjerojatnosti (adicijski teorem)

Za međusobno isključive događaje (tj. one događaje koji ne mogu nastupiti istodobno, odnosno čiji su pripadni skupovi elementarnih događaja disjunktni), vjerojatnost realizacije jednog **ili** drugog događaja jednaka je zbroju vjerojatnosti realizacije jednog i vjerojatnosti realizacije drugog događaja.

$$P(A\,ili\,B) = P(A) + P(B)\,. \tag{1.1.26}$$

Za događaje koji se međusobno ne isključuju vjerojatnost realizacije jednog **ili** drugog događaja je:

$$P(A\,ili\,B) = P(A) + P(B) - P(AB)\,, \tag{1.1.27}$$

gdje je:

$P(AB)$ -vjerojatnost da slučajni događaji A i B nastupe istovremeno.

Primjer 1.1.8.

U pakiranju od 50 proizvoda ima 10 crvenih i 10 bijelih. Kolika je vjerojatnost da će jedan slučajno uzet proizvod biti ili crven ili bijeli?

Izbor crvenog proizvoda isključuje bijeli i obratno. Znači da se događaji međusobno isključuju pa treba primijeniti adicijski teorem:

$$P(AiliB) = P(A) + P(B) = \frac{10}{50} + \frac{10}{50} = 0{,}2 + 0{,}2 = 0{,}4\ .$$

Vjerojatnost iznosi 0,4 pa događaj nije vjerojatan.

1.1.2.2. Teorem o množenju vjerojatnosti (multiplikacijski teorem)

Ako se događaji A i B **međusobno ne isključuju**, tj. vjerojatnost realizacije jednog ne zavisi o vjerojatnosti realizacije drugog događaja (nezavisni događaji), tada je vjerojatnost istovremene realizacije događaja A i B jednaka produktu vjerojatnosti realizacije događaja A i događaja B:

$$P(A\ i\ B) = P(A) \cdot P(B) . \qquad (1.1.28)$$

Primjer 1.1.9.

U kutiji se nalazi 10 proizvoda, od kojih su 2 neispravna. Kolika je vjerojatnost da će u 2 uzastopna slučajna izvlačenja (bez vraćanja izvučenog proizvoda u kutiju) izvući samo ispravni proizvodi?

Primijenit će se multiplikacijski teorem za zavisne događaje:

$$P(AiB) = P(A) \cdot P(B/A) = \left(\frac{8}{10}\right) \cdot \left(\frac{7}{9}\right) = \frac{56}{90} = 0{,}62 \ .$$

Vjerojatnost da su dva odabrana proizvoda ispravna iznosi 0,62 pa je takav izbor vjerojatan.

1.1.2.3. Uvjetna (kondicionalna) vjerojatnost (Bayesov teorem)

Ako je realizacija događaja A uvjetovana prethodnom realizacijom događaja B vrijedi da je **uvjetna ili kondicionalna vjerojatnost** događaja A:

$$P(A/B) = \frac{P(AB)}{P(B)}, \qquad P(B) > 0 , \qquad (1.1.29)$$

odnosno:

$$P(B/A) = \frac{P(AB)}{P(A)}, \qquad P(A) > 0 . \qquad (1.1.30)$$

Ako **su događaji nezavisni**, vrijedi da je:

$$P(AB) = P(A) \cdot P(B) \Rightarrow P(A/B) = \frac{P(A) \cdot P(B)}{P(A)} = P(A) \tag{1.1.31}$$

Bayesov teorem:

Ako se događaj B realizira istovremeno kada nastupi jedan od n disjunktnih događaja $A_1, A_2, ..., A_n$ za koje je $\sum_{i=1}^{n} P(A_i) = 1$ onda se vjerojatnost događaja B dobiva po formuli potpune vjerojatnosti:

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B/A_i) \tag{1.1.32}$$

$$P(A_k / B) = \frac{P(A_k) \cdot P(B/A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(B/A_k)}, \quad \sum_{i=1}^{n} P(A_i) = 1. \tag{1.1.33}$$

Primjer 1.1.10.

Dvije su kutije: A i B. U kutiji A 8 je crvenih proizvoda i 2 zelena, a u kutiji B 4 su crvena proizvoda i 6 zelenih. Ako se kutija bira slučajno i iz nje slučajno izabere proizvod kolika je vjerojatnost da je izabran zeleni proizvod i da je iz kutije A?

Vjerojatnost izbora zelenog proizvoda uvjetovana je prethodnom vjerojatnosti izbora kutije, radi se dakle o uvjetnoj vjerojatnosti.

Vjerojatnost da će biti izabrana kutija A iznosi 0,5: $P(A) = \frac{1}{2} = 0{,}5$.

Jednaka je vjerojatnost da će biti izabrana kutija B: $P(B) = \frac{1}{2} = 0{,}5$.

Vjerojatnost da će izabrani proizvod iz kutije A biti zelen iznosi 0,2: $P(Z) = \frac{2}{10} = 0{,}2$,

a vjerojatnost da će izabrani proizvod iz kutije B biti zelen je: $P(Z) = \frac{6}{10} = 0{,}6$.

S obzirom da u drugom koraku biramo zeleni proizvod, a da pri tom ne znamo koju smo kutiju izabrali u prvom koraku, koristit će se Bayesov teorem:

$$P(A/Z)=\frac{P(A)\cdot P(Z/A)}{P(A)\cdot P(Z/A)+P(B)\cdot P(Z/B)}=\frac{0{,}5\cdot 0{,}2}{(0{,}5\cdot 0{,}2)+(0{,}5\cdot 0{,}6)}=\frac{0{,}1}{0{,}1+0{,}3}=0{,}25.$$

Vjerojatnost iznosi 0,25 pa događaj nije vjerojatan.

1.2. Slučajna varijabla i njezina svojstva

Slučajna varijabla može biti:

- **diskretna ili diskontinuirana slučajna varijabla** i
- **kontinuirana slučajna varijabla.**

1.2.1. Diskretna slučajna varijabla

Diskontinuirana ili diskretna slučajna varijabla je takva varijabla koja na slučajan način može poprimiti najviše prebrojivo mnogo vrijednosti i svaku od njih s određenom vjerojatnošću.

Zadovoljava uvjete:

- normativnost: $\sum_{i=1}^{\infty} P(x_i)=1$,
- nenegativnost: $P(x_i)\geq 0,\ \forall i$.

Skup svih uređenih parova vrijednosti slučajne varijable i njoj pripadajućih vjerojatnosti naziva se **distribucija slučajne varijable** X:

$$\{[x_i; P(x_i)],\ i=1,2,...,\infty\}\ . \tag{1.2.1}$$

Zakon po kojem svakoj vrijednosti slučajne varijable *X* pripada vjerojatnost $P(x_i)$ naziva se **zakon vjerojatnosti slučajne varijable** *X*.

Funkcija distribucije slučajne varijable *X* je funkcija koja daje vjerojatnost da će slučajna varijabla *X* poprimiti vrijednost jednaku ili manju od nekog realnog broja x_k:

$$F(x_k) = P(X \leq x_k) = \sum_{i=1}^{k} P(x_i) \tag{1.2.2}$$

Može se reći da je to **kumulativ vjerojatnosti** od $F(X=0)$ do $F(X=x_k)$. Funkcija distribucije slučajne varijable X je monotono neopadajuća funkcija.

Parametri distribucije:

- **Očekivanje slučajne varijable** je zbroj umnožaka vrijednosti varijable X i njoj pripadajućih odgovarajućih vjerojatnosti $P(x_i)$:

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot P(x_i) = \mu \, . \tag{1.2.3}$$

- **Varijanca slučajne varijable** je očekivanje kvadratnog odstupanja vrijednosti varijable X od njenog očekivanja:

$$V(X) = E(X^2) - [E(X)]^2 = E(X-\mu)^2 = \sum_{i=1}^{\infty} x_i^2 \cdot P(x_i) - \mu^2 \, . \tag{1.2.4}$$

1.2.2. Kontinuirana slučajna varijabla

Kontinuirana slučajna varijabla X je takva varijabla koja može poprimiti neprebrojivo beskonačno mnogo vrijednosti. Zato se za kontinuiranu slučajnu varijablu ne računa vjerojatnost u određenoj točki, nego nad određenim intervalom.

Funkcija vjerojatnosti ili funkcija gustoće vjerojatnosti kontinuirane slučajne varijable X ima svojstva:

- $f(x) \geq 0, \ \forall x$,
- $\int_{-\infty}^{+\infty} f(x)dx = 1$, (površina ispod krivulje funkcije vjerojatnosti je 1),
- $P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)dx, \ [x_2 > x_1]$, (**vjerojatnost** da slučajna varijabla X poprimi vrijednost veću od x_1 i manju ili jednaku x_2, **jednaka je** određenom integralu funkcije vjerojatnosti $f(x)$ na tom intervalu, odnosno

površini koju krivulja funkcije vjerojatnosti na tom intervalu zatvara s pozitivnim smjerom osi x).

Funkcija distribucije slučajne varijable X je funkcija koja daje vjerojatnost da će slučajna varijabla X poprimiti vrijednost jednaku ili manju od nekog realnog broja x:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx \tag{1.2.5}$$

Parametri distribucije:

- **Očekivanje slučajne varijable** (ako integral konvergira):

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx = \mu \,. \tag{1.2.6}$$

- **Varijanca slučajne varijable**:

$$V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)dx \,. \tag{1.2.7}$$

1.3. Odabrane diskretne i kontinuirane distribucije vjerojatnosti

1.3.1. Teorijske distribucije diskontinuirane slučajne varijable

1.3.1.1. Binomna distribucija

Ako je vjerojatnost da se dogodi neki događaj poznata, unaprijed utvrđena i konstantna tijekom cijelog istraživanja (iznosi p) kaže se da se diskontinuirana slučajna varijabla X ravna prema tzv. **binomnoj distribuciji**.

Vjerojatnost da se neki slučajni događaj X u n pokusa realizira x puta prema **binomnom zakonu vjerojatnosti** je:

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = \binom{n}{x} \cdot p^x \cdot q^{n-x}, \qquad x = 0,1,2,...,n \qquad (1.3.1)$$

gdje je:

n - broj pokusa,

p - vjerojatnost realizacije slučajnog događaja,

q - vjerojatnost da se slučajni događaj ne realizira,

x - broj povoljnih ishoda u n pokusa,

$n-x$ - broj nepovoljnih ishoda u n pokusa.

Očekivanje slučajne varijable X kod binomne distribucije:

$$E(X) = n \cdot p \qquad (1.3.2)$$

Varijanca slučajne varijable X kod binomne distribucije:

$$\sigma^2 = n \cdot p \cdot q \qquad (1.3.3)$$

Koeficijent asimetrije slučajne varijable X kod binomne distribucije:

$$\alpha_3 = \frac{q-p}{\sqrt{n \cdot p \cdot q}} \qquad (1.3.4)$$

Koeficijent zaobljenosti slučajne varijable X kod binomne distribucije:

$$\alpha_4 = 3 + \frac{1 - 6 \cdot p \cdot q}{n \cdot p \cdot q} \qquad (1.3.5)$$

Općenito se kaže da se **slučajna varijabla X ravna po binomnoj distribuciji**, koja je određena parametrima n i p:

$$X \sim B(n, p). \qquad (1.3.6)$$

Primjer 1.3.1.

Pregledano je 50 kutija po 10 proizvoda i dobivena je distribucija prikazana u tablici 1.1.

Pretpostavlja se da se zadani skup ravna po binomnoj distribuciji. Zadatak je izračunati pripadne vjerojatnosti slučajne varijable x i obilježja ove binomne distribucije.

Tablica 1.1.

Broj oštećenih proizvoda (x)	Broj pregledanih kutija (f_i)	$f_i \cdot x$	Vjerojatnost $P(x)$
0	17	0	0,35
1	22	22	0,39
2	6	12	0,19
3	4	12	0,06
4	1	4	0.01
Ukupno	**50**	**50**	**1,00**

Binomna distribucija određena je parametrima n i p.

n je broj pokusa ili veličina uzoraka ($n = 10$);

$$\overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{50}{50} = 1$$

Prosječan broj oštećenih proizvoda po kutiji iznosi 1.

p je vjerojatnost ostvarenja slučajnog događaja: $p = \dfrac{\overline{X}}{n} = \dfrac{1}{10} = 0{,}1$.

Konstantna vjerojatnost da će u slučajno izabranoj kutiji jedan proizvod biti oštećen iznosi 0,1.

Konstantna vjerojatnost da će u slučajno izabranoj kutiji proizvod biti ispravan, protivna je vjerojatnost vjerojatnosti da bude oštećen i iznosi 0,9:

$q = 1 - p = 1 - 0{,}1 = 0{,}9.$

Pripadajuće vjerojatnosti vrijednostima slučajne varijable x prikazane su u posljednjem stupcu tablice 1.1, a računane su binomnom jednadžbom:

$$P(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x};$$

$$P(0) = \binom{n}{0} \cdot p^0 \cdot q^{n-0} = q^n = 0{,}9^{10} = 0{,}35;$$

$$P(1) = \binom{10}{1} \cdot 0{,}1^1 \cdot 0{,}9^{10-1} = 10 \cdot 0{,}1 \cdot 0{,}387 = 0{,}39; \text{ itd.}$$

Najveća je vjerojatnost da će u kutiji od 10 proizvoda 1 biti oštećen i ona iznosi 0,39, a to znači da je u 39% slučajeva moguće naići na oštećeni proizvod u kutiji.

Obilježja ove binomne distribucije jesu:

1. **Očekivanje** slučajne varijable

$$E(x)=n\cdot p=10\cdot 0{,}1=1\,.$$

U prosjeku se očekuje 1 oštećeni proizvod po kutiji.

2. **Varijanca** slučajne varijable

$$\sigma^2=n\cdot p\cdot q=10\cdot 0{,}1\cdot 0{,}9=0{,}9\,.$$

3. **Koeficijent asimetrije** slučajne varijable

$$\alpha_3=\frac{q-p}{\sqrt{n\cdot p\cdot q}}=\frac{0{,}9-0{,}1}{\sqrt{0{,}9}}=0{,}84\,.$$

4. **Koeficijent zaobljenosti** slučajne varijable

$$\alpha_4=3+\frac{1-6\cdot p\cdot q}{n\cdot p\cdot q}=3+\frac{1-6\cdot 0{,}1\cdot 0{,}9}{10\cdot 0{,}1\cdot 0{,}9}=3{,}36\,.$$

1.3.1.2. Poissonova distribucija

Ako je vjerojatnost da se dogodi neki događaj poznata, unaprijed utvrđena, konstantna i jako mala tijekom cijelog istraživanja (iznosi p), te broj pokusa teži u beskonačnost ($n\to\infty$) umjesto binomne distribucije, koristi se **Poissonova distribucija**.

U praksi vrijedi da ako je:

$$n\geq 50 \quad i \quad p\leq 0{,}10 \qquad (1.3.7)$$

koristi se Poissonova distribucija. Može se reći da je ova distribucija definirana za **rijetke događaje**, odnosno one događaje koji imaju veliki uzorak i malu vjerojatnost.

Zakon vjerojatnosti po Poissonovoj distribuciji je:

$$P(X=x)=\frac{(np)^x\cdot e^{-np}}{x!}=\frac{(\mu)^x\cdot e^{-\mu}}{x!},\quad x=0,1,2,...,\infty\,, \tag{1.3.8}$$

jer je: $\mu = n\cdot p$;

gdje je:

n - broj pokusa,

p - vjerojatnost realizacije slučajnog događaja,

x - broj povoljnih ishoda u n pokusa.

Očekivanje slučajne varijable X kod Poissonove distribucije:

$$E(X)=n\cdot p=\mu\,. \tag{1.3.9}$$

Varijanca slučajne varijable X kod Poissonove distribucije:

$$\sigma^2 = n\cdot p=\mu\,. \tag{1.3.10}$$

Koeficijent asimetrije slučajne varijable X kod Poissonove distribucije:

$$\alpha_3=\frac{1}{\sqrt{\mu}}\,. \tag{1.3.11}$$

Koeficijent zaobljenosti slučajne varijable X kod Poissonove distribucije:

$$\alpha_4=3+\frac{1}{\mu}\,. \tag{1.3.12}$$

Općenito se kaže da se **slučajna varijabla X ravna po Poissonovoj distribuciji**, koja je određena parametrom μ :

$$X\sim P(\mu)\,. \tag{1.3.13}$$

Primjer 1.3.2.

Izvršen je pokus od 2226 promatranja u kojem je slučajna varijabla x poprimila vrijednosti od 0 do 15.

Rezultati promatranja prikazani su u tablici 1.2.

Zadatak je, za zadanu Poissonovu distribuciju, izračunati pripadne vjerojatnosti slučajne varijable x i obilježja distribucije.

Tablica 1.2.

Vrijednosti slučajne varijable (x_i)	Broj promatranja (f_i)	$f_i \cdot x$	Vjerojatnost $P(x)=\frac{(\mu)^x \cdot e^{-\mu}}{x!}$
0	1050	0	0,36788
1	526	251	0,36788
2	325	650	0,18394
3	188	564	0,06131
4	56	224	0,01533
5	33	165	0,00307
6	18	108	0,00051
7	10	70	0,00007
8	6	48	0,00001
9	5	45	0,00000
10	3	30	0,00000
11	2	22	0,00000
12	1	12	0,00000
13	1	13	0,00000
14	1	14	0,00000
15	1	15	0,00000
Ukupno	**2226**	**2231**	**1,00000**

Poissonova distribucija određena je parametrom μ.

$$\mu = n \cdot p = E(x) = \overline{X} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{2231}{2226} = 1 .$$

Očekuje da će slučajna varijabla prosječno poprimiti vrijednost 1.

Pripadajuće vjerojatnosti vrijednostima slučajne varijable x prikazane su u posljednjem stupcu gornje tabele.

Obilježja ove Poissonove distribucije jesu:

1. **Očekivanje** slučajne varijable

   $E(x) = n \cdot p = \mu = 1 .$

2. **Varijanca** slučajne varijable

$\sigma^2 = n \cdot p = \mu = 1$.

3. **Koeficijent asimetrije** slučajne varijable

$$\alpha_3 = \frac{1}{\sqrt{\mu}} = \frac{1}{\sqrt{1}} = 1 .$$

4. **Koeficijent zaobljenosti** slučajne varijable

$$\alpha_4 = 3 + \frac{1}{\mu} = 3 + \frac{1}{1} = 4 .$$

1.3.1.3. Jednolika distribucija

Ako slučajna varijabla X može poprimiti n različitih vrijednosti i svaku od njih s uvijek istom vjerojatnosti realizacije kaže se da se diskontinuirana slučajna varijabla X ravna prema tzv. **jednolikoj distribuciji**.

Zakon vjerojatnosti po jednolikoj distribuciji je:

$$P(X = x_i) = \frac{1}{n}, \qquad \forall x_i . \tag{1.3.14}$$

Primjer za jednoliku distribuciju je kod bacanja kocke. Vjerojatnost P da kocka padne na bilo koju stranu: 1,2,3,4,5 ili 6, je uvijek ista i iznosi: $P = 1/6$.

1.3.1.4. Dvodimenzionalna diskretna distribucija

Diskontinuirana slučajna varijabla X može poprimiti vrijednosti $x_1, x_2, x_3, ..., x_n$, a diskontinuirana slučajna varijabla Y može istovremeno poprimiti vrijednosti $y_1, y_2, y_3, ..., y_m$.

Vjerojatnost da slučajna varijabla X poprimi vrijednost x_i, a istovremeno slučajna varijabla Y poprimi vrijednost y_j je:

$$P(X = x_i, Y = y_j) = P(x_i, y_j) \qquad (1.3.15)$$

Za svaku distribuciju vjerojatnosti mora biti ispunjen uvjet normativnosti:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) = 1 \qquad (1.3.16)$$

Skup svih uređenih parova $\{(x_i, y_j); P(x_i, y_j)\}$ je dvodimenzionalna distribucija slučajnih varijabli (X, Y).

1.3.1.5. Marginalne distribucije diskretne slučajne varijable

Marginalna distribucija slučajne varijable X je skup uređenih parova odgovarajućih vrijednosti varijable X i njima pripadajućih vjerojatnosti da slučajna varijabla X poprimi neku vrijednost x_i, bez obzira na to koju će vrijednost poprimiti slučajna varijabla Y.

Marginalna distribucija slučajne varijable Y je skup uređenih parova odgovarajućih vrijednosti varijable Y i njima pripadajućih vjerojatnosti da slučajna varijabla Y poprimi neku vrijednost y_j, bez obzira na to koju će vrijednost poprimiti slučajna varijabla X.

Marginalna vjerojatnost slučajne varijable X je:

$$P(x_i \cdot) = \sum_{j=1}^{m} P(x_i, y_j), \qquad (1.3.17)$$

i obično se u dvostrukoj statističkoj tablici nalazi u zbirnom stupcu (ako se slučajna varijabla X mijenja po redcima).

Marginalna vjerojatnost slučajne varijable Y je:

$$P(\cdot y_j) = \sum_{i=1}^{n} P(x_i, y_j), \qquad (1.3.18)$$

i obično se u dvostrukoj statističkoj tablici nalazi u zbirnom retku (ako se slučajna varijabla Y mijenja po stupcima).

1.3.1.6. Uvjetne distribucije diskretne slučajne varijable

Vjerojatnost da slučajna varijabla X poprimi neku vrijednost x_i pod uvjetom da je varijabla Y poprimila neku vrijednost y_j je uvjetna ili kondicionalna vjerojatnost:

$$P(x_i / y_j) = \frac{P(x_i, y_j)}{P(y_j)}, \ P(y_j) > 0 . \tag{1.3.19}$$

Slično, vjerojatnost da slučajna varijabla Y poprimi neku vrijednost y_j pod uvjetom da je varijabla X poprimila neku vrijednost x_i je uvjetna ili kondicionalna vjerojatnost:

$$P(y_j / x_i) = \frac{P(x_i, y_j)}{P(x_i)}, \ P(x_i) > 0 . \tag{1.3.20}$$

Ako su **slučajne varijable nezavisne** primjenjuje se multiplikacijski teorem:

$$P(x_i, y_j) = P(x_i) \cdot P(y_j) , \tag{1.3.21}$$

pa vrijedi da je:

$$P(x_i / y_j) = P(x_i) , \tag{1.3.22}$$

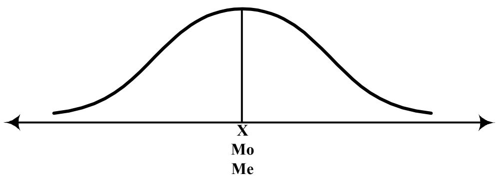
$$P(y_j / x_i) = P(y_j) . \tag{1.3.23}$$

1.3.2. Teorijske distribucije kontinuirane slučajne varijable

1.3.2.1. Normalna distribucija

Normalna distribucija je simetrična ($\alpha_3 = 0$), unimodalna i ima oblik zvona. Normalno je zaobljena ($\alpha_4 = 3$). Na sredini distribucije u istoj točki nalaze se centralne mjere tendencije: aritmetička sredina, medijan i mod. To je prikazano na slici 4.2..

Slika 1.2.



Kod binomne distribucije ako je:

- $p = q$,
- $n \to \infty$,

binomna distribucija teži normalnoj distribuciji.

Funkcija vjerojatnosti ili funkcija gustoće vjerojatnosti kod normalne distribucije je:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \quad x \in < -\infty, +\infty > . \tag{1.3.24}$$

Općenito se kaže da se **kontinuirana slučajna varijabla X ravna po normalnoj distribuciji**, koja je određena parametrima μ (očekivanje) i σ^2 (varijanca):

$$X \sim N(\mu, \sigma^2) . \tag{1.3.25}$$

Za standardiziranu varijablu *Z*:

$$Z = \frac{x-\mu}{\sigma} , \tag{1.3.26}$$

funkcija vjerojatnosti je:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} . \tag{1.3.27}$$

Ova normalna distribucija se naziva **standardizirana ili jedinična normalna distribucija** s očekivanjem 0 i jediničnom varijancom, pa vrijedi da je:

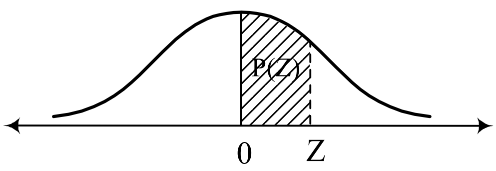
$$X \sim N(0,1) . \tag{1.3.28}$$

Jedinična normalna distribucija ima uvijek iste vrijednosti parametara, pa se vrijednosti za intervale vrijednosti varijable Z, od 0 do z, (u statističkoj literaturi ima i drugačije koncipiranih tablica), mogu prikazati u jedinstvenoj **tablici površina (vjerojatnosti) ispod normalne krivulje**.

Na primjer:

$$P(x_1 < X \le x_2) = P_2\left(\frac{x_2-\mu}{\sigma}\right) - P_1\left(\frac{x_1-\mu}{\sigma}\right) = P_2(Z_2) - P_1(Z_1) \qquad (1.3.29)$$

Slika 1.3.



Primjer 1.3.3.

Kontrolom kvalitete utvrđeno je da životni vijek žarulja u grafoskopu slijedi normalnu raspodjelu s prosječnim vijekom trajanja *μ=2000 sati* i standardnom devijacijom *σ=200 sati*. Kolika je vjerojatnost da će slučajno odabrana žarulja trajati više od prosjeka, a manje od 2300 sati?

Slika 1.3. pokazuje krivulju vjerojatnosti, a osjenčano područje ispod krivulje odgovara intervalu "od 2000 do 2300" sati. Donja granica intervala je prosječni vijek trajanja žarulja (*μ=2000*) tj. aritmetička sredina distribucije stoga ima standardiziranu vrijednost *Z=0.* Gornja granica intervala odgovara vrijednosti od 2300 sati pa je standardizirana vrijednost slučajne varijable:

X=2300; $\qquad Z = \frac{X-\mu}{\sigma} = \frac{2300-2000}{200} = +1{,}5$.

Vrijednost *Z=+1,5* pokazuje da je 2300 sati 1,5 standardnih devijacija iznad prosjeka od 2000 sati.

Površina (vjerojatnost) ispod normalne distribucije za *Z=+1,5* iznosi:

$$f(Z) = \frac{1}{\sqrt{2\pi}} \int_0^Z e^{-\frac{Z^2}{2}} dZ = 0{,}4332 .$$

Stoga je *P(2000 ≤ X ≤ 2300) = 0,4332.*

Vjerojatnost da će žarulja trajati više od prosjeka (*μ=2000*), a manje od 2300 sati iznosi 0,43.

1.3.2.2. Studentova distribucija

Studentova distribucija se još naziva t-distribucija. Ima svoju primjenu u statistici kod procjene parametara osnovnog skupa i kod testiranja hipoteza na osnovu uzorka.

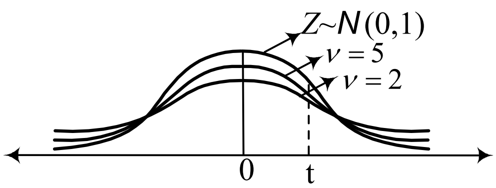
Varijabla "t" je definirana na području: $< -\infty, +\infty >$. Studentova distribucija je simetrična s obzirom na $t = 0$.

Ako je $Z \sim N(0,1)$ i χ^2 varijabla tzv. hi-kvadrat ili gama-distribucije s υ (ili df - degree of freedom) stupnjevima slobode i ako su one nezavisne slučajne varijable, tada je:

$$t_\upsilon = \frac{z}{\sqrt{\frac{\chi^2}{\upsilon}}}, \qquad (1.3.30)$$

t - varijabla Studentove distribucije sa stupnjevima slobode υ (ili df).

Slika 1.4.



Na slici 1.4. se vidi da se s povećanjem stupnjeva slobode υ, studentova distribucija približava normalnoj distribuciji. Već za $\upsilon = 30$ u praksi se umjesto t - distribucije upotrebljava normalna distribucija (pogreška aproksimacije u tom je slučaju manja od 0,03.

1.3.2.3. Hi-kvadrat distribucija

Hi-kvadrat distribucija se još naziva gama-distribucija. Ako su $X_1, X_2, ..., X_n$ nezavisne normalne varijable koje imaju jednaka očekivanja,

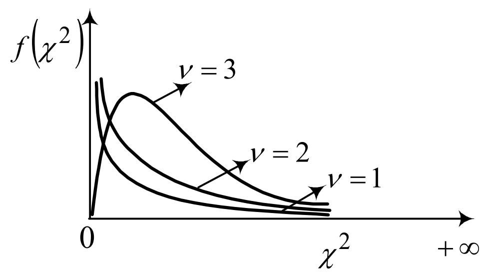
$E(X_1) = E(X_2) = ... = E(X_n) = \mu$ i jednake varijance
$V(X_1) = V(X_2) = ... = V(X_n) = \sigma^2$, tada je:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2, \tag{1.3.31}$$

χ^2 - gama varijabla sa stupnjem slobode $\upsilon = n$.

Varijabla "χ^2" je definirana na području: $< 0, +\infty >$, i za različite stupnjeve slobode distribucija ima različit oblik, što se jasno vidi na slici 1.5.

Slika 1.5.



Slika 1.5. prikazuje da za broj stupnjeva slobode $\upsilon \geq 3$ krivulja gama distribucije ima pozitivnu ili desnostranu asimetriju. Kako broj stupnjeva slobode υ raste, to se gama distribucija približava obliku normalne distribucije.

Varijabla χ^2:

$$\chi^2 = \sum_{i=1}^{r} \left(\frac{f_i - f_{ti}}{f_{ti}} \right)^2, \tag{1.3.32}$$

pripada χ^2 distribuciji sa stupnjevima slobode: $\upsilon = r - 1$ (gdje je r - broj nepoznatih parametara u pretpostavljenoj distribuciji).

Kada je broj stupnjeva slobode $\upsilon > 30$, upotrebljava se aproksimacija normalnom distribucijom:

$$\chi^2 = \frac{1}{2} \left(z + \sqrt{2 \cdot \upsilon - 1} \right)^2, \tag{1.3.33}$$

gdje je z - varijabla iz jedinične normalne distribucije.

1.3.2.4. F - distribucija

F-distribucija je određena s dva parametra υ_1 i υ_2 koji predstavljaju stupnjeve slobode. Ako su χ_1^2 i χ_2^2 dvije nezavisne hi-kvadrat (gama) distribucije sa stupnjevima slobode υ_1 i υ_2 tada varijabla F:

$$F = \frac{\chi_1^2 / \upsilon_1}{\chi_2^2 / \upsilon_2}, \tag{1.3.34}$$

pripada F - distribuciji sa stupnjevima slobode υ_1 i υ_2.

2. METODA UZORKA

2.1. Osnove statističkog istraživanja

Deskriptivna ili opisna statistika temelji se na potpunom obuhvatu statističkog skupa. Kada se statističke metode i tehnike primjenjuju na **čitav statistički skup**, dakle ako su istraživanjem obuhvaćeni svi elementi skupa oni tvore **statističku populaciju**. Takva istraživanja, koja obuhvaćaju čitav statistički skup, često zahtijevaju puno vremena i skupa su. Stoga se kod mnogih istraživanja u praksi iz osnovnog skupa odabire podskup, tj. uzorak.

Inferencijalna statistika temelji se na dijelu (uzorku) jedinica izabranih iz cjelovitog statističkog skupa, pomoću kojeg se uz primjenu odgovarajućih statističkih metoda i tehnika donose zaključci o čitavom statističkom skupu. Uvijek je prisutan odgovarajući stupanj rizika kada se koriste rezultati iz uzorka, za kojeg je poželjno da bude izabran na slučajan način i da bude reprezentativan.

Dakle, inferencijalna statistika pripada skupini **induktivnih metoda**, kojima se izvode zaključci polazeći **od posebnoga prema općem.**

Osnovne **faze statističkog istraživanja** jesu:

a) statističko promatranje,

b) grupiranje (tabelarno i grafičko prikazivanje statističkih podataka),

c) statistička analiza i interpretacija rezultata provedene analize.

Statističko promatranje je organizirano prikupljanje statističkih podataka. Pojedinačne metode statističkog promatranja su:

a) mjerenje,
b) brojanje,
c) ocjenjivanje,
d) evidentiranje i
e) anketiranje.

Mjeri se na primjer urod pšenice po jedinici poljoprivredne površine. Može se mjeriti težina proizvoda, kao i visina stanovništva nekog područja.

Brojanjem se može doći do podataka o broju zaposlenih u pojedinim organizacijskim jedinicama poduzeća, o broju upisanih studenata na fakultete. Broje se i noćenja turista u turističkoj sezoni, pa se na taj način dobivaju podaci relevantni za analizu uspješnosti sezone.

Ocjenjivanjem kao metodom prikupljanja statističkih podataka određuje se kvaliteta provođenja određenih radnji, pa se stoga taj postupak obično veže za redoslijedna obilježja statističkog skupa. Ocjenjuju se i rangiraju studenti na testu iz statistike, ocjenjuje se usluga nekog hotelskog poduzeća i slično.

Evidentiranje podataka podrazumijeva kontinuirano praćenje kretanja neke pojave u duljem ili kraćem vremenskom razdoblju. Za potrebe evidencije često se uređuju odgovarajući obrasci. Na primjer pri promatranju izostanaka i broja ostvarenih radnih sati djelatnika u nekom poduzeću postoji obrazac za evidenciju gdje zaposlenici upisuju: ime i prezime, vrijeme dolaska i odlaska s radnog mjesta, vrijeme izostanka, razlog izostanka. Evidentiranje se može vršiti i tehničkim uređajima. To su umrežena računala, optički čitači, uređaji za brojenje npr. putnika i slično.

Anketa i/ili intervju je metoda kojom se prikupljaju podaci uz pomoć unaprijed pripremljenih upitnika, na kojima ispitanici svojim odgovorima daju informacije o promatranim obilježjima statističkog skupa. Pri provođenju ankete pristup ispitaniku, odnosno jedinici statističkog skupa može izravan i neizravan.

Izravan pristup ostvaruje se kada osoba ili tim koji provodi anketu izlazi na teren i u direktnom kontaktu s ispitanicima prikuplja odgovore na pitanja iz upitnika. **Neizravan pristup** ostvaruje se putem pošte, telefonom, elektroničkom poštom i web-om. Na ovaj način smanjuju se troškovi prikupljanja podataka (npr. putni troškovi osoba koje provode anketu tj. anketara). Iako se i na ovaj način ispitanicima prezentira tko provodi i koja je svrha istraživanja, praksa je pokazala da je ovim neizravnim pristupom anketiranja prisutan velik postotak neodaziva, kao i često velik postotak nevaljano i nepotpuno popunjenih upitnika. Naravno, razloge treba tražiti u činjenici da ispitaniku nije na raspolaganju osoba koja će mu pojasniti nejasna pitanja.

Ovisno o obimu istraživanja, **organizaciju prikupljanja podataka može provoditi jedna osoba, skupina istraživača ili osoblje specijalizirane ustanove(,)** kojoj je to osnovna djelatnost. Ako se radi o manjem istraživanju, u prikupljanju podataka će sudjelovati manji broj istraživača, dok će veće istraživanje zahtijevati rad većeg broja istraživača.

2.2. Pojam uzorka

Uzorak je podskup osnovnog skupa, a uzima se u svrhu ispitivanja obilježja elemenata osnovnog skupa (ili populacije).

Uzorak **treba biti reprezentativan.**

Poželjno je da uzorak bude **što veći,** ali u konkretnim istraživanjima može se dogoditi:

- da povećavanjem uzorka istraživanje postaje sve skuplje
- da se katkada uzorci uništavaju (npr. kemijska analiza nekih prehrambenih artikala).

Kad god je moguće poželjno je izabrati **slučajni uzorak**. U takvom uzorku svaka jedinica populacije (osnovnog skupa) ima jednaku vjerojatnost da bude izabrana. Uzorak **ne smije biti selekcioniran** (npr. potrebno je obuhvatiti istraživanjem i gradsko i seosko stanovništvo, zatim u nekom medicinskom istraživanju nije cilj samo analizirati i donositi zaključke za «dobrovoljce»).

Ako neki članovi populacije imaju veću šansu od drugih da budu izabrani, takav uzorak se naziva **pristrani uzorak** (biased sample).

Slučajan uzorak sastavlja se prema određenim načelima koji odgovaraju zakonu slučaja. Najbolji način je upotreba «**tablice slučajnih brojeva**» ili korištenje računalnog sustava slučajnog izbora (generator slučajnih brojeva).

Uzorak može biti i **sistematski** ako se jedinice iz osnovnog skupa biraju sistematski. Na primjer, ako se po redu u uzorak bira svaki 10. element iz osnovnog skupa.

Stratificirani ili slojeviti uzorak je takav uzorak koji se dobije tako da se populacija podijeli u slojeve ili stratume prema nekim karakteristikama te da se iz svake od grupa uzme slučajni uzorak. Na primjer, u sociološkim istraživanjima stratumi se mogu birati prema dobnim skupinama.

Klaster uzorci su lošija varijanta slučajnog uzorka i upotrebljavaju se u velikim tržišnim, ekonomskim ili političkim istraživanjima. Na primjer, pri ispitivanju mišljenja stanovnika nekoga grada o nekoj problematici, grad se može prema planu podijeliti na 50-ak blokova, odnosno kvartova. Tada se na slučajan način biraju neki blokovi u kojima anketari *detaljno* intervjuiraju sve stanovnike. Čak se vraćaju na adrese dok ne dobiju intervju od svakog stanovnika u odabranim blokovima.

Kvotni uzorci su još lošiji jer predstavljaju neslučajni stratificirani uzorak. Upotrebljavaju se kod ad hoc organiziranih istraživanja za tržišne potrebe, za prikupljanje mišljenja građana o nekom problemu i slično. Istraživač unaprijed izabere broj ljudi (kvotu) svakog pojedinog stratuma koje mora intervjuirati, stoga se ti uzorci nazivaju kvotni.

Prigodni uzorak je onaj koji se «nađe pri ruci» jer je drugi nedostupan. Na primjer, dostupni bolesnici na odjelu u bolnici, prisutni studenti neke godine studija i slično.

Pri istraživanju se mogu dogoditi **pogreške** koje mogu biti:

a) **pogreške izvan uzorka** (zbog pogrešno odabrane metode istraživanja, zbog pogrešne obrade rezultata i slično)

b) **pogreške uzorka**.

Veličinu uzorka nije moguće općenito definirati jer to ovisi o varijabilnosti pojave koja se mjeri i o preciznosti kojom se pojava želi izmjeriti. Za uzorak je najvažnije da bude reprezentativan.

Frakcija odabiranja (*f*) omjer je jedinica u uzorku i broja jedinica u osnovnom skupu:

$$f = \frac{n}{N} \tag{2.1}$$

gdje je:

n - broj jedinica u uzorku

N - broj jedinica u osnovnom skupu.

Korak izbora je recipročna vrijednost frakcije odabiranja ($1/f$) i upotrebljava se kod sistematskog izbora jedinica u uzorak. To znači da ako je korak izbora jednak 20, tada se u uzorak iz osnovnog skupa odabire svaki 20. element.

Broj svih mogućih uzoraka (bez ponavljanja) veličine n iz osnovnog skupa veličine N jednak je broju kombinacija bez ponavljanja n-tog razreda:

$$K = \binom{N}{n}. \tag{2.2}$$

Pomoću uzorka procjenjuju se određeni parametri osnovnog skupa i testiraju se hipoteze o nepoznatim parametrima osnovnog skupa.

2.3. Sampling distribucije

Ako se iz osnovnog skupa veličine N izaberu svi mogući uzorci veličine n, te se za svaki uzorak izračuna neki odgovarajući parametar, distribucija tih parametara naziva se **sampling distribucija.**

Pristranost je **razlika** između očekivane vrijednosti nekog parametra iz sampling distribucije i toga istog parametra iz osnovnog skupa:

$$E(\hat{\Theta}) \neq \Theta, \tag{2.3}$$

pri čemu je $E(\hat{\Theta})$ očekivana vrijednost parametra iz sampling distribucije, a Θ je parametar iz osnovnog skupa.

Ako između očekivane vrijednosti nekog parametra iz sampling distribucije i toga istog parametra iz osnovnog skupa **ne postoji razlika,** onda se to svojstvo naziva **nepristranost**:

$$E(\hat{\Theta}) = \Theta. \tag{2.4}$$

Standardna devijacija sampling distribucije parametra naziva se **standardna pogreška**.

Može se zaključiti da su sampling distribucije teorijske distribucije vjerojatnosti. U praksi se sampling distribucije ne formiraju, već se koriste rezultati teorijske statistike o njihovim poznatim oblicima.

Primjer 2.3.1.

Proizvod X (numerička varijabla) prodaje se u 5 različitih trgovina. Osnovni skup (N=5) sačinjavaju cijene u kunama po kojima te trgovine prodaju proizvod X i one iznose: 15, 17, 18, 20 i 20. Svaki uzorak (n=2) ima jednaku vjerojatnost izbora, a izbor se provodi bez ponavljanja.

a) Izračunajte aritmetičku sredinu i standardnu devijaciju osnovnog skupa.
b) Izdvojite sve moguće uzorke veličine 2 elementa iz zadanog osnovnog skupa i za svaki uzorak izračunajte aritmetičku sredinu te formirajte sampling-distribuciju sredina.
c) Izračunajte očekivanu vrijednost i standardnu devijaciju sampling-distribucije sredina.
(Napomena: u praksi se mali osnovni skupovi ne ispituju metodom uzoraka!).

Rješenje:

a) Aritmetička sredina osnovnog skupa iznosi:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{5}(15 + 17 + 18 + 20 + 20) = 18\,\text{kn.}$$

Standardna devijacija osnovnog skupa jednaka je:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} x_1^2}{N} - \mu^2} = \sqrt{327{,}6 - 324} = 1{,}89737\ \text{kn.}$$

b) Broj mogućih formiranja uzoraka veličine n iz osnovnog skupa od N *elemenata* jednak je:

$$k = \binom{N}{n} = \frac{N!}{n!(N-n)!},$$

stoga uz uvjete: n=2 i zadani osnovni skup (N=8) moguće je oformiti:

$$k = \binom{5}{2} = \frac{5!}{2!3!} = 10\ \text{uzoraka.}$$

Ako se cijene proizvoda X označe slovima: A, B, C, D i E onda je moguće oformiti sljedeće uzorke:

Uzorci	Vrijednosti u uzorku	Aritmetičke sredine uzoraka
AB	15, 17	16,0
AC	15, 18	16,5
AD	15, 20	17,5
AE	15, 20	17,5
BC	17, 18	17,5
BD	17, 20	18,5
BE	17, 20	18,5
CD	18, 20	19,0
CE	18, 20	19,0
DE	20, 20	20,0

Sampling-distribucija sredina uzoraka:

Aritmetičke sredine uzoraka	$p(\bar{x}_i)$
16,0	0,1
16,5	0,1
17,5	0,3
18,5	0,2
19,0	0,2
20,0	0,1

c) Očekivana vrijednost sampling-distribucije iznosi:

$$E\left[\overline{X}\right] = \sum_{i=1}^{k} \bar{x}_i \, p(\bar{x}_i) = 16 \cdot 0{,}1 + 16{,}5 \cdot 0{,}1 + 17{,}5 \cdot 0{,}3 + 18{,}5 \cdot 0{,}2 + 19 \cdot 0{,}2 + 20 \cdot 0{,}1 = 18 = \overline{X}.$$

Varijanca sampling-distribucije jednaka je:

$$\sigma_{\overline{X}}^2 = \sum_{i=1}^{k} \bar{x}_i^2 \, p(\bar{x}_i) - \overline{X}^2 = (25{,}6 + 27{,}225 + 91{,}875 + 68{,}45 + 72{,}2 + 40) - 18^2 = 325{,}35 - 324 = 1{,}35.$$

Standardna devijacija (standardna pogreška) je: $\sigma_{\overline{X}} = \sqrt{\sigma_{\overline{X}}^2} = \sqrt{1{,}35} = 1{,}1619.$

3. PROCJENJIVANJE PARAMETARA

3.1. Procjenjivanje brojem i intervalom aritmetičke sredine jedne populacije

Na temelju podataka iz uzorka procjenjuje se aritmetička sredina za jednu populaciju, tj. statistički skup. Aritmetička sredina se može procijeniti brojem i intervalom.

Ako veličina uzorka teži prema beskonačno **sampling distribucija aritmetičkih sredina** teži **normalnom obliku**. Kod malih uzoraka sampling distribucija aritmetičkih sredina ima oblik Studentove ili t-distribucije. Stoga vrijedi da ako je:

$n > 30$ $\Rightarrow$ koristi se normalna distribucija i

$n \leq 30$ $\Rightarrow$ koristi se Studentova ili t-distribucija.

Standardna pogreška aritmetičke sredine (standardna devijacija sampling distribucije aritmetičke sredine) uz odgovarajuću frakciju odabiranja f je[2]:

$$Se(\hat{\bar{X}}) = \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \text{ako je } f \leq 0{,}05 \text{ i} \tag{3.1}$$

$$Se(\hat{\bar{X}}) = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \Rightarrow \quad \text{ako je } f > 0{,}05\,. \tag{3.2}$$

Ako nije poznata standardna devijacija osnovnog skupa σ, računa se **nepristrana** (točkasta) **ocjena varijance osnovnog skupa na osnovi uzorka** S^2 :

$$S^2 = \hat{\sigma}^2 \cdot \left(\frac{n}{n-1}\right) \quad \Rightarrow \quad \text{ako je } n \leq 30 \text{ i} \tag{3.3}$$

[2] **Frakcija odabiranja** (f) omjer je jedinica u uzorku i broja jedinica u osnovnom skupu:

$f = \frac{n}{N}$, gdje je: n - broj jedinica u uzorku; N - broj jedinica u osnovnom skupu.

$$S^2 = \hat{\sigma}^2 \qquad \Rightarrow \qquad \text{ako je } n > 30\ . \tag{3.4}$$

Točkasta procjena aritmetičke sredine osnovnog skupa na osnovi uzorka je sama aritmetička sredina uzorka, za kojeg se naravno pretpostavlja da je dovoljne veličine i slučajan, tj. reprezentativan: $\hat{\overline{X}}$.

Intervalna procjena aritmetičke sredine osnovnog skupa na osnovi uzorka uz odgovarajuću razinu pouzdanosti procjene je:

$$\Pr\left\{\hat{\overline{X}} - Z \cdot Se(\hat{\overline{X}}) < \overline{X} < \hat{\overline{X}} + Z \cdot Se(\hat{\overline{X}})\right\} = 1 - \alpha\ , \tag{3.5}$$

gdje je:

$\hat{\overline{X}}$ - aritmetička sredina izračunata na osnovi uzorka,

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako je ($n \le 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n-1]}$),

$Se(\hat{\overline{X}})$ - standardna pogreška aritmetičke sredine,

$1-\alpha$ - razina pouzdanosti procjene (ako nije određeno drugačije, najčešće se uzima de je $1-\alpha = 95\%$, a u nekim slučajevima 99%).

Veličina uzorka uz zadanu razinu pouzdanosti i maksimalnu pogrešku uz odgovarajuću frakciju odabiranja f određuje se na sljedeći način:

$$n' = \left[\frac{Z \cdot \sigma}{pogreška}\right]^2 \qquad \Rightarrow \qquad \text{ako je } f \le 0{,}05 \text{ ili} \tag{3.6}$$

$$n = \frac{n'}{1 + \frac{n'}{N}} \qquad \Rightarrow \qquad \text{ako je } f > 0{,}05\ , \tag{3.7}$$

gdje je:

Z - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

σ - standardna devijacija iz osnovnog skupa, *pogreška* - maksimalna pogreška.

Ako je **maksimalna pogreška** određena **u relativnom izrazu** umjesto σ (standardne devijacije) veličina uzorka se računa pomoću V (koeficijenta varijacije):

$$V = \frac{\sigma}{\overline{X}} \cdot 100. \tag{3.8}$$

Primjer 3.1.1.

Procijenite aritmetičku sredinu osnovnog skupa na razini pouzdanosti 95%, ako je: N=10.000 članova, n=300, aritmetička sredina uzorka je 33, a standardna devijacija uzorka iznosi 5.

Rješenje:
Procjenjuje se aritmetička sredina osnovnog skupa intervalom:

$$\Pr\left\{\hat{\overline{X}} - Z \cdot Se(\hat{\overline{X}}) < \overline{X} < \hat{\overline{X}} + Z \cdot Se(\hat{\overline{X}})\right\} = 1 - \alpha .$$

Parametri:
$N = 10.000$
$n = 300$
$\hat{\overline{X}} = 33$
$\hat{\sigma} = 5.$

Ocjena standardne devijacije osnovnog skupa (S) na osnovi standardne devijacije uzorka ($\hat{\sigma}$):
$n = 300 > 30 \Rightarrow S = \hat{\sigma} = 5.$

Izračun standardne pogreške aritmetičke sredine:

$$f = \frac{n}{N} = \frac{300}{10.000} = 0{,}03 < 0{,}05 \Rightarrow Se(\hat{\overline{X}}) = \frac{S}{\sqrt{n}} = \frac{5}{\sqrt{300}} = 0{,}28868.$$

Određivanje koeficijenta pouzdanosti (Z):
Obzirom da se radi o velikom uzorku (n>30) koeficijent pouzdanosti se ravna po jediničnoj normalnoj distribuciji, stoga je: (1-α) = 0,95, α = 0,05, α/2 = 0,025, tj. $Z_{\alpha/2} = Z_{0,025}$. U polju Tablice površine ispod normalne krivulje nalaze se površine između 0 i Z. Površina između 0 i $Z_{0,025}$ iznosi 0,50000-0,02500=0,47500. Stoga u polju

tablice treba pronaći vrijednost 0,47500 (ili njoj najbližu). Vrijednosti 0,47500 odgovara Z vrijednost 1,96. Prema tome: $Z_{0,025} = 1,96$.

Interval procjene aritmetičke sredine iznosi:

$$\Pr\left\{\hat{\overline{X}} - Z \cdot Se(\hat{\overline{X}}) < \overline{X} < \hat{\overline{X}} + Z \cdot Se(\hat{\overline{X}})\right\} = 1 - \alpha$$

$$\Pr\{33 - 1,96 \cdot 0,28868 < \overline{X} < 33 + 1,96 \cdot 0,28868\} = 0,95$$

$$\Pr\{32,43 < \overline{X} < 33,57\} = 0,95.$$

Zaključak:
Na razini pouzdanosti od 95%, procjenjuje se da je aritmetička sredina osnovnog skupa veća od 32,43 i manja od 33,57.

Primjer 3.1.2.

Koliko treba biti velik uzorak, ako osnovni skup sačinjava 1.211 radnika i procjenjuju se njihova prosječna mjesečna primanja? Standardna devijacija osnovnog skupa iznosi 150 kn, tolerira se pogreška od 32,2, a procjena se vrši s 95% pouzdanosti.

Rješenje:

Projektira se uzorak za procjenu aritmetičke sredine osnovnog skupa.

Parametri:
$N = 1.211$
$\sigma = 150$
d (ukupna pogreška) = 32,2
$Z_{0,025} = 1,96$.

Određivanje veličine uzorka:

$$n' = \left[\frac{Z \cdot \sigma}{d}\right]^2 = \left[\frac{1,96 \cdot 150}{32,2}\right]^2 = 83,4 \approx 84.$$

Provjera frakcije izbora:

$$f = \frac{n}{N} = \frac{84}{1211} = 0{,}06936 > 0{,}05 \Rightarrow n = \frac{n'}{1+\frac{n'}{N}} = \frac{84}{1+\frac{84}{1211}} = 78{,}6 \approx 79.$$

Zaključak:
Kako bi se procijenila prosječna mjesečna primanja u osnovnom skupu, uz dane uvjete u uzorak treba izabrati 79 radnika.

Primjer 3.1.3.

Koliki treba biti uzorak, ako za osnovni skup od 1.000 vrećica slanih štapića procijenjeni koeficijent varijacije iznosi 5%, a procjenjuje se njihova prosječna težina? Tolerira se relativna pogreška od 0,02, dok se procjena vrši s 99% pouzdanosti.

Rješenje:

Projektira se uzorak za procjenu aritmetičke sredine osnovnog skupa.

Parametri:
$N = 1.000$
$V = 5\% = 0{,}05$
d_r (ukupna relativna pogreška) = 0,02.

Određivanje koeficijenta pouzdanosti (Z):
Obzirom da se radi o velikom uzorku (n>30) koeficijent pouzdanosti se ravna po jediničnoj normalnoj distribuciji, stoga je: (1-α) = 0,99, α = 0,01, α/2 = 0,005, tj. $Z_{\alpha/2} = Z_{0,005}$. U polju Tablice površine ispod normalne krivulje nalaze se površine između 0 i Z. Površina između 0 i $Z_{0,005}$ iznosi 0,50000-0,00500=0,49500. Stoga u polju tablice treba pronaći vrijednost 0,49500, odnosno njoj najbližu. Najbliža joj je vrijednost 0,49506, kojoj odgovara Z vrijednost 2,58. Prema tome: $Z_{0,005} = 2{,}58$.

Određivanje veličine uzorka:

$$n = \left[\frac{Z \cdot V}{d_r}\right]^2 = \left[\frac{2{,}58 \cdot 0{,}05}{0{,}02}\right]^2 = 41{,}6 \approx 42.$$

Provjera frakcije izbora: $f = \dfrac{n}{N} = \dfrac{42}{1.000} = 0{,}042 < 0{,}05 \Rightarrow n = 42.$

Zaključak:
Kako bi se procijenila prosječna težina vrećica slanih štapića u uzorak treba izabrati 42 vrećice.

Primjer 3.1.4.

Iz pošiljke od 5.000 tablica čokolada kontrolor je slučajnim izborom uzeo 200 tablica u uzorak. Prosječna težina tablica u uzorku je 98,6 grama sa standardnom devijacijom od 8,64 grama. Na temelju rezultata uzorka s 99% pouzdanosti procijenite prosječnu težinu tablica čokolade u cjelokupnoj pošiljci.

Rješenje:

Procjenjuje se aritmetička sredina osnovnog skupa intervalom:

$$\Pr\left\{\hat{\overline{X}} - Z \cdot Se(\hat{\overline{X}}) < \overline{X} < \hat{\overline{X}} + Z \cdot Se(\hat{\overline{X}})\right\} = 1 - \alpha.$$

Parametri:
$N = 5.000$
$n = 200$
$\hat{\overline{X}} = 98{,}6$
$\hat{\sigma} = 8{,}64.$

Ocjena standardne devijacije osnovnog skupa (S) na osnovi standardne devijacije uzorka ($\hat{\sigma}$):
$n = 200 > 30 \Rightarrow S = \hat{\sigma} = 8{,}64.$

Izračun standardne pogreške aritmetičke sredine:
$$f = \frac{n}{N} = \frac{200}{5.000} = 0{,}04 < 0{,}05 \Rightarrow Se(\hat{\overline{X}}) = \frac{S}{\sqrt{n}} = \frac{8{,}64}{\sqrt{200}} = 0{,}61094.$$

Obzirom da se radi o velikom uzorku (n>30) koeficijent pouzdanosti se ravna po jediničnoj normalnoj distribuciji i uz 99% pouzdanosti Z iznosi: $Z_{0,005} = 2{,}58$.

Interval procjene aritmetičke sredine iznosi:

$$\Pr\left\{\hat{\bar{X}} - Z \cdot Se(\hat{\bar{X}}) < \bar{X} < \hat{\bar{X}} + Z \cdot Se(\hat{\bar{X}})\right\} = 1 - \alpha$$

$$\Pr\left\{98{,}6 - 2{,}58 \cdot 0{,}61094 < \bar{X} < 98{,}6 + 2{,}58 \cdot 0{,}61094\right\} = 0{,}99$$

$$\Pr\left\{97{,}02 < \bar{X} < 100{,}18\right\} = 0{,}99.$$

Zaključak:
Na razini pouzdanosti od 99% procjenjuje se da je prosječna težina tablica čokolada u cjelokupnoj pošiljci veća od 97,02 i manja od 100,18 grama.

Primjer 3.1.5.

Iz nove serije od 150 žarulja izabran je slučajni uzorak od 25 žarulja. Ispitivanjem uzorka utvrđeno je da je prosječno trajanje žarulja 1000 sati sa standardnom devijacijom od 58 sati. Procijenite s 95% pouzdanosti prosječno trajanje žarulja ove nove serije.

Rješenje:

Procjenjuje se aritmetička sredina osnovnog skupa intervalom (mali uzorak):

$$\Pr\left\{\hat{\bar{X}} - t \cdot Se(\hat{\bar{X}}) < \bar{X} < \hat{\bar{X}} + t \cdot Se(\hat{\bar{X}})\right\} = 1 - \alpha.$$

Parametri:
$N = 150$
$n = 25$
$\hat{\bar{X}} = 1000$
$\hat{\sigma} = 58.$

Ocjena standardne devijacije osnovnog skupa (S) na osnovi standardne devijacije uzorka ($\hat{\sigma}$):

$$n = 25 < 30 \Rightarrow S = \hat{\sigma} \cdot \sqrt{\frac{n}{n-1}} = 58 \cdot \sqrt{\frac{25}{25-1}} = 59{,}2.$$

Izračun standardne pogreške aritmetičke sredine:

$$f = \frac{n}{N} = \frac{25}{150} = 0{,}16667 > 0{,}05 \Rightarrow Se(\hat{\bar{X}}) = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{59{,}2}{\sqrt{25}}\sqrt{\frac{150-25}{150-1}} = 10{,}84.$$

Određivanje koeficijenta pouzdanosti (t):

Obzirom da se radi o malom uzorku (n<30) koeficijent pouzdanosti se ravna po Studentovoj (t) distribuciji. Za procjenu aritmetičke sredine pomoću malog uzorka koeficijent pouzdanosti t u tablici Studentove distribucije nalazi se u retku s brojem stupnjeva slobode (n-1) i u stupcu t_α. Stoga je tablična vrijednost koeficijenta pouzdanosti jednaka:
$n = 25$, $(1-\alpha) = 0{,}95$, ss = $(n-1) = (25-1) = 24$, $\alpha = 0{,}05$, $t_{0,05}(24) = 3{,}745$.

Interval procjene aritmetičke sredine iznosi:

$$\Pr\left\{\hat{\overline{X}} - t \cdot Se(\hat{\overline{X}}) < \overline{X} < \hat{\overline{X}} + t \cdot Se(\hat{\overline{X}})\right\} = 1-\alpha$$

$$\Pr\left\{1000 - 3{,}745 \cdot 10{,}84 < \overline{X} < 1000 + 3{,}745 \cdot 10{,}84\right\} = 0{,}95$$

$$\Pr\left\{959{,}4 < \overline{X} < 1040{,}6\right\} = 0{,}95.$$

Zaključak:
Na razini pouzdanosti od 95% procjenjuje se da je prosječno trajanje žarulja u novoj seriji veće od 959,4 i manje od 1040,6 sati.

3.2. Procjenjivanje brojem i intervalom totala jedne populacije

Na temelju podataka iz uzorka procjenjuje se total za jednu populaciju, tj. statistički skup. Total se može procijeniti brojem i intervalom.

Točkasta procjena totala osnovnog skupa na osnovu uzorka je sam total uzorka, za kojeg se opet naravno pretpostavlja da je dovoljne veličine i slučajan, tj. da je reprezentativan:

$$\sum_{i=1}^{N} x_i = N \cdot \hat{\overline{X}} . \qquad (3.9)$$

Intervalna procjena totala osnovnog skupa na osnovi uzorka uz odgovarajuću razinu pouzdanosti procjene je:

$$\Pr\left\{N \cdot \hat{\overline{X}} - Z \cdot N \cdot Se(\hat{\overline{X}}) < T < N \cdot \hat{\overline{X}} + Z \cdot N \cdot Se(\hat{\overline{X}})\right\} = 1-\alpha ,$$
(3.10)

gdje je:

$N \cdot \hat{\bar{X}}$ - točkasta procjena totala izračunata na osnovi uzorka,

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako je ($n \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n-1]}$),

$N \cdot Se(\hat{\bar{X}})$ - standardna pogreška totala,

$1-\alpha$ - razina pouzdanosti procjene (ako nije određeno drugačije, najčešće se uzima de je $1-\alpha = 95\%$, a u nekim slučajevima 99%).

Primjer 3.2.1.

Osnovni skup sastoji se od 15.000 trgovina na području A. Pomoću uzorka od 250 trgovina ocjenjuje se ukupan broj zaposlenih. Prosječan broj zaposlenih po trgovini u uzorku iznosi 5, sa standardnom devijacijom 1,5 zaposlenih. Procijenite ukupan broj zaposlenih s 95% pouzdanosti.

Rješenje:

Procjenjuje se total osnovnog skupa intervalom:

$$\Pr\left\{N \cdot \hat{\bar{X}} - Z \cdot N \cdot Se(\hat{\bar{X}}) < T < N \cdot \hat{\bar{X}} + Z \cdot N \cdot Se(\hat{\bar{X}})\right\} = 1-\alpha.$$

Parametri:
N = 15.000
n = 250
$\hat{\bar{X}} = 5$
$\hat{\sigma} = 1{,}5$
$Z_{0,025}$ = 1,96.

Ocjena standardne devijacije osnovnog skupa (S) na osnovi standardne devijacije uzorka ($\hat{\sigma}$):
$n = 250 > 30 \Rightarrow S = \hat{\sigma} = 1{,}5.$

Izračun standardne pogreške aritmetičke sredine:

$$f = \frac{n}{N} = \frac{250}{15.000} = 0{,}01667 < 0{,}05 \Rightarrow Se(\hat{\bar{X}}) = \frac{S}{\sqrt{n}} = \frac{1{,}5}{\sqrt{250}} = 0{,}09487.$$

Procjena totala osnovnog skupa je:

$\hat{T} = N \cdot \hat{\bar{X}} = 15.000 \cdot 5 = 75.000.$

Interval procjene aritmetičke sredine iznosi:

$$\Pr\left\{N \cdot \hat{\bar{X}} - Z \cdot N \cdot Se(\hat{\bar{X}}) < T < N \cdot \hat{\bar{X}} + Z \cdot N \cdot Se(\hat{\bar{X}})\right\} = 1 - \alpha$$

$$\Pr\{75.000 - 1{,}96 \cdot 15.000 \cdot 0{,}09487 < T < 75.000 + 1{,}96 \cdot 15.000 \cdot 0{,}09487\} = 0{,}95$$

$$\Pr\{72.211 < T < 77.789\} = 0{,}95.$$

Zaključak:
Na razini pouzdanosti od 95%, procjenjuje se da je ukupan broj zaposlenih na području A veći od 72.211 i manji od 77.789 osoba.

Primjer 3.2.2.

Procjenjuje se total osnovnog skupa koji ima 5.000 jedinica. Standardna devijacija osnovnog skupa iznosi 5, a tolerira se ukupna pogreška procjene totala od 3.000. Koliko treba biti velik uzorak ako se procjena vrši s 99% pouzdanosti.

Rješenje:

Projektira se uzorak za procjenu totala osnovnog skupa.

Parametri:
$N = 5.000$
$\sigma = 5$
d (ukupna pogreška) = 3.000
$Z_{0,005} = 2{,}58.$

Određivanje veličine uzorka:

$$n' = \left[N \frac{Z \cdot \sigma}{d}\right]^2 = \left[5.000 \frac{2{,}58 \cdot 5}{3.000}\right]^2 = 462{,}3 \approx 463.$$

Provjera frakcije izbora: $f = \frac{n}{N} = \frac{463}{5.000} = 0{,}0928 > 0{,}05 \Rightarrow n = \frac{n'}{1 + \frac{n'}{N}} = \frac{463}{1 + \frac{463}{5.000}} = 424.$

Zaključak:
Uz dane uvjete u uzorak treba izabrati 424 jedinice da bi se mogao procijeniti total osnovnog skupa.

3.3. Procjenjivanje brojem i intervalom proporcije jedne populacije

Na temelju podataka iz uzorka procjenjuje se proporcija za jednu populaciju, tj. statistički skup. Proporcija se također može procijeniti brojem i intervalom.

Ako veličina uzorka teži prema beskonačno **sampling distribucija proporcija** teži **normalnom obliku**. Kod malih uzoraka sampling distribucija proporcija ima oblik Studentove ili t-distribucije. Stoga vrijedi da ako je:

$n > 30 \quad \Rightarrow$ koristi se normalna distribucija i

$n \leq 30 \quad \Rightarrow$ koristi se Studentova ili t-distribucija.

Standardna pogreška proporcije (standardna devijacija sampling distribucije proporcije) uz odgovarajuću frakciju odabiranja i veličinu uzorka je:

- ako je $f \leq 0{,}05$:

$$Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \quad \Rightarrow \quad \text{ako je } n > 30 \text{ i} \qquad (3.11)$$

$$Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n-1}} \quad \Rightarrow \quad \text{ako je } n \leq 30. \qquad (3.12)$$

- ako je $f > 0{,}05$:

$$Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \Rightarrow \quad \text{ako je } n > 30 \text{ i} \qquad (3.13)$$

$$Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n-1}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \Rightarrow \quad \text{ako je } n \le 30\,. \qquad (3.14)$$

Točkasta procjena proporcije osnovnog skupa na osnovi uzorka je proporcija uzorka, za kojeg se opet naravno pretpostavlja da je dovoljno velik i slučajan, tj. da je reprezentativan: $\hat{p} = \frac{m}{n}$.

Intervalna procjena proporcije osnovnog skupa na osnovi uzorka uz odgovarajuću razinu pouzdanosti procjene je:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha\,, \qquad (3.15)$$

gdje je:

$\hat{p}$ - proporcija izračunata na osnovi uzorka; ($\hat{q} = 1 - \hat{p}$),

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako je ($n \le 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n-1]}$),

$Se(\hat{p})$ - standardna pogreška proporcije,

$1-\alpha$ - razina pouzdanosti procjene (ako nije određeno drugačije, najčešće se uzima de je $1-\alpha = 95\%$, a u nekim slučajevima 99%).

Veličina uzorka uz zadanu razinu pouzdanosti i maksimalnu pogrešku uz odgovarajuću frakciju odabiranja f određuje se na sljedeći način:

$$n' = \left[\frac{Z \cdot \sqrt{P \cdot Q}}{greška}\right]^2 \quad \Rightarrow \quad \text{ako je } f \le 0{,}05 \text{ ili} \qquad (3.16)$$

$$n = \frac{n'}{1 + \frac{n'}{N}} \quad \Rightarrow \quad \text{ako je } f > 0{,}05\,, \qquad (3.17)$$

gdje je:

Z - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

$P \cdot Q$ - varijanca osnovnog skupa,

pogreška - maksimalna pogreška.

Ako je **maksimalna pogreška** određena **u relativnom izrazu** umjesto $\sqrt{P \cdot Q}$ (standardne devijacije) veličina uzorka se računa pomoću V (koeficijenta varijacije osnovnog skupa):

$$V = \sqrt{\frac{Q}{P}}, \tag{3.18}$$

pa vrijedi da je:

$$n' = \left[\frac{Z \cdot \sqrt{\frac{Q}{P}}}{greška} \right]^2. \tag{3.19}$$

Primjer 3.3.1.

Od 250 studenata nazočnih na predavanju anketirano ih je 45, od kojih su 4 negativno ocijenila kvalitetu održanog predavanja. Procijenite proporciju studenata nezadovoljnih održanim predavanjem s 95% pouzdanosti.

Rješenje:

Procjenjuje se proporcija osnovnog skupa intervalom:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha.$$

Parametri:
$N = 250$
$n = 45$
$m = 4$ (elementi uzorka s određenim oblikom obilježja)
$Z_{0,025} = 1,96$ (95% pouzdanosti).

Ocjena proporcija uzorka ($\hat{p}$):

$$\hat{p} = \frac{m}{n} = \frac{4}{45} = 0,08889 \qquad \hat{q} = 1 - \hat{p} = 1 - 0,08889 = 0,91111.$$

Izračun standardne pogreške proporcije:

$$f = \frac{n}{N} = \frac{45}{250} = 0{,}18 > 0{,}05 \text{ i } (n > 30) \Rightarrow Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

$$Se(p) == \sqrt{\frac{0{,}08889 \cdot 0{,}91111}{45}} \cdot \sqrt{\frac{250-45}{250-1}} = 0{,}03849.$$

Interval procjene proporcije iznosi:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha$$

$$\Pr\{0{,}08889 - 1{,}96 \cdot 0{,}03849 < P < 0{,}08889 + 1{,}96 \cdot 0{,}03849\} = 0{,}95$$

$$\Pr\{0{,}01345 < P < 0{,}16433\} = 0{,}95.$$

Zaključak:
Na razini pouzdanosti od 95% procjenjuje se da je proporcija studenata nezadovoljnih održanim predavanjem osnovnog skupa veća od 1,35% i manja od 16,43%.

Primjer 3.3.2.

Trgovac građevinskim materijalom treba preuzeti pošiljku od 1.500 komada parketa. Između 50 parketa, izabranih iz pošiljke na slučajan način, pronađeno je 8 komada druge klase. Uz pouzdanost od 99% procijenite proporciju parketa druge klase u cjelokupnoj pošiljci.

Rješenje:

Procjenjuje se proporcija osnovnog skupa intervalom:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha.$$

Parametri:
N = 1.500
n = 50
m = 8 (elementi uzorka s određenim oblikom obilježja)
$Z_{0,005}$ = 2,58 (99% pouzdanosti).

Ocjena proporcija uzorka ($\hat{p}$):

$$\hat{p} = \frac{m}{n} = \frac{8}{50} = 0{,}16 \qquad \hat{q} = 1 - \hat{p} = 1 - 0{,}16 = 0{,}84.$$

Izračun standardne pogreške proporcije:

$$f = \frac{n}{N} = \frac{50}{1.500} = 0{,}033 < 0{,}05 \ \text{ i } \ n > 30 \Rightarrow Se(p) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = \sqrt{\frac{0{,}16 \cdot 0{,}84}{50}} = 0{,}05185.$$

Interval procjene proporcije iznosi:

$$\Pr\{\hat{p} - Z \cdot Se(\hat{p}) < P < \hat{p} + Z \cdot Se(\hat{p})\} = 1 - \alpha$$

$$\Pr\{0{,}16 - 2{,}58 \cdot 0{,}05185 < P < 0{,}16 + 2{,}58 \cdot 0{,}05185\} = 0{,}99.$$

Zaključak:
Na razini pouzdanosti od 99% procjenjuje se da je proporcija parketa druge klase u cjelokupnoj pošiljci veća od 2,62% i manja od 29,38%.

Primjer 3.3.3.

U tvornici trikotaže ispituje se kvaliteta proizvodnje čarapa. Proporcija škarta je 5%, a tolerira se pogreška od 0,03 s 95% pouzdanosti. Koliko čarapa treba uzeti u uzorak da bi se procijenila proporcija škarta? Pretpostavlja se da je $n << N$.

Rješenje:

Projektira se uzorak za procjenu proporcije osnovnog skupa.

Parametri:
P = 5% = 0, 05
d (ukupna pogreška) = 0,03
$Z_{0,025}$ = 1,96 (95% pouzdanosti)
$Q = 1 - P$ = 1- 0,05 = 0,95.

Određivanje veličine uzorka:

$$n = \left[\frac{Z \cdot \sqrt{P \cdot Q}}{d}\right]^2 = \left[\frac{1{,}96 \cdot \sqrt{0{,}05 \cdot 0{,}95}}{0{,}03}\right]^2 = 202{,}8 \approx 203.$$

Zaključak:
Uz dane uvjete u uzorak treba izabrati 203 čarape da bi se procijenila proporcija škarta.

Primjer 3.3.4.

Na području grada A zaposleno je 30.000 osoba. Pomoću uzorka procjenjuje se proporcija zaposlenih žena. Koliko zaposlenih treba izabrati u uzorak ako se u procjeni s 99% pouzdanosti tolerira relativna pogreška od 0,05

Rješenje:

Projektira se uzorak za procjenu proporcije osnovnog skupa.

Parametri:
$N = 30.000$
d_r (ukupna relativna pogreška) = 0,05
$Z_{0,005} = 2,58$ (99% pouzdanosti).

Pogreška procjene je relativna pa se veličine uzorka određuje izrazom:

$$n = \left[\frac{Z \cdot \sqrt{\frac{Q}{P}}}{d_r} \right]^2 .$$

Obzirom da nije poznata standardna devijacija, odnosno koeficijent varijacije uzima se da je P=Q=0,5 pa je koeficijent varijacije 1:

$$n' = \left[\frac{2,58 \cdot \sqrt{1}}{0,05} \right]^2 = 2.662,6 \approx 2.663.$$

Provjera frakcije izbora:

$$f = \frac{n}{N} = \frac{2.663}{30.000} = 0,08877 > 0,05 \Rightarrow n = \frac{n'}{1 + \frac{n'}{N}} = \frac{2.663}{1 + \frac{2.663}{30.000}} = 2.446.$$

Zaključak:

Da bi se procijenila proporcija zaposlenih žena u uzorak treba izabrati 2.446 zaposlenih.

3.4. Procjenjivanje razlike aritmetičkih sredina dviju populacija

Na temelju podataka iz uzorka procjenjuje se razlika aritmetičkih sredina dviju populacija, tj. statističkih skupova. Razlika aritmetičkih sredina se može procijeniti brojem i intervalom.

Ako veličine uzoraka teže prema beskonačno **sampling distribucija razlike aritmetičkih sredina** teži **normalnom obliku**. Kod malih uzoraka sampling distribucija razlike aritmetičkih sredina ima oblik Studentove ili t-distribucije. Stoga vrijedi da ako je:

$n_1 + n_2 - 2 > 30 \quad \Rightarrow$ koristi se normalna distribucija i

$n_1 + n_2 - 2 \leq 30 \quad \Rightarrow$ koristi se Studentova ili t-distribucija.

Standardna pogreška razlike aritmetičkih sredina može se izračunati ovisno o slučaju:

- Ako su σ osnovnog skupa poznate i jednake za oba skupa:

$$Se(\overline{X}_1 - \overline{X}_2) = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,. \qquad (3.20)$$

- Ako je uzorak mali, tj. $n_1 + n_2 - 2 \leq 30$:

$$Se(\overline{X}_1 - \overline{X}_2) = \sqrt{\left(\frac{n_1 \cdot \hat{\sigma}_1^2 + n_2 \cdot \hat{\sigma}_2^2}{n_1 + n_2 - 2}\right) \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}\,. \qquad (3.21)$$

- Ako je uzorak veliki, tj. $n_1 + n_2 - 2 > 30$:

$$Se(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\,. \qquad (3.21)$$

gdje su:

σ - standardna devijacija osnovnog skupa (jednaka za oba skupa),

n_1 - veličina jednog uzorka,

n_2 - veličina drugog uzorka,

$\hat{\sigma}_1^2$ - varijanca jednog uzorka,

$\hat{\sigma}_2^2$ - varijanca drugog uzorka,

S_1^2 - nepristrana ocjena varijance jednog osnovnog skupa,

S_2^2 - nepristrana ocjena varijance drugog osnovnog skupa.

Točkasta procjena razlike aritmetičkih sredina dvaju osnovnih skupova na osnovi uzorka je razlika aritmetičkih sredina iz odabranih uzoraka, za koje se naravno pretpostavlja da su reprezentativni: $(\hat{\bar{X}}_1 - \hat{\bar{X}}_2)$.

Intervalna procjena razlike aritmetičkih sredina dvaju osnovnih skupova na osnovi uzorka uz odgovarajuću razinu pouzdanosti procjene je:

$$\Pr\left\{(\hat{\bar{X}}_1 - \hat{\bar{X}}_2) - Z \cdot Se(\bar{X}_1 - \bar{X}_2) < (\bar{X}_1 - \bar{X}_2) < (\hat{\bar{X}}_1 - \hat{\bar{X}}_2) + Z \cdot Se(\bar{X}_1 - \bar{X}_2)\right\} = 1 - \alpha, \quad (3.22)$$

gdje je:

$(\hat{\bar{X}}_1 - \hat{\bar{X}}_2)$ - razlika aritmetičkih sredina izračunata na osnovi uzorka,

Z - ako je ($n_1 + n_2 - 2 > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako je ($n_1 + n_2 - 2 \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df = n_1 + n_2 - 2]}$),

$Se(\bar{X}_1 - \bar{X}_2)$ - standardna pogreška razlike aritmetičkih sredina,

$1 - \alpha$ - nivo pouzdanosti procjene (ako nije određeno drugačije, najčešće se uzima da je $1 - \alpha = 95\%$, a u nekim slučajevima 99%).

Primjer 3.4.1.

Procjenjuje se razlika u prosječnim ocjenama na ispitu iz statistike studenata na smjeru «Ekonomija» i studenata na smjeru «Poslovna ekonomija». Prosječna ocjena 80 studenata smjera «Ekonomije» iznosi 3,3 s varijancom 1,1. Ispitu je pristupilo 150 studenata smjera Poslovne ekonomije, koji su ostvarili prosječnu ocjenu 3,0 s varijancom 0,9. Procjenu izvršite s 95% pouzdanosti.

Rješenje:

Procjenjuje se razlika aritmetičkih sredina dvaju osnovnih skupova intervalom:

$$\Pr\left\{(\hat{\overline{X}}_1 - \hat{\overline{X}}_2) - Z \cdot Se(\overline{X}_1 - \overline{X}_2) < (\overline{X}_1 - \overline{X}_2) < (\hat{\overline{X}}_1 - \hat{\overline{X}}_2) + Z \cdot Se(\overline{X}_1 - \overline{X}_2)\right\} = 1 - \alpha.$$

Parametri:

$n_1 = 80$

$n_2 = 150$

$\hat{\overline{X}}_1 = 3{,}3$

$\hat{\overline{X}}_2 = 3{,}0$

$\hat{\sigma}_1^2 = 1{,}1 = S_1^2$ jer je $n_1 + n_2 - 2 = 80 + 150 - 2 = 228 > 30$

$\hat{\sigma}_2^2 = 0{,}9 = S_2^2$ jer je $n_1 + n_2 - 2 = 80 + 150 - 2 = 228 > 30$

$Z_{0,025} = 1{,}96$ (95% pouzdanosti).

Izračun standardne pogreške razlike aritmetičkih sredina dvaju osnovnih skupova:

$$n_1 + n_2 - 2 > 30 \Rightarrow Se(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{1{,}1}{80} + \frac{0{,}9}{150}} = 0{,}14.$$

Intervalna procjena razlike aritmetičkih sredina dvaju osnovnih skupova iznosi:

$$\Pr\left\{(\hat{\overline{X}}_1 - \hat{\overline{X}}_2) - Z \cdot Se(\overline{X}_1 - \overline{X}_2) < (\overline{X}_1 - \overline{X}_2) < (\hat{\overline{X}}_1 - \hat{\overline{X}}_2) + Z \cdot Se(\overline{X}_1 - \overline{X}_2)\right\} = 1 - \alpha$$

$$\Pr\left\{(3{,}3 - 3{,}0) - 1{,}96 \cdot 0{,}14 < (\overline{X}_1 - \overline{X}_2) < (3{,}3 - 3{,}0) + 1{,}96 \cdot 0{,}14\right\} = 0{,}95$$

$$\Pr\left\{0{,}0256 < (\overline{X}_1 - \overline{X}_2) < 0{,}5744)\right\} = 0{,}95\ .$$

Zaključak:

Na razini pouzdanosti od 95% procjenjuje se da je razlika prosječnih ocjena na ispitu iz statistike studenata na smjeru «Ekonomija» i studenata na smjer «Poslovna ekonomija» veća od 0,0256 i manja od 0,5744.

3.5. Procjenjivanje razlike proporcija dviju populacija

Na temelju podataka iz uzorka procjenjuje se razlika proporcija dviju populacija, tj. statističkih skupova. Razlika proporcija se može procijeniti brojem i intervalom.

Ako veličine uzoraka teže prema beskonačno **sampling distribucija razlike proporcija** teži **normalnom obliku**. Kod malih uzoraka sampling distribucija razlike proporcija ima oblik Studentove ili t-distribucije. Stoga vrijedi da ako je:

$n_1 + n_2 - 2 > 30 \quad \Rightarrow$ koristi se normalna distribucija i

$n_1 + n_2 - 2 \leq 30 \quad \Rightarrow$ koristi se Studentova ili t-distribucija.

Standardna pogreška razlike proporcija se računa na sljedeći način:

$$Se(P_1 - P_2) = \sqrt{\hat{\hat{p}} \cdot \left(1 - \hat{\hat{p}}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \tag{3.23}$$

gdje je:

$$\hat{\hat{p}} = \frac{m_1 + m_2}{n_1 + n_2}, \tag{3.24}$$

$\hat{\hat{p}}$ - prosječna proporcija (relativna frekvencija) za oba uzorka zajedno,

m_1 - broj «povoljnih ishoda» jednog uzorka,

m_2 - broj «povoljnih ishoda» drugog uzorka,

n_1 - veličina jednog uzorka,

n_2 - veličina drugog uzorka.

Točkasta procjena razlike proporcija dvaju osnovnih skupova na osnovu uzorka je razlika proporcija iz odabranih uzoraka, za koje se pretpostavlja da su reprezentativni: $(\hat{p}_1 - \hat{p}_2)$.

Intervalna procjena razlike proporcija dvaju osnovnih skupova na osnovi uzorka uz odgovarajuću razinu pouzdanosti procjene je:

$$\Pr\{(\hat{p}_1 - \hat{p}_2) - Z \cdot Se(P_1 - P_2) < (P_1 - P_2) < (\hat{p}_1 - \hat{p}_2) + Z \cdot Se(P_1 - P_2)\} = 1 - \alpha, \tag{3.25}$$

gdje je:

$(\hat{p}_1 - \hat{p}_2)$ - razlika proporcija izračunata na osnovu uzoraka,

Z - ako je $(n_1 + n_2 - 2 > 30)$ računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako je $(n_1 + n_2 - 2 \leq 30)$ umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df = n_1 + n_2 - 2]}$),

$Se(P_1 - P_2)$ - standardna pogreška razlike proporcija,

$1-\alpha$ - razina pouzdanosti procjene (ako nije određeno drugačije, najčešće se uzima da je $1-\alpha = 95\%$, a u nekim slučajevima 99%).

Primjer 3.5.1.

Procjenjuje se razlika u proporciji muškaraca i žena koji nisu zadovoljni prehranom u tvorničkom restoranu. Od 64 anketiranih žena 16 nije zadovoljno prehranom, a od 95 anketiranih muškaraca 10 ih se izjasnio da nije zadovoljno kvalitetom hrane. Procjenu izvršite s 95% pouzdanosti.

Rješenje:

Procjenjuje se razlika proporcija dvaju osnovnih skupova intervalom:

$$\Pr\{(\hat{p}_1 - \hat{p}_2) - Z \cdot Se(P_1 - P_2) < (P_1 - P_2) < (\hat{p}_1 - \hat{p}_2) + Z \cdot Se(P_1 - P_2)\} = 1 - \alpha.$$

Parametri:

$n_1 = 64$

$n_2 = 95$

$m_1 = 16$

$m_2 = 10$

$Z_{0,025} = 1,96$ (95% pouzdanosti).

Izračun proporcija uzoraka:

$$\hat{p}_1 = \frac{m_1}{n_1} = \frac{16}{64} = 0,25 \qquad \hat{p}_2 = \frac{m_2}{n_2} = \frac{10}{95} = 0,105\,.$$

Izračun prosječne proporcije:

$$\hat{\hat{p}} = \frac{m_1 + m_2}{n_1 + n_2} = \frac{16 + 10}{64 + 95} = 0{,}1635.$$

Izračun standardne pogreške razlike proporcija dvaju osnovnih skupova:

$$Se(P_1 - P_2) = \sqrt{\hat{\hat{p}} \cdot \left(1 - \hat{\hat{p}}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0{,}1635 \cdot (1 - 0{,}1635)\left(\frac{1}{64} + \frac{1}{95}\right)} = 0{,}05981.$$

Intervalna procjena razlike aritmetičkih sredina dvaju osnovnih skupova iznosi:

$$\Pr\left\{(\hat{p}_1 - \hat{p}_2) - Z \cdot Se(P_1 - P_2) < (P_1 - P_2) < (\hat{p}_1 - \hat{p}_2) + Z \cdot Se(P_1 - P_2)\right\} = 1 - \alpha$$

$$\Pr\left\{(0{,}25 - 0{,}105) - 1{,}96 \cdot 0{,}05981 < (P_1 - P_2) < (0{,}25 - 0{,}105) + 1{,}96 \cdot 0{,}05981\right\} = 0{,}95$$

$$\Pr\left\{0{,}028 < (P_1 - P_2) < 0{,}262\right\} = 0{,}95 .$$

Zaključak:
Na razini pouzdanosti od 95% procjenjuje se da je razlika proporcija muškaraca i žena koji nisu zadovoljni prehranom u tvorničkom restoranu veća od 0,028 i manja od 0,262.

4. TESTIRANJE HIPOTEZA

Istraživanja koja se provode na cijelom osnovnom skupu često su skupa i zahtijevaju mnogo vremena, pa se u praksi često, koristeći metode i tehnike inferencijalne statistike, na temelju podataka iz uzorka donose zaključci vezani za osnovni skup. Pri takvom zaključivanju postavljaju se različite pretpostavke ili hipoteze. Potrebno je razlikovati **znanstvene hipoteze** i **statističke hipoteze**.

Znanstvene hipoteze predstavljaju nagađanje, naslućivanje i pretpostavke koje motiviraju istraživača. Iz znanstvene hipoteze, tj. hipoteze istraživača (koja je najčešće afirmativna) izvodi se statistička hipoteza.

Statističke hipoteze postavljaju se na način da može biti vrednovana statističko-analitičkim postupcima. Ona je u stvari matematički izraz koji predstavlja polaznu osnovu na kojoj se temelji kalkulacija statističkog testa.

Testiranje hipoteza je statistički postupak kojim se određuje jesu li i koliko pouzdano raspoloživi podaci iz reprezentativnog uzorka podupiru pretpostavljenu pretpostavku.

Pri testiranju hipoteza **potrebno je**:

- postaviti *nultu ili početnu hipotezu* (H_0) i *alternativnu hipotezu* (H_1);
- izabrati *razinu značajnosti ili signifikantnosti* (α);
- prikupiti primjerene podatke na *reprezentativnom uzorku;*
- *izračunati vrijednost rezultata statističkog testa* (*empirijska vrijednost testa* iz uzorka) specifičnog za nultu hipotezu (H_0);
- *usporediti empirijsku vrijednost testa* s vrijednosti iz poznate distribucije vjerojatnosti (*s tabličnom vrijednosti testa*) specifičnog za nultu hipotezu (H_0);
- *interpretirati* rezultat statističkog testa *u terminima vjerojatnosti (signifikantnosti).*

Nulta hipoteza, H_0 (eng. null hypothesis) pretpostavka je o izostanku efekta, tj. ne postoji razlika među uzorcima u promatranoj populaciji. Ta početna hipoteza u stvari je ona pretpostavka koja se testira, tj. hipoteza da nema razlike (eng. hypothesis of no difference). Postavlja se najčešće u svrhu odbacivanja.

Alternativna hipoteza, H_1 (eng. alternative hypothesis) vrijedi ako nul-hipoteza nije istinita. Ona se najčešće direktno odnosi na teorijsku pretpostavku koja se želi istražiti, tj. može se reći da je alternativna hipoteza ustvari hipoteza istraživača.

Kada se ne može unaprijed sa sigurnošću odrediti smjer neke razlike, a ona postoji, primjenjuje se **dvosmjerni test** (eng. two-tailed test).

Jednosmjerni test (eng. one-tailed test) primjenjuje se kada je smjer razlike specificiran u alternativnoj hipotezi (H_1).

S obzirom da se zaključivanje provodi na temelju informacija o uzorku, moguće je pogriješiti i donijeti krivi zaključak.

Veličina signifikantnosti ili značajnosti testa α je u literaturi poznata kao **Greška tipa I**, odnosno kao vjerojatnost da se odbaci nulta hipoteza premda je ona istinita. U praktičnim istraživanjima najčešće se uzima da je $\alpha = 5\%$.

Greška tipa II (β) predstavlja vjerojatnost da se prihvati nulta hipoteza premda ona nije istinita. **Snaga testa** $(1-\beta)$ je vjerojatnost da se ne prihvati lažna nulta hipoteza. Navedene **vjerojatnosti** prikazane su u tablici 4.1.

Tablica 4.1.

Vjerojatnosti prihvaćanja/odbacivanja lažne/istinite H_0 hipoteze

	H_0 **prihvaćena**	H_0 **odbačena**
H_0 **istinita**	$1-\alpha$	α
H_0 **lažna**	β	$1-\beta$

Izvor: Konstrukcija autorica prema teorijskim postavkama.

4.1. Testiranje hipoteze o nepoznatoj aritmetičkoj sredini osnovnog skupa

Postavljaju se **hipoteze za dvosmjerno testiranje,** gdje nulta hipoteza glasi da je aritmetička sredina osnovnog skupa $\overline{X}$ jednaka nekoj pretpostavljenoj vrijednosti $\overline{X}_0$:

$$H_0 : ....\overline{X} = \overline{X}_0$$
$$H_1 : ....\overline{X} \neq \overline{X}_0.$$

Interval prihvaćanja hipoteze H_0 glasi:

$$\overline{X}_0 \pm Z \cdot Se(\overline{X}), \tag{4.1}$$

gdje je:

$\overline{X}_0$ - neka pretpostavljena aritmetička sredina,

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako je ($n \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n-1]}$),

$Se(\overline{X})$ - standardna pogreška aritmetičke sredine (standardna devijacija sampling distribucije aritmetičke sredine):

$$Se(\hat{\overline{X}}) = \frac{\sigma}{\sqrt{n}} \qquad \Rightarrow \qquad \text{ako je } f \leq 0{,}05 \text{ i} \tag{4.2}$$

$$Se(\hat{\overline{X}}) = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \qquad \Rightarrow \qquad \text{ako je } f > 0{,}05. \tag{4.3}$$

Ako nije poznata standardna devijacija osnovnog skupa σ, računa se **nepristrana** (točkasta) **ocjena varijance osnovnog skupa na osnovi uzorka** S^2:

$$S^2 = \hat{\sigma}^2 \cdot \left(\frac{n}{n-1}\right) \qquad \Rightarrow \qquad \text{ako je } n \leq 30 \text{ i} \tag{4.4}$$

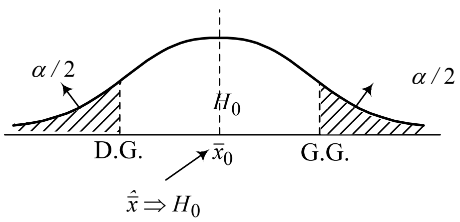
$$S^2 = \hat{\sigma}^2 \qquad \Rightarrow \qquad \text{ako je } n > 30 \text{ .} \tag{4.5}$$

α - razina signifikanosti ili značajnosti testa (ako nije određeno drukčije, najčešće se uzima de je $\alpha = 5\%$).

Zaključak o prihvaćanju ili odbacivanju nulte H_0 hipoteze donosi se na osnovi $\hat{\overline{X}}$ aritmetičke sredine iz uzorka (prema slici 4.1.).

Slika 4.1.

Odluka o prihvaćanju hipoteza kod dvosmjernog testiranja o nepoznatoj aritmetičkoj sredini osnovnog skupa



Izvor: Konstrukcija autorica

Prema slici 4.1., ako se aritmetička sredina iz uzorka $\hat{\overline{X}}$ nalazi između donje (D.G.) i gornje granice (G.G.) intervala prihvaćanja hipoteze H_0, ta se hipoteza prihvaća kao istinita uz odgovarajuću razinu signifikantnosti testa (α). U suprotnom se ta hipoteza odbacuje.

Testiranje se može izvesti Z testom (ako je $n > 30$) ili t testom (ako je $n \leq 30$):

$$Z^* = \frac{\hat{\overline{X}} - \overline{X}_0}{Se(\overline{X})}; \qquad Z_{tab\left[\frac{1-\alpha}{2}\right]}, \tag{4.6}$$

gdje je:

Z^* - Z empirijska vrijednost izračunata na osnovi uzorka,

Z_{tab} - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

$t*$ - t empirijska vrijednost izračunata na osnovi uzorka (ako je $n \leq 30$),

t_{tab} - tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n-1]}$).

Zaključak se donosi na način da ako je:

$|Z*| < Z_{tab} \quad \Rightarrow H_0$; odnosno $|t*| < t_{tab} \quad \Rightarrow H_0$,

dok se u suprotnom slučaju ta početna hipoteza odbacuje.

Testiranje se može izvršiti i izračunavanjem granične signifikantnosti $\alpha*$ pomoću $Z*$ ili t^* (Tablica A ili B): ako je $\alpha* > 5\%$ ili $1\% \Rightarrow H_0$, dok se u suprotnom slučaju hipoteza H_0 odbacuje.

Hipoteze se mogu postaviti i za **jednosmjerno testiranje** na donju ili gornju granicu.

4.1.1. Testiranje na donju granicu

Testiranjem na donju granicu testira se hipoteza da je aritmetička sredina osnovnog skupa $\overline{X}$ manja od neke pretpostavljene vrijednosti $\overline{X}_0$**:**

$$H_0 : ....\overline{X} \geq \overline{X}_0$$
$$H_1 : ....\overline{X} < \overline{X}_0.$$

Ako iz određenih razloga postoji sigurnost da aritmetička sredina osnovnoga skupa ne može poprimiti vrijednost veću od $\overline{X}_0$, tada se nulta hipoteza u jednosmjernom testiranju može postaviti kao i kod dvosmjernoga testiranja (znakom jednakosti).

Donja granica prihvaćanja hipoteze H_0 glasi:

$$D.G. = \overline{X}_0 - Z \cdot Se(\overline{X}), \tag{4.7}$$

gdje je:

$\overline{X}_0$ - neka pretpostavljena aritmetička sredina,

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-2\alpha}{2}}$),

t - ako je ($n \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha}_{[df=n-1]}$),

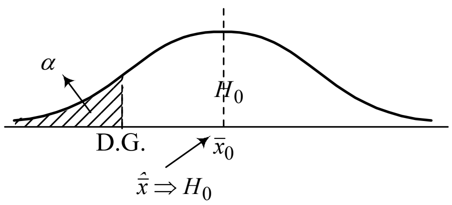
$Se(\overline{X})$ - standardna pogreška aritmetičke sredine,

α - razina signifikanosti ili značajnosti testa (ako nije određeno drukčije, najčešće se uzima de je $\alpha = 5\%$).

Zaključak o prihvaćanju ili odbacivanju nulte H_0 hipoteze donosi se na osnovi $\hat{\overline{X}}$ aritmetičke sredine iz uzorka (prema slici 4.2).

Slika 4.2.

Odluka o prihvaćanju hipoteza kod jednosmjernog testiranja na «donju granicu» o nepoznatoj aritmetičkoj sredini osnovnog skupa



Izvor: Konstrukcija autorica

Prema slici 4.2., ako se aritmetička sredina iz uzorka $\hat{\overline{X}}$ nalazi iznad donje (D.G.) granice prihvaćanja hipoteze H_0, ta se hipoteza prihvaća kao istinita uz odgovarajuću razinu signifikantnosti testa (α). U suprotnom se ta hipoteza odbacuje.

Testiranje se može izvesti Z testom (ako je $n > 30$) ili t testom (ako je $n \leq 30$):

$$Z^* = \frac{\hat{\overline{X}} - \overline{X}_0}{Se(\overline{X})} \; ; \qquad Z_{tab\left[\frac{1-2\alpha}{2}\right]}, \tag{4.8}$$

gdje je:

Z^* - Z empirijska vrijednost izračunata na osnovi uzorka,

Z_{tab} - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-2\alpha}{2}}$),

t^* - t empirijska vrijednost izračunata na osnovi uzorka (ako je $n \leq 30$),

t_{tab} - tablice površina studentove ili t-distribucije (pomoću $t^{\alpha}_{[df=n-1]}$).

Zaključak se donosi na način da ako je:

$Z^* > -Z_{tab} \Rightarrow H_0$; odnosno $t^* > -t_{tab} \Rightarrow H_0$,

dok se u suprotnom ta hipoteza odbacuje.

4.1.2. Testiranje na gornju granicu

Testiranjem na gornju granicu testira se hipoteza da je aritmetička sredina osnovnog skupa $\overline{X}$ veća od neke pretpostavljene vrijednosti $\overline{X}_0$:

$$H_0 : ....\overline{X} \leq \overline{X}_0$$
$$H_1 : ....\overline{X} > \overline{X}_0.$$

Ako iz određenih razloga postoji sigurnost da aritmetička sredina osnovnoga skupa ne može poprimiti vrijednost manju od $\overline{X}_0$, tada se nulta hipoteza u jednosmjernom testiranju može postaviti kao i kod dvosmjernoga testiranja (znakom jednakosti).

Gornja granica prihvaćanja hipoteze H_0 glasi:

$$G.G. = \overline{X}_0 + Z \cdot Se(\overline{X}), \tag{4.9}$$

gdje je:

$\overline{X}_0$ - neka pretpostavljena aritmetička sredina,

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-2\alpha}{2}}$),

t - ako je ($n \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha}_{[df=n-1]}$),

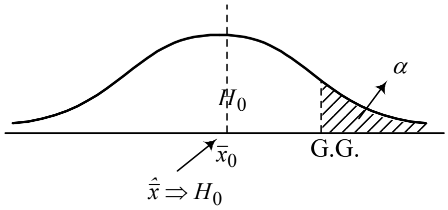
$Se(\overline{X})$ - standardna pogreška aritmetičke sredine,

α - razina signifikanosti ili značajnosti testa (ako nije određeno drukčije, najčešće se uzima de je $\alpha = 5\%$).

Zaključak o prihvaćanju ili odbacivanju nulte H_0 hipoteze donosi se na osnovi $\hat{\overline{X}}$ aritmetičke sredine iz uzorka (prema slici 4.3).

Slika 4.3.

Odluka o prihvaćanju hipoteza kod jednosmjernog testiranja na «gornju granicu» o nepoznatoj aritmetičkoj sredini osnovnog skupa



Izvor: Konstrukcija autorica

Prema slici 4.3., ako se aritmetička sredina iz uzorka $\hat{\overline{X}}$ nalazi ispod gornje (G.G.) granice prihvaćanja hipoteze H_0, ta se hipoteza prihvaća kao istinita uz odgovarajuću razinu signifikantnosti testa (α). U suprotnom se ta hipoteza odbacuje.

Testiranje se može izvesti Z testom (ako je $n > 30$) ili t testom (ako je $n \leq 30$):

$$Z^* = \frac{\hat{\overline{X}} - \overline{X}_0}{Se(\overline{X})}; \qquad Z_{tab\left[\frac{1-2\alpha}{2}\right]}, \tag{4.10}$$

gdje je:

Z^* - Z empirijska vrijednost izračunata na osnovi uzorka,

Z_{tab} - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-2\alpha}{2}}$),

t^* - t empirijska vrijednost izračunata na osnovi uzorka (ako je $n \leq 30$),

t_{tab} - tablice površina studentove ili t-distribucije (pomoću $t^{\alpha}_{[df=n-1]}$).

Zaključak se donosi na način da ako je:

$$Z^* < Z_{tab} \Rightarrow H_0 \text{; odnosno } t^* < t_{tab} \Rightarrow H_0,$$
dok se u suprotnom ta hipoteza odbacuje.

Primjer 4.1.1.

Prosječna trajnost baterija iznosi 20 sati sa standardnim odstupanjem od 1,5 sata. Uz frakciju odabiranja manju od 5% izabran je uzorak od 45 baterija. Prosječna trajnost baterija u uzorku iznosi 19,6 sati. Može li se prihvatiti pretpostavka da je uzorak izabran iz osnovnog skupa kojemu je aritmetička sredina jednaka 20 sati? Testirate na razini signifikantnosti 5% (pogreška tipa I).

Rješenje:

Testira se hipoteza o nepoznatoj aritmetičkoj sredini osnovnog skupa:
$$H_0 : \overline{X} = \overline{X}_0 \qquad H_1 : \overline{X} \neq \overline{X}_0.$$
Parametri:
$\overline{X}_0 = 20$ sati
$\sigma = 1{,}5$ sati
$f < 5\%$
$n = 45$
$\hat{\overline{X}} = 19{,}6$ sati
$Z_{0,025} = 1{,}96$ (95% pouzdanosti ili 5% signifikantnosti).

Izračun standardne pogreške aritmetičke sredine:
$$f < 0{,}05 \Rightarrow Se(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{1{,}5}{\sqrt{45}} = 0{,}2236.$$

Interval prihvaćanja hipoteze H_0 je:
$$\overline{X}_0 \mp Z \cdot Se(\overline{X}) = 20 \mp 1{,}96 \cdot 0{,}2236$$
$19{,}16 < \hat{\overline{X}} = 19{,}6 < 20{,}04 \Rightarrow H_0$ se prihvaća kao istinita.

Z test prihvaćanja hipoteze H_0 glasi:
$$Z^* = \frac{\hat{\overline{X}} - \overline{X}_0}{Se(\overline{X})} = \frac{19{,}6 - 20}{0{,}2236} = -1{,}789$$

$|Z^*| = |1{,}789| < Z_{0{,}025} = 1{,}96 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 95%, odnosno uz mogućnost pogreške tipa I, prihvaća se pretpostavka da je uzorak izabran iz osnovnog skupa u kojemu prosječna trajnost baterija iznosi 20 sati.

Primjer 4.1.2.

Tvornica trikotaže namjerava nabaviti novi stroj. Procjenjuje se da je nabavka novog stroja isplativa samo ako stroj proizvodi više od 50 komada trikotaže na sat. Sprovedena su 35 mjerenja i izračunan je prosječni broj od 51 proizvedenog komada trikotaže na sat sa standardnom devijacijom od 5 komada po satu. Isplati li se tvornici nabaviti novi stroj. Testiranje izvršite uz razinu signifikantnosti od 5%.

Rješenje:

Testira se hipoteza da je aritmetička sredina osnovnog skupa $\overline{X}$ veća od pretpostavljene vrijednosti $\overline{X}_0$ pa se provodi jednosmjerno testiranje i to na gornju granicu:

$$H_0 : \overline{X} \leq \overline{X}_0 \qquad H_1 : \overline{X} > \overline{X}_0.$$

Parametri:
$\overline{X}_0 = 50$
$n = 35$
$\hat{\overline{X}} = 51$
$\hat{\sigma} = 5$
$Z_{0{,}05} = 1{,}65$ (uz razinu 5% signifikantnosti za jednosmjerno testiranje).

Ocjena standardne devijacije osnovnog skupa (S) na osnovi standardne devijacije uzorka ($\hat{\sigma}$):
$n = 35 > 30 \Rightarrow S = \hat{\sigma} = 5.$

Izračun standardne pogreške aritmetičke sredine:

$$f < 0{,}05 \Rightarrow Se(\overline{X}) = \frac{S}{\sqrt{n}} = \frac{5}{\sqrt{35}} = 0{,}8452.$$

Interval prihvaćanja hipoteze H_0 je:
$G.G. = \overline{X}_0 + Z \cdot Se(\overline{X}) = 50 + 1{,}65 \cdot 0{,}8452 = 51{,}39$
$\hat{\overline{X}} = 51 < G.G. = 51{,}39 \Rightarrow H_0$ se prihvaća kao istinita.

Z test prihvaćanja hipoteze H_0 glasi:

$$Z^* = \frac{\hat{\overline{X}} - \overline{X}_0}{Se(\overline{X})} = \frac{51 - 50}{0{,}8452} = 1{,}18$$

$Z^* = 1{,}18 < Z_{0,05} = 1{,}65 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 95%, odnosno signifikantnosti od 5%, može se zaključiti da se ne isplati nabavka novog tkalačkog stroja jer novi stroj ne proizvodi više od 50 komada trikotaže na sat, a što je pretpostavljena granica isplativosti.

Primjer 4.1.3.

U radionici se razmatra mogućnost promjene procesa rada. Promjena je isplativa ako se troškovi proizvodnje smanje ispod trenutnih 80 kn prosječno po proizvodu. U probnom procesu proizvedeno je 164 proizvoda s prosječnom cijenom proizvodnje od 76 kn i standardnom devijacijom od 21 kn. Uz 99% pouzdanosti procijenite isplativost promjene procesa rada.

Rješenje:

Žele se smanjiti troškovi proizvodnje pa se testira hipoteza da je aritmetička sredina osnovnog skupa $\overline{X}$ manja od pretpostavljene vrijednosti $\overline{X}_0$. Znači da se provodi jednosmjerno testiranje i to na donju granicu:

$$H_0 : \overline{X} \geq \overline{X}_0 \qquad H_1 : \overline{X} < \overline{X}_0 .$$

Parametri:
$\overline{X}_0 = 80$
$n = 164$
$\hat{\overline{X}} = 76$

$\hat{\sigma} = 21$
$Z_{0,01} = 2{,}33$ (uz razinu 5% signifikantnosti za jednosmjerno testiranje).

Ocjena standardne devijacije osnovnog skupa (*S*) na osnovi standardne devijacije uzorka ($\hat{\sigma}$):
$n = 164 > 30 \Rightarrow S = \hat{\sigma} = 21.$

Izračun standardne pogreške aritmetičke sredine:

$$f < 0{,}05 \Rightarrow Se(\overline{X}) = \frac{S}{\sqrt{n}} = \frac{21}{\sqrt{164}} = 1{,}64.$$

Interval prihvaćanja hipoteze H_0 je:

$$D.G. = \overline{X}_0 - Z \cdot Se(\overline{X}) = 80 - 2{,}33 \cdot 1{,}64 = 76{,}2$$

$\hat{\overline{X}} = 76 < D.G. = 76{,}2 \Rightarrow H_1$ se prihvaća kao istinita, a H_0 odbacuje kao neistinita.

Z test prihvaćanja hipoteze H_0 glasi:

$$Z^* = \frac{\hat{\overline{X}} - \overline{X}_0}{Se(\overline{X})} = \frac{76 - 80}{1{,}64} = -2{,}43$$

$Z^* = -2{,}43 < -Z_{0,01} = -2{,}33 \Rightarrow H_1$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 99%, odnosno signifikantnosti od 1%, može se zaključiti da se isplati promjena procesa rada jer novi način proizvodnje smanjuje troškove proizvodnje ispod 80 kn prosječno po proizvodu, što je granica isplativosti.

Primjer 4.1.4.

Proizvodnjom vina bavi se 350 proizvođača u regiji. Pretpostavlja se da oni prosječno zapošljavaju po 5 radnika. U slučajnom uzorku od 20 proizvođača vina prosječan broj zaposlenih iznosio je 4,8 sa standardnom devijacijom od 1,9 radnika. Testirajte s 5% signifikantnosti.

Rješenje:

Testira se hipoteza o nepoznatoj aritmetičkoj sredini osnovnog skupa:

$$H_0: \overline{X} = \overline{X}_0 \qquad H_1: \overline{X} \neq \overline{X}_0.$$

Parametri:
$N = 350$
$\overline{X}_0 = 5$
$n = 20$
$\hat{\overline{X}} = 4{,}8$
$\hat{\sigma} = 1{,}9.$

Ocjena standardne devijacije osnovnog skupa (S) na osnovi standardne devijacije uzorka ($\hat{\sigma}$):

$$n = 20 < 30 \Rightarrow S = \hat{\sigma} \cdot \sqrt{\frac{n}{n-1}} = 1{,}9 \cdot \sqrt{\frac{20}{20-1}} = 1{,}949.$$

Izračun standardne pogreške aritmetičke sredine:

$$f = \frac{n}{N} = \frac{20}{350} = 0{,}057 > 0{,}05 \Rightarrow Se(\overline{X}) = \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{1{,}949}{\sqrt{20}} \cdot \sqrt{\frac{350-20}{350-1}} = 0{,}424$$

t test (mali uzorak) prihvaćanja hipoteze H_0 glasi:

$$t^* = \frac{\hat{\overline{X}} - \overline{X}_0}{Se(\overline{X})} = \frac{4{,}8-5}{0{,}424} = -0{,}47$$

t_{tab} iz tablice površina studentove ili t-distribucije, uz 19 stupnjeva slobode (*df=n-1=20-1=19*) i 5% značajnosti ($t_{19}^{0,05}$) $\Rightarrow t = 1{,}729$

$|t^*| = |0{,}47| < t_{19}^{0,05} = 1{,}729 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini signifikantnosti od 5%, odnosno uz mogućnost pogreške tipa I, prihvaća se pretpostavka da proizvođači vina u toj regiji prosječno zapošljavaju po 5 radnika.

4.2. Testiranje hipoteze o pretpostavljenoj proporciji (relativnoj frekvenciji) osnovnog skupa

Postavljaju se **hipoteze** da je proporcija (relativna frekvencija) osnovnog skupa P jednaka nekoj pretpostavljenoj vrijednosti P_0:

$$H_0 : ....P = P_0$$
$$H_1 : ....P \neq P_0.$$

Interval prihvaćanja hipoteze H_0 glasi:

$$P_0 \pm Z \cdot Se(p) \qquad (4.11)$$

gdje je:

P_0 - neka pretpostavljena proporcija (relativna frekvencija),

Z - ako je ($n > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako je ($n \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t_{[df=n-1]}^{\alpha/2}$),

$Se(p)$ - standardna pogreška proporcije (relativne frekvencije),

α - razina signifikanosti ili značajnosti testa (ako nije određeno drukčije, najčešće se uzima de je $\alpha = 5\%$).

Standardna pogreška proporcije (relativne frekvencije) kod ovog testiranja hipoteza računa se:

$$Se(p) = \sqrt{\frac{P_0 \cdot Q_0}{n}} \quad \Rightarrow \quad \text{ako je } n > 30 \text{ i} \qquad (4.12)$$

$$Se(p) = \sqrt{\frac{P_0 \cdot Q_0}{n-1}} \quad \Rightarrow \quad \text{ako je } n \leq 30 \qquad (4.13)$$

gdje su:

P_0 - neka pretpostavljena proporcija (relativna frekvencija),

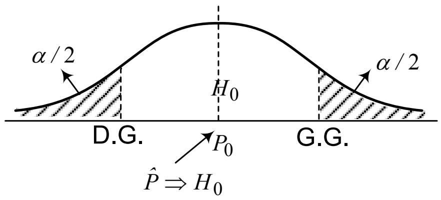
Q_0 = $1 - P_0$,

n - veličina uzorka.

Zaključak o prihvaćanju ili odbacivanju nulte H_0 hipoteze donosi se na osnovi $\hat{P}$ proporcije (relativne frekvencije) iz uzorka (prema slici 4.4).

Slika 4.4.

Odluka o prihvaćanju hipoteza kod testiranja hipoteze o nepoznatoj proporciji (relativnoj frekvenciji) osnovnog skupa



Izvor: Konstrukcija autorica

Prema slici 4.4., ako se proporcija (relativna frekvencija) iz uzorka $\hat{P}$ nalazi između donje (D.G.) i gornje granice (G.G.) intervala prihvaćanja hipoteze $H_{0,}$ ta se hipoteza prihvaća kao istinita uz odgovarajuću razinu signifikantnosti testa (α). U suprotnom se ta hipoteza odbacuje.

Testiranje se može izvesti Z testom (ako je $n > 30$):

$$Z^* = \frac{\hat{P} - P_0}{Se(P)}; \qquad Z_{tab\left[\frac{1-\alpha}{2}\right]} \tag{4.14}$$

gdje je:

Z^* - Z empirijska vrijednost izračunata na osnovi uzorka,

Z_{tab} - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$).

Zaključak se donosi na način da ako je:

$|Z^*| < Z_{tab} \Rightarrow H_0$ dok se u suprotnom ta hipoteza odbacuje.

Testiranje se može izvršiti i izračunavanjem granične signifikantnosti α pomoću Z^* (Tablica A): ako je $\alpha^* > 5\% \Rightarrow H_0$, dok se u suprotnom slučaju hipoteza H_0 odbacuje.

Testovi se i ovdje mogu postaviti jednosmjerno, kao i kod hipoteza o nepoznatoj aritmetičkoj sredini osnovnog skupa.

Primjer 4.2.1.

Pretpostavlja se da je proporcija odličnih studenata na Ekonomskom fakultetu Sveučilišta u Rijeci 10%. Uz frakciju izbora manju od 5% u uzorak je izabrano 150 studenata među kojima je bilo 18 odličnih. Pretpostavku testirate na razini 95% pouzdanosti.

Rješenje:

Testira se hipoteza o pretpostavljenoj proporciji osnovnog skupa:

$$H_0 : P = P_0 \qquad H_1 : P \neq P_0.$$

Parametri:

$P_0 = 10\% = 0{,}1$

$f < 5\%$

$n = 150$

$m = 18$

$Z_{0,025} = 1{,}96$ (95% pouzdanosti ili 5% signifikantnosti).

Izračun proporcije uzorka:

$$\hat{p} = \frac{m}{n} = \frac{18}{150} = 0{,}12.$$

Izračun protivne proporcije pretpostavljenoj:

$$Q_0 = 1 - P_0 = 1 - 0{,}1 = 0{,}9.$$

Izračun standardne pogreške proporcije:

$$n > 30 \Rightarrow \sqrt{\frac{P_0 \cdot Q_0}{n}} = \sqrt{\frac{0{,}1 \cdot 0{,}9}{150}} = 0{,}02449.$$

Interval prihvaćanja hipoteze H_0 je:

$P_0 \pm Z \cdot Se(p)$

$0{,}1 \mp 1{,}96 \cdot 0{,}02449$

$0{,}052 < \hat{p} = 0{,}12 < 0{,}148 \Rightarrow H_0$ se prihvaća kao istinita.

Z test prihvaćanja hipoteze H_0 glasi:

$$Z^* = \frac{\hat{p} - P_0}{Se(p)} = \frac{0{,}12 - 0{,}1}{0{,}02449} = 0{,}82$$

$Z^* = 0{,}82 < Z_{0{,}025} = 1{,}96 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 95%, odnosno uz mogućnost pogreške tipa I, prihvaća se pretpostavka da proporcija odličnih studenata na Ekonomskom fakultetu Sveučilišta u Rijeci iznosi 10%.

Primjer 4.2.2.

Proizvod A plasira na tržište veći broj proizvođača. Proizvođač iz Rijeke pretpostavlja da će pomoću propagande povećati prodaju tako da će proizvod A njegove proizvodnje kupovati najmanje svaki drugi kupac. Nakon provedene promidžbene kampanje u uzorak je slučajnim izborom izabrano 500 kupaca (f<5%) od kojih je 270 kupilo proizvod A proizvođača iz Rijeke. Testirajte je li promidžbena kampanja bila uspješna na razini 5% signifikantnosti.

Rješenje:

Proizvođač pretpostavlja da će proporcija prodaje biti veća od pretpostavljene (P_0) pa se provodi jednosmjerno testiranje i to na gornju granicu:

$$H_0: P \leq P_0 \qquad H_1: P > P_0 .$$

Parametri:
$P_0 = 1/2 = 0{,}5$
$f < 5\%$
$n = 500$
$m = 270$
$Z_{0{,}05} = 1{,}65$ (uz razinu 5% signifikantnosti za jednosmjerno testiranje).

Izračun proporcije uzorka:

$$\hat{p} = \frac{m}{n} = \frac{270}{500} = 0{,}54.$$

Izračun protivne proporcije pretpostavljenoj:
$Q_0 = 1 - P_0 = 1 - 0{,}54 = 0{,}46.$

Izračun standardne pogreške proporcije:
$$n > 30 \Rightarrow \sqrt{\frac{P_0 \cdot Q_0}{n}} = \sqrt{\frac{0{,}54 \cdot 0{,}46}{500}} = 0{,}02229.$$

Z test prihvaćanja hipoteze H_0 glasi:
$$Z^* = \frac{\hat{p} - P_0}{Se(p)} = \frac{0{,}54 - 0{,}5}{0{,}02229} = 1{,}79$$

$Z^* = 1{,}79 > Z_{0{,}05} = 1{,}65 \Rightarrow H_1$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 95%, odnosno signifikantnosti od 5%, može se zaključiti da je promidžbena kampanja bila uspješna, tj. da je pomoću propagande povećana prodaja proizvoda A iz proizvodnje proizvođača iz Rijeke tako da ga kupuje najmanje svaki drugi kupac.

Primjer 4.2.3.

Iz pošiljke od 5.000 olovaka izabran je uzorak uz frakciju odabiranja od 4% među kojima je pronađeno 12 oštećenih olovaka. Hoće li kupac prihvatiti ovu pošiljku ako je ugovoreno da će se tolerirati najviše 3% škarta? Testirajte uz 95% pouzdanosti.

Rješenje:

Kupac će prihvatiti ponudu ako je proporcija škarta manja od pretpostavljene (P_0). Znači da se provodi jednosmjerno testiranje i to na donju granicu:
$$H_0: P \geq P_0 \qquad H_1: P < P_0.$$

Parametri:
$N = 5.000$
$$f = \frac{n}{N} = 4\% \Rightarrow n = 200$$
$m = 12$

$P_0 = 3\% = 0{,}03$
$Z_{0,05} = 1{,}65$ (uz razinu 5% signifikantnosti za jednosmjerno testiranje).

Izračun proporcije uzorka:
$$\hat{p} = \frac{m}{n} = \frac{12}{200} = 0{,}06.$$

Izračun protivne proporcije pretpostavljenoj:
$$Q_0 = 1 - P_0 = 1 - 0{,}03 = 0{,}97.$$

Izračun standardne pogreške proporcije:
$$n > 30 \Rightarrow \sqrt{\frac{P_0 \cdot Q_0}{n}} = \sqrt{\frac{0{,}03 \cdot 0{,}97}{200}} = 0{,}012.$$

Z test prihvaćanja hipoteze H_0 glasi:
$$Z^* = \frac{\hat{p} - P_0}{Se(p)} = \frac{0{,}06 - 0{,}03}{0{,}012} = 2{,}5$$

$Z^* = 2{,}5 > -Z_{0,05} = -1{,}65 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 95%, odnosno signifikantnosti od 5%, može se zaključiti da kupac neće prihvatiti prispjelu pošiljku jer je proporcija škarta viša od ugovorenih 3%.

4.3. Testiranje hipoteze o razlici aritmetičkih sredina dvaju nezavisnih osnovnih skupova

Postavlja se početna ili nulta hipoteza da su aritmetičke sredine dvaju nezavisnih osnovnih skupova $\overline{X}_1$ i $\overline{X}_2$ jednake tj. da je njihova razlika nula. Suprotna ili alternativna hipoteza pretpostavlja da razlika između aritmetičkih sredina dvaju osnovnih skupova postoji:

$H_0 : ....\overline{X}_1 - \overline{X}_2 = 0$
$H_1 : ....\overline{X}_1 - \overline{X}_2 \neq 0$

Interval prihvaćanja hipoteze H_0 glasi:

$$0 \pm Z \cdot Se(\overline{X}_1 - \overline{X}_2) \tag{4.15}$$

gdje je:

$\overline{X}_1$; $\overline{X}_2$ - aritmetičke sredine dvaju nezavisnih osnovnih skupova,

Z - ako su veličine uzoraka ($n_1 + n_2 - 2 > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako su veličine uzoraka ($n_1 + n_2 - 2 \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df = n_1 + n_2 - 2]}$),

α - razina signifikanosti ili značajnosti testa (ako nije određeno drukčije, najčešće se uzima de je $\alpha = 5\%$).

$Se(\overline{X}_1 - \overline{X}_2)$ - **standardna pogreška razlike aritmetičkih sredina** koja se računa:

- Ako su σ osnovnog skupa poznate i jednake za oba skupa:

$$Se(\overline{X}_1 - \overline{X}_2) = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \,. \tag{4.16}$$

- Ako je uzorak mali, tj. $n_1 + n_2 - 2 \leq 30$:

$$Se(\overline{X}_1 - \overline{X}_2) = \sqrt{\left(\frac{n_1 \cdot \hat{\sigma}_1^2 + n_2 \cdot \hat{\sigma}_2^2}{n_1 + n_2 - 2} \right) \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2} \right)} \,. \tag{4.17}$$

- Ako je uzorak veliki, tj. $n_1 + n_2 - 2 > 30$:

$$Se(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \tag{4.18}$$

gdje su:

σ - standardna devijacija osnovnog skupa (jednaka za oba skupa),

n_1 - veličina jednog uzorka,

n_2 - veličina drugog uzorka,

$\hat{\sigma}_1^2$ - varijanca jednog uzorka,

$\hat{\sigma}_2^2$ - varijanca drugog uzorka,

S_1^2 - nepristrana ocjena varijance jednog osnovnog skupa,

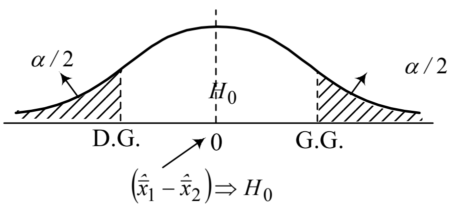
S_2^2 - nepristrana ocjena varijance drugog osnovnog skupa.

Zaključak o prihvaćanju ili odbacivanju nulte H_0 hipoteze donosi se na osnovi ($\hat{\bar{X}}_1 - \hat{\bar{X}}_2$) razlike aritmetičkih sredina iz promatranih uzoraka (prema slici 4.5).

Prema slici 4.5, ako se razlika aritmetičkih sredina iz uzorka ($\hat{\bar{X}}_1 - \hat{\bar{X}}_2$) nalazi između donje (D.G.) i gornje granice (G.G.) intervala prihvaćanja hipoteze $H_{0,}$ ta se hipoteza prihvaća kao istinita uz odgovarajuću razinu signifikantnosti testa (α). U suprotnom se navedena hipoteza odbacuje.

Slika 4.5.

Odluka o prihvaćanju hipoteza kod testiranja o razlici aritmetičkih sredina dvaju nezavisnih osnovnih skupova



Izvor: Konstrukcija autorica

Testiranje se može izvesti Z-testom (ako je: $n_1 + n_2 - 2 > 30$) ili t-testom (ako je: $n_1 + n_2 - 2 \leq 30$):

$$Z^* = \frac{\hat{\bar{X}}_1 - \hat{\bar{X}}_2}{Se(\bar{X}_1 - \bar{X}_2)}; \qquad Z_{tab\left[\frac{1-\alpha}{2}\right]} \tag{4.19}$$

gdje je:

Z^* - Z empirijska vrijednost izračunata na osnovi uzoraka,

Z_{tab} - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

$t*$ - t empirijska vrijednost izračunata na osnovi uzorka (ako je $n_1 + n_2 - 2 \leq 30$),

t_{tab} - tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n_1+n_2-2]}$).

Zaključak se donosi na način da ako je:

$|Z*| < Z_{tab}$; $\Rightarrow H_0$ odnosno $|t*| < t_{tab}$ $\Rightarrow H_0$,

dok se u suprotnom odbacuje navedena hipoteza.

Testiranje se može izvršiti i izračunavanjem granične signifikantnosti $\alpha*$ pomoću $Z*$ ili t^* (Tablica A ili B): ako je $\alpha* > 5\%$ ili $1\% \Rightarrow H_0$, dok se u suprotnom ta hipoteza odbacuje.

Primjer 4.3.1.

U trgovačkom poduzeću ispituje se efikasnost rada po trgovinama. U tu svrhu izabrano je 37 trgovaca iz trgovine A i 39 trgovaca iz trgovine B. Jednomjesečnim praćenjem prometa izabranih trgovaca utvrđeno je da je u trgovini A prosječna prodaja iznosila 57 proizvoda po trgovcu, dok je u trgovini B u prosjeku prodano 54 proizvoda po trgovcu. Varijanca trgovine A procijenjena je na 9,4, a varijanca trgovine B procijenjena je na 8,1. Na razini 1% značajnosti usporedite efikasnost po trgovinama.

Rješenje:

Testira se hipoteza o razlici aritmetičkih sredina dvaju nezavisnih osnovnih skupova:

$$H_0: \overline{X}_1 - \overline{X}_2 = 0 \qquad H_1: \overline{X}_1 - \overline{X}_2 \neq 0.$$

Parametri:

$n_1 = 37$

$n_2 = 39$

$\hat{\overline{X}}_1 = 57$

$\hat{\overline{X}}_2 = 54$

$S_1^2 = 9,4$

$S_2^2 = 8,1$

$Z_{0,005} = 2,58$ (99% pouzdanosti ili 1% signifikantnosti).

Izračun standardne pogreške razlike aritmetičkih sredina za veliki uzorak:

$$n_1 + n_2 - 2 = 37 + 39 - 2 = 74 > 30 \Rightarrow Se(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{9{,}4}{37} + \frac{8{,}1}{39}} = 0{,}6795.$$

Z test prihvaćanja hipoteze H_0 glasi:

$$Z^* = \frac{\hat{\overline{X}}_1 - \hat{\overline{X}}_2}{Se(\overline{X}_1 - \overline{X}_2)} = \frac{57 - 54}{0{,}6795} = 4{,}42$$

$Z^* = 4{,}42 > Z_{0{,}005} = 2{,}58 \Rightarrow H_1$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 99%, odnosno uz mogućnost pogreške tipa II, odbacuje se nulta hipoteza da je prosječna prodaja po trgovcu u dvije trgovine jednaka i prihvaća se alternativna hipoteza, tj. da efikasnost rada po trgovinama u tom poduzeću nije jednaka.

Primjer 4.3.2.

Udruga potrošača ispituje razinu cijena jagoda u dva grada. U gradu A izabran je uzorak od 12 prodavača, dok je u gradu B izabrano 15 prodavača kod kojih je zabilježena cijena jagoda. U prvom uzorku grada A prosječna cijena jagoda iznosila je 18,4 kn uz prosječno odstupanje od 3,5 kn, dok je u drugom uzorku u gradu B prosječna cijena bila 19,2 kn uz prosječno odstupanje od 4,7 kn. Do kojih se zaključaka može doći analizirajući prikupljene podatke s 95% pouzdanosti?

Rješenje:

Testira se hipoteza o jednakosti aritmetičkih sredina dvaju nezavisnih osnovnih skupova:

$$H_0 : \overline{X}_1 - \overline{X}_2 = 0 \qquad H_1 : \overline{X}_1 - \overline{X}_2 \neq 0.$$

Parametri:

$n_1 = 12$

$n_2 = 15$

$\hat{\overline{X}}_1 = 18{,}4$

$\hat{\bar{X}}_2 = 19{,}2$
$\hat{\sigma}_1^2 = 3{,}5$
$\hat{\sigma}_2^2 = 4{,}7.$

Izračun standardne pogreške razlike aritmetičkih sredina za mali uzorak $(n_1 + n_2 - 2 = 25 < 30)$:

$$Se(\bar{X}_1 - \bar{X}_2) = \sqrt{\left(\frac{n_1 \cdot \hat{\sigma}_1^2 + n_2 \cdot \hat{\sigma}_2^2}{n_1 + n_2 - 2}\right) \cdot \left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)} = \sqrt{\left(\frac{12 \cdot 3{,}5 + 15 \cdot 4{,}7}{12 + 15 - 2}\right) \cdot \left(\frac{12 + 15}{12 \cdot 15}\right)} =$$
0,82.

t test prihvaćanja hipoteze H_0 glasi:

$$t^* = \frac{\hat{\bar{X}}_1 - \hat{\bar{X}}_2}{Se(\bar{X}_1 - \bar{X}_2)} = \frac{18{,}4 - 19{,}2}{0{,}82} = -0{,}98$$

t_{tab} iz tablice površina studentove ili *t*-distribucije, uz 25 stupnjeva slobode $(df = n_1 + n_2 - 2)$ i 5% značajnosti $(t_{25}^{0,05}) \Rightarrow t = 2{,}060$

$|t^*| = |0{,}98| < t_{25}^{0,05} = 2{,}060 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 95%, odnosno uz mogućnost pogreške tipa I, prihvaća se pretpostavka da nema značajne razlike u prosječnoj cijeni jagoda u dva analizirana grada.

4.4. Testiranje hipoteze o razlici aritmetičkih sredina dvaju zavisnih osnovnih skupova

Kod ovog testiranja postavljaju se hipoteze i donose zaključci o njihovu prihvaćanju na jednak način **kao kod testiranja hipoteze o razlici aritmetičkih sredina dvaju nezavisnih osnovnih skupova**. Postavlja se početna ili nulta hipoteza da su aritmetičke sredine dvaju zavisnih osnovnih skupova $\bar{X}_1$ i $\bar{X}_2$ jednake tj. da je njihova razlika nula. Suprotna ili alternativna hipoteza pretpostavlja da razlika između aritmetičkih sredina dvaju osnovnih skupova postoji:

$$H_0 :.........\overline{X}_1 - \overline{X}_2 = 0$$
$$H_1 :.........\overline{X}_1 - \overline{X}_2 \neq 0.$$

Interval prihvaćanja hipoteze H_0 glasi:

$$0 \pm Z \cdot Se(\overline{X}_1 - \overline{X}_2), \tag{4.20}$$

gdje je:

$\overline{X}_1$; $\overline{X}_2$ - aritmetičke sredine dvaju nezavisnih osnovnih skupova,

Z - ako su veličine uzoraka ($n_1 + n_2 - 2 > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako su veličine uzoraka ($n_1 + n_2 - 2 \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n_1+n_2-2]}$),

α - razina signifikanosti ili značajnosti testa (ako nije određeno drukčije, najčešće se uzima de je $\alpha = 5\%$).

$Se(\overline{X}_1 - \overline{X}_2)$ - **standardna pogreška razlike aritmetičkih sredina** koja **se računa**:

$$Se(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} - 2 \cdot r_{1,2} \cdot \sqrt{\frac{S_1^2}{n_1}} \cdot \sqrt{\frac{S_2^2}{n_2}}}, \tag{4.21}$$

gdje je:

$r_{1,2}$ - Pearsonov koeficijent linearne korelacije između dvaju mjerenja iste slučajne varijable na istom uzorku (zavisni skupovi),

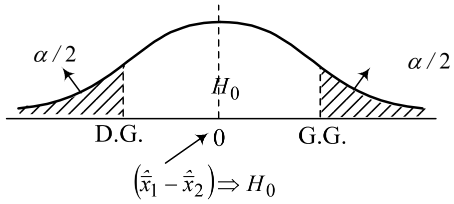
$\sqrt{\frac{S_1^2}{n_1}}$ - standardna pogreška aritmetičke sredine jednog uzorka ($Se(\overline{X}_1)$),

$\sqrt{\frac{S_2^2}{n_2}}$ - standardna pogreška aritmetičke sredine drugog uzorka ($Se(\overline{X}_2)$).

Zaključak o prihvaćanju ili odbacivanju nulte H_0 hipoteze donosi se na osnovi ($\hat{\bar{X}}_1 - \hat{\bar{X}}_2$) razlike aritmetičkih sredina iz promatranih zavisnih uzoraka (prema slici 4.6.).

Slika 4.6.

Odluka o prihvaćanju hipoteza kod testiranja o razlici aritmetičkih sredina dvaju zavisnih osnovnih skupova



Izvor: Konstrukcija autorica

Prema slici 4.6., ako se razlika aritmetičkih sredina iz uzorka ($\hat{\bar{X}}_1 - \hat{\bar{X}}_2$) nalazi između donje (D.G.) i gornje granice (G.G.) intervala prihvaćanja hipoteze H_0, ta se hipoteza prihvaća kao istinita uz odgovarajuću razinu signifikantnosti testa (α). U suprotnom se navedena hipoteza odbacuje.

Testiranje se može izvesti Z-testom (ako je: $n_1 + n_2 - 2 > 30$) ili t-testom (ako je: $n_1 + n_2 - 2 \leq 30$):

$$Z^* = \frac{\hat{\bar{X}}_1 - \hat{\bar{X}}_2}{Se(\bar{X}_1 - \bar{X}_2)}; \qquad Z_{tab\left[\frac{1-\alpha}{2}\right]} \tag{4.22}$$

gdje je:

Z^* - Z empirijska vrijednost izračunata na osnovi uzoraka,

Z_{tab} - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t^* - t empirijska vrijednost izračunata na osnovi uzorka (ako je $n_1 + n_2 - 2 \leq 30$),

t_{tab} - tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df = n_1 + n_2 - 2]}$).

Zaključak se donosi na način da ako je:

$|Z^*| < Z_{tab} \quad \Rightarrow H_0$; odnosno $|t^*| < t_{tab} \quad \Rightarrow H_0$, dok se u suprotnom odbacuje navedena hipoteza.

Testiranje se može izvršiti i izračunavanjem granične signifikantnosti $\alpha*$ pomoću $Z*$ ili t^* (Tablica A ili B): ako je $\alpha* > 5\%\,\text{ili}\,1\% \Rightarrow H_0$, dok se u suprotnom ta hipoteza odbacuje.

4.5. Testiranje hipoteze o razlici proporcija (relativnih frekvencija) dvaju nezavisnih osnovnih skupova

Postavlja se početna ili nulta hipoteza da su proporcije (relativne frekvencije) dvaju nezavisnih osnovnih skupova P_1 i P_2 jednake tj. da je njihova razlika nula. Suprotna ili alternativna hipoteza pretpostavlja da razlika između proporcija (relativnih frekvencija) dvaju osnovnih skupova postoji:

$$H_0 : ....P_1 - P_2 = 0$$

$$H_1 : ....P_1 - P_2 \neq 0.$$

Interval prihvaćanja hipoteze H_0 glasi:

$$0 \pm Z \cdot Se(P_1 - P_2) \tag{4.23}$$

gdje je:

P_1; P_2 - proporcije (relativne frekvencije) dvaju nezavisnih osnovnih skupova,

Z - ako su veličine uzoraka ($n_1 + n_2 - 2 > 30$) računa se vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t - ako su veličine uzoraka ($n_1 + n_2 - 2 \leq 30$) umjesto (Z) računa se vrijednost iz tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n-1]}$),

α - razina signifikanosti ili značajnosti testa (ako nije određeno drukčije, najčešće se uzima de je $\alpha = 5\%$),

$Se(P_1 - P_2)$ **- standardna pogreška razlike proporcija (relativnih frekvencija)** koja se računa:

$$Se(P_1 - P_2) = \sqrt{\hat{\bar{p}} \cdot \left(1 - \hat{\bar{p}}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \tag{4.24}$$

gdje je:

$$\hat{\bar{p}} = \frac{m_1 + m_2}{n_1 + n_2}, \tag{4.25}$$

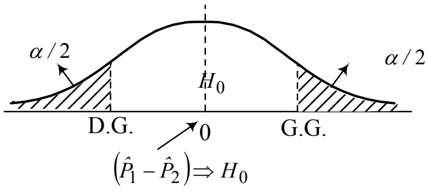
$\hat{\bar{p}}$ - prosječna proporcija (relativna frekvencija) za oba uzorka zajedno,

n_1 - veličina jednog uzorka,

n_2 - veličina drugog uzorka.

Slika 4.7.

Odluka o prihvaćanju hipoteza kod testiranja o razlici proporcija dvaju nezavisnih osnovnih skupova



Izvor: Konstrukcija autorica

Zaključak o prihvaćanju ili odbacivanju nulte H_0 hipoteze donosi se na osnovi ($\hat{P}_1 - \hat{P}_2$) razlike proporcija (relativnih frekvencija) iz promatranih uzoraka (prema slici 4.7.). Ako se razlika proporcija (relativnih frekvencija) iz uzorka ($\hat{P}_1 - \hat{P}_2$) nalazi između donje (D.G.) i gornje granice (G.G.) intervala prihvaćanja hipoteze $H_{0,}$ ta se hipoteza prihvaća kao istinita uz odgovarajuću razinu signifikantnosti testa (α). U suprotnom se navedena hipoteza odbacuje.

Testiranje se može izvesti Z-testom (ako je: $n_1 + n_2 - 2 > 30$) ili t-testom (ako je: $n_1 + n_2 - 2 \le 30$):

$$Z^* = \frac{\hat{P}_1 - \hat{P}_2}{Se(P_1 - P_2)};\ Z_{tab\left[\frac{1-\alpha}{2}\right]}, \tag{4.26}$$

gdje je:

Z^* - Z empirijska vrijednost izračunata na osnovi uzoraka,

Z_{tab} - vrijednost iz tablice površina normalne distribucije (pomoću $Z_{\frac{1-\alpha}{2}}$),

t^* - t empirijska vrijednost izračunata na osnovi uzorka (ako je $n_1 + n_2 - 2 \leq 30$),

t_{tab} - tablice površina studentove ili t-distribucije (pomoću $t^{\alpha/2}_{[df=n_1+n_2-2]}$).

Zaključak se donosi na način da ako je:

$|Z^*| < Z_{tab} \quad \Rightarrow H_0$; odnosno $|t^*| < t_{tab} \quad \Rightarrow H_0$,

dok se u suprotnom odbacuje navedena hipoteza.

Testiranje se može izvršiti i izračunavanjem granične signifikantnosti α^* pomoću Z^* ili t^* (Tablica A ili B): ako je $\alpha^* > 5\% \Rightarrow H_0$, dok se u suprotnom ta hipoteza odbacuje.

Primjer 4.5.1.

U Rijeci je anketirano 220 domaćinstava i proporcija njihovih izdataka za hranu i piće iznosila je 45,3%. U Splitu je anketirano 285 domaćinstava, a proporcija njihovih izdataka za hranu i piće bila je 49,2%. Je proporcija izdataka za hranu i piće u dva grada jednaka? Testirate uz 95% pouzdanosti.

Rješenje:

Testira se hipoteza o razlici proporcija dvaju osnovnih skupova:

$$H_0 : P_1 - P_2 = 0 \qquad H_1 : P_1 - P_2 \neq 0.$$

Parametri:

$n_1 = 220$

$n_2 = 285$

$\hat{p}_1 = 45,3\% = 0,453$

$\hat{p}_2 = 49,2\% = 0,492$

$Z_{0,025} = 1,96$ (95% pouzdanosti ili 5% signifikantnosti).

Izračun prosječne proporcije za oba uzorka zajedno:

$$\hat{\hat{p}} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2} = \frac{0,453 \cdot 220 + 0,492 \cdot 285}{220 + 285} = 0,475.$$

Izračun standardne pogreške razlike proporcija:

$$Se(P_1 - P_2) = \sqrt{\hat{\hat{p}}(1 - \hat{\hat{p}})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0,475 \cdot 0,525 \cdot \left(\frac{1}{220} + \frac{1}{285}\right)} = 0,0448.$$

Z test prihvaćanja hipoteze H_0 glasi:

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2}{Se(P_1 - P_2)} = \frac{0,453 - 0,492}{0,0448} = -0,87$$

$|Z^*| = |0,87| < Z_{0,025} = 1,96 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini pouzdanosti od 95%, odnosno uz mogućnost pogreške tipa I, prihvaća se pretpostavka da je proporcija izdataka za hranu i piće u Rijeci i Splitu jednaka.

4.6. Hi - kvadrat test

Hi-kvadrat test ne pretpostavlja oblik distribucije i svrstava se u neparametrijske testove.

Zasniva se na rasporedu frekvencija unutar tablice kontigence (tablica po kojoj se izračunava empirijska vrijednost Hi-kvadrata). To znači da je zbroj originalnih apsolutnih frekvencija i očekivanih teorijskih frekvencija (koje se izračunavaju uz pretpostavku nulte hipoteze H_0) uvijek jednak, a pri donošenju zaključka bitan je njihov raspored u distribuciji. Ako je razlika originalnih i teorijskih frekvencija velika, početna hipoteza H_0 se odbacuje, a ako njihova razlika statistički nije značajna ta se hipoteza prihvaća kao istinita. Stoga se Hi-kvadrat test u literaturi naziva "frequency based statistic".

Prema časopisu ***Science*** ovaj test nalazi se između dvadeset najvažnijih znanstvenih otkrića dvadesetog stoljeća.

4.6.1. Testiranje hipoteze o jednakosti proporcija triju ili više populacija

Pomoću Hi-kvadrat testa može se testirati **hipoteza o jednakosti proporcija triju ili više populacija**. Ako su sve proporcije u jednoj distribuciji jednake radi se o jednolikoj razdiobi, pa ovaj test u stvari **predstavlja testiranje hipoteze ima li zadana distribucija oblik jednolike distribucije**.

Postavljaju se hipoteze za testiranje **jednolike distribucije**:

$H_0 :........P_1 = P_2 = ... = P_k = P$

$H_1 :........\exists P_i \neq P$ (barem 1 proporcija različita od *P*).

Empirijska vrijednost Hi-kvadrat testa za ovo testiranje je:

$$\chi^2 * = \sum_{i=1}^{k} \frac{(m_i - e_i)^2}{e_i}, \qquad (4.27)$$

gdje su:

m_i - originalne frekvencije (iz distribucije uzorka),

e_i - teorijske (očekivane, eng. expected) frekvencije koje se izračunavaju pod pretpostavkom početne ili nulte hipoteze, tj. da zadana distribucija ima jednoliki oblik.

Teorijske frekvencije su:

$$e_i = p(x_i) \cdot n_i, \qquad (4.28)$$

gdje je:

n_i - ukupni ishodi, tj. veličine pojedinih uzoraka,

$p(x_i)$ - vjerojatnost da slučajna varijabla X poprimi vrijednost x_i prema jednolikom zakonu, koji se pretpostavlja u nultoj H_0 hipotezi.

Vjerojatnost prema **jednolikoj distribuciji**, tj. zajednička, jednolika proporcija na osnovi uzorka je:

$$p(x_i) = \frac{\sum_{i=1}^{k} m_i}{\sum_{i=1}^{k} n_i}, \qquad (4.29)$$

gdje je:

$\sum_{i=1}^{k} m_i$ - zbroj povoljnih ishoda (iz distribucije uzorka),

$\sum_{i=1}^{k} n_i$ - zbroj ukupnih ishoda (iz distribucije uzorka).

Tablična vrijednost testa se traži iz tablica Hi-kvadrat distribucije (Tablice C1 i C2): $\chi^2_{tab}[\alpha, df]$, uz odgovarajuću razinu signifikantnosti α i stupnjeve slobode df, koji se za jednoliku distribuciju računaju: $df = k - 1$,
gdje "k" predstavlja broj frekvencija.

Zaključak se donosi na način da se usporedi Hi-kvadrat empirijska i tablična vrijednost:

$\chi^2 * < \chi^2_{tab}$ $\Rightarrow H_0$; što znači da zadana distribucija ima oblik jednolike distribucije, dok se u suprotnom se ta početna hipoteza odbacuje.

Testiranje se može izvršiti i izračunavanjem granične signifikantnosti $\alpha *$ pomoću $\chi *$ (Tablice C1 i C2): ako je $\alpha * > 5\%$ ili $1\% \Rightarrow H_0$, dok se u suprotnom ta hipoteza odbacuje.

Primjer 4.6.1.

Želi se provjeriti tvrdnja da je udio škarta kod tri proizvođača jednak. Od svakog proizvođača uzet je uzorak, a rezultati su prikazani u tablici:

Proizvođač	Broj proizvoda (n_i)	Oštećeni proizvodi (m_i)
A	98	7
B	86	6
C	103	9

Testirate gornju tvrdnju uz 1% značajnosti.

Rješenje:

Testira se hipoteza o razlici aritmetičkih sredina triju osnovnih skupova, tj. testira se hipoteza ima li zadana distribucija oblik jednolike distribucije:

$$H_0: P_1 = P_2 = P_3 = P \qquad H_1: \exists P_i \neq P.$$

Empirijska vrijednost Hi-kvadrat testa za ovo testiranje je:

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

Teorijske frekvencije su: $e_i = p(x_i) \cdot n_i$.

Vjerojatnost, tj. zajednička proporcija na osnovi uzorka jednaka je:

$$p(x_i) = \frac{\sum_{i=i}^{k} m_i}{\sum_{i=1}^{k} n_i} = \frac{7+6+9}{98+86+103} = 0{,}07666.$$

Izračun Hi-kvadrat testa prikazan je u tablici:

Proizvođač	Broj proizvoda (n_i)	Oštećeni proizvodi (m_i)	Teorijske frekvencije $e_i = p(x_i) \cdot n_i$	$\frac{(m_i - e_i)^2}{e_i}$
A	98	7	7,51	0,0346
B	86	6	6,59	0,0528
C	103	9	7,90	0,1532
Ukupno	287	22	22,0	0,2406

Znači da je $\chi^2 * = 0{,}2406$

Tablična vrijednost, uz 1% značajnosti i 2 stupnja slobode (*df = k-1*), je $\chi^2 = 9{,}21$

$\chi^2 * = 0{,}24065 < \chi^2 = 9{,}21 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:

Na razini pouzdanosti od 99%, odnosno uz mogućnost pogreške tipa II, prihvaća se nulta hipoteza da je udio škarta kod sva tri ispitivana proizvođača jednak.

Primjer 4.6.2.

Regionalni distributer klima sustava podijelio je regiju na 4 područja. Pretpostavlja se da će interes za ugradnjom opreme na svim područjima biti podjednak. Iz podataka poduzeća za prethodnu godinu slučajno je izabrano 40 ugradnji, a njihova teritorijalna raspodjela prikazana je u tablici:

	Područje				Ukupno
	A	B	C	D	
Broj instaliranih klima (m_i)	9	11	12	8	40

Testirate pretpostavku uz 5% značajnosti.

Rješenje:

Testira se hipoteza ima li zadana distribucija oblik jednolike distribucije:

$$H_0 : P_1 = P_2 = P_3 = P_4 = P \qquad H_1 : \exists P_i \neq P .$$

Empirijska vrijednost Hi-kvadrat testa je:

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

Izračun Hi-kvadrat testa prikazan je u tablici:

Područje	Broj instaliranih klima uređaja (m_i)	Teorijske frekvencije (e_i) (jednolika distribucija)	$\frac{(m_i - e_i)^2}{e_i}$
A	9	40:4 = 10	0,1
B	11	40:4 = 10	0,1
C	12	40:4 = 10	0,4
D	8	40:4 = 10	0,4
Ukupno	40	40	1,0

Znači da je $\chi^2 * = 1,0$

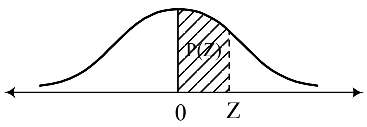
Tablična vrijednost, uz 5% značajnosti i 3 stupnja slobode (*df = k-1*), je $\chi^2 = 7,81$

$\chi^2{*} = 1,0 < \chi^2 = 7,81 \Rightarrow H_0$ se prihvaća kao istinita.

Zaključak:
Na razini 5% signifikantnosti prihvaća se H_0 kao moguća, tj. prihvaća se moguća pretpostavka da interes za ugradnjom klima uređaja na svim područjima podjednak

5. TABLICE ODABRANIH STATISTIČKIH DISTRIBUCIJA

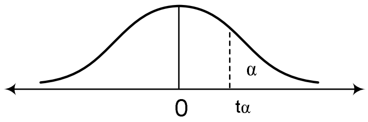
A. Površine ispod normalne krivulje



	0	1	2	3	4	5	6	7	8	9
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34849	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42785	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49896	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
4,0	499968	499970	499971	499972	499973	499974	499975	499976	499977	499978
4,5	499997	499997	499997	499997	499997	499997	499997	499998	499998	499998
5,0	500000	500000	500000	500000	500000	500000	500000	500000	500000	500000

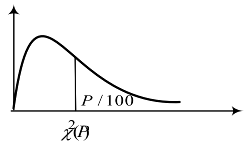
Napomena: Ispred svakog broja u polju tablice je decimalna točka.

B. Kritične vrijednosti t, Studentove distribucije



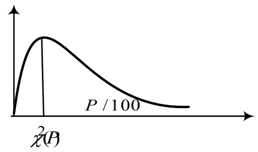
df/P	40	30	25	20	15	10	5	2,5	1	0,5	0,1	0,05
1	0,325	0,727	1,000	1,376	1,963	3,078	6,314	12,71	31,82	31,83	318,3	636,6
2	0,289	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	22,33	31,60
3	0,277	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	10,21	12,92
4	0,271	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	5,894	6,869
6	0,265	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,262	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,257	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,256	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,256	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,421	3,689
28	0,256	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,396	3,660
30	0,256	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	0,255	0,530	0,682	0,853	1,054	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	0,255	0,529	0,682	0,852	1,052	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	0,255	0,529	0,681	0,852	1,052	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	0,255	0,529	0,681	0,851	1,051	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	0,255	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678	3,261	3,496
60	0,254	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,232	3,460
100	0,254	0,526	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626	3,174	3,390

C1. Kritične vrijednosti Hi-kvadrat distribucije (za $P \leq 0{,}05$)



df	0,05	0,1	0,5	1	2	5	10	20	30	40	50
1	12,12	10,83	7,88	6,63	5,02	3,84	2,706	1,642	1,0742	0,7083	0,4549
2	15,20	13,82	10,60	9,21	7,38	5,99	4,605	3,219	2,4079	1,8326	1,3863
3	17,73	16,27	12,84	11,34	9,35	7,81	6,251	4,642	3,665	2,946	2,366
4	20,00	18,47	14,86	13,28	11,14	9,49	7,779	5,989	4,878	4,045	3,357
5	22,11	20,51	16,75	15,09	12,83	11,07	9,236	7,289	6,064	5,132	4,351
6	24,10	22,46	18,55	16,81	14,45	12,59	10,645	8,558	7,231	6,211	5,348
7	26,02	24,32	20,28	18,48	16,01	14,07	12,017	9,803	8,383	7,283	6,346
8	27,87	26,12	21,95	20,09	17,53	15,51	13,362	11,030	9,524	8,351	7,344
9	29,67	27,88	23,59	21,67	19,02	16,92	14,684	12,242	10,656	9,414	8,343
10	31,42	29,59	25,19	23,21	20,48	18,31	15,987	13,442	11,781	10,473	9,342
11	33,14	31,26	26,76	24,73	21,92	19,68	17,28	14,63	12,90	11,53	10,34
12	34,82	32,91	28,30	26,22	23,34	21,03	18,55	15,81	14,01	12,58	11,34
13	36,48	34,53	29,82	27,69	24,74	22,36	19,81	16,98	15,12	13,64	12,34
14	38,11	36,12	31,32	29,14	26,12	23,68	21,06	18,15	16,22	14,69	13,34
15	39,72	37,70	32,80	30,58	27,49	25,00	22,31	19,31	17,32	15,73	14,34
16	41,31	39,25	34,27	32,00	28,85	26,30	23,54	20,47	18,42	16,78	15,34
17	42,88	40,79	35,72	33,41	30,19	27,59	24,77	21,61	19,51	17,82	16,34
18	44,43	42,31	37,16	34,81	31,53	28,87	25,99	22,76	20,60	18,87	17,34
19	45,97	43,82	38,58	36,19	32,85	30,14	27,20	23,90	21,69	19,91	18,34
20	47,50	45,31	40,00	37,57	34,17	31,41	28,41	25,04	22,77	20,95	19,34
21	49,01	46,80	41,40	38,93	35,48	32,67	29,62	26,17	23,86	21,99	20,34
22	50,51	48,27	42,80	40,29	36,78	33,92	30,81	27,30	24,94	23,03	21,34
23	52,00	49,73	44,18	41,64	38,08	35,17	32,01	28,43	26,02	24,07	22,34
24	53,48	51,18	45,56	42,98	39,36	36,42	33,20	29,55	27,10	25,11	23,34
25	54,95	52,62	46,93	44,31	40,65	37,65	34,38	30,68	28,17	26,14	24,34
26	56,41	54,05	48,29	45,64	41,92	38,89	35,56	31,79	29,25	27,18	25,34
27	57,86	55,48	49,65	46,96	43,19	40,11	36,74	32,91	30,32	28,21	26,34
28	59,30	56,89	50,99	48,28	44,46	41,34	37,92	34,03	31,39	29,25	27,34
29	60,73	58,30	52,34	49,59	45,72	42,56	39,09	35,14	32,46	30,28	28,34
30	62,16	59,70	53,67	50,89	46,98	43,77	40,26	36,25	33,53	31,32	29,34
40	76,10	73,40	66,77	63,69	59,34	55,76	51,81	47,27	44,16	41,62	39,34
50	89,56	86,66	79,49	76,15	71,42	67,50	63,17	58,16	54,72	51,89	49,33
60	102,70	99,61	91,95	88,38	83,30	79,08	74,40	68,97	65,23	62,13	59,33
100	153,16	149,45	140,17	135,81	129,56	124,34	118,50	111,67	106,91	102,95	99,33

C2. Kritične vrijednosti Hi-kvadrat distribucije (za $P > 0{,}05$)



df/P	60	70	75	80	90	95	97,5	99	99,5
1	0,275	0,148	0,102	0,064	0,016	0,004	0,00098	0,00016	0,00004
2	1,022	0,713	0,575	0,446	0,211	0,103	0,05064	0,02010	0,01002
3	1,869	1,424	1,213	1,005	0,584	0,352	0,216	0,115	0,072
4	2,753	2,195	1,923	1,649	1,064	0,711	0,484	0,297	0,207
5	3,656	3,000	2,675	2,343	1,610	1,145	0,831	0,554	0,412
6	4,570	3,828	3,455	3,070	2,204	1,635	1,237	0,872	0,676
7	5,493	4,671	4,255	3,822	2,833	2,167	1,690	1,239	0,989
8	6,423	5,527	5,071	4,594	3,490	2,733	2,180	1,647	1,344
9	7,357	6,393	5,899	5,380	4,168	3,325	2,700	2,088	1,735
10	8,295	7,267	6,737	6,179	4,865	3,940	3,247	2,558	2,156
11	9,237	8,148	7,584	6,989	5,578	4,575	3,816	3,053	2,603
12	10,182	9,034	8,438	7,807	6,304	5,226	4,404	3,571	3,074
13	11,129	9,926	9,299	8,634	7,041	5,892	5,009	4,107	3,565
14	12,078	10,821	10,165	9,467	7,790	6,571	5,629	4,660	4,075
15	13,030	11,721	11,037	10,307	8,547	7,261	6,262	5,229	4,601
16	13,983	12,624	11,912	11,152	9,312	7,962	6,908	5,812	5,142
17	14,937	13,531	12,792	12,002	10,085	8,672	7,564	6,408	5,697
18	15,893	14,440	13,675	12,857	10,865	9,390	8,231	7,015	6,265
19	16,850	15,352	14,562	13,716	11,651	10,117	8,907	7,633	6,844
20	17,809	16,266	15,452	14,578	12,443	10,851	9,591	8,260	7,434
21	18,768	17,182	16,344	15,445	13,240	11,591	10,283	8,897	8,034
22	19,729	18,101	17,240	16,314	14,041	12,338	10,982	9,542	8,643
23	20,690	19,021	18,137	17,187	14,848	13,091	11,689	10,196	9,260
24	21,652	19,943	19,037	18,062	15,659	13,848	12,401	10,856	9,886
25	22,616	20,867	19,939	18,940	16,473	14,611	13,120	11,524	10,520
26	23,579	21,792	20,843	19,820	17,292	15,379	13,844	12,198	11,160
27	24,544	22,719	21,749	20,703	18,114	16,151	14,573	12,878	11,808
28	25,509	23,647	22,657	21,588	18,939	16,928	15,308	13,565	12,461
29	26,475	24,577	23,567	22,475	19,768	17,708	16,047	14,256	13,121
30	27,442	25,508	24,478	23,364	20,599	18,493	16,791	14,953	13,787
40	37,134	34,872	33,660	32,345	29,051	26,509	24,433	22,164	20,707
50	46,864	44,313	42,942	41,449	37,689	34,764	32,357	29,707	27,991
60	56,620	53,809	52,294	50,641	46,459	43,188	40,482	37,485	35,534
100	95,808	92,129	90,133	87,945	82,358	77,929	74,222	70,065	67,328

6. LITERATURA

1. Aczel, A.D. (1999), Complete Business Statistics, 4th edt., Irving McGraw-Hill, NewYork.
2. Allen, M. V. and Myddelton, D. R. (1992), Esential Management Accounting, Prentice Hall, London.
3. Anderson, D. R., et al. (1981), Statistics for Business and Economics, 2nd edt., West, St. Paul.
4. Anderson, D. R., et al. (1998), Quantitative Methods for Business, 7th edt., South-Western College Publishing.
5. Anderson, R. E., Black, W., Hair, J. F. and Tatham, R. L. (1998), Multivariate Data Analysis, Prentice Hall.
6. Anderson, T. W. (2003), An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics), Wiley-Interscience.
7. Arnold, J. A. and Hope, A. J. B. (1990), Accounting for Management Decisions, Prentice Hall, London.
8. Babbie, E. Halley, F. Zaino, J. (2000), Adventures in Social Research, Data Analysis Using SPSS^{X} for Windows 95./98., Pine Forge Press, USA.
9. Basilevsky, A. T. (1994), Statistical Factor Analysis and Related Methods, Theory and Applications, John Wiley an Sons Inc..
10. Bešter, M.i Bregar, L. (1986), Ekonomska statistika, Ekonomska fakulteta, Ljubljana.
11. Biljan-August, M, Pivac, S. and Štambuk, A. (2006), Upotreba statistike u ekonomiji, Sveučilište u Rijeci, Ekonomski fakultet Rijeka, (www.efri.hr).
12. Blejec, M. (1976), Statističke metode za ekonomiste, Ekonomska fakulteta, Ljubljana.
13. Brooks, C. (1997), GARCH Modelling in Finance, A Review of the Software Options, Economics Journal, 107 (443), pp. 1271-1276.
14. Brooks, C. (2002), Introductory econometrics for finance, University Press, Cambridge.
15. Canavos, G. C. and Miller, D. M. (1993), An Introduction to Modern Business Statistics, Belmont, Wadsworth.

16. Charniak, E. (1993), Statistical Language Learning, A Bradford Book, The MIT Press, Cambridge and Massachusetts.
17. Chiang, A. C. (1994), Osnovne metode matematičke ekonomije, Mate, Zagreb.
18. Cochran, W. G. (1997), Sampling Techniques, Wiley, New York.
19. Conover, W. J. (1980), Practical Nonparametric Statistics, Wiley, New York.
20. Dillon, W. R. and Goldstein, M. (1984), Multivariate Analysis, Methods and Applications, Wiley, New York.
21. Dodge, M., Kinata, C. i Stison, C. (1997), Kako koristiti Microsoft Excel 97, Znak, Zagreb.
22. Everit, B. S. (1996), Making Sense of Statistics in Psychology, Oxford University Press, Oxford.
23. Feller, W. (1971), An Introduction to Probability Theory and its Applications, Wiley, New York, Vol. I and Vol. II.
24. Fisher, R. A. (1950), Statistical Methods for Research Workers, 11th edt., Oliver and Boyd, Edinburgh.
25. Fisher, R. A. (1993), Statistical Methods, Eksperimental Design and Scientific Inference, Oxford University Press, Oxford.
26. Frye, C. (2003), Microsoft Excel 2002 Korak po korak, Algoritam, Zagreb.
27. Fulton, J. (1997), Vodič kroz Excel 97, Znak, Zagreb.
28. Hadživuković, S. (1975), Tehnika metoda uzorka, Naučna knjiga, Beograd.
29. Halmi, A. (2003), Multivarijatna analiza u društvenim znanostima, Alineja, Zagreb.
30. Heij, C. et al. (2004), Econometrisc Methods with Applications in Business and Economics, Oxford University Press, New York.
31. Horvatić, K. (2004), Linearna algebra, Golden marketing-Tehnička knjiga, Zagreb.
32. Johnson, R. A. and Wichern, D. W. (2002), Applied Multivariate Statistical Analysis, Prentice Hall, London.
33. Jolliffe, I.T. (2002), Principal Component Analysis, Berlin, Springer, New York.
34. Kaplan, R. S. (1984), The Evoulation of Management Accounting, The Accounting Review, 59 (3), pp. 95-101.

35. Karatzas, I. and Shreve, S E. (2001), Methods of Mathematical Finance, Berlin, Springer, New York.
36. Kendall, M. and Stuart, A. (1973), The Advanced Theory of Statistics, Griffin, London, Vol. I.
37. Kendall, M. and Stuart, A. (1974), The Advanced Theory of Statistics,. Griffin, London, Vol. II.
38. Kendall, M. and Stuart, A. (1976), The Advanced Theory of Statistics,. Griffin, London, Vol. III.
39. Kish, L. (1965), Survey Sampling, Wiley, New York.
40. Kmenta, J. (1997), Počela ekonometrije, MATE d.o.o., Zagreb.
41. Kolesarić, V. i Petz. B. (2003), Statistički rječnik, Tumač statističkih pojmova, Naklada Slap, Zagreb.
42. de Levie, R. (2004), Advanced Excel for Scientific Data Analysis, Oxford University Press., Oxford.
43. Lucey, T. (1989), Management Information Systems, DP Publication LTD, London.
44. Martić, Lj. (1986), Mjere nejednakosti i siromaštva, Birotehnika, Zagreb.
45. Mason, D. R. and Lind, D. A.. (1993), Statistical Techniques in Business and Economics, Boston, IRVIN Publishing, Masachussetts.
46. Metre, J. G. and Gilbreath, G. H. (1983), Statistics for Business and Economics, Plano, Business Publications, Texas.
47. Milton, J. S. et al. (1986), Introduction to Statistics, D. C. Heath,, Lexington.
48. Momirović, K. (1998), Uvod u analizu nominalnih varijabli, Metodološke sveske 2, JUS, Ljubljana.
49. Mood, A. M. and Graybill, F. P. (1963), Introduction to the Theory of Statistics, McGraw-Hill, New York.
50. Mott, G. (1991), Management Accounting for Decision Makers, Pitman Publishing. London.
51. Neter, J. et al. (1993), Applied Statistics, Allyn and Bacon, Boston, 4th Edt.
52. Newbold, P. (1991), Statistics for Business and Economics, Prentice-Hall, Englewood Cliffs.
53. Pauše, Ž. (1993), Uvod u matematičku statistiku, Školska knjiga, .Zagreb.

54. Pauše, Ž. (2003), Vjerojatnost, informacija, stohastički procesi, Školska knjiga, Zagreb.
55. Pavlić, I. (1971), Statistička teorija i primjena, Tehnička knjiga, Zagreb.
56. Petz, B. (2004), Osnovne statističke metode za nematematičare, Naklada Slap, Zagreb.
57. Pivac, S. and Jurun, E. (2005), Parameter Estimation in Excel, Proceedings of the 28th Internationl Convention MIPRO 2005, Computers in Education, Opatija, pp. 168-173.
58. Pivac, S. and Rozga, A. (2006), Statistika za sociološka istraživanja, Sveučilište u Splitu, Filozofski fakultet, Split.
59. Pivac S., Bodrožić I. and Jurun E. (2007), Chi-Square Versus Proportions Testing - Case Study on Tradition in Croatian Brand, Proceedings of the 9th International Symposium on Operational Research SOR'07, Slovenia, pp. 415-421.
60. Pivac, S. and Rozga, A. (2008), Statističke analize socioloških istraživanja, Redak, Split.
61. Pivac, S. and Šego, B. (2005), Statistika, udžbenik sa zbirkom zadataka za IV. razred srednje Ekonomske škole, Alkascript., Zagreb.
62. Rozga, A. (2003), Statistika za ekonomiste, Sveučilište u Splitu, Ekonomski fakultet Split.
63. Rozga, A. i Grčić, B. (2003), Poslovna statistika, Ekonomski fakultet Split.
64. Seddighi, H. R., Lawler, K. A. and Katos, A. V. (2006), Econometrics, A practical approach, Routledge, London and New York.
65. Seplaki, L. (1991), Atorneys' Dictionary and Handbook of Economics and Statistics, New York, Professional Horizons Press.
66. Serdar, V. i Šošić, I. (1994), Uvod u statistiku, Školska knjiga, Zagreb.
67. Siegel, A. F. (1994), Practical Business Statistics. Boston, IRVIN Publishing, Masachussetts.
68. Studenmund, A. H. (2006), Using Econometrics, A Practical Guide. Boston, Pearson International Edition, New York.
69. Šošić, I. (1983), Metode statističke analize, Sveučilišna naklada Liber, Zagreb.
70. Šošić, I. (1985), Zbirka zadataka iz osnova statistike, Sveučilišna naklada Liber, Zagreb.

71. Šošić, I. and Serdar, V. (2002), Uvod u statistiku, Školska knjiga, XII. izdanje, Zagreb.

72. Šošić, I. (2004), Primijenjena statistika, Školska knjiga, Zagreb.

73. Tenjović, L. (2002), Statistika u psihologiji, Centar za primenjenu psihologiju, Beograd.

74. Terrell, D. (1992), Business Statistics for Management and Economics, Hoghton Mifflin Company, Boston.

75. Vujković, T. (1976), Ekonometrijske metode i tehnike, Informator, Zagreb.

76. Žarković, S. S. (1965), Sampling Methods and Censuses, FAO, Rim.

77. Weisberg, H.F., Krosnick, J.A. and Bowen, B.D. (2005), Survey research, Polling and Data Analysis, Ohio State University. Assessment Systems Corporation, 3rd edt., USA.

78. Whigham, D. (1998), Quantitative Business Methods Using Excel, Oxford University Press.

79. Wonnacott, T. H. and Wonnacott, R. J. (1990), Introductory Statistics for Business and Economics, 4th edt., Wiley, New York.

80. The Central Bureau of Statistics. Republic of Croatia, URL: http:// www.dzs.hr

PRIVITAK PRIMJENA PROGRAMSKOG PAKETA *STATISTICA* U STATISTIČKOJ ANALIZI

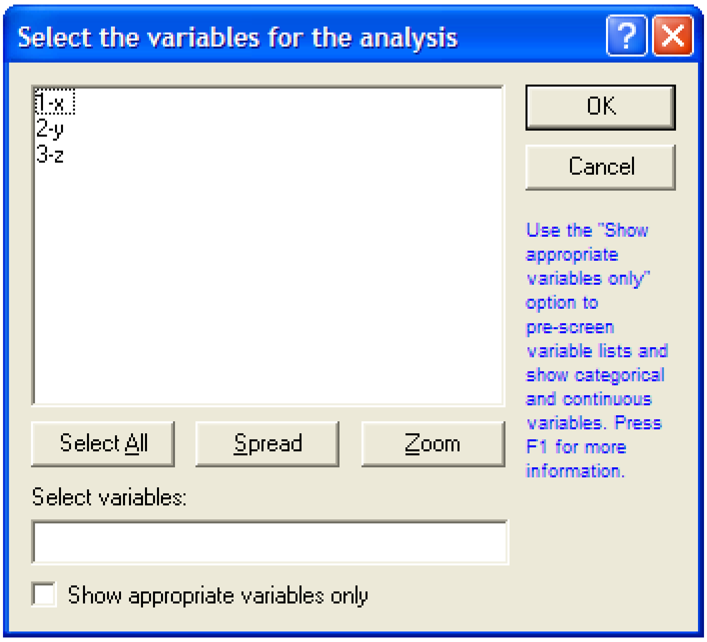
1. Uvod

S razvojem računala, statistički programski paketi postali su osnovni alat statističke analize. Neki od programa koji su namijenjeni statističkoj obradi su *Statistica*, *SPSS*, *SAS* i razni drugi programski paketi. U nastavku je objašnjena uporaba programskog sustava *Statistica*.

Većina zadataka u *Statistici* može se riješiti na više načina, a naveden je u pravilu najjednostavniji način. U zadacima u kojima rješenje ovisi o razini značajnosti ostavljena je razina signifikantnosti 0,05, koja je postavljena kao pretpostavka u *Statistici*. Vrijednost je moguće promijeniti.

U *Statistici* se zadaci rješavaju otvaranjem odgovarajućih prozora u kojima se izabiru potrebne varijable i izračuni. Kada se u prozoru označi sve potrebno izabere se ponuđena tipka, u pravilu **SUMMARY** ili **OK** nakon čega se dobiju rezultati ili se otvara novi prozor.

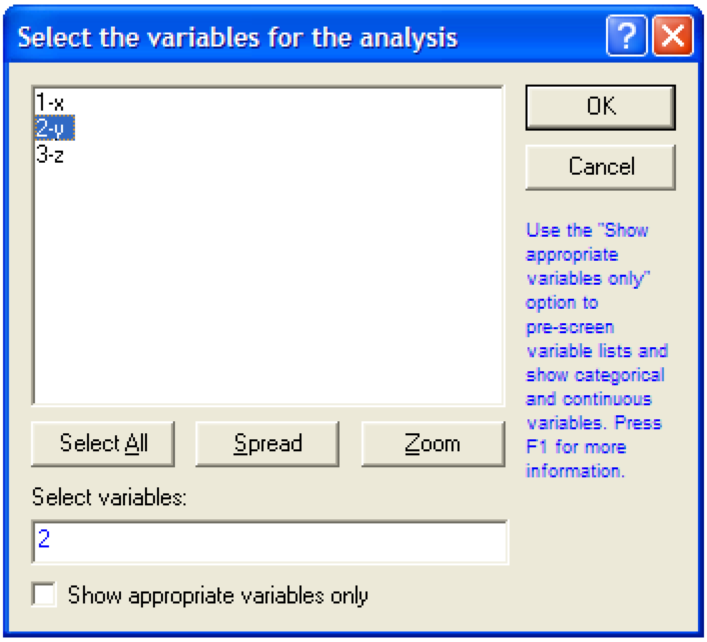
Kada se biraju varijable otvara se prozor s ponuđenim varijablama



Za svaku varijablu piše redni broj i kratki naziv varijable. Varijabla se izabire klikom miša, a za varijablu se upiše redni broj varijable u polju:

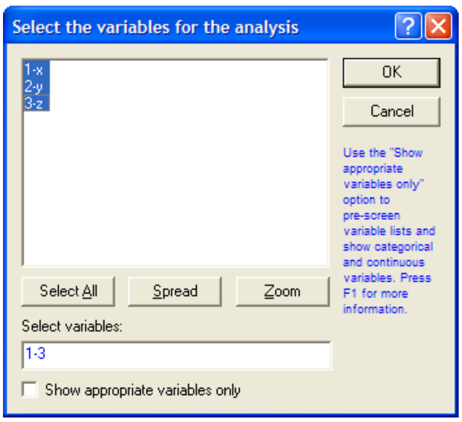
Select variables:

Primjer: izabrana je varijabla y, pod rednim brojem 2.

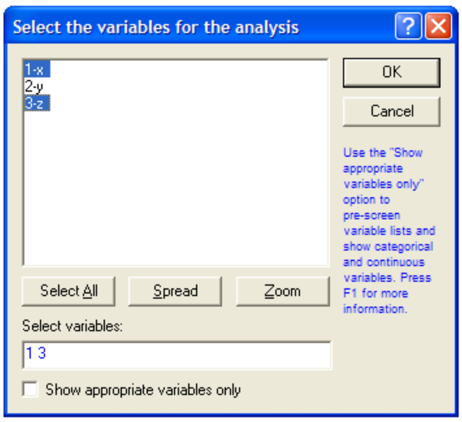


Ako se želi izabrati više varijabli koje su poredane jedna ispod druge, tada se označi prva varijabla, pritisne se tipka **SHIFT** , a nakon toga označi se posljednja varijabla. Ukoliko se žele izabrati varijable koje nisu jedna iza druge, tada se prije izbora sljedeće varijable izabere tipka **CTRL**.

Primjer: izbor varijabli od 1 do 3



Primjer: izbor varijabli 1 i 3



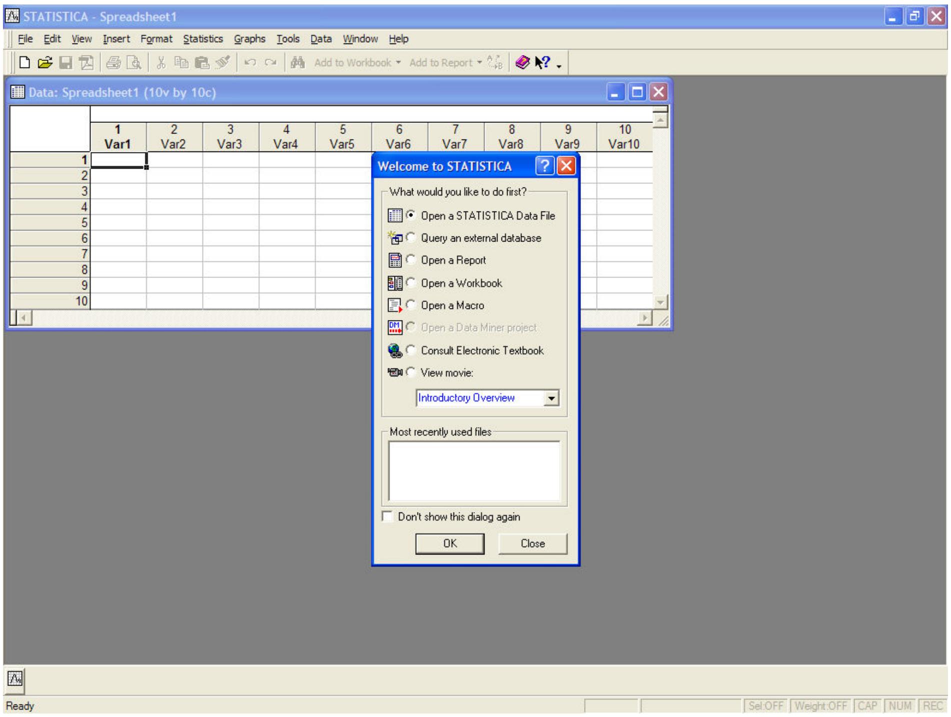
1.1. Pokretanje programa *Statistica*

Program se pokreće klikom na tipku Start te izborom odgovarajućih menija kojim se dolazi do programa *Statistica*. Najčešće se nalazi u meniju ***Programs*** i pod meniju ***Statistica** 7*.

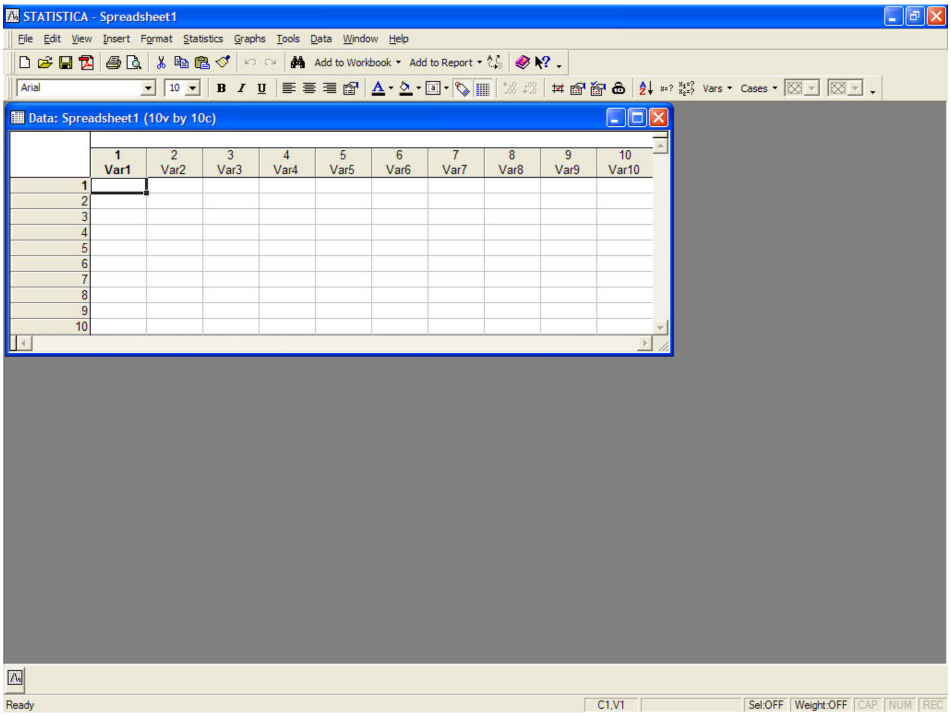
➢ **START / PROGRAMS / STATISTICA 7 / STATISTICA**

Drugi način pokretanja programa je pomoću ikone: StatSoft koja se može nalaziti na radnoj površini (desktopu) ili radnoj traci (task baru).

Kada se *Statistica* pokrene pokazuju se dva prozora:



Manji prozor nudi izbor koji se dio sustava *Statistica* želi otvoriti. Izborom predložene opcije ***Open a STATISTICA Data File***, otvara se posljednji korišteni dokument.



Radni list (Spreadsheet) namijenjen je unosu, uređivanju, prikazu i pohrani podataka, a ima oblik tablice. Osnovni element radnog lista je **polje** (cell) u koje se unosi pojedini podatak. Polje se može odrediti kao presjek **stupca** koji predstavlja varijablu (variable) i **retka** koji predstavlja pojedini slučaj (case). Polje se selektira klikom na lijevu tipku miša ili navigacijom pomoću tastature. Polje koje je selektirano je aktivno polje, a koordinate tog polja prikazane su u statusnoj traci (status baru). Adresa polja označava se s C i rednim brojem retka te s V i rednim brojem stupca, tako npr. adresa (C3, V8) označava polje u osmoj varijabli i trećem retku. Varijabla odnosno redak selektira se klikom na **zaglavlje varijable** odnosno na odgovarajuće polje u **pred stupcu**.

U naslovnoj traci nalazi se naziv dokumenta (radnog lista), a u zagradi je naveden broj varijabli (v) i redaka (c).

Prilikom unosa podataka iz tablice naslov se zapisuje u **informativno polje** (info box) i u **zaglavlje** (header). Klikom na informativno polje selektira se cijeli radni list, a dvostrukim klikom selektira se polje i omogućuje unos i promjenu podatka. Zaglavlje je moguće selektirati jednim klikom miša, a podaci se, također, mogu unositi nakon dvostrukog klika lijevom tipkom miša.

U pojedino polje moguće je unijeti niz znakova; dokumente sa zvučnim ili video zapisima, grafovima, animacijama ili bilo koje dokumente kojima je ActiveX kompatibilan.

1.2. Stvaranje novog dokumenta

Primjer 1.1.

Starost i spol izabranih studenata upisanih na prvu godinu studija, akademske godine 2008./2009., na Ekonomski fakultet u Rijeci

Student/ica	***Godine***	***Spol***
A	18	Ženski
B	19	Muški
C	18	Ženski
D	25	Ženski
E	20	Muški

Izvor: Informacijski sustav visokih učilišta (simulirani podaci)

Treba kreirati novi dokument za unos podataka iz gornje tablice.

- **FILE / NEW…**

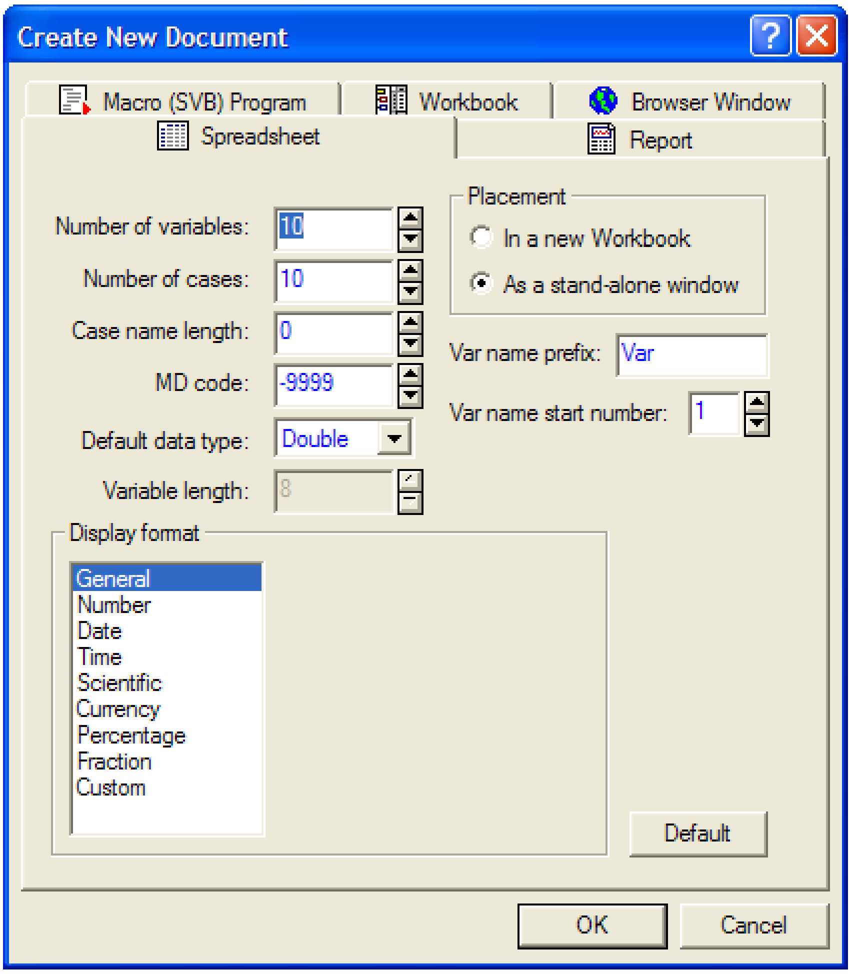
ili

- **<Crtl> + N**

ili

- **Ikona**

Nakon toga otvara se sljedeći prozor:



Ponuđeno je staranje novog **radnog lista** (Spreadsheet), **izvješća** (Report), **makroa** (Macro (SVB) Program), **Radnih zapisa** (Workbook) i **otvaranje internetske stranice** (Browser Window).

Radni list kreira se izborom opcije:

- **Spreadsheet**

- ***Number of variables*** – **Broj varijabli**. Unosi se broj stupaca, može se predvidjeti veći broj stupaca u kojima će biti rezultati izračuna.
- ***Number of cases*** – **Broj redaka**. Upisuje se broj redaka u tablici.

Može se ostaviti ponuđeni broj stupaca i redaka (10×10), a broj varijabli i broj redaka može se na jednostavan način promijeniti tijekom rada.

1.3. Unos podataka

Podatke iz primjera 1.1. treba unijeti u dokument.

- U **zaglavlje** dokumenta upisuje se naziv tablice:

  Starost i spol izabranih studenata upisanih na prvu godinu studija, akademske godine 2008./2009., na Ekonomski fakultet u Rijeci

- U **informativno polje** upisuje se naziv pred stupca:

  Student/ica

Prije početka unosa podataka potrebno je definirati varijable. Definiranje varijabli moguće je izvesti na više načina. U *Statistici* se većina zadataka može izraditi na više načina, a svejedno je koji se način koristi.

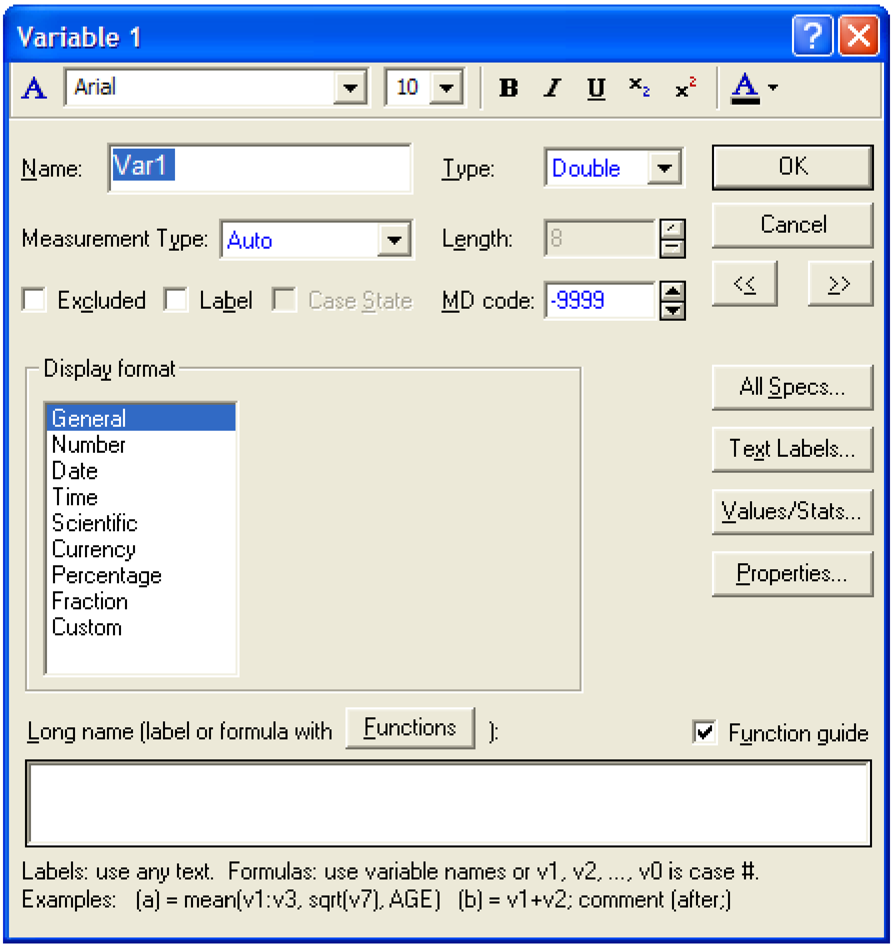
Definiraju se dvije varijable: godine i spol izabranih studenata. Nakon pozicioniranja na polje u određenoj varijabli može se izabrati:

- **DATA / VARIABLE SPECS...**

ili

- **2 puta kliknuti na zaglavlje varijable**

čime se otvara prozor za specifikaciju varijable:



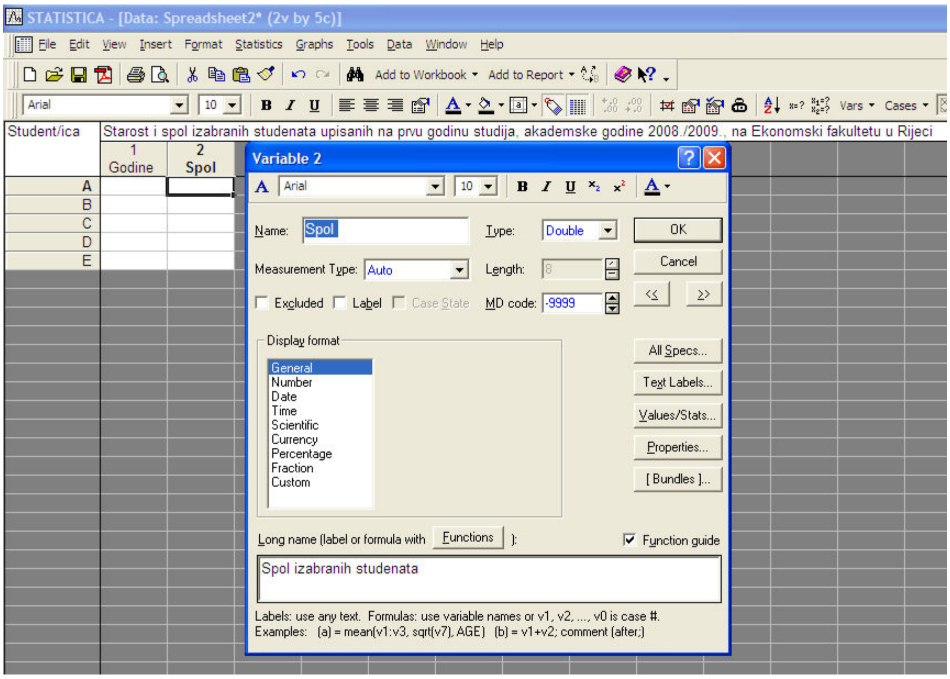
- ***Name*** – **Kratki naziv varijable** – Praktično je ako je kratko ime. Treba izbjegavati „**-**„ i **razmak** te rabiti samo slova engleske abecede. Premda program prihvaća i druge znakove, neke operacije ne mogu se izvoditi, ako se upotrebljavaju ti znakovi.
  - 1. varijabla: upisati ***Godine***
  - 2. varijabla: upisati ***Spol***

- ***Long name (label or formula with*** **Functions** ***)*** **–Dugi naziv varijable (ili formula)** – Unosi se dugi naziv, nema ograničenja dužine, stoga treba dati što bolji opis
  - 1. varijabla: upisati ***Godine starosti izabranih studenata***
  - 2. varijabla: upisati ***Spol izabranih studenata***

U ovu rubriku može se upisati i formula kojom se izračunavaju vrijednosti varijable. Formula započinje znakom jednakosti (=). Dugi naziv varijable u tom slučaju dolazi nakon formule, a između se stavlja znak (;) točka sa zarezom.

- ***Type*** **– Tip varijable**

Predviđeni tip je *Double*. To je jedini tip varijable koji podržava decimale. Moguće je unositi brojeve, slova i ostale znakove. Najbolje je ostaviti ovaj tip varijable.



Umjesto otvaranja prozora za specifikaciju svake varijable posebno, moguće je izabrati opciju za specificiranje svih varijabli na sljedeće načine:

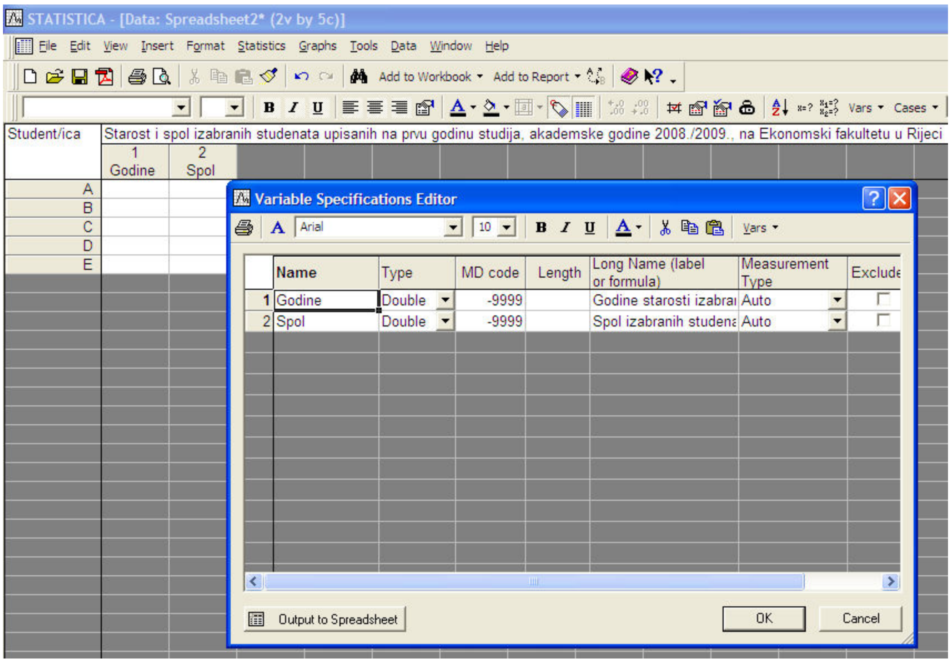
Na već otvorenom prozoru za specifikaciju varijable izabere se:

- **ALL SPECS...**

ili se preko sustava menija izabere

- **DATA / ALL VARIABLE SPECS…**

Nakon čega se otvara prozor za specifikaciju svih varijabli.

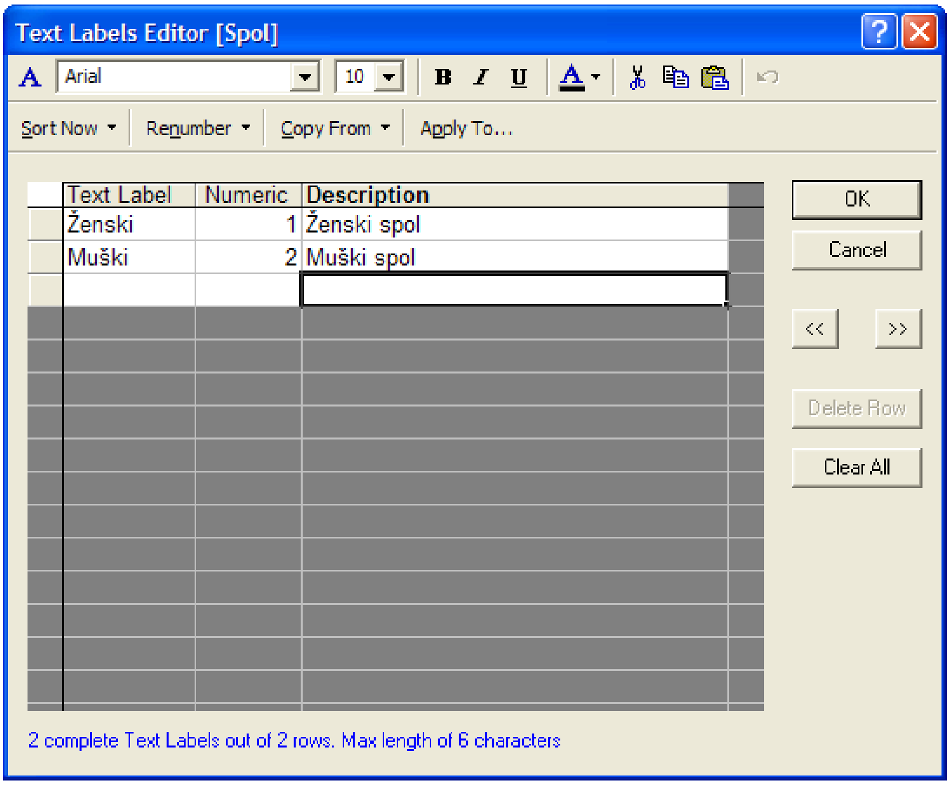


Ukoliko varijabla može poprimiti neke određene vrijednosti modaliteta, tada se mogu kodirati vrijednosti. Varijabla Spol može poprimiti dvije vrijednosti: *Ženski* i *Muški*.

- Pozicionirati se na drugoj varijabli i otvoriti prozor za specifikaciju varijable ukoliko već nije otvoren
    - **DATA / VARIABLE SPECS...**

- Izabrati tipku
    - **Text Labels…**

te se otvara prozor za kodiranje varijabli.

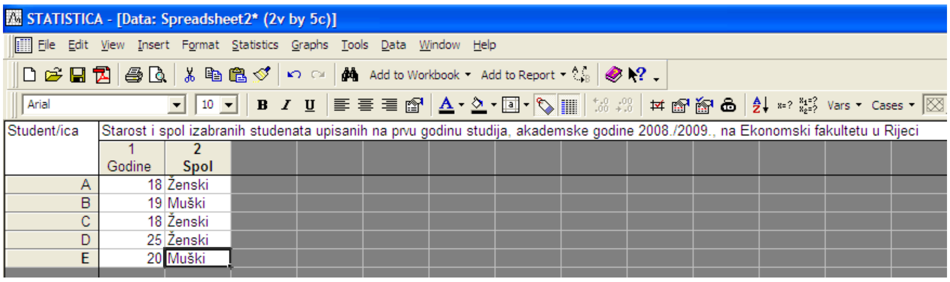


- ➢ ***Text label*** – **Modalitet varijable** – unosi se vrijednost koju varijabla može poprimiti

- ➢ ***Numeric*** – **Brojčana šifra** – izabire se broj, brojčana šifra, koja predstavlja vrijednost modaliteta

- ➢ ***Description*** – **Opis modaliteta** – ovdje se može staviti dugi naziv ili opis kodiranog modaliteta, nije obavezan unos

Nakon definiranja varijabli unose se vrijednosti. Kod unosa pred stupca i varijable godine prepisuju se vrijednosti zadane u tablici, a kod unosa spola unose se šifre: *1* i *2*, a vidi se *Ženski* i *Muški*.



1.4. Spremanje dokumenta

Kod prvog spremanja

- **FILE / SAVE ili SAVE AS…**

ili

- **<Crtl> + S**

ili

- **Ikona**

Ako se dokument snima prvi put, otvara prozor ***Save As*** u koji se unosi naziv dokumenta i izabere direktorij u koji će biti spremljen. Radni list ima nastavak *.sta*, izvješće ima nastavak *.str*, nastavak makroa je *.svb*, radni zapisi označavaju se s *.stw*.

Ukoliko se u rubrici ***Save as type*** izabere neki drugi nastavak, tada se dokument snima u nekom drugom formatu, npr. Excelu, SPSS-u i sl.

Kad je dokument jednom snimljen, za spremanje izmjena više se ne koristi ***Save As***, nego samo ***Save***. Promjene se tada snimaju u već postojeću datoteku.

1.5. Otvaranje dokumenta

Postojeći dokument otvara se:

- **FILE / OPEN…**

ili

- **<Crtl> + O**

ili

- **Ikona** 

Prikazuje se prozor u kojem se izabere direktorij i dokument koji ima neki od ponuđenih nastavaka. U rubrici ***Files of Type*** može se izabrati neki drugi tip dokumenta.

1.6. Prikazivanje rezultata obrade

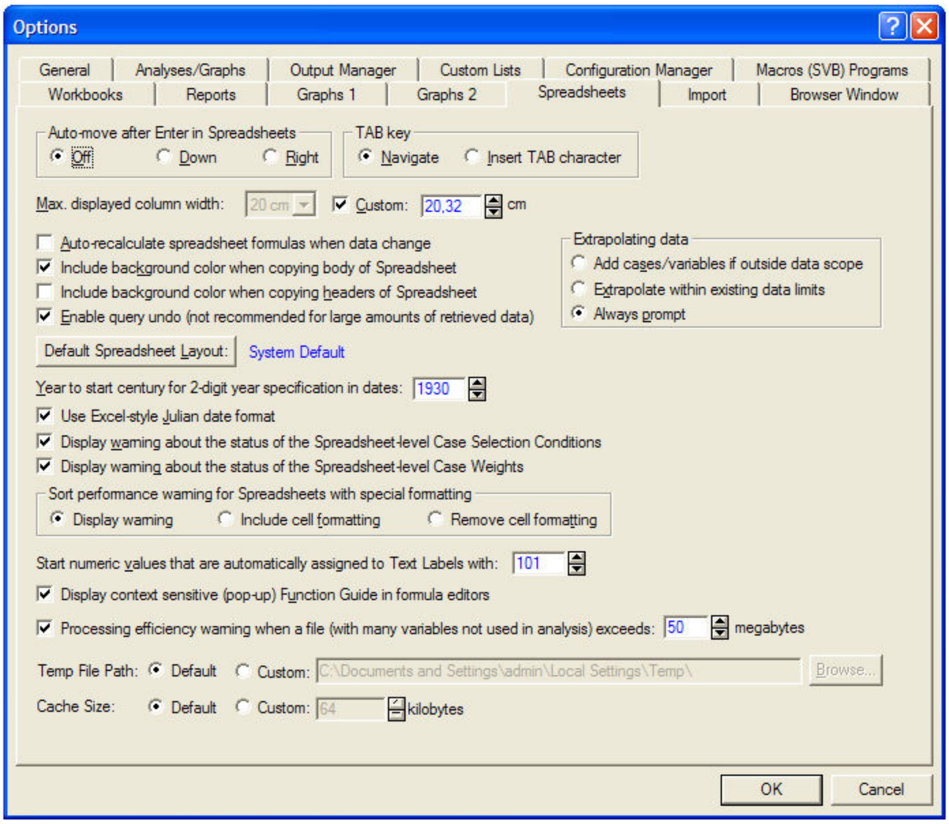
Rezultati statističkih izračuna prikazuju se u prozoru radnih zapisa (Workbook), a moguće ih je dobiti i u obliku izvješća.

Da bi se pokrenulo zapisivanje rezultata u izvješće, potrebno je putem menija izabrati:

- **TOOLS / OPTIONS...**

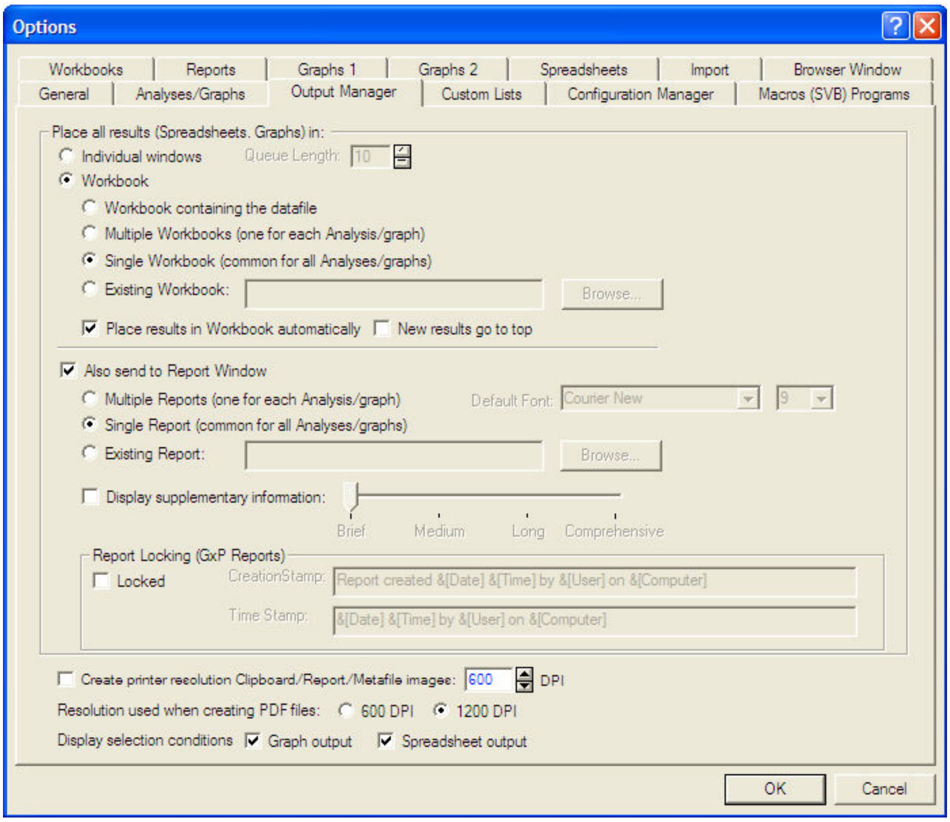
Nakon otvaranja prozora izabere se rubrika:

- **OUTPUT MANAGER**



i označi se:

- ➢ **ALSO SEND TO REPORT WINDOW**



Zapisivanje rezultata u izvješće može se izvršiti na više načina:

- ***Multiple Reports (one for each Analysis/graph)*** – stvaranje više prozora, po jedan prozor za svaku analizu ili grafički prikaz

- ***Single Report (common for all Analysis/graphs)*** – kreira se jedno izvješće, zajedničko za sve analize i grafikone

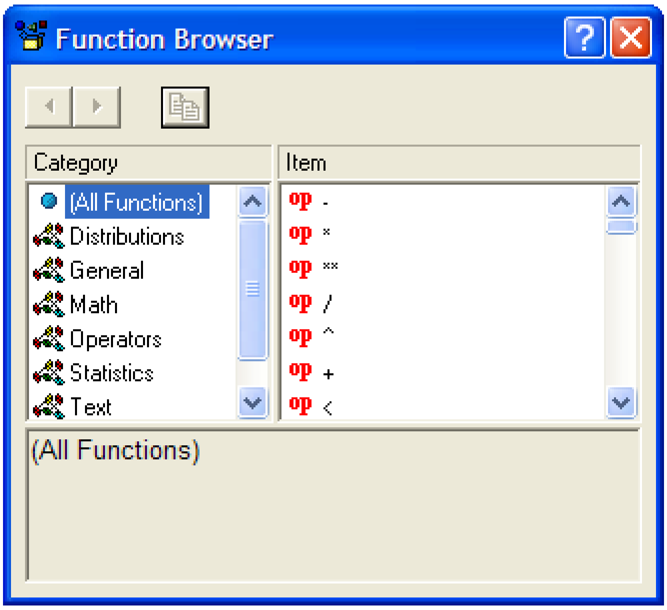
- ***Existing Report*** – rezultati se unose u neko već postojeće izvješće

Uz to moguće je izabrati i:

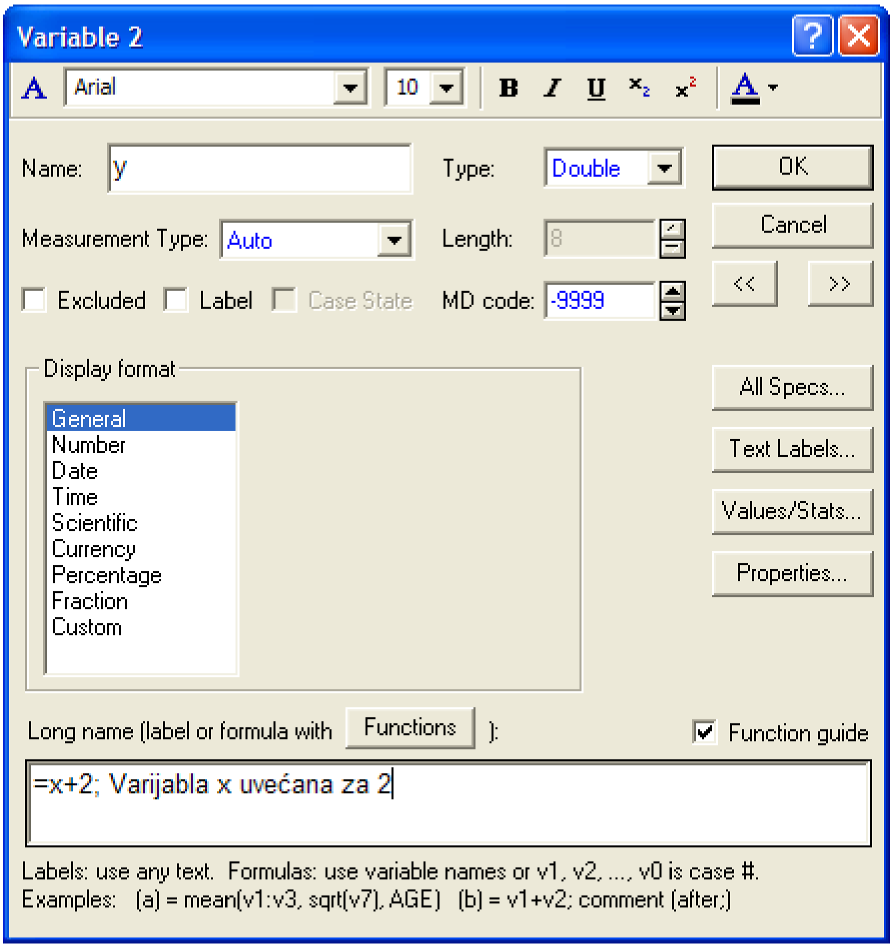
- ***Display supplementary information:*** – prikazuju se dodatne informacije
  - ***Brief*** – kratke
  - ***Medium*** – srednje
  - ***Long*** – duge
  - ***Comprehensive*** – iscrpne

1.7. Definiranje varijable formulom

Varijabla se može definirati formulom. Tada se u rubrici dugi naziv varijable, upisuje znak =, a nakon toga slijedi formula. Moguće je koristiti standardne matematičke operacije ili neku od ponuđenih funkcija koja se dobije pritiskom na tipku **Functions**.



Ako se nakon formule upiše ; može se u nastavku upisati dugi naziv varijable.

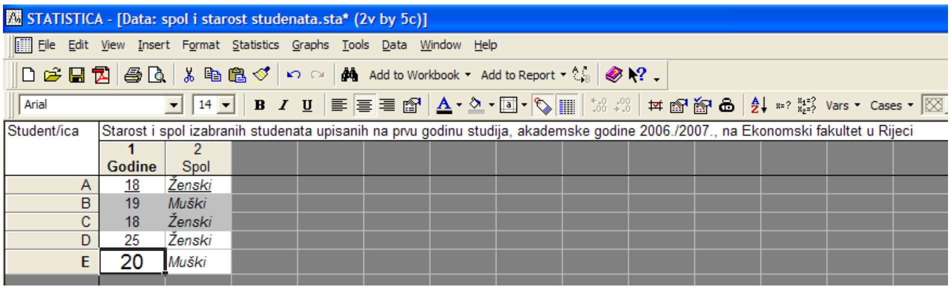


1.8. Uređivanje

Dio dokumenta koji se želi uređivati mora se prvo selektirati. Polje se selektira mišem ili navigacijom pomoću tastature. Varijabla se selektira klikom miša na zaglavlje varijable, redak se selektira klikom miša na naziv retka u pred stupcu. Blok podataka se izabere klikom na gornji lijevi kut bloka podataka i pritiskom na lijevu tipku miša dok se miš povlači do donjeg desnog kuta.

Nakon selektiranja izabranog dijela može se pristupiti uređivanju.

Na slici je prikazan dokument u kojem su vrijednosti varijable "Godine" centrirane, vrijednosti varijable "Spol" napisane su kurzivom (Italic), podaci u retku A su podcrtani, blok podataka od C2V1 do C3V2 je osjenčan, a polje C5V1 ima veću veličinu slova.

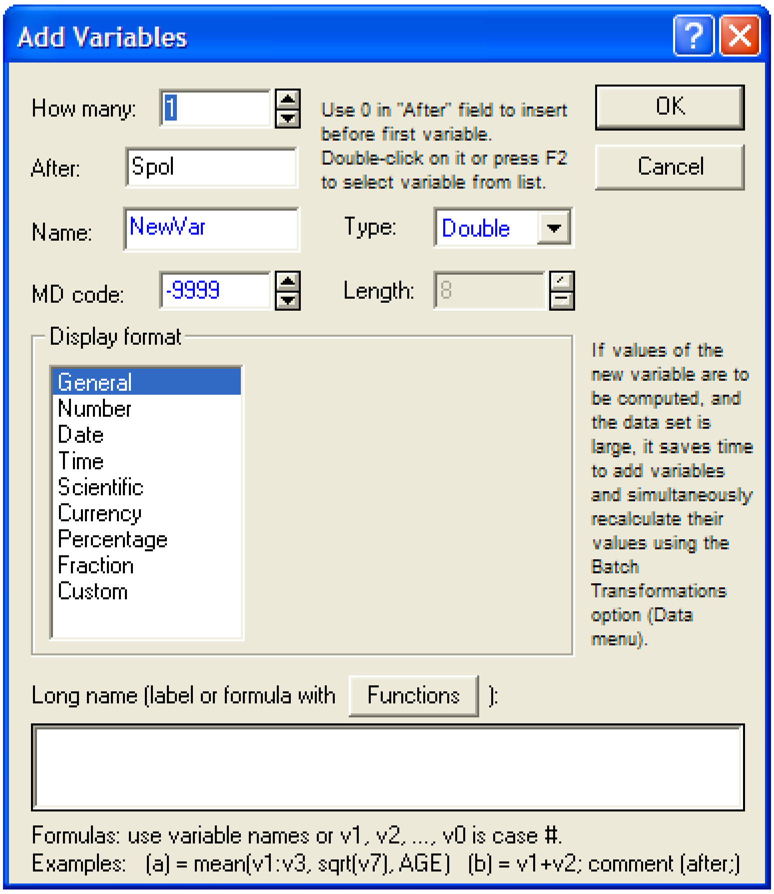


Osim mijenjanja izgleda, moguće je i dodavanje ili brisanje pojedinog stupca ili retka.

Dodavanje novih stupaca:

➢ **INSERT / ADD VARIABLES…**

Otvara se prozor za dodavanje varijabli:

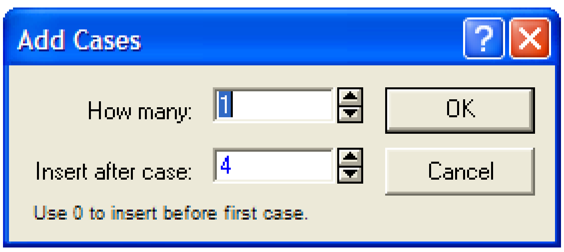


- ***How many*** – broj novih stupaca
- ***After*** – nakon koje varijable se unose nove

Dodavanje novih redaka:

- **INSERT / ADD CASES…**

Otvara se prozor za dodavanje redaka:

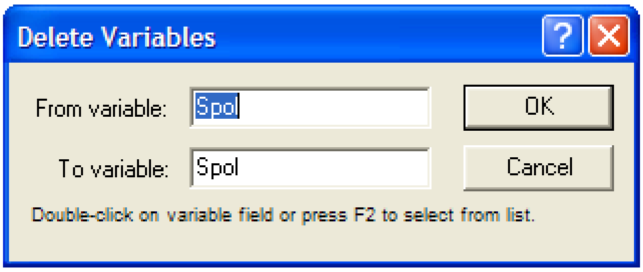


- ***How many*** – broj novih redaka
- ***Insert after case*** – nakon kojeg retka se unose novi

Brisanje stupaca:

- **EDIT / DELETE / VARIABLES…**

Otvara se prozor za brisanje varijabli:



- ***From variable*** – od koje varijable
- ***To variable*** – do koje varijable

Brisanje redaka:

- **EDIT / DELETE / CASES…**

Otvara se prozor za brisanje redaka:

- ***From cases*** – od kojeg retka
- ***To cases*** – do kojeg retka

2. Izabrane distribucije vjerojatnosti kontinuirane slučajne varijable

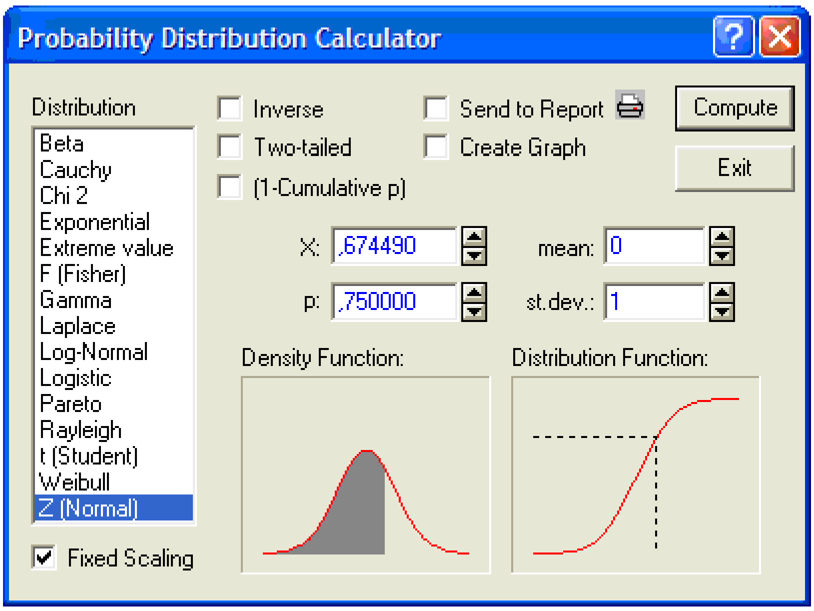
Prilikom računanja vjerojatnosti kontinuirane slučajne varijable *Statistica* ima mogućnost rada na dva načina: upotrebom kalkulatora raspodjele vjerojatnosti (Probability Distribution Calculator) ili upotrebom funkcija.

I. način pomoću kalkulatora raspodjele vjerojatnosti

Kalkulator vjerojatnosti poziva se iz menija

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**

Otvara se prozor:



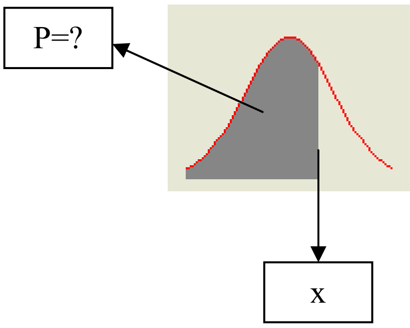
U lijevom dijelu prozora u rubrici ***Distribution*** ponuđene su podržane raspodjele vjerojatnosti:

- ***Beta***
- ***Cauchy***
- ***Chi 2*** – $\chi 2$
- ***Exponential*** **– eksponencijalna**
- ***Extreme value*** **– ekstremna vrijednost**
- ***F (Fisher)***

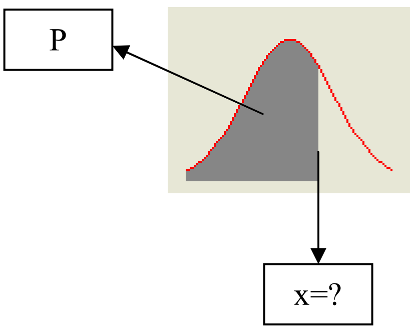
- ***Gamma* – gama**
- ***Laplace***
- ***Log-normal***
- ***Logistic* – logistička**
- ***Pareto***
- ***Rayleigh***
- ***t (Student)***
- ***Weibull***
- ***Z (Normal)***

Za izabranu raspodjelu otvore se odgovarajuća polja.

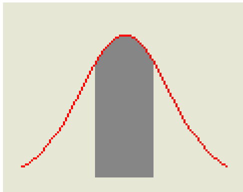
Ako se traži vjerojatnost da će slučajna varijabla poprimiti vrijednost manju od "x" (zadan je x, traži se P = P(X < x)), tada se u polje **X** unese zadana vrijednost, u ostala polja unesu se zadani parametri: broj stupnjeva slobode, prosjek, standardna devijacija,… te se pritiskom na tipku **Compute** ispiše rezultat u polju **p**.



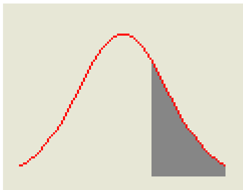
Ukoliko se izabere ponuđena mogućnost ***Inverse***, tada se zadaje vjerojatnost da će slučajna varijabla poprimiti vrijednost manju od "x" (zadana je "P", traži se "x"), a rezultat je "x". Unosi se vrijednost "p", te ostali potrebni parametri. Čim se unese vrijednost "p", automatski se aktivira funkcija Inverse. Ako se unese vrijednost u rubriku "x", isključuje se inverzna funkcija.



Funkcija $\boxed{\textit{\textbf{Two-tailed}}}$, daje rezultat na dva kraka krivulje. Za zadani "x" daje vjerojatnost da će slučajna varijabla poprimiti vrijednost od –x do x. (-x < P < x)

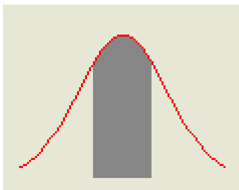


Suprotna vjerojatnost, tj. vjerojatnost da će varijabla poprimiti vrijednost veću od "x" (zadan je x traži se Q = P(X>x)), računa se izborom funkcije $\boxed{\textit{\textbf{1-Cumulative p}}}$.

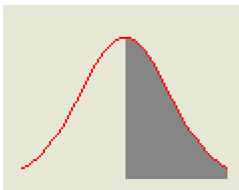


Funkcije se mogu kombinirati:

Inverse i Two- tailed

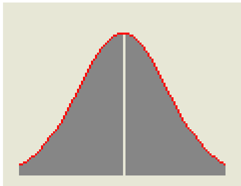


Inverse i 1-Cumulative P



Two-tailed i 1-Cumulative p,

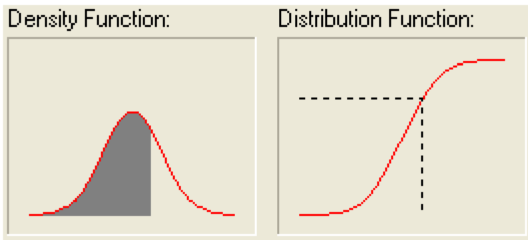
Inverse, Two-tailed i 1-Cumulative p



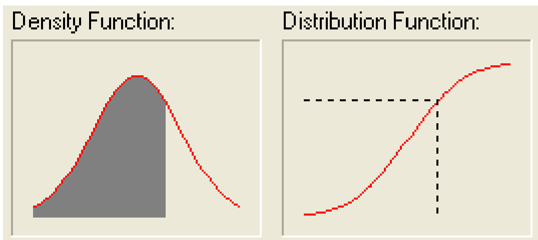
Izborom funkcije ***Send to Report***, rezultati se ispisuju u obliku izvješća.

U kalkulatoru raspodjele vjerojatnosti uvijek se prikazuju dva grafikona:

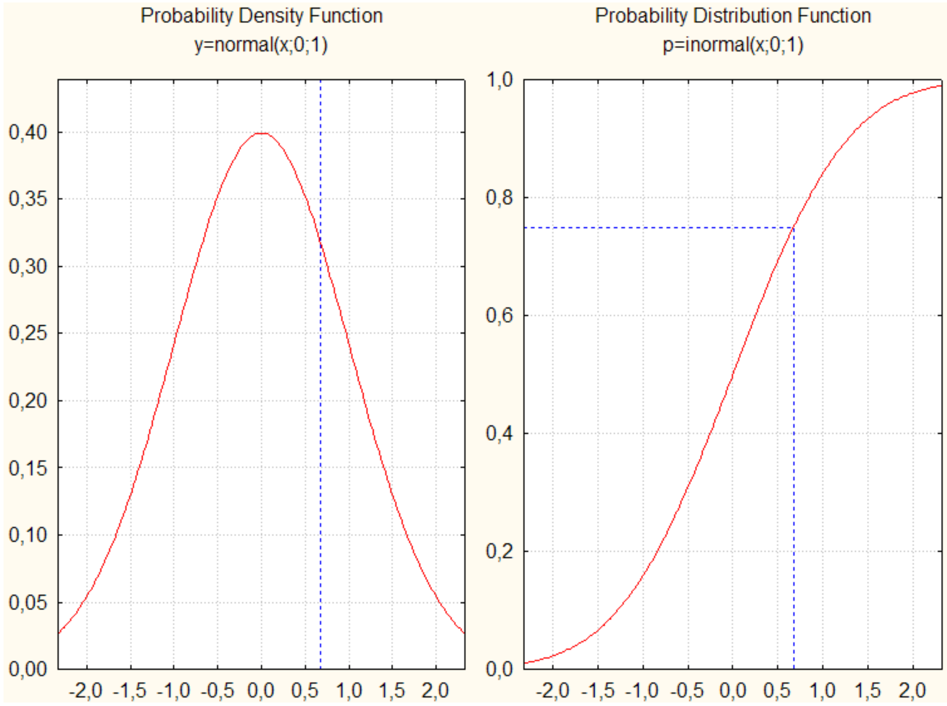
- ***Density function* – funkcija gustoće vjerojatnosti**
- ***Distribution function* – funkcija distribucije**



Isključenjem funkcije ***Fixed scaling***, dimenzije slike prilagođuju se prostoru za grafikon.



Izborom funkcije ***Create Graph***, grafikon će se prikazati u prozoru radnih zapisa.



II. način upotrebom funkcija u radnom listu

Isti rezultati, ali bez gotovih grafikona mogu se dobiti i rabljenjem funkcija u radnom listu. U tom slučaju rezultati se mogu izračunati za više zadanih vrijednosti odjedanput.

II. a) Vjerojatnost i suprotna vjerojatnost

Ako su zadane vrijednosti slučajne varijable, x1, x2, x3,.. a traži se vjerojatnost, da će slučajna varijabla poprimiti vrijednost, manju (P) ili veću (Q) od zadanih vrijednosti x1, x2, x3,…tada je potrebno:

- **Definirati stupac – "xi"**
- **Unijeti zadane vrijednosti:**

xi
x1
x2
x3
…

- **Definirati novi stupac – "Pi"**
- **Long name (label or formula with Functions):**

  - U rubrici dugi naziv unosi se funkcija koja se može utipkati ili izabrati pritiskom na tipku ***Functions***. Izabere se funkcija koja ima prefiks ***I*** i nakon toga naziv raspodjele, npr. ***INormal***, ***IStudent***,…Unose se traženi parametri funkcije: broj stupnjeva slobode, prosjek, standardna devijacija i sl., a varijabla "***xi***" predstavlja parametar x.

  - Rezultat je nova varijabla s vrijednostima varijable "Pi" koje predstavljaju vjerojatnosti da će slučajna varijabla poprimiti vrijednost manju od zadanog x_i-a.

- **Definirati novi stupac – Qi**
- **Long name (label or formula with Functions): =1-Pi**

  - Izračunat će se suprotna vjerojatnost, tj. vjerojatnost da će slučajna varijabla poprimiti vrijednost veću od zadanog "x_i"-a (Isto što i *1-Cumulative p* u kalkulatoru vjerojatnosti)

  - Qi moguće je direktno dobiti tako da se unese ***1-Inormal***, 1 – ***Istudent*** i sl.

II. b) Razlika

Ukoliko se traži vjerojatnost da će slučajna varijabla poprimiti vrijednost između "x_1" i "x_2", tada se računa vjerojatnost da će slučajna varijabla biti manja od tih vrijednosti $P(x_1)$ i $P(x_2)$ te se dobivene vrijednosti oduzmu. Oduzimanje je moguće izvršiti u novom stupcu ili izvan programa *Statistica*.

II. c) Inverz funkcije

Ako su zadane vjerojatnosti P1, P2, P3,…, a traže se x1, x2, x3,… tada treba:

- **Definirati stupac – "Pi"**
- **Unijeti zadane vrijednosti:**

Pi
P1
P2
P3
…

- **Definirati novi stupac – "Xi"**
- **Long name (label or formula with Functions):**

  - Izabere se funkcija koja ima prefiks *V* i nakon toga naziv raspodjele, npr. ***VNormal***, ***VStudent***,…Unose se traženi parametri funkcije: broj stupnjeva slobode, prosjek, standardna devijacija i sl., a varijabla "***Pi***" predstavlja parametar "x".

  - Ova funkcija analogna je funkciji ***Inverse*** u kalkulatoru vjerojatnosti.

  Primjeri dani u nastavku riješeni su na oba načina.

2.1. Normalna raspodjela

Primjer 2.1.

Empirijska raspodjela ima oblik normalne raspodjele i njene karakteristike su:

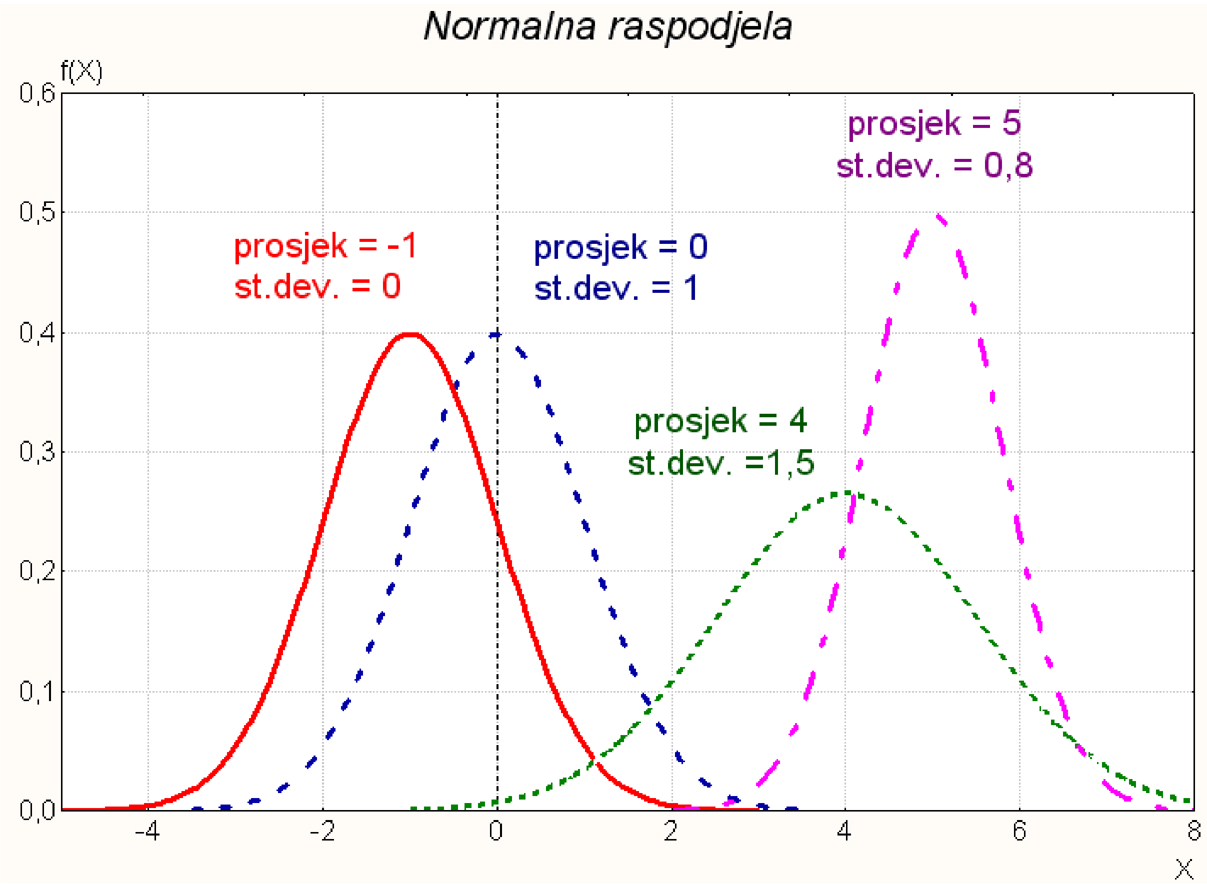
$\mu = 12$

$\sigma^2 = 4.$

Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost:

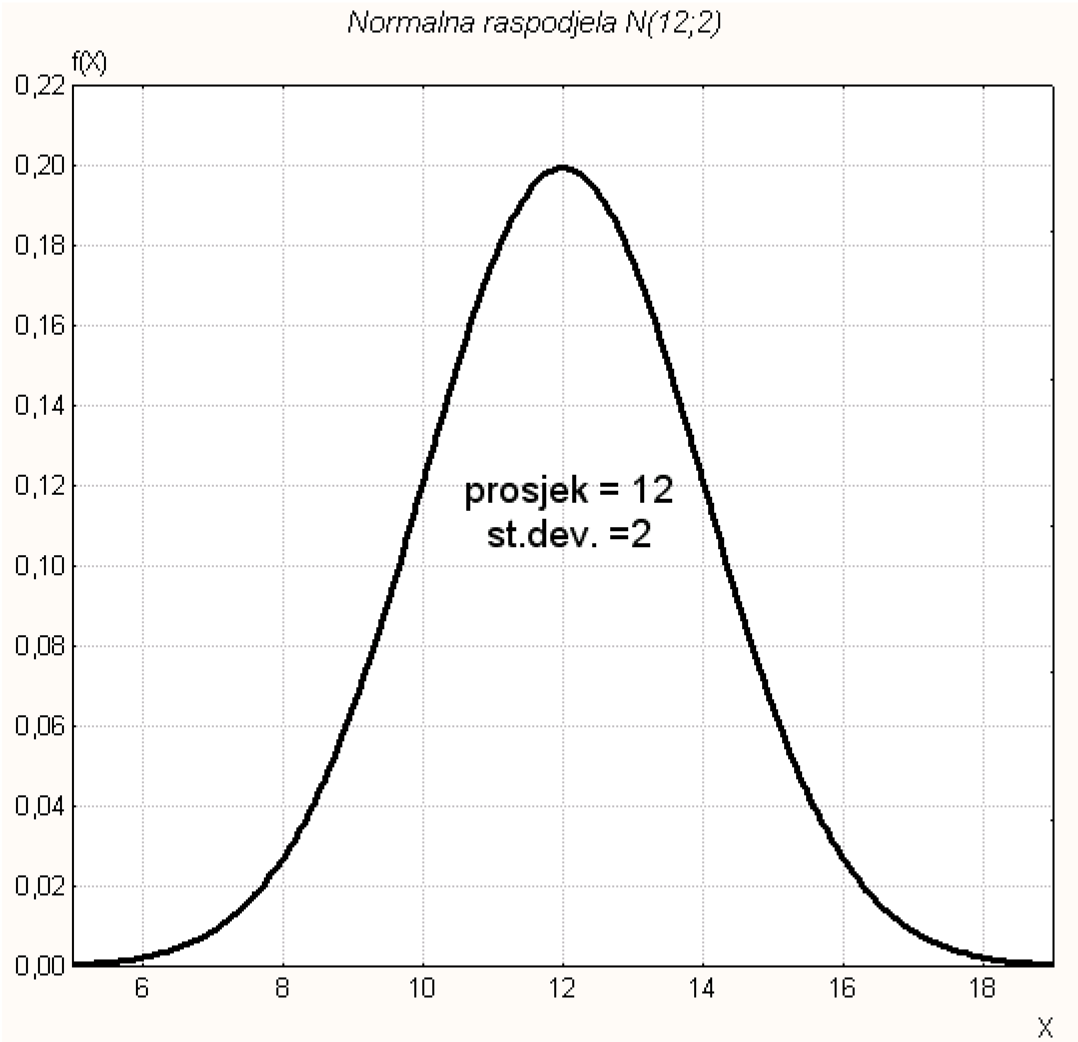
a) manju od 15
b) manju od 8
c) veću od 15
d) veću od 8
e) između 8 i 15?

Postoji beskonačno mnogo normalnih raspodjela:

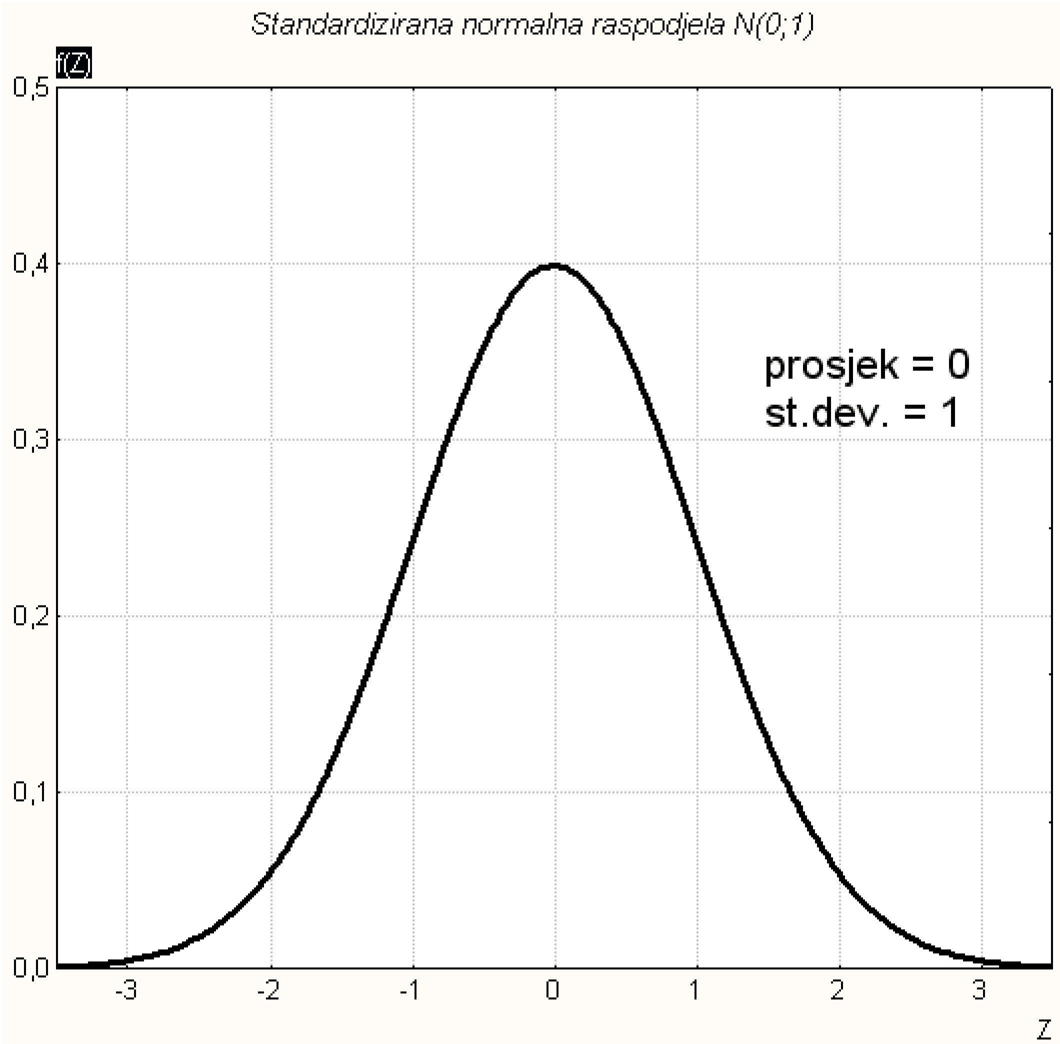


Normalna raspodjela jednoznačno je određena parametrima μ i σ^2, stoga se označava $N(\mu; \sigma^2)$. U ovom zadatku raspodjela se može označiti s $N(12;4)$.

U programskom paketu *Statistica* za normalnu raspodjelu ne rabi se parametar σ^2, nego σ. Parametar σ u zadatku iznosi 2 ($\sigma = \sqrt{4} = 2$).



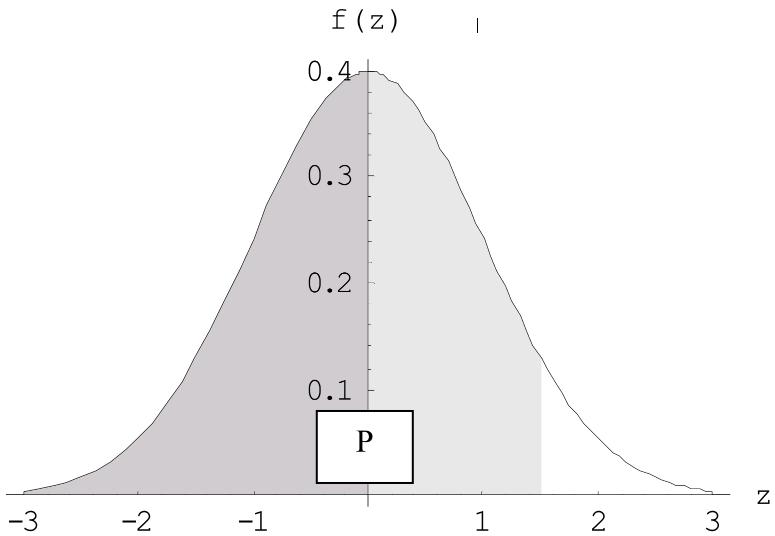
Svaka normalna raspodjela može se svesti na standardiziranu normalnu raspodjelu koja ima: $\boxed{\mu = 0 \text{ i } \sigma^2 = 1}$.
Umjesto slučajne varijable X tada se koristi standardizirana slučajna varijabla $Z = \dfrac{x-\mu}{\sigma}$.



$x_1 = 8$	$Z_1 = \dfrac{x_1 - \mu}{\sigma} = \dfrac{8-12}{2} = -2$	$Z_1 = -2$
$x_2 = 15$	$Z_2 = \dfrac{x_2 - \mu}{\sigma} = \dfrac{15-12}{2} = 1{,}5$	$Z_2 = 1{,}5$

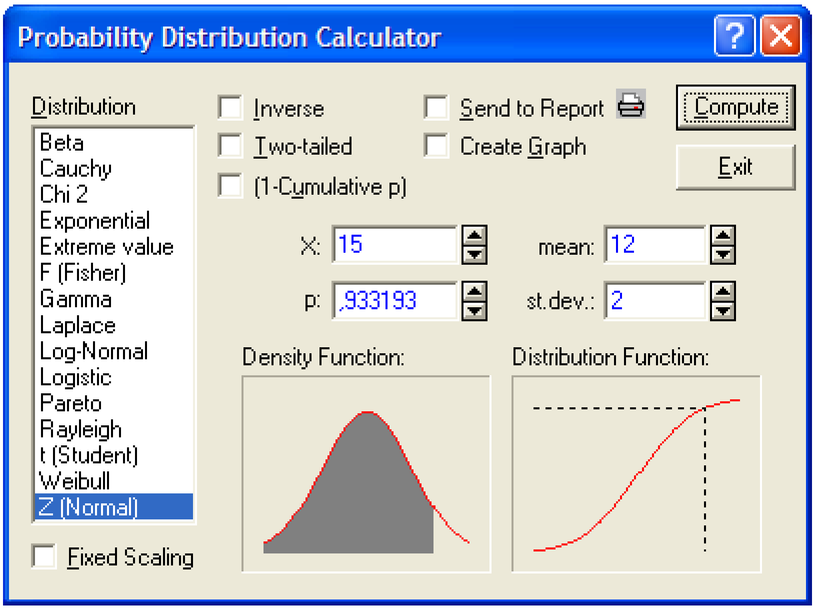
a) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost manju od 15?

To je površina osjenčanog dijela:



♦ **I. način pomoću kalkulatora raspodjele vjerojatnosti**
A) Uporabom zadanih vrijednosti

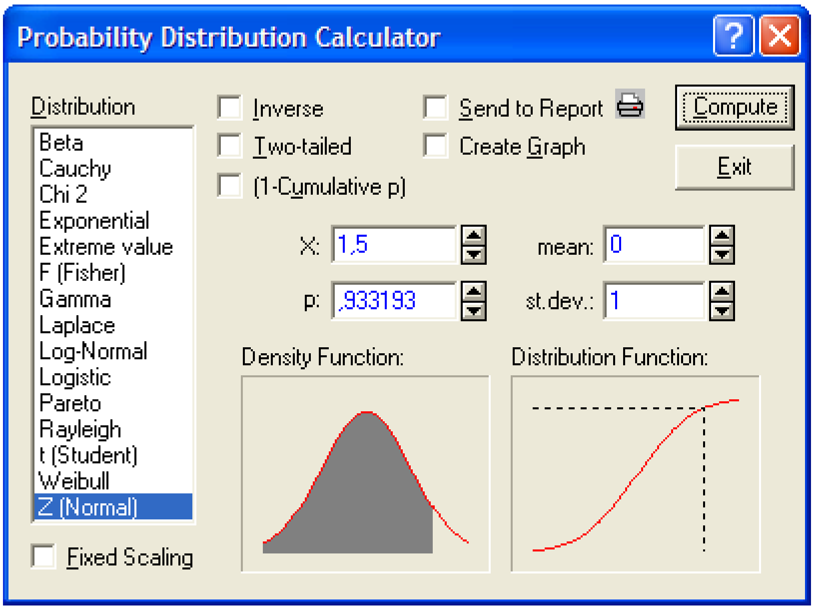
- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Z (Normal)***
  - **Fixed Scaling - odznačiti**
    - **X:** ***15***
    - **mean:** ***12***
    - **st.dev.:** ***2***
- ***Compute***
  - rezultat = 0,933193 = 93,3%



Vjerojatnost da slučajna varijabla poprimi vrijednost manju od 15 je 0,933 ili 93,3%.

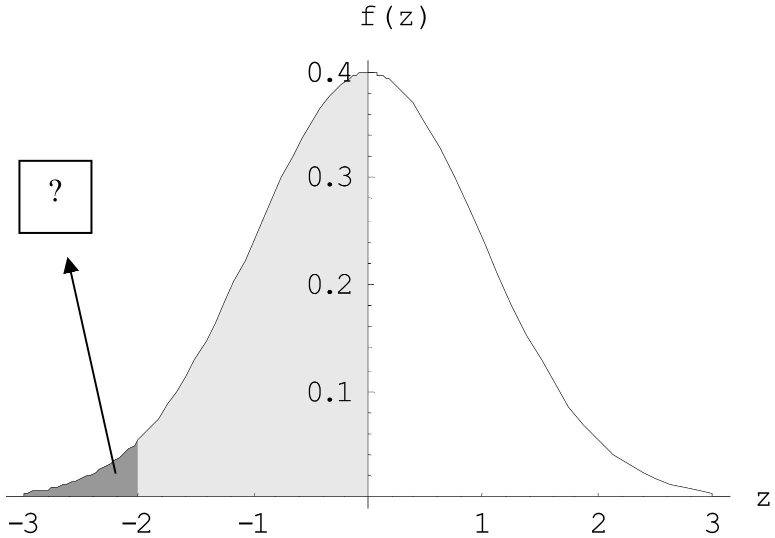
B) Pomoću standardizirane raspodjele

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution: *Z (Normal)***
    - **X: *1,5***
    - **mean: *0***
    - **st.dev.: *1***



- rezultat je isti: 0,933193 = 93,3%

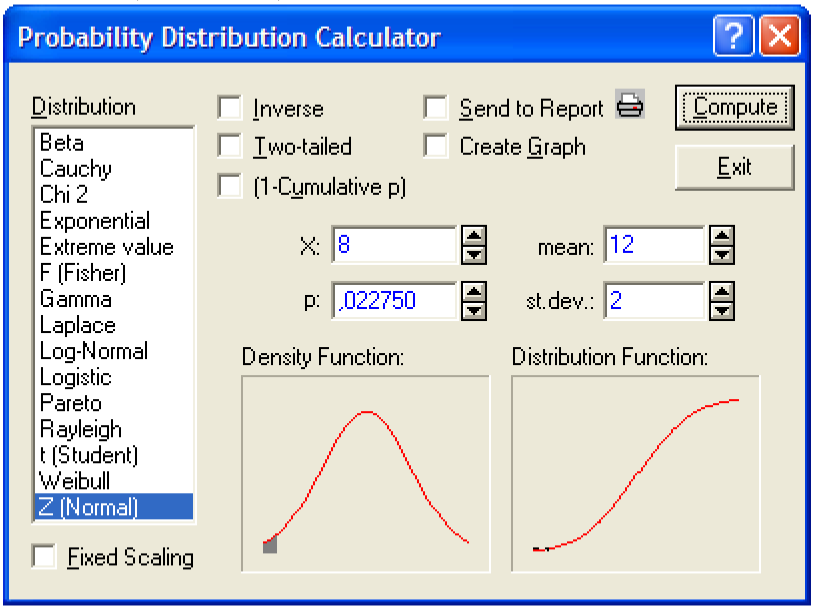
b) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost manju od 8?



- **I. način pomoću kalkulatora raspodjele vjerojatnosti**
  A) Uporabom zadanih vrijednosti

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Z (Normal)***
  - **Fixed Scaling - odznačiti**
    - **X:** ***8***
    - **mean:** ***12***
    - **st.dev.:** ***2***

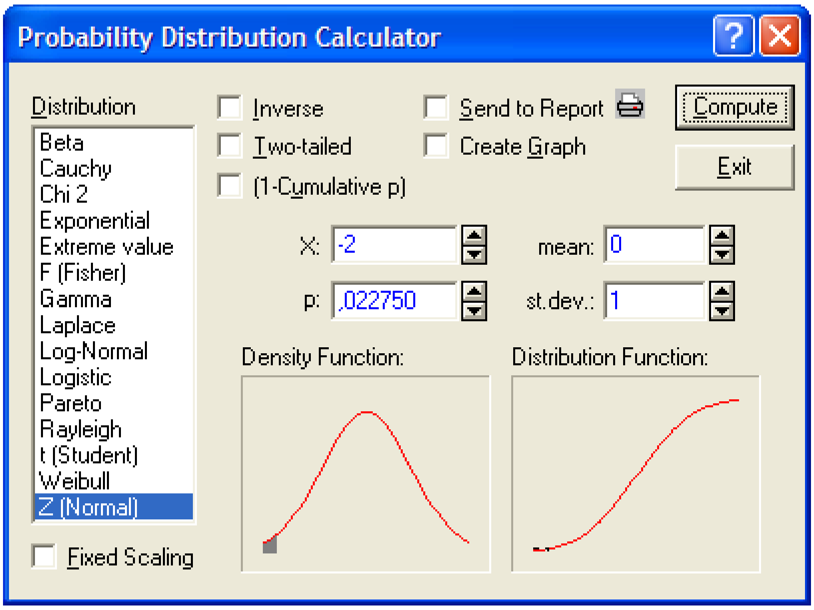
rezultat = 0,022750 = 2,3%



Vjerojatnost da slučajna varijabla poprimi vrijednost manju od 8 iznosi 0,023 ili 2,3%.

B) Pomoću standardizirane raspodjele

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution: *Z (Normal)***
    - **X: *-2***
    - **mean: *0***
    - **st.dev.: *1***



rezultat je isti: 0,02275 = 2,3%

II. način upotrebom funkcija u radnom listu

Općenito, kada se traži vjerojatnost da je slučajna varijabla manja od neke vrijednosti:

- zatvoriti prozor: Probability Distribution Calculator
- izabrati prozor s podacima:

A) Uporabom zadanih vrijednosti:

- **Definirati stupac – "xi"**
  - **Unijeti vrijednosti:**
    - *8*
    - *15*

- **Definirati novi stupac – "Pxi"**
  - **Long name:**

  Koristi se funkcija INormal (x; mu; sigma). Objašnjenje parametara je:

  - ***x*** **– vrijednost slučajne varijable xi, unosi se xi**
  - ***mu*** **– μ, unosi se 12**
  - ***sigma*** **– σ, unosi se 2**

- **=INormal (x; mu; sigma)**
- **=INormal (xi; μ; σ)**
- ***=INormal (xi; 12; 2)***

xi	*Pxi*
8	0,02275
15	0,933193

B) Pomoću standardizirane raspodjele"

- **Definirati varijablu – "zi"**
  - **Long name:** ***=(xi-12)/2***

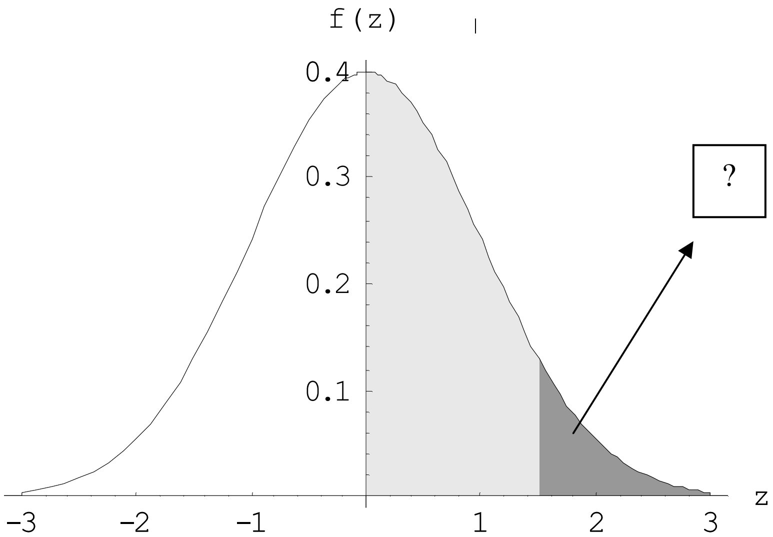
xi	*zi*
8	-2
15	1,5

- **Definirati novu varijablu – "Pzi"**
  - **Long name:**
    - **=INormal (x; mu; sigma)**
    - **=INormal (zi; μ; σ)**
    - ***=INormal (zi; 0; 1)***

xi	*Pzi*
8	0,02275
15	0,933193

Rezultati su jednaki.

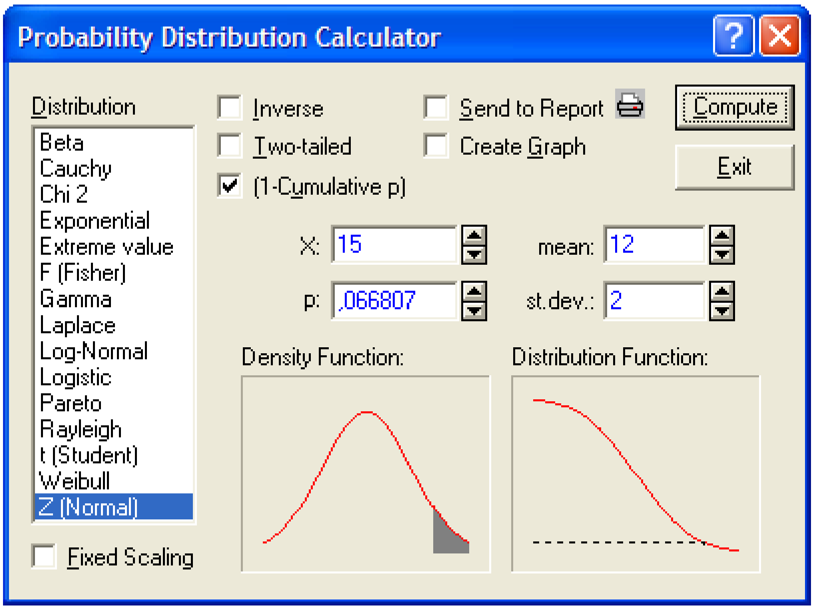
c) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost veću od 15?



- **I. način pomoću kalkulatora raspodjele vjerojatnosti**

  A) Uporabom zadanih vrijednosti:

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Z (Normal)***
  - **Fixed Scaling – ukloniti oznaku**
  - ***(1 – Cumulative p)***
    - **X:** ***15***
    - **mean:** ***12***
    - **st.dev.:** ***2***
  - rezultat = 0,066807 = 6,7%.

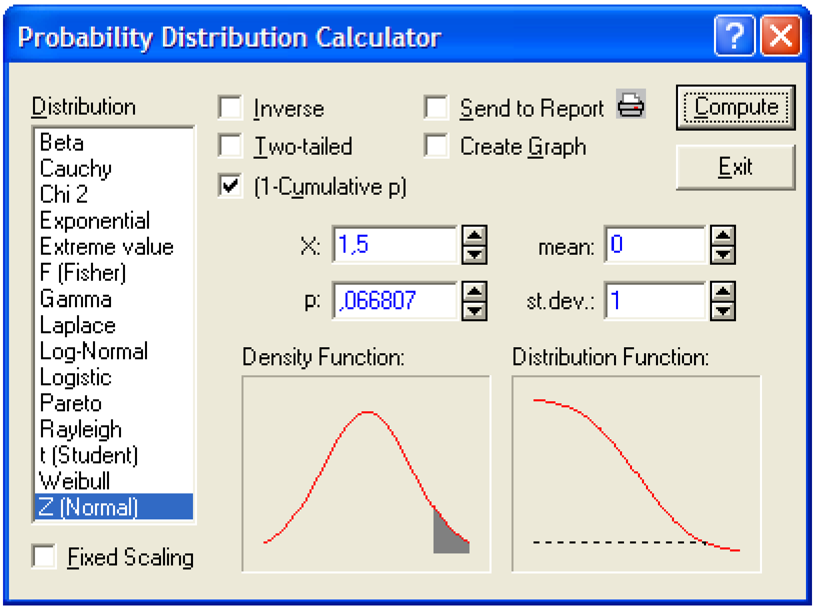


Vjerojatnost da slučajna varijabla poprimi vrijednost veću od 15 je 0,067 ili 6,7%.

B) Pomoću standardizirane raspodjele:

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Z (Normal)***
  - **Fixed Scaling – ukloniti oznaku**
  - ***(1 – Cumulative p)***
    - **X:** ***1,5***
    - **mean:** ***0***
    - **st.dev.:** ***1***

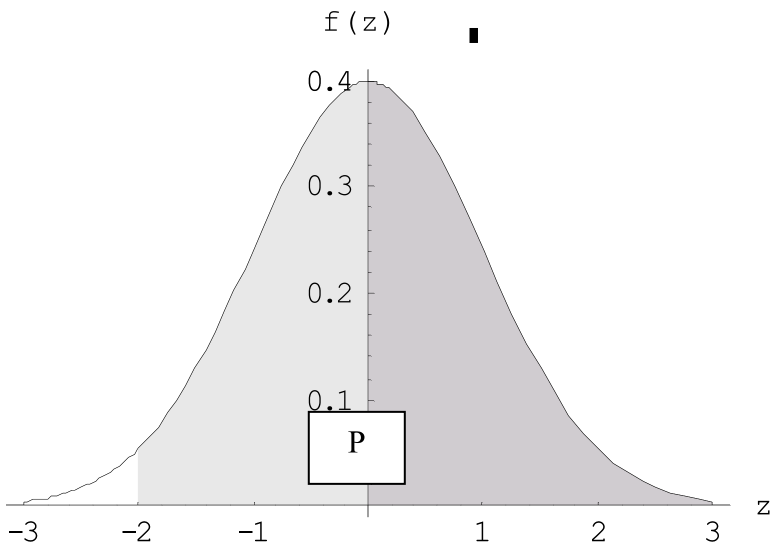
\- rezultat = 0,066807 = 6,7%



Rezultat je jednak: 0,067 ili 6,7%.

d) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost veću od 8?

To je površina osjenčanog dijela.

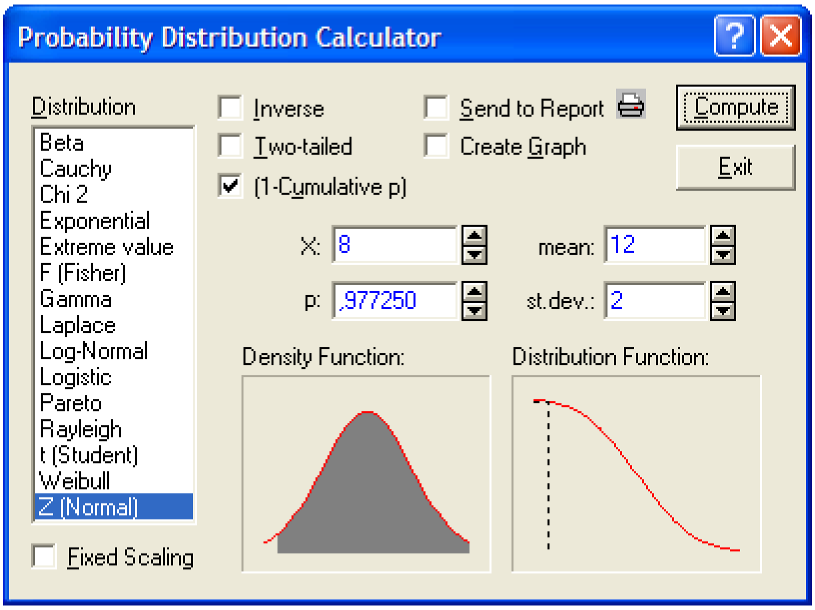


A) Uporabom zadanih vrijednosti

♦ **I. način pomoću kalkulatora raspodjele vjerojatnosti**

➢ **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - ➢ **Distribution:** ***Z (Normal)***
  - ➢ **Fixed Scaling – ukloniti oznaku**
  - ➢ ***(1 – Cumulative p)***
    - ➢ **X:** ***8***
    - ➢ **mean:** ***12***
    - ➢ **st.dev.:** ***2***

- rezultat = 0,977250 = 97,7%

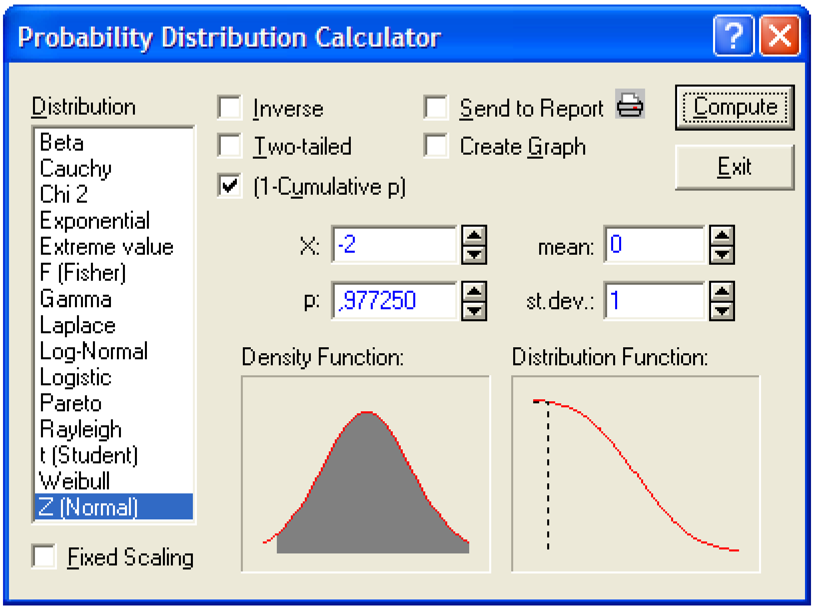


Vjerojatnost da slučajna varijabla poprimi vrijednost veću od 8 iznosi 0,977 ili 97,7%.

B) Pomoću standardizirane raspodjele:

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution: *Z (Normal)***
  - **Fixed Scaling – ukloniti oznaku**
  - ***(1 – Cumulative p)***
    - **X: *-2***
    - **mean: *0***
    - **st.dev.: *1***

- rezultat = 0,977250 = 97,7%



Rezultat je isti i iznosi 0,977 ili 97,7%.

♦ **II. Način upotrebom funkcija u radnom listu**

➢ **Definirati novi stupac – "Qi"**

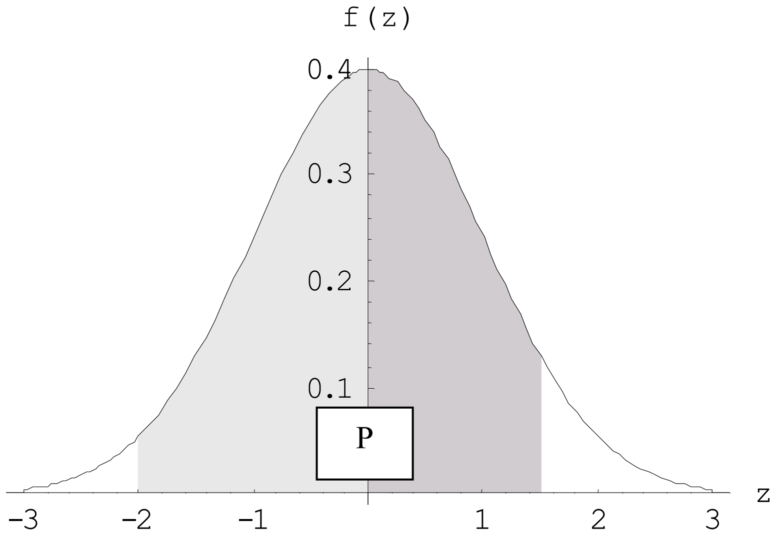
➢ **Long name:** $\boxed{= 1 - Pxi}$

xi	*Qi*
8	0,97725
15	0,066807

---

e) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost između 8 i 15?

To je površina osjenčanog dijela:



- **Definirati novu varijablu – P_razlika**
  - **Long name:** ***=0,933193 - 0,02275***

- rezultat = 0,910443 = 91,0%

- ili
  - **Long name:**

    =INormal(15;12;2)-INormal(8;12;2)

Vjerojatnost da slučajna varijabla poprimi vrijednost između 8 i 15 je 0,910 ili 91,0%.

Ista stvar može se izračunati:
A) pomoću zadanih vrijednosti i
B) pomoću standardiziranih vrijednosti:

- ***=INormal (xi; 12; 2)* -> *Pxi* - A)**

- **Standardizirana raspodjela ima prosjek = 0 i standardnu devijaciju =1:**
- ***=INormal (zi; 0; 1)* -> *Pzi* -B)**

Ako nas ne zanima "P", nego samo "Q" onda
- ***=1 - INormal (xi; 12; 2)* -> *Qxi* - A)**
- ***=1 - INorlmal (zi; 0; 1)* -> *Qzi* -B)**

Primjer 2.2.

Vjerojatnost da slučajna varijabla poprimi vrijednost manju od "x" iznosi:
a) 0,02275
b) 0,933193.
Koliki je "x"? Koliki je "z"?

Inverzna funkcija:
I. način pomoću kalkulatora raspodjele vjerojatnosti

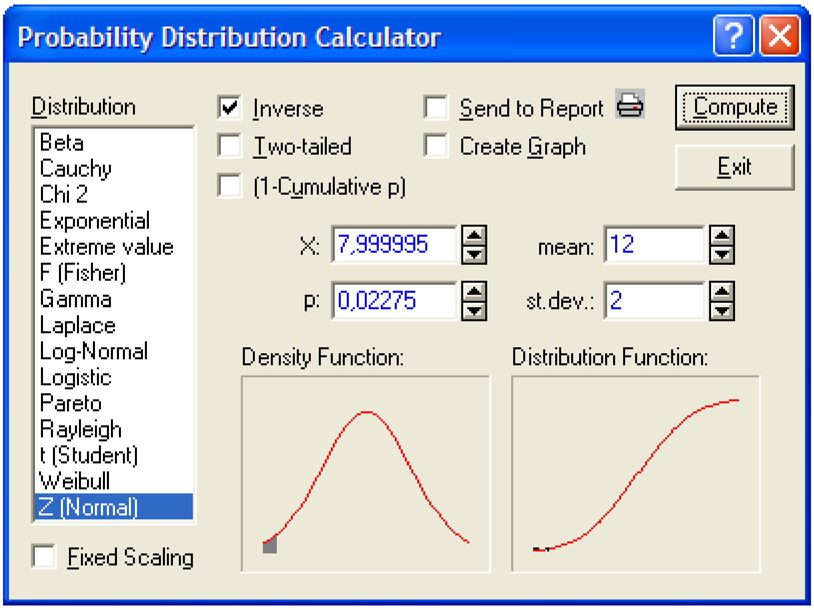
A) Uporabom zadanih vrijednosti (**x**)

a)

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Z (Normal)***
  - **Distribution:** ***Inverse***
  - **Fixed Scaling - ukloniti oznaku**
    - **p:** ***0,02275***
    - **mean:** ***12***
    - **st.dev.:** ***2***

rezultat = 7,999995 = 8

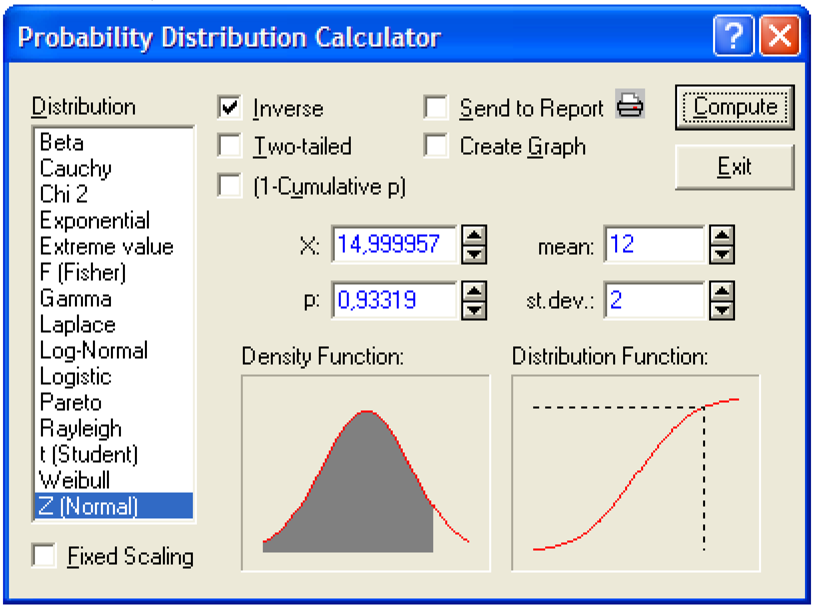
X = 8.



b)

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** *Z (Normal)*
  - **Distribution:** *Inverse*
  - **Fixed Scaling - ukloniti oznaku**
    - **p:** *0,933193*
    - **mean:** *12*
    - **st.dev.:** *2*

rezultat = 15,000003 = 15



X = 15.

B) Pomoću standardizirane raspodjele (**z**)**:**

a)

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Z (Normal)***
  - **Distribution:** ***Inverse***
  - **Fixed Scaling - ukloniti oznaku**
    - **p:** ***0,02275***
    - **mean:** ***0***
    - **st.dev.:** ***1***

rezultat = -2,000002 = -2

z = -2.

b)

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Z (Normal)***
  - **Distribution:** ***Inverse***
  - **Fixed Scaling - ukloniti oznaku**
    - **p:** ***0,933193***
    - **mean:** ***0***
    - **st.dev.:** ***1***

rezultat = 1,500002 = 1,5

z = 1,5.

II. način upotrebom funkcija u radnom listu

A) Uporabom zadanih vrijednosti (**x**)

- **Definirati novi stupac – "x"**
  - **Long name:**
    - **=VNormal (x; mu; sigma)**
    - **=VNormal (Pi; μ; σ)**
    - ***=VNormal (Pxi; 12; 2)***

x
8
15

B) Pomoću standardizirane raspodjele (**z**):

- **Definirati novi stupac – "z"**
  - **Long name:**
    - **=VNormal (x; mu; sigma)**
    - **=VNormal (Pi; μ; σ)**
    - ***=VNormal (Pzi; 0; 1)***

z
-2
1,5

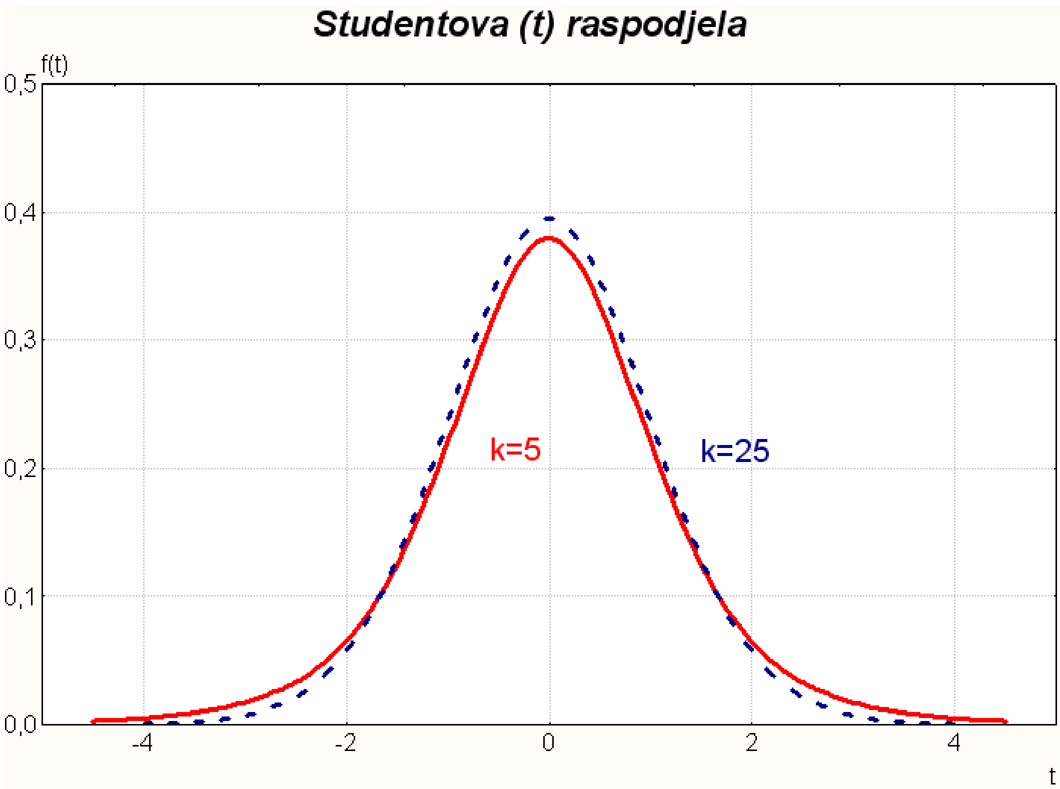
2.2. Studentova (t) raspodjela

Primjer 2.3.

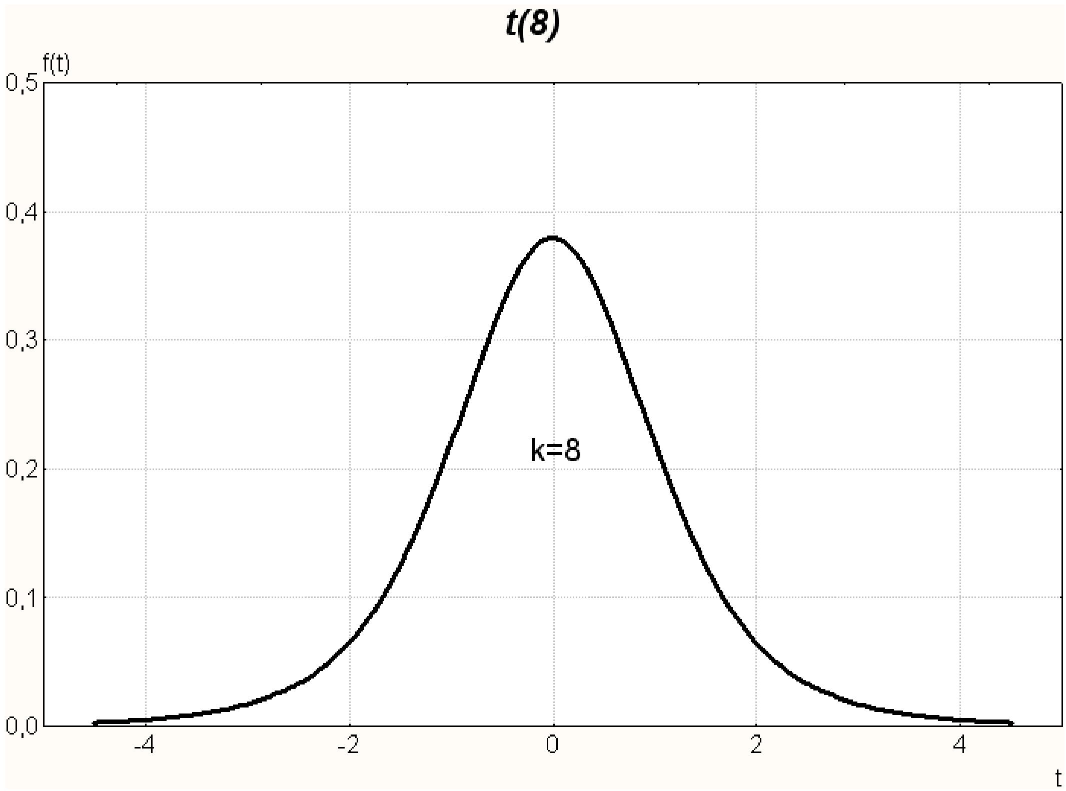
Slučajna varijabla t distribuirana je po Studentovoj raspodjeli s 8 stupnjeva slobode. Odredite vjerojatnost da je slučajna varijabla:

a) manja od 1,397
b) manja od 1,367
c) manja od -1,397
d) veća od 1,397
e) veća od 1,367
f) veća od - 1,397
g) između 1,397 do 1,860.

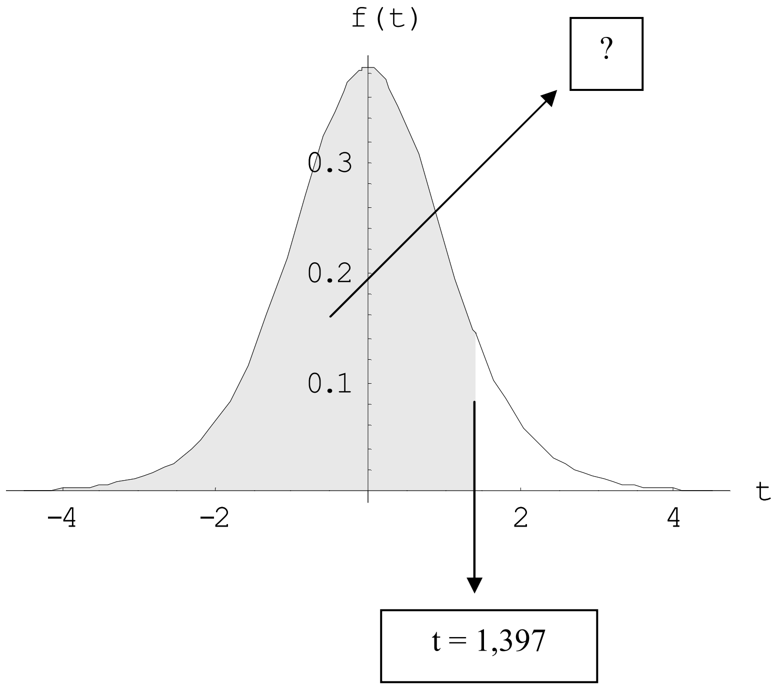
Kod Studentove raspodjele slučajna varijabla se označava "***t***", a kod svih ostalih raspodjela označava se "$\boldsymbol{x_i}$".



Studentova raspodjela jednoznačno je određena brojem ***stupnjeva slobode k*** (= ν = df = degrees of freedom). Studentova raspodjela označava se s ***t(k)***. Zadana raspodjela može se označiti **t(8)**.

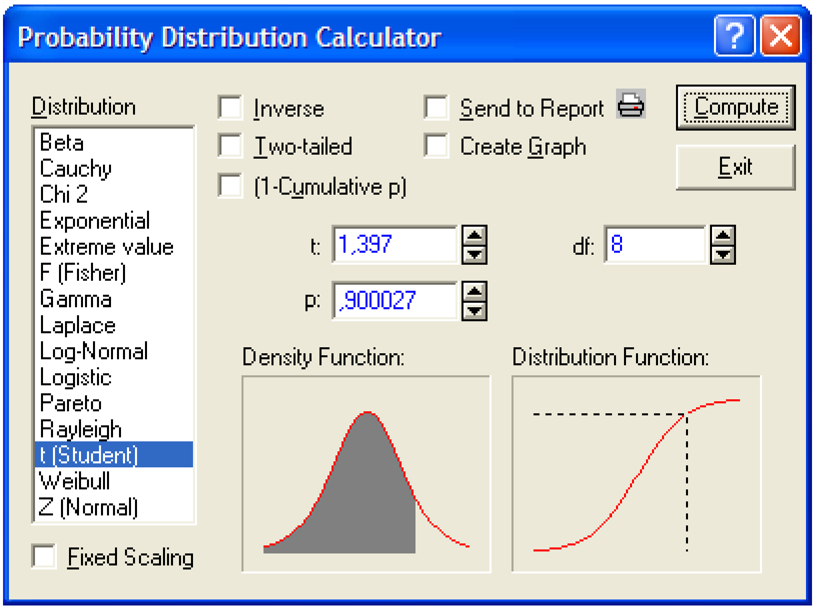


a) Kolika je vjerojatnost da je slučajna varijabla manja od 1,397?



I. način pomoću kalkulatora vjerojatnosti

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***t (Student)***
    - **t:** ***1,397***
    - **df:** ***8***

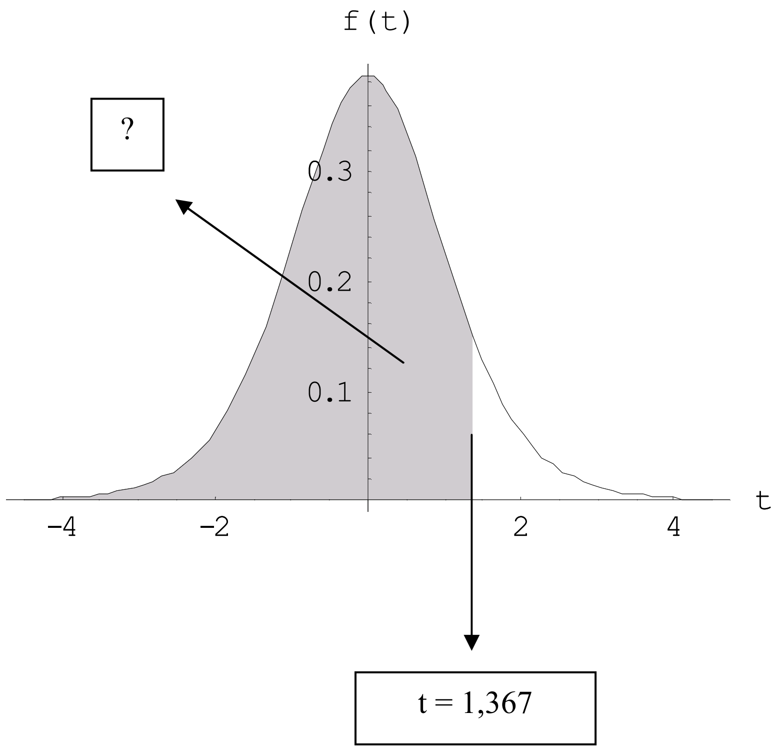


➢ **Fixed Scaling – ukloniti oznaku (za cjeloviti grafički prikaz)**

- rezultat = p = 0,900027 = 90,0%

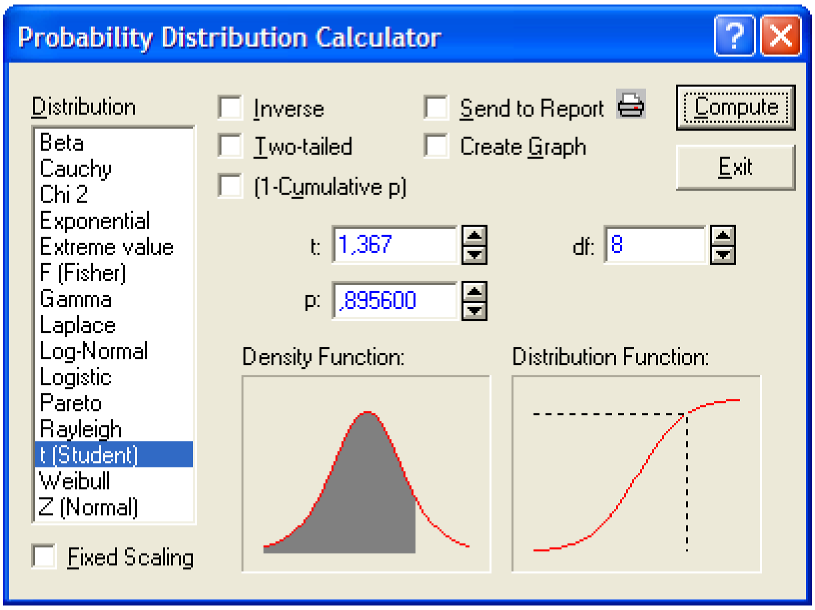
Vjerojatnost da je slučajna varijabla manja od 1,397 iznosi 90,0%.

b) Kolika je vjerojatnost da je slučajna varijabla manja od 1,367?



Promijeni se vrijednost t:

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***t (Student)***
    - **t:** ***1,367***
    - **df:** ***8***



- rezultat = 0,895600= 89,6%

Vjerojatnost da je slučajna varijabla manja od 1,367 iznosi 89,6%.

II. način upotrebom funkcija u radnom listu

Općenito kada se traži vjerojatnost da je slučajna varijabla manja od neke vrijednosti treba:
- zatvoriti prozor: Probability Distribution Calculator
- otići na prozor s podacima:

- **Definirati stupac – "t"**
- **Unijeti vrijednosti:**

t
1,397
1,367
-1,397
1,860

- **Definirati novi stupac – "Pi"**
- **Long name:**
  - **=IStudent (x; df)**
  - **=IStudent (t; k)**
  - ***=IStudent (t; 8)***

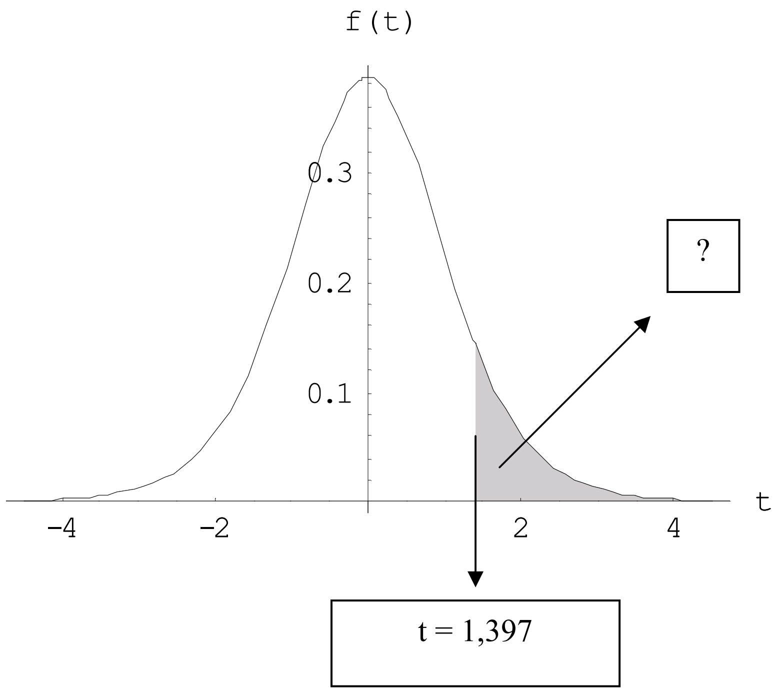
t	*Pi*
1,397	0,900027
1,367	0,8956
-1,397	0,099973
1,860	0,950035

c) Kolika je vjerojatnost da je slučajna varijabla manja od -1,397?

Već je izračunato
- 0,099973 = 10,0 %

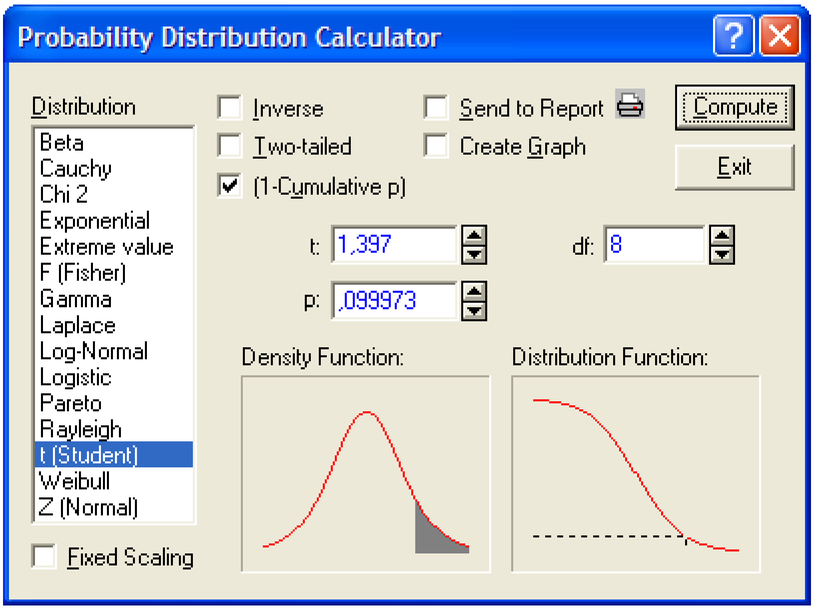
Vjerojatnost da je slučajna varijabla manja od -1,397 iznosi 10%.

d) Kolika je vjerojatnost da je slučajna varijabla veća od 1,397?



- **I. način pomoću kalkulatora raspodjele vjerojatnosti**

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***t (Student)***
  - ***(1 – Cumulative p)***
    - **t:** ***1,397***
    - **df:** ***8***



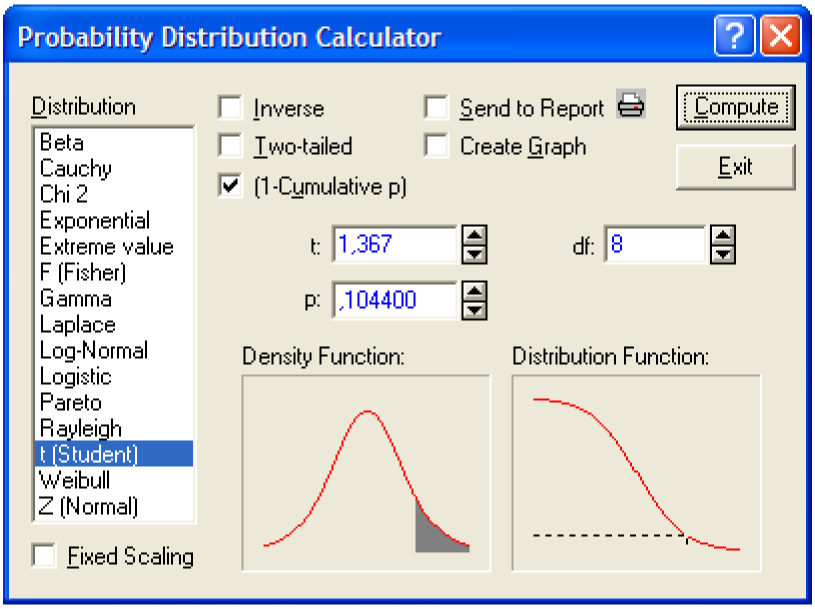
- rezultat = 0,099973 = 10,0%

Vjerojatnost da je slučajna varijabla veća od 1,397 iznosi 10%.

e) Kolika je vjerojatnost da je slučajna varijabla veća od 1,367?

➢ **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**

- ➢ **Distribution:** ***t (Student)***
- ➢ ***(1 – Cumulative p)***
  - ➢ **t:** ***1,367***
  - ➢ **df:** ***8***



- rezultat = 0,104400= 10,4%

Vjerojatnost da je slučajna varijabla veća od 1,367 iznosi 10,4%.

♦ **II. način upotrebom funkcija u radnom listu**

➢ **Definirati novi stupac – "Qi"**

➢ **Long name:** ***= 1 - Pi***

t	*Qi*
1,397	0,099973
1,367	0,1044
-1,397	0,900027

1,860	0,049965

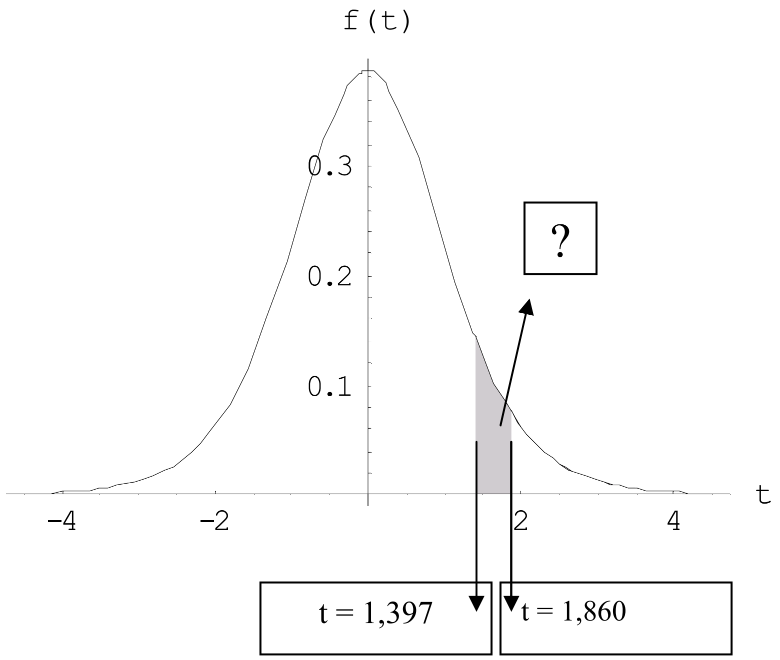
f) Kolika je vjerojatnost da je slučajna varijabla veća od -1,397?

- 0,900 = 90,0 %

Vjerojatnost da je slučajna varijabla veća od - 1,397 iznosi 90,0%.

Qi se može direktno izračunati kao ***1 - IStudent (t,8)***. Izračunati i usporediti.

g) Kolika je vjerojatnost da je slučajna varijabla u rasponu od 1,397 do 1,860?



- **Definirati novu varijablu – "P_razlika"**
- **Long name:**

  =IStudent(1,860;8)-IStudent(1,397;8)

- rezultat = 0,050008 = 5,00%

Vjerojatnost da će slučajna varijabla poprimiti vrijednost između 1,397 i 1,860 iznosi 5%.

Primjer 2.4.

Slučajna varijabla "t" distribuirana je po Studentovoj raspodjeli s 8 stupnjeva slobode. Vjerojatnost da je slučajna varijabla manja od "t" iznosi:

a) 0,900027
b) 0,8956
c) 0, 099973
d) 0,950035.

Odredite t.

I. način pomoću kalkulatora vjerojatnosti

a)

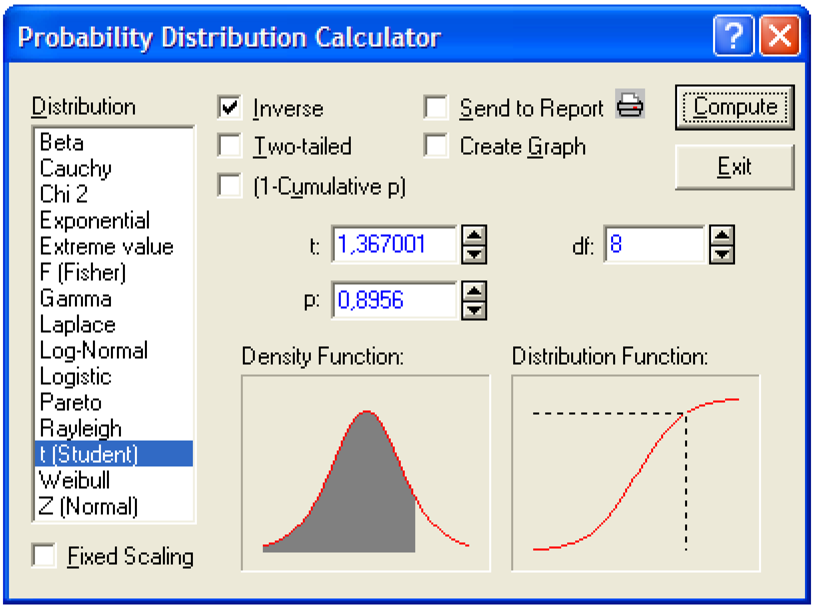
- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***t (Student)***
  - ***Inverse***
    - **p:** ***0,900027***
    - **df:** ***8***

rezultat = 1,397002

t = 1,397002.

b)

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***t (Student)***
  - ***Inverse***
    - **p:** ***0,8956***
    - **df:** ***8***



rezultat = 1,367001

t = 1,367001.

c)

rezultat = -1,397002

t = -1,397.

d)

rezultat = 1,860004

t = 1,860.

II. način upotrebom funkcija u radnom listu

- **Definirati novi stupac – “t1”**
- **Long name:**
  - **=VStudent (x; df)**
  - **=VStudent (Pi; k)**
  - ***=VStudent (Pi; 8)***

t1
1,397
1,367
-1,397
1,86

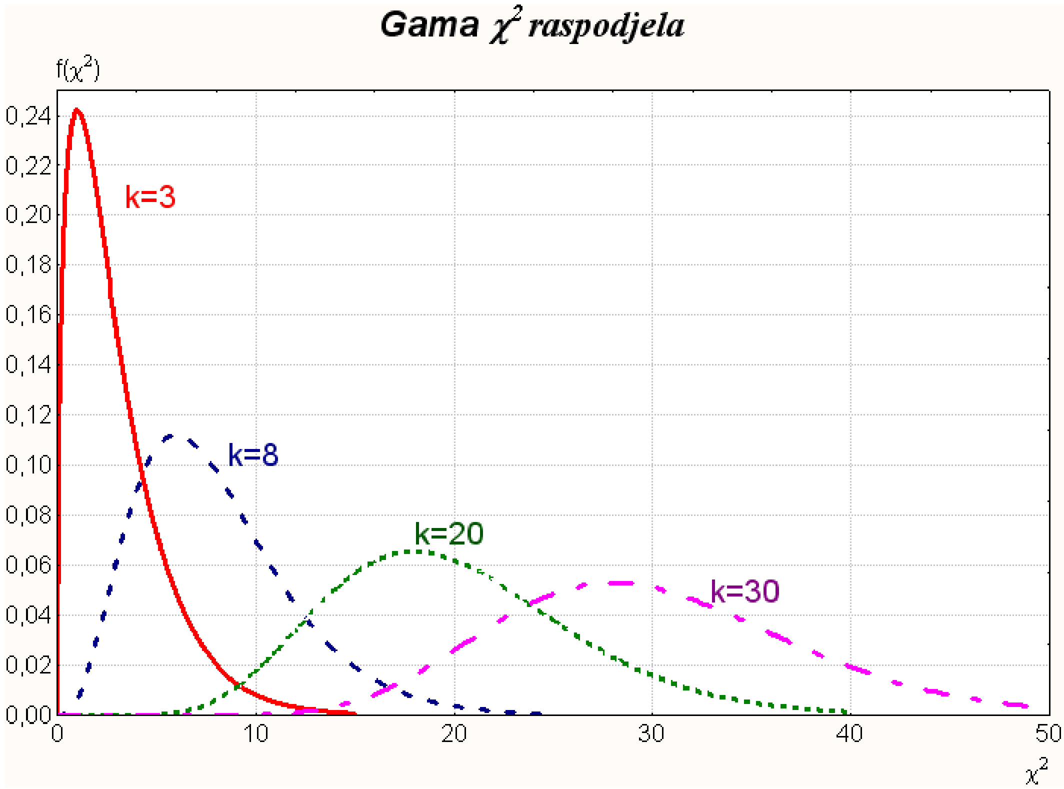
a) t = 1,397
b) t = 1,367
c) t = -1,397
d) t = 1,860.

2.3. Hi-kvadrat raspodjela

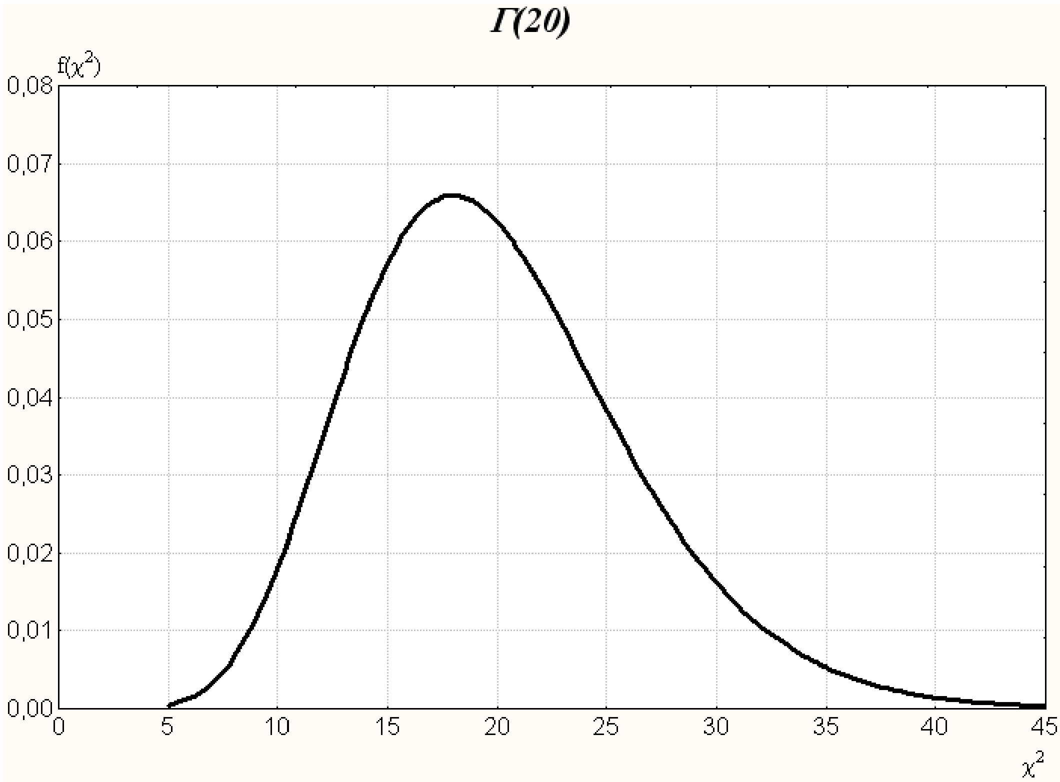
Primjer 2.5.

Slučajna varijabla “X” pripada $\chi 2$ raspodjeli s 20 stupnjeva slobode. Zadatak je odrediti vjerojatnosti da je slučajna varijabla:

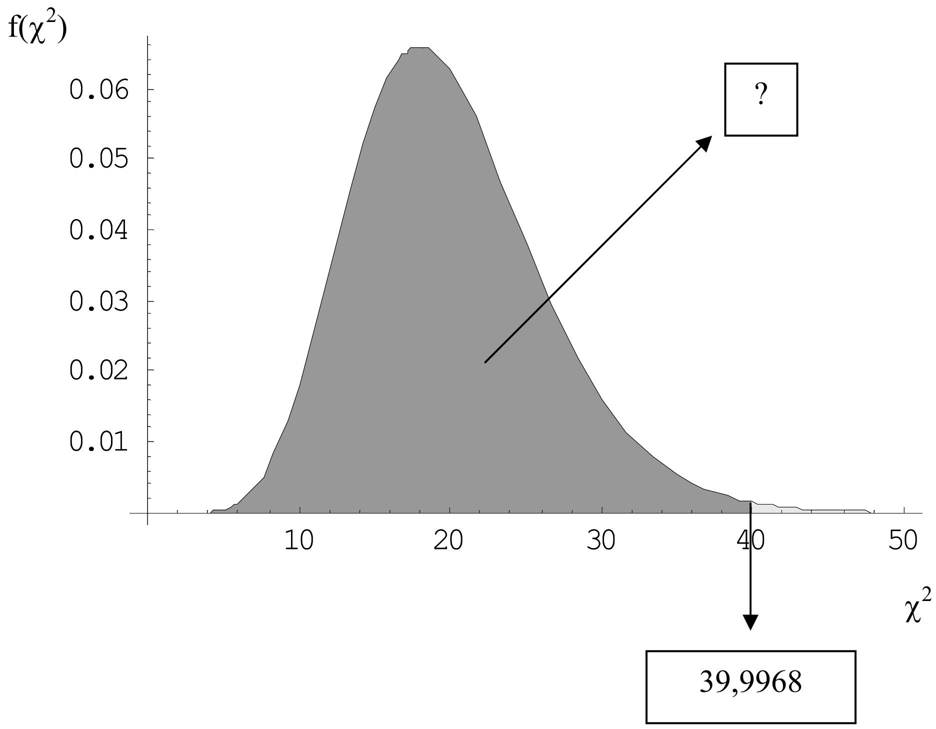
a) manja od 39,9968
b) manja od 34,1696
c) manja od 10,8508
d) manja od 7,43386
e) veća od 39,9968
f) veća od 34,1696
g) veća od 10,8508
h) veća od 7,43386
i) u rasponu od 34,1696 do 39,9968.



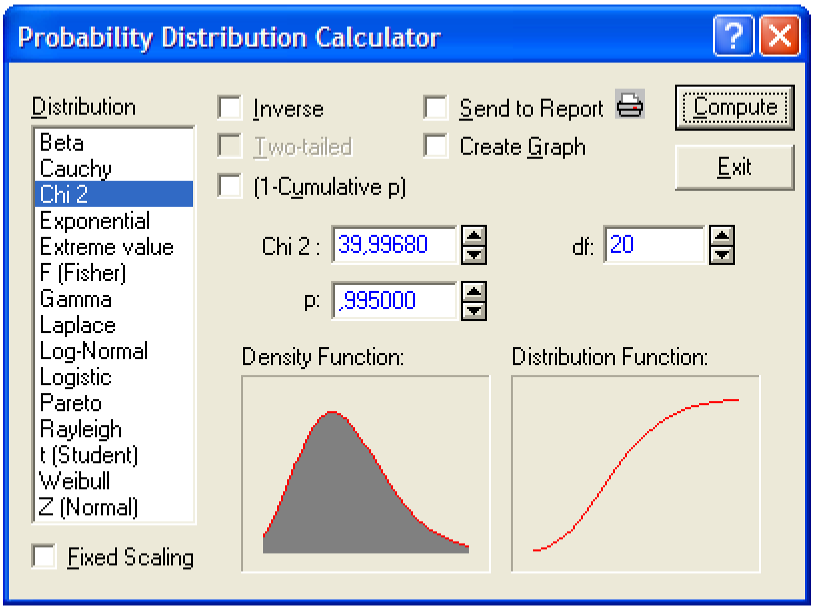
χ^2 raspodjela jednoznačno je određena brojem ***stupnjeva slobode "υ" ili «k»***. Označava se Γ***(k)***. Zadana raspodjela može se označiti **Γ(20)**.


Γ(20)
f(χ²)
0,08
0,07
0,06
0,05
0,04
0,03
0,02
0,01
0,00
0
5
10
15
20
25
30
35
40
45
χ²

a) Odrediti vjerojatnost da je slučajna varijabla manja od 39,9968.



- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution: *Chi 2***
  - **Fixed Scaling - ukloniti oznaku**
    - **Chi 2: *39,99680***
    - **df: *20***

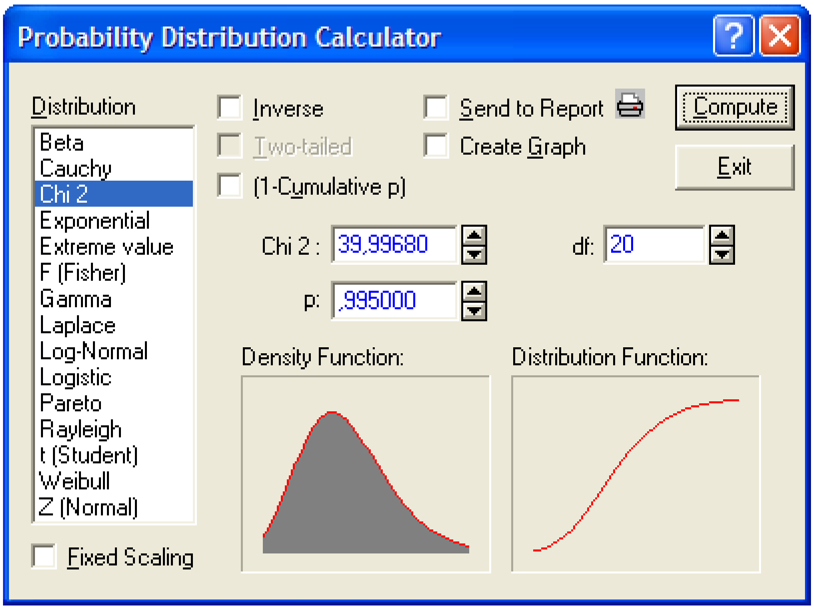


- rezultat = 0,995000 = 99,5%

Vjerojatnost da je slučajna varijabla manja od 39,9968 iznosi 99,5%.

b) Odrediti vjerojatnost da je slučajna varijabla manja od 34,1696.

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Chi 2***
  - **Fixed Scaling - ukloniti oznaku**
    - **Chi 2:** ***34,1696***
    - **df:** ***20***



- rezultat = 0,975000 = 97,5%

Vjerojatnost da je slučajna varijabla manja od 34,1696 iznosi 97,5%.

Općenito kada se traži vjerojatnost da je slučajna varijabla manja od neke vrijednosti treba:

- zatvoriti prozor: Probability Distribution Calculator
- izabrati prozor s podacima:

- **Definirati stupac – "xi"**
  - dugi naziv: "***HI2***"

- **Unijeti vrijednosti:**

xi
39,9968
34,1696
10,8508
7,43386

- **Definirati novi stupac – "Pi"**
- **Long name:**
  - **=IChi2 (x; nu)**
  - **=IChi2 (xi; k)**
  - ***=IChi2 (xi; 20)***

xi	*Pi*
39,9968	0,995
34,1696	0,975
10,8508	0,05
7,43386	0,005

---

c) Odrediti vjerojatnost da je slučajna varijabla manja od 10,8508.

- izračunato
- 0,050 = 5,0 %

Vjerojatnost da je slučajna varijabla manja od 10,8508 iznosi 5,0%.

---

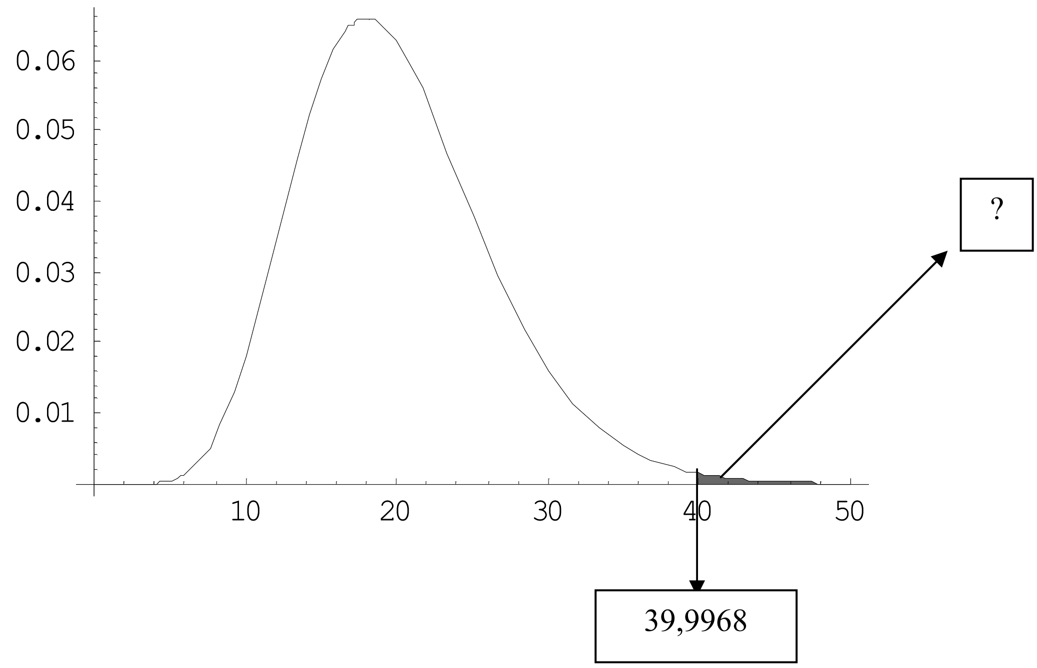
d) Odrediti vjerojatnost da je slučajna varijabla manja od 7,43386.

- izračunato
- 0,005 = 0,5 %

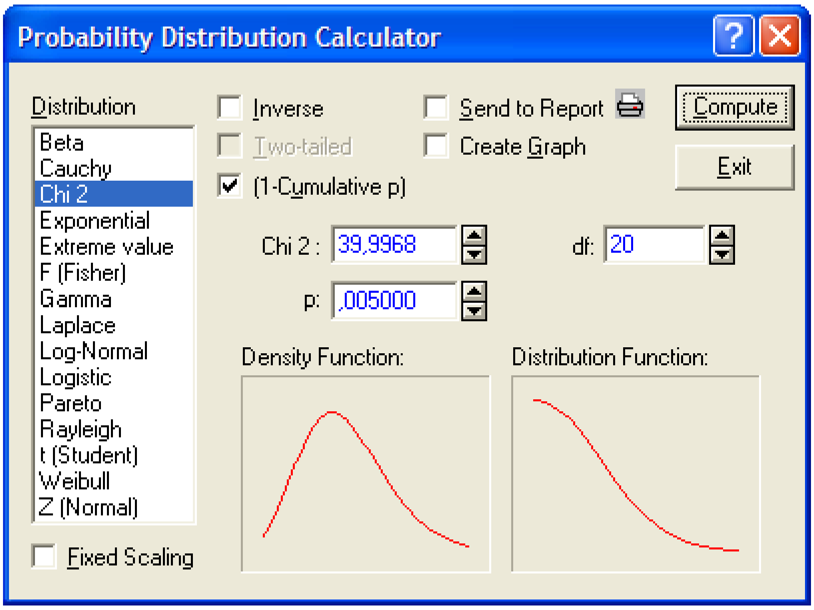
Vjerojatnost da je slučajna varijabla manja od 7,43386 iznosi 0,5%.

---

e) Odrediti vjerojatnost da je slučajna varijabla veća od 39,9968.



- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Chi 2***
  - **Fixed Scaling - ukloniti oznaku**
    - ***(1 – Cumulative p)***
    - **Chi 2:** ***39,9968***
    - **df:** ***20***



- rezultat = 0,005 = 0,5%

Vjerojatnost da je slučajna varijabla veća od 39,99680 iznosi 0,5%.

- **Definirati novi stupac – Qi**
- **Long name:** ***= 1 - Pi***

xi	*Qi*
39,9968	0,005
34,1696	0,025
10,8508	0,95
7,43386	0,995

Vjerojatnost da je slučajna varijabla veća od 39,99680 iznosi 0,5%.

f) Odrediti vjerojatnost da je slučajna varijabla veća od 34,1696.

- 0,025 = 2,50 %

Vjerojatnost da je slučajna varijabla veća od 34,1696 iznosi 2,5%.

g) Odrediti vjerojatnost da je slučajna varijabla veća od 10,8508.

- 0,950 = 95 %

Vjerojatnost da je slučajna varijabla veća od 10,8508 iznosi 95,0%.

---

h) Odrediti vjerojatnost da je slučajna varijabla veća od 7,43386.

- 0,995 = 99,5 %

Vjerojatnost da je slučajna varijabla veća od 7,43386 iznosi 99,5%.

Dobiveni rezultati mogu se prikazati i u obliku postotka:

- Selektirati varijablu

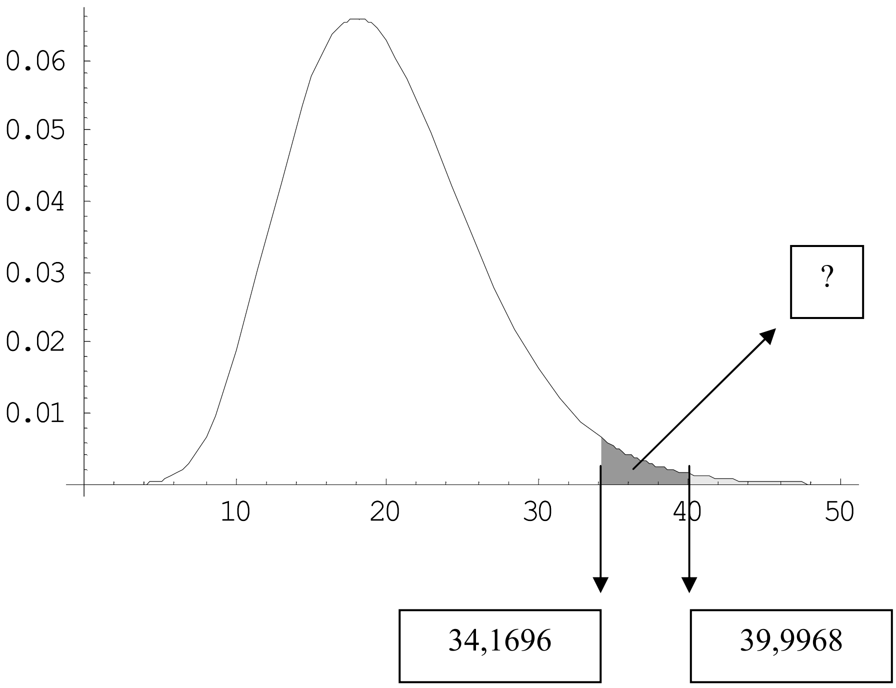
- Definirati varijable
  - **Display format** $\boxed{\textit{Percentage}}$
    - **Decimal places:** $\boxed{\textit{1}}$

xi	*Qi*
39,9968	0,5%
34,1696	2,5%
10,8508	95,0%
7,43386	99,5%

“Qi” se može izravno izračunati kao ***1 – IChi2(x,nu)***.

---

i) Odrediti vjerojatnost da je slučajna varijabla u rasponu od 34,1696 do 39,9968.



- **Definirati novu varijablu – P_razlika**
- **Long name:**

  =IChi2(39,9968;20)-IChi2(34,1696;20)

- rezultat = 0,02 = 2%

Vjerojatnost da je slučajna varijabla između 34,1696 i 39,9968 iznosi 2,0%.

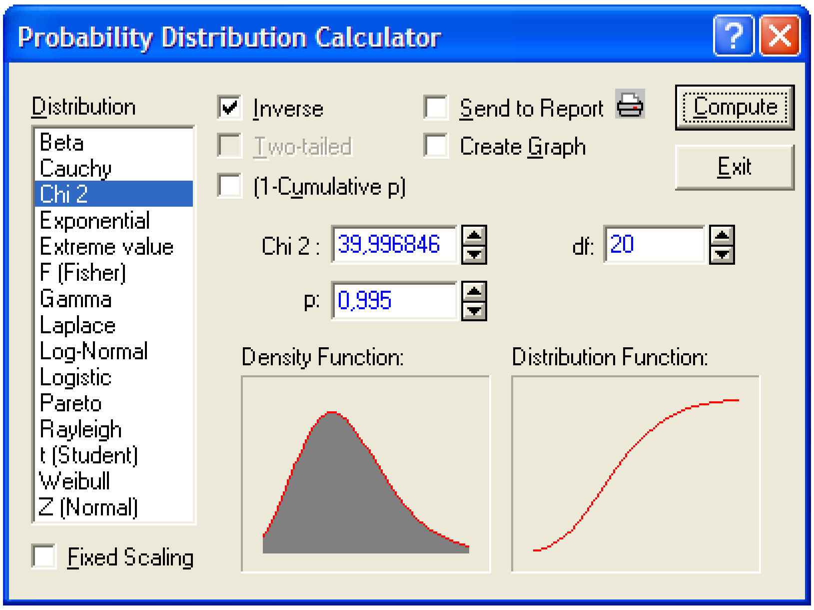
Primjer 2.6.

Slučajna varijabla "X" pripada χ^2 raspodjeli s 20 stupnjeva slobode. Vjerojatnost da je slučajna varijabla manja od X iznosi:

a) 0,995
b) 0,975
c) 0,05
d) 0,005.

Koliki je "X"?

- **STATISTICS / PROBABILITY CALCULATOR / DISTRIBUTIONS**
  - **Distribution:** ***Chi 2***
  - **Fixed Scaling – ukloniti oznaku**
  - ***Inverse***
    - **df:** ***20***
    - **P:** ***0,995***



- rezultat - 39,996846 = 39,9968

- **Definirati novi stupac – "x"**
- **Long name:**
  - **=VChi2 (x; nu)**
  - **=VChi2 (Pi; k)**
  - ***=VChi2 (Pi; 20)***

xi
39,9968
34,1696
10,8508
7,43386

a) 39,9968
b) 34,1696
c) 10,8508
d) 7,43386.

3. Izabrane distribucije vjerojatnosti diskretne slučajne varijable

Kod računanja vjerojatnosti diskretne slučajne varijable moguće je upotrebljavati funkcije u radnom listu, a kalkulator vjerojatnosti nije na raspolaganju. Kao i kod kontinuiranih raspodjela rezultati se mogu istovremeno izračunati za više zadanih vrijednosti.

Ako su zadane vrijednosti x1, x2, x3,…diskretne slučajne varijable koje su upisane u stupac xi:

xi
x1
x2
x3
…

može se odrediti vjerojatnost da će:

a) slučajna varijabla poprimiti zadanu vrijednost P = P(X = x)

- **Definirati novi stupac – "Pi"**
- **Long name (label or formula with Functions):**

  - U rubrici dugi naziv izabere se funkcija nazvana prema raspodjeli i nema nikakvog prefiksa u nazivu, npr.: ***Binom***, ***Poisson***,…Unose se traženi parametri funkcije: vjerojatnost, broj elemenata, aritmetička sredina…, a varijabla "***x***" predstavlja parametar x.

b) slučajna varijabla poprimiti vrijednost manju ili jednaku zadanoj vrijednosti F = P(X ≤ x)

- **Definirati novi stupac – "Fi"**
- **Long name (label or formula with Functions):**

  - Izabere se funkcija koja ima prefiks ***I***: ***IBinom***, ***IPoisson***,…

c) slučajna varijabla ne će poprimiti zadanu vrijednost Q = P(X ≠ x)

- **Definirati novi stupac – "Qi"**
- **Long name (label or formula with Functions): =1-Pi**

  - Izračunat će se suprotna vjerojatnost, tj. vjerojatnost da će slučajna varijabla poprimiti vrijednost različitu od zadanog "x"-a.
  - Qi moguće je direktno dobiti unoseći ***1-Bionm***, ***1-Poisson*** i sl.

d) slučajna varijabla poprimiti vrijednost veću od zadane vrijednosti QF = P(X > x)

- **Definirati novi stupac – “QFi”**
- **Long name (label or formula with $\boxed{\text{Functions}}$): $\boxed{\text{=1-Qi}}$**

  - Izračunat će se vjerojatnost da će slučajna varijabla poprimiti vrijednost veću od zadanog “x”-a.
  - Qi moguće je direktno dobiti unoseći ***1-IBionm***, ***1-IPoisson*** i sl.

    e) slučajna varijabla poprimiti vrijednost jednaku x1 ili x2 ili x3…

  - Zbroje se vjerojatnosti P(x1)+P(x2)+P(x3)+… Ovo se može napraviti u novom stupcu ili izvan programa *Statistica*

3.1. Binomna raspodjela

Binomna raspodjela označava se s: $B(n; p)$

n - broj pokusa, veličina uzorka
p - vjerojatnost nastupa događaja A

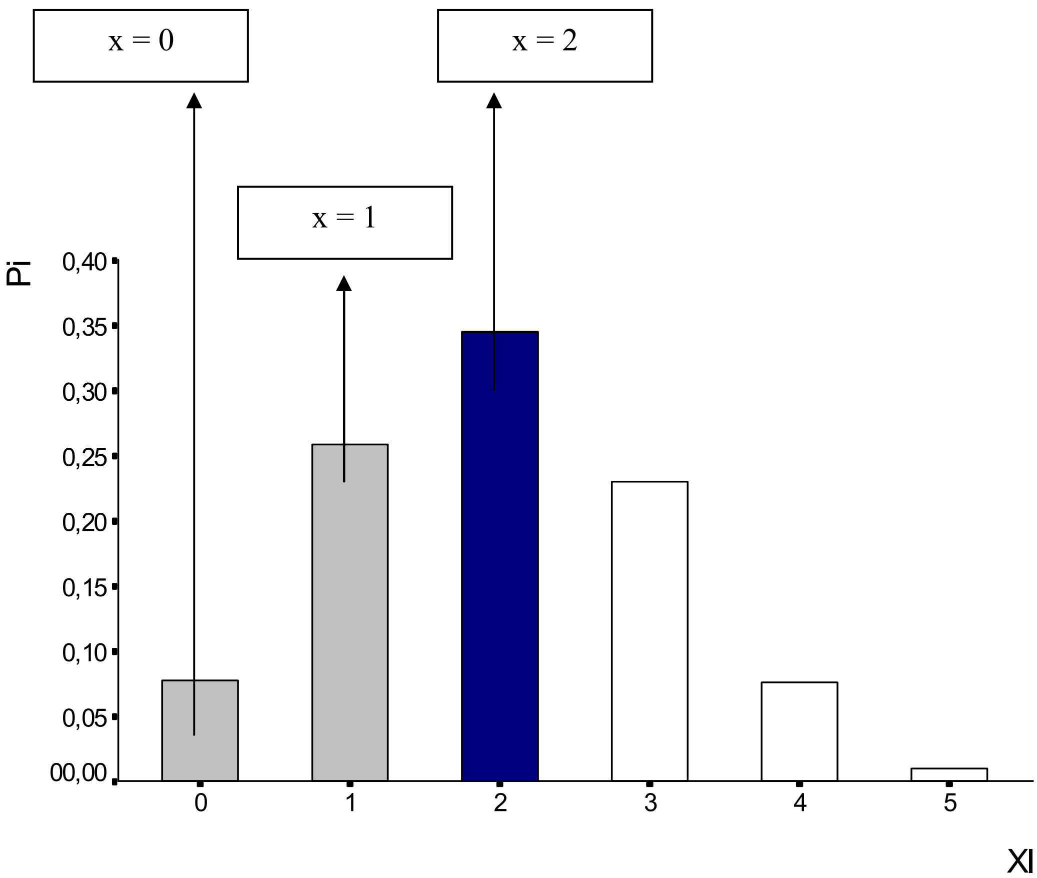
Primjer 3.1.

Slučajna varijabla “X” ravna se po zakonima binomne raspodjele B(5; 0,4). Kolika je vjerojatnost da slučajna varijabla poprimi vrijednosti:

a) $x = 0$, $x = 1$, $x = 2$, $x = 3$, $x = 4$, $x = 5$
b) $x \leq 0$, $x \leq 1$, $x \leq 2$, $x \leq 3$, $x \leq 4$, $x \leq 5$
c) $x \neq 0$, $x \neq 1$, $x \neq 2$, $x \neq 3$, $x \neq 4$, $x \neq 5$
d) $x > 0$, $x > 1$, $x > 2$, $x > 3$, $x > 4$, $x > 5$

a) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost:
$x = 0$, $x = 1$, $x = 2$, $x = 3$, $x = 4$, $x = 5$?

Slučajna varijabla može poprimiti samo vrijednosti 0, 1, 2, 3, 4 ili 5.



- **Definirati varijablu – "xi"**
- **Unijeti vrijednosti:**

xi
0
1
2
3
4
5

- **Definirati novi stupac – "Pi"**
- **Long name:**
  - **=Binom (x; p; n)**
  - **=Binom (xi; p; n)**
  - ***=Binom (xi; 0,4; 5)***

xi	*Pi*
0	0,07776
1	0,2592
2	0,3456
3	0,2304
4	0,0768
5	0,01024

Neka slučajna varijabla X predstavlja broj neispravnih proizvoda u skupu od ukupno 5 proizvoda.

Vjerojatnost da je x = 0 ili vjerojatnost da nemamo neispravnih proizvoda ili da su svi proizvodi ispravni iznosi 7,8%.

Vjerojatnost da je x = 1 ili vjerojatnost da je točno jedan proizvod neispravan iznosi 25,9%.

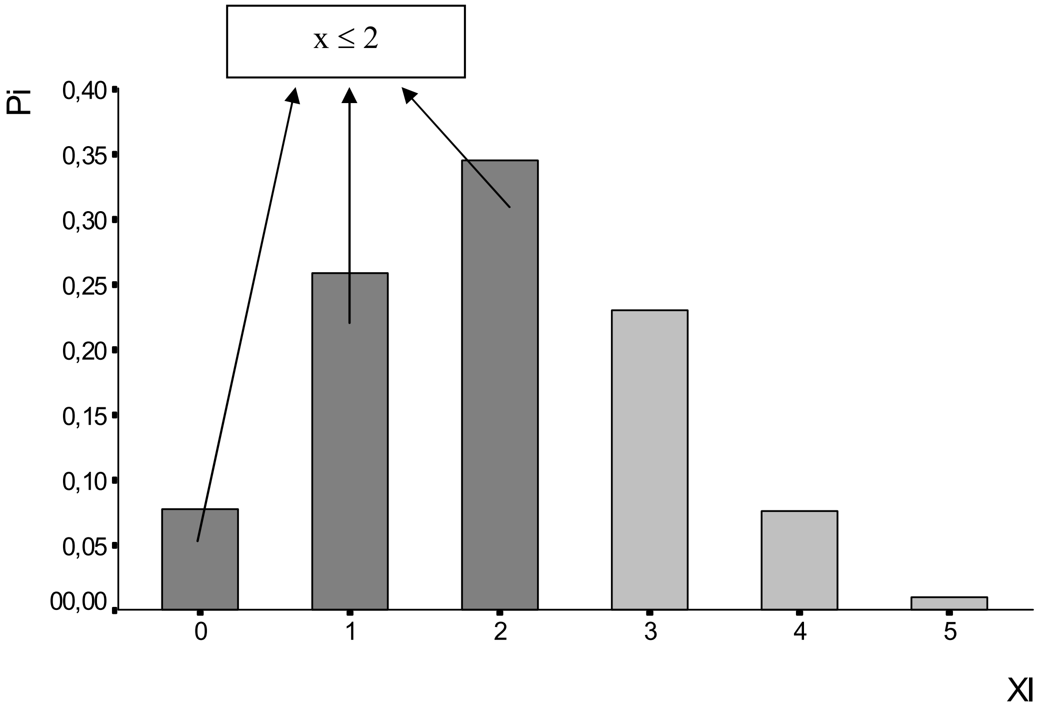
Vjerojatnost da su dva proizvoda neispravna je (x = 2) iznosi 34,6%.

Vjerojatnost da su neispravna 3 proizvoda od ukupno 5 (x = 3) iznosi 23,0%.

Vjerojatnost da imamo 4 neispravna proizvoda (x = 4) iznosi 7,7%.

Vjerojatnost da je x = 5 tj. vjerojatnost da je 5 proizvoda neispravno ili vjerojatnost da su svi proizvodi neispravni, odnosno, vjerojatnost da ne postoji ispravan proizvod iznosi 1,0%

b) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost: $x \leq 0$, $x \leq 1$, $x \leq 2$, $x \leq 3$, $x \leq 4$, $x \leq 5$?



- **Definirati novi stupac – "Fi"**
- **Long name:**
  - **=IBinom (x; p; n)**
  - **=IBinom (xi; p; n)**
  - ***=IBinom (xi; 0,4; 5)***

xi	***Fi***
0	0,07776
1	0,33696
2	0,68256
3	0,91296
4	0,98976
5	1

Vjerojatnost da je $x \leq 0$ iznosi 0,07776 ili 7,8%.

Vjerojatnost da je $x \leq 0$ je vjerojatnost da je broj neispravnih proizvoda manji ili jednak nuli, tj. vjerojatnost da su svi proizvodi ispravni, iznosi 7,8%.

Vjerojatnost da je $x \leq 1$, tj. vjerojatnost da je 1 ili manje (1 ili 0) neispravnih proizvoda, iznosi 33,7%.

Vjerojatnost da je $x \leq 2$, tj. vjerojatnost da je 2 ili manje (2 ili 1 ili 0) neispravnih proizvoda, iznosi 68,3%.

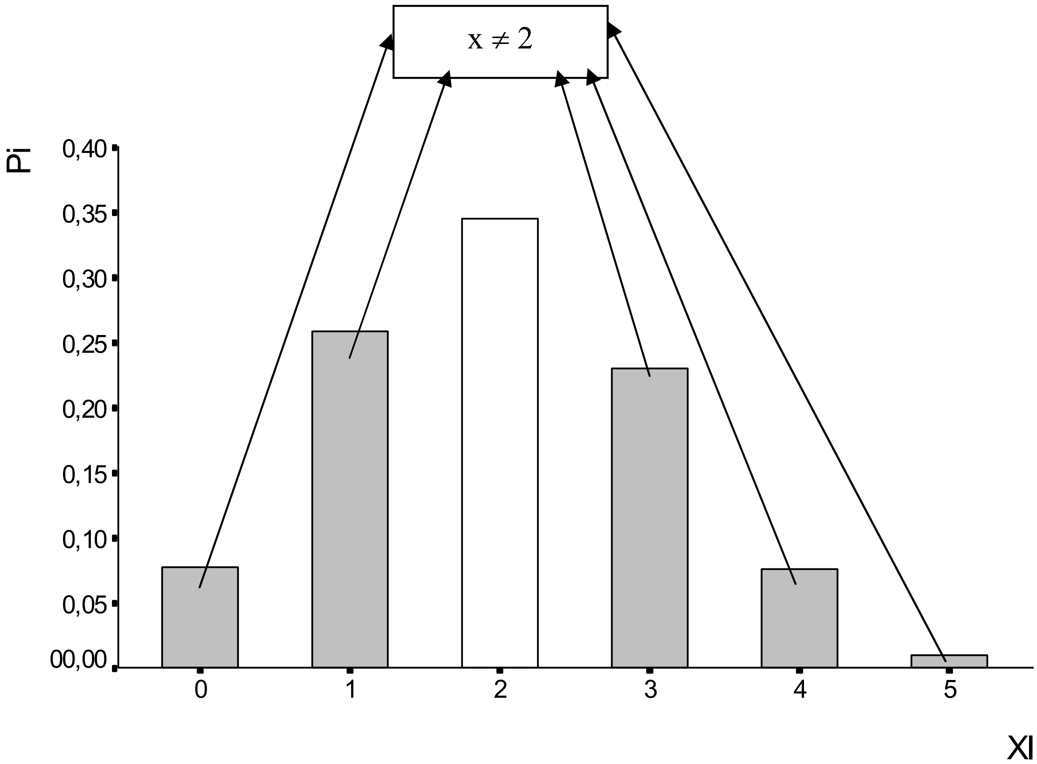
Vjerojatnost da je $x \leq 3$, tj. vjerojatnost da je najviše 3 (3, 2, 1 ili 0) neispravna proizvoda, iznosi 91,3%.

Vjerojatnost da je $x \leq 4$, tj. vjerojatnost da je 4 ili manje (4, 3, 2, 1 ili 0) neispravnih proizvoda, iznosi 99,0%.

Vjerojatnost da je $x \leq 5$, tj. vjerojatnost da je 5 ili manje (5, 4, 3, 2, 1 ili 0) neispravnih proizvoda, iznosi 100%.

---

c) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost:
$x \neq 0$, $x \neq 1$, $x \neq 2$, $x \neq 3$, $x \neq 4$, $x \neq 5$?



Vjerojatnost da je $x \neq 2$ je suprotna vjerojatnosti da je $x = 2$.

- **Definirati novi stupac – "Qi"**
- **Long name:** $\boxed{= 1 - Pi}$

xi	*Qi*
0	0,92224
1	0,7408
2	0,6544
3	0,7696
4	0,9232
5	0,98976

Vjerojatnost da je x ≠ 0, tj. vjerojatnost da nema točno 0 neispravnih proizvoda ili vjerojatnost da nisu svi proizvodi ispravni ili vjerojatnost da je bar jedan proizvod neispravan, iznosi 92,2%.

Vjerojatnost da je x ≠ 1 tj. da nije točno jedan proizvod neispravan iznosi 74,1%.

Vjerojatnost da je x≠2 ili vjerojatnost da nemamo točno 2 neispravna proizvoda ili vjerojatnost da nema točno 3 ispravna proizvoda, iznosi 65,4%.

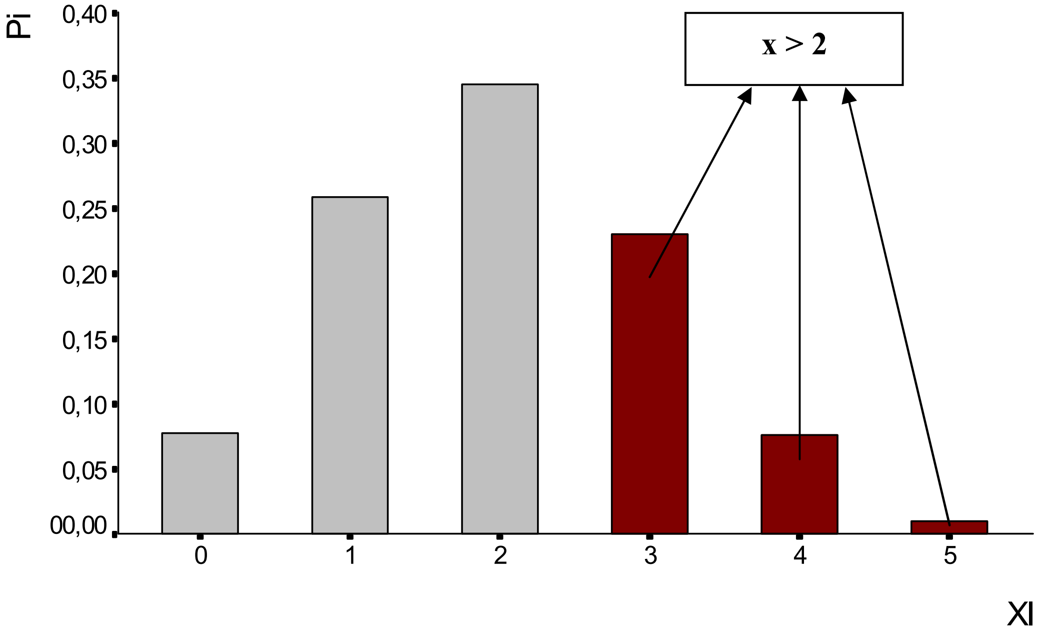
Vjerojatnost da nema točno 3 neispravna proizvoda (x≠3, x=0, 1, 2, 4 ili 5), iznosi 77,0%.

Vjerojatnost da nema točno 4 neispravna proizvoda (x ≠ 4), tj. vjerojatnost da nije točno 1 ispravan proizvod, iznosi 92,3%.

Vjerojatnost da nema točno 5 neispravnih proizvoda (x ≠ 5), tj. vjerojatnost da nisu svi proizvodni neispravni ili vjerojatnost da je bar 1 proizvod ispravan, iznosi 98,0%.

Qi se može izravno izračunati kao ***1 – Binom (x;p;n)***.

d) Kolika je vjerojatnost da slučajna varijabla poprimi vrijednost:
$x > 0$, $x > 1$, $x > 2$, $x > 3$, $x > 4$, $x > 5$?



Vjerojatnost da je $x > 2$ je suprotna vjerojatnosti da je $x \leq 2$.

- **Definirati novi stupac – "QFi"**
- **Long name:** ***= 1 - Fi***

xi	*QFi*
0	0,92224
1	0,66304
2	0,31744
3	0,08704
4	0,01024
5	0

Vjerojatnost da je $x > 0$ ili da je neispravnih proizvoda više od nula ili da postoje neispravni proizvodi ili da nisu svi proizvodi ispravni je 92,2%.

Vjerojatnost da je $x > 1$ ili da je neispravnih proizvoda više od 1 iznosi 66,3%.

Vjerojatnost da je $x > 2$ ili da je $x = 3$, 4 ili 5 iznosi 31,7%.

Vjerojatnost da je $x > 3$, tj. da je neispravnih proizvoda više od 3 ili da je neispravnih proizvoda 4 ili više iznosi 8,7%.

Vjerojatnost da je $x > 4$ ili da je neispravnih proizvoda više od 4 ili da je neispravnih proizvoda 5 ili da su svi proizvodi neispravni ili da nema ispravnih proizvoda iznosi 1,0%.

Vjerojatnost da je $x > 5$ ili da je više od 5 proizvoda neispravno iznosi 0 (ne može biti 6 neispravnih proizvoda ako je ukupno 5 proizvoda).

QFi se može izravno izračunati izračunati kao ***1 – IBinom (x;p;n)***.

PITANJA ZA VJEŽBU

Pi	Fi	Qi	QFi
=	≤	≠	>

1. Kolika je vjerojatnost da su 4 proizvoda neispravna?
   - 7,7% (Pi)
2. Kolika je vjerojatnost da nema točno 4 neispravna proizvoda?
   - 92,3% (Qi)
3. Kolika je vjerojatnost da je 4 ili više neispravnih proizvoda?
   - 8,7% ($x > 3$, Qfi)
4. Kolika je vjerojatnost da je najmanje 3 proizvoda ispravno?
   - Ispravnih 3, 4 , 5
   - *Neispravnih 2, 1, 0*
   - $x \leq 2$
   - 68,3% (Fi)
5. Kolika je vjerojatnost da je najviše 3 proizvoda ispravno?
   - I = 0, 1, 2, 3
   - *N = 5, 4, 3, 2*
   - $x > 1$
   - 66,3% (QFi)
6. Kolika je vjerojatnost da je najviše 2 proizvoda neispravno?
   - $x \leq 2$
   - 68,3% (Fi)

7. Kolika je vjerojatnost da je najviše 2 proizvoda ispravno?
    - I = 0, 1, 2
    - *N = 5, 4, 3*
    - x > 2
    - 31,7% (QFi)
8. Kolika je vjerojatnost da su barem 2 proizvoda ispravna?
    - I = 2, 3, 4, 5
    - *N = 3, 2, 1, 0*
    - x ≤ 3
    - 91,3% (Fi)
9. Kolika je vjerojatnost da je više od 2 proizvoda ispravno?
    - I = 3, 4, 5
    - *N = 2, 1, 0*
    - x ≤ 2
    - 68,3% (Fi)
10. Kolika je vjerojatnost da ima manje od 3 ispravna proizvoda?
    - I = 2, 1, 0
    - *N = 3, 4, 5*
    - x >2
    - 31,7% (QFi)

Primjer 3.2.

Trgovina u dogovoru s proizvođačem daje jednogodišnje jamstvo za fotografske aparate. Prema podacima iz ranijeg razdoblja 18% kupaca javlja se u servis za popravak u jamstvenom roku. Ako je u jednom danu prodano 6 fotografskih aparata i ako je varijabla "X" broj fotografskih aparata prijavljenih u jamstvenom roku, odredite vjerojatnosti slučajne varijable "X".

---

n = 6 fotografskih aparata
p = 18% = 0,18 (vjerojatnost popravka u jamstvenom roku)

Kolika je vjerojatnost da će u jamstvenom roku biti popravljano 0, 1, 2, 3, 4, 5, 6 fotografskih aparata. Kolika je vjerojatnost da je X = 0, 1, 2, 3, 4, 5, 6?

- **Definirati varijablu – "xi"**
- **Unijeti vrijednosti:**

xi
0
1
2
3
4
5
6

- **Definirati novi stupac – "Pi"**
- **Long name:**
  - **=Binom (x; p; n)**
  - **=Binom (xi; p; n)**
  - ***=Binom (xi; 0,18; 6)***

xi	*Pi*
0	0,304007
1	0,400399
2	0,219731
3	0,064312
4	0,010588
5	0,00093
6	0,000034

Vjerojatnost da je x = 0, tj. da niti jedan fotografski aparat neće biti popravljan u jamstvenom roku iznosi 30,4%

Vjerojatnost da je x = 1, tj. da će se u jamstvenom roku popravljati samo 1 od 6 prodanih fotografskih aparata iznosi 40,0% ...

3.2. Poissonova raspodjela

Poissonova raspodjela označava se: **P(μ)**

μ - aritmetička sredina

Primjer 3.3.

Varijabla X ravna se po Poissonovoj raspodjeli s parametrom μ = 2. Kolika je vjerojatnost da slučajna varijabla poprimi vrijednosti:

a) x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 i više
b) x ≤ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
c) x između 2 i 4 (uključivo)
d) x ≠ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
e) x > 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13

a) Kolika je vjerojatnost da je x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 i više

- **Definirati varijablu – "xi"**
- **Unijeti vrijednosti:**

xi
0
1
2
3
4
5
6
7
8
9
10
11
12
13

- **Definirati novi stupac – "Pi"**
- **Long name:**
  - **=Poisson (x; lambda)**
  - **=Poisson (xi; μ)**
  - ***=Poisson (xi; 2)***

- **Staviti na 5 decimala**
  - **Data / Variable Specification**
    - **Display format** $\boxed{\textit{Number}}$
      - **Decimal places** $\boxed{5}$

xi	*Pi*
0	0,13534
1	0,27067
2	0,27067
3	0,18045
4	0,09022
5	0,03609
6	0,01203
7	0,00344
8	0,00086
9	0,00019
10	0,00004
11	0,00001
12	0,00000
13	0,00000

Vjerojatnost da će x biti 0 iznosi 0,13534 ili 13,53%.
Vjerojatnost da će x biti 1 je 27,07%.
...
Vjerojatnost da će x biti 11 je 0,001%.
Vjerojatnost da će x biti 12 je 0%.
Vjerojatnost da će x biti 13 je 0%.

Vjerojatnost da će x biti 11 ili više je 0,00001 + 0,00000 + 0,00000 = 0,00001 = 0,001%.

b) Kolika je vjerojatnost da je x ≤ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13

- **Definirati novi stupac – "Fi"**
- **Long name:**
  - **=IPoisson (x; lambda)**
  - **=IPoisson (xi; μ)**
  - ***=IPoisson (xi; 2)***

- **Staviti na 5 decimala**
  - **Data / Variable Specification**
    - **Display format** *Number*
      - **Decimal places** *5*

xi	***Fi***
0	0,13534
1	0,40601
2	0,67668
3	0,85712
4	0,94735
5	0,98344
6	0,99547
7	0,99890
8	0,99976
9	0,99995
10	0,99999
11	1,00000
12	1,00000
13	1,00000

Vjerojatnost da će x biti (manje ili) jednak 0 iznosi 0,13534 ili 13,53%.
Vjerojatnost da će x biti manji ili jednak 1 (1 ili 0) je 40,60%.
...
Vjerojatnost da će x biti manji ili jednak 11 je 100%.
Vjerojatnost da će x biti manji ili jednak 12 je 100%.
Vjerojatnost da će x biti manji ili jednak 13 je 100%.

---

c) Kolika je vjerojatnost da je x između 2 i 4 (uključivo)?

x = 2 ili 3 ili 4
P(x) = 27% + 18% +9% = 54%.

---

d) Kolika je vjerojatnost da je x ≠ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13?

- **Definirati novi stupac –"Qi"**
- **Long name:** $\boxed{= 1 - Pi}$
- **Staviti na 5 decimala**

xi	*Qi*
0	0,86466
1	0,72933
2	0,72933
3	0,81955
4	0,90978
5	0,96391
6	0,98797
7	0,99656
8	0,99914
9	0,99981
10	0,99996
11	0,99999
12	1,00000
13	1,00000

Vjerojatnost da će x biti različit od 0 iznosi 0,86466 ili 86,47%.
Vjerojatnost da x ne će biti 1 je 72,93%.
…
Vjerojatnost da x ne će biti 11 je 99,999%.
Vjerojatnost da x ne će biti 12 je 100%.
Vjerojatnost da x ne će biti 13 je 100%.

Vjerojatnost da x nije 11 ili više je 99,999%.

---

e) Kolika je vjerojatnost da je x > 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13?

- **Definirati novi stupac – "QFi"**
- **Long name:** $\boxed{= 1 - Fi}$
- **Staviti na 5 decimala**

xi	***QFi***
0	0,86466
1	0,59399
2	0,32332
3	0,14288
4	0,05265
5	0,01656
6	0,00453
7	0,00110
8	0,00024
9	0,00005
10	0,00001
11	0,00000
12	0,00000
13	0,00000

Vjerojatnost da će x biti veći od 0 iznosi 0,86466 ili 86,47%.
Vjerojatnost da će x biti veći od 1 je 59,40%.
...
Vjerojatnost da će x biti veći od 10 je 0,001%.
Vjerojatnost da će x biti veći od 11 je 0%.
Vjerojatnost da će x biti veći od 12 je 0%.
Vjerojatnost da će x biti veći od 13 je 0%.

Vjerojatnost da će x bit veći od 10 (11 ili više) je 0,001%.

Primjer 3.4.

Prema podacima pošte u prosjeku 5 stranaka na sat šalje paket. Ako se pretpostavi da stranke dolaze u poštu neovisno, i s istom vjerojatnosti po satima u radnom vremenu, kolika je vjerojatnost da će u isti sat doći:

a) 2 stranke
b) više od 8 stranaka

a) Kolika je vjerojatnost da u isti sat dođu 2 stranke?

- **Definirati varijablu – "xi"**
- **Unijeti vrijednosti:**

xi
2
8

- **Definirati novi stupac – "Pi"**
- **Long name:**
  - **=Poisson (x; lambda)**
  - **=Poisson (xi; μ)**
  - ***=Poisson (xi; 5)***

Vjerojatnost da će u isti sat doći 2 stranke je 8,42%.

b) Kolika je vjerojatnost da će u isti sat doći više od 8 stranaka?

- **Definirati novi stupac – "QFi"**
- **Long name:** ***(1 – Fi) = 1 – IPoisson (xi;5)***

Vjerojatnost da će u isti sat doći više od 8 stranaka je 6,81%.

4. Procjenjivanje parametara

4.1. Procjenjivanje aritmetičke sredine jedne populacije

Primjer 4.1.

U tvornici automobilskih guma „AG" ispituje se novi model guma. Utvrđuje se koliko kilometara se može prijeći s tim modelom dok se ne unište. Slučajno su izabrane gume koje su postavljene na 10 automobila koji su voženi dok se gume nisu uništile. Podaci u tisućama kilometara zapisani su u sljedećoj tablici:

R. br.	**Prijeđeni put (u tisućama km)**
1.	34
2.	53
3.	39
4.	46
5.	49
6.	41
7.	33
8.	51
9.	45
10.	41

Izvor: Podaci tvornice automobilskih guma "AG" (simulirani podaci)

Zadatak je procjeniti koliko se prosječno kilometara može voziti s tim modelom guma. Procjenu izvršiti na razini 95% pouzdanosti.

$\overline{X} = ?$

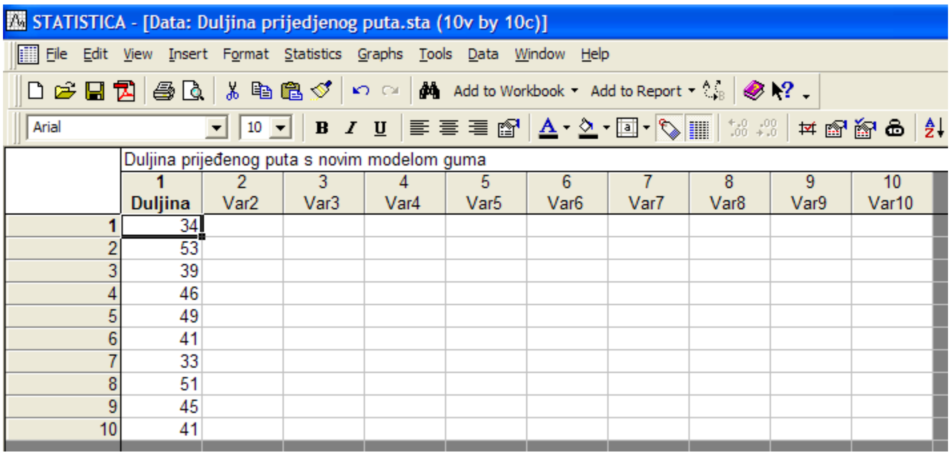
Postupak za rješavanje zadatka:

- U **zaglavlje** dokumenta upisati:

Duljina prijeđenog puta s novim modelom guma

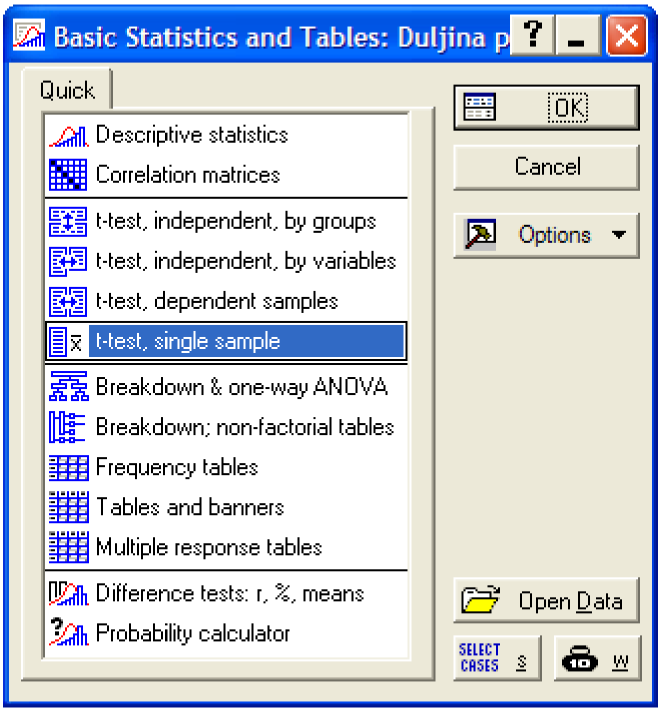
- Definirati varijablu:
  - Kratki naziv: ***Duljina***, dugi naziv: ***Duljina prijeđenog puta u km***

- Unijeti vrijednosti

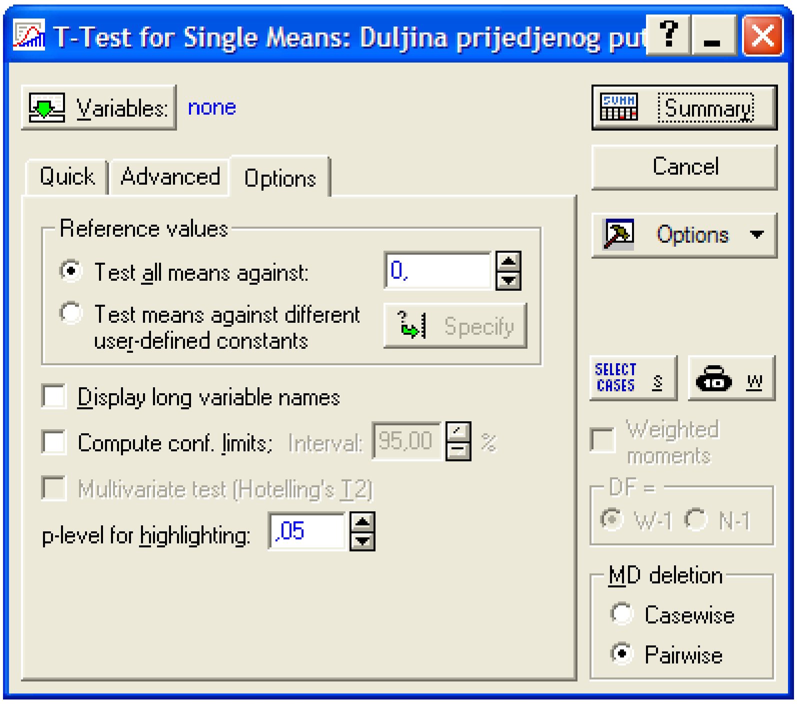


	1 Duljina	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1	34									
2	53									
3	39									
4	46									
5	49									
6	41									
7	33									
8	51									
9	45									
10	41									

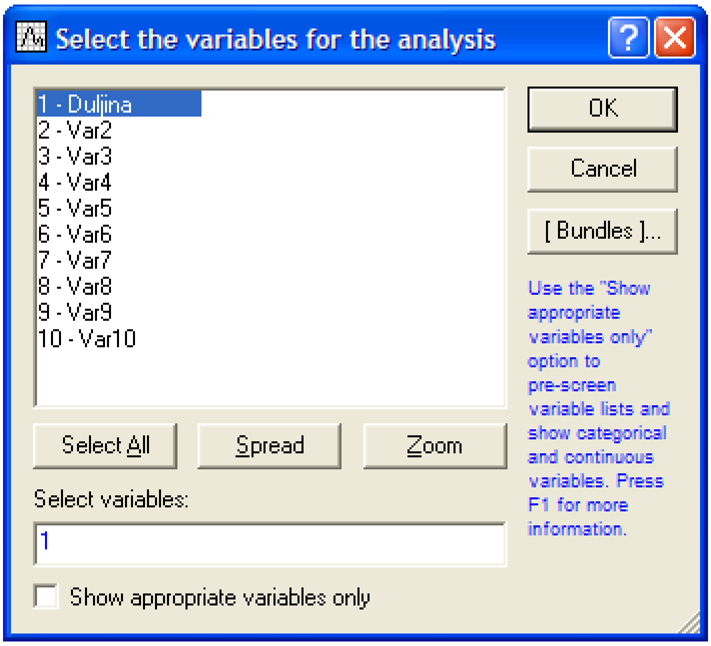
- **STATISTICS / BASIC STATATISTICS/TABLES**



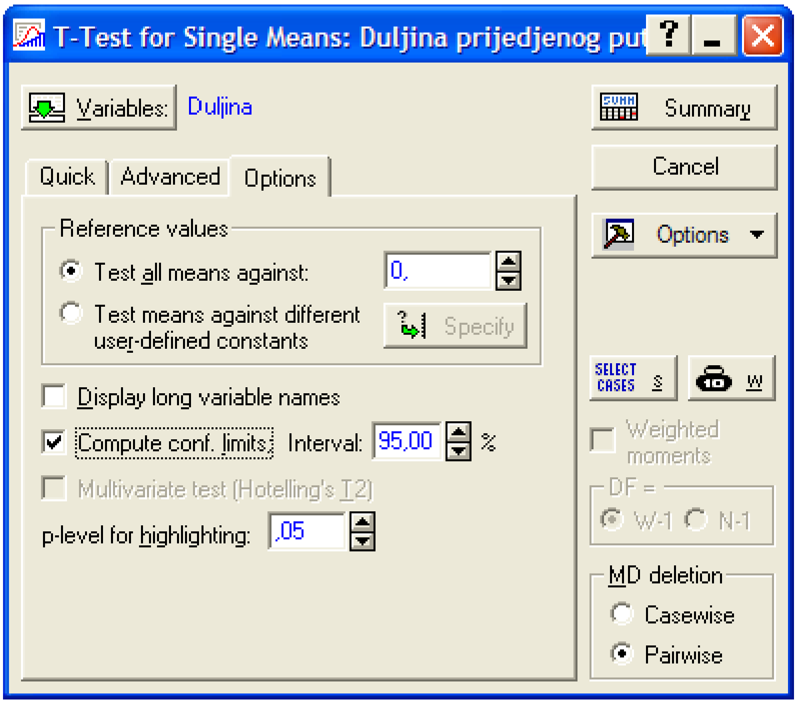
  - **QUICK: ξ t-test, single sample**



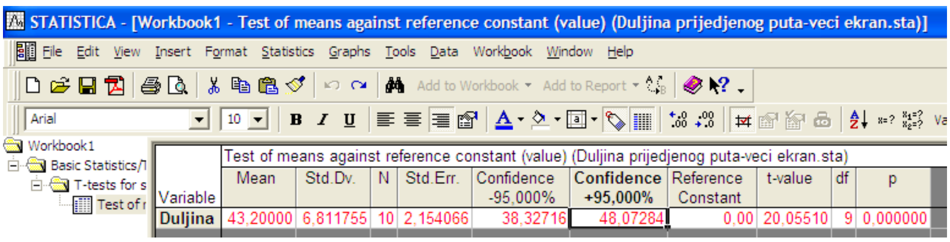
➢ **Variables:** ***1-Duljina***



- **OPTIONS**
  - **Compute conf. limits; Interval: *95,00* (Razina pouzdanosti)**



Rezultati se ispisuju u prozoru radnih zapisa:



Interval procjene aritmetičke sredine izračunat je u poljima *Confidence -95,000%* i *Confidence +95,000%.*

Variable	Test of means against reference constant (value) (Duljina prijedjenog p					
	Mean	Std.Dv.	N	Std.Err.	Confidence -95,000% *Donja granica*	Confidence +95,000% *Gornja granica*
Duljina	43,20000	6,811755	10	2,154066	*38,32716*	*48,07284*

Interval procjene aritmetičke sredine iznosi: $\Pr\{38,327 < \overline{X} < 40,073\} = 0,95$

Zaključak:
Na razini 95% pouzdanosti procjenjuje se da je prosiječan vijek trajanja novog modela guma između 38 327 km i 48 073 km.

5. Testiranje hipoteza

5.1. Testiranje hipoteze o nepoznatoj aritmetičkoj sredini osnovnog skupa

Primjer 5.1.

Proizvođač namještaja „A“ je za jednu vrstu stolice koju proizvodi preporučio prodajnu cijenu od 154 kn. Da bi utvrdio stanje na terenu proizvođač je u slučajno izabranim trgovinama pogledao cijene svoje stolice. Cijene su zapisane u sljedećoj tablici:

R. br.	Maloprodajna cijena u kn
1.	149
2.	156
3.	156
4.	163
5.	161
6.	167
7.	171
8.	147
9.	169
10.	164

Izvor: Podaci tvornice namještaja “A”

Zadatak je odrediti, (drže) pridržavaju li se trgovci preporučene cijene? Test izvršiti na razini 5% signifikantnosti.

$$H_0 : \overline{X} = \overline{X}_0$$
$$H_1 : \overline{X} \neq \overline{X}_0$$

$$\overline{X}_0 = 154.$$

Postupak za rješavanje zadatka:

- U **zaglavlje** dokumenta upisati:

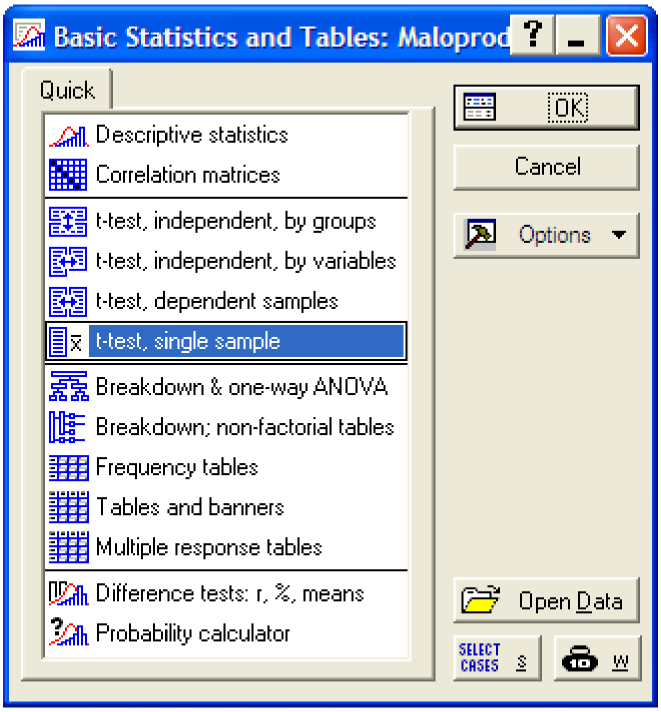
  Maloprodajna cijena stolice proizvođača „A“ u izabranim trgovinama u kn

- Definirati varijablu:
  - Kratki naziv: ***Cijena***, dugi naziv: ***Cijena stolice u kn***

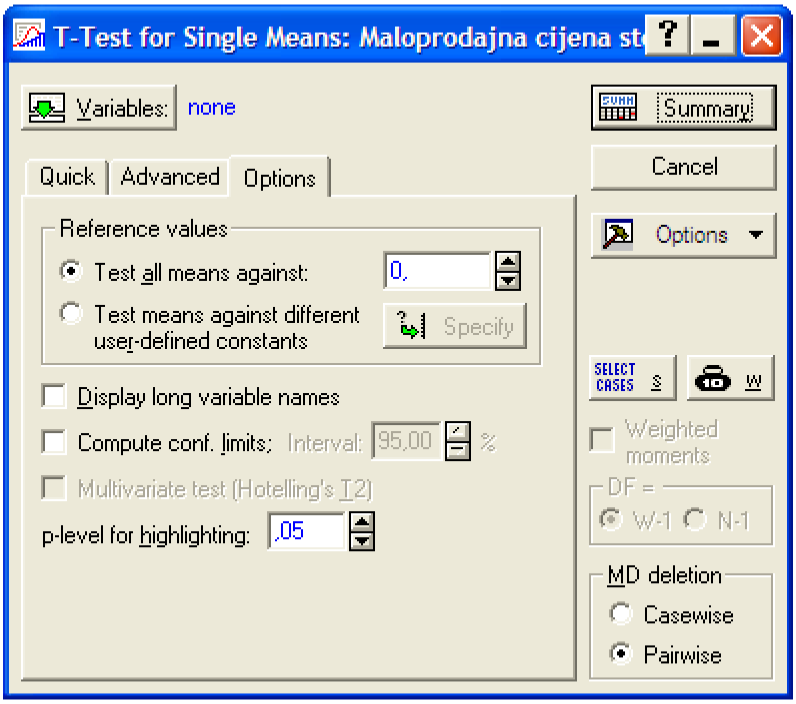
- Unijeti vrijednosti:

Cijena
149
156
156
163
161
167
171
147
169
164

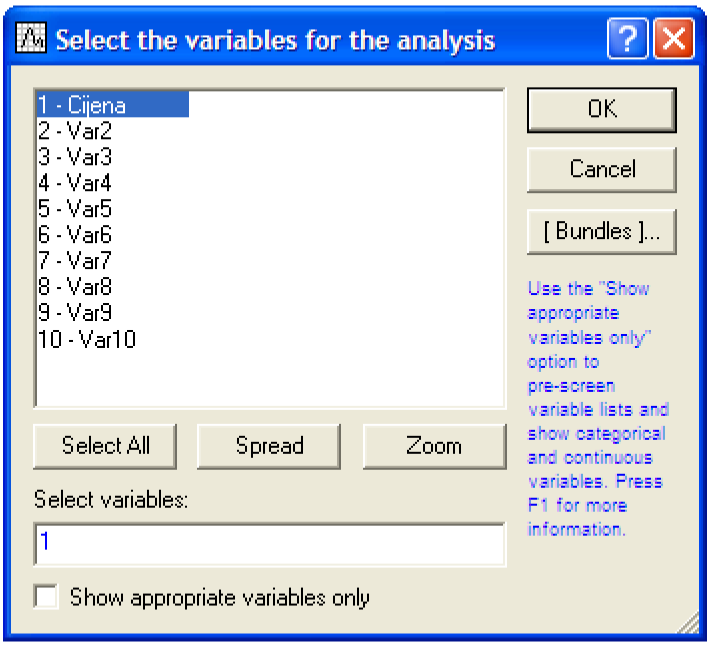
- **STATISTICS / BASIC STATATISTICS/TABLES**



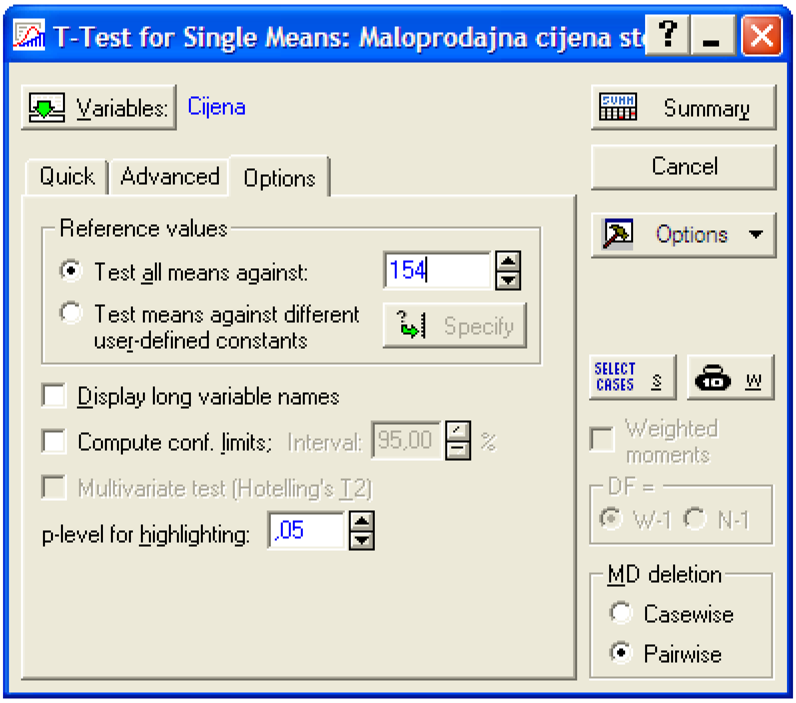
- **QUICK: $\boxed{\text{ξ t-test, single sample}}$**



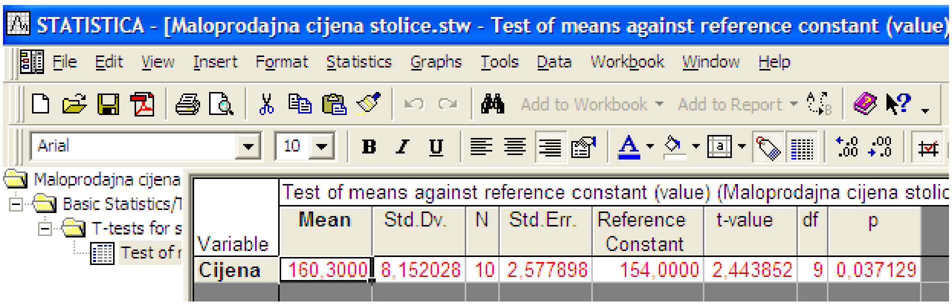
➢ **Variables:** ***1-Cijena***



- **OPTIONS**
  - **Test all means against:** ***154***
  - **Compute conf. limits; -** ***ukloniti oznaku***
  - **p-level for highlighting:** ***0,05*** **(razina signifikantnosti)**



Rezultati:



Nulta hipoteza testira se pomoću vrijednosti *p* na sljedeći način:
Ako je signifikantnost ***p* > zadane razine signifikantnosti** (5%, 1%,…), **prihvaća se pretpostavka da je aritmetička sredina** osnovnog skupa **jednaka** pretpostavljenoj aritmetičkoj sredini **(H_0)**.

Ako je signifikantnost ***p* < zadane razine signifikantnosti** (5%, 1%,…), **prihvaća se pretpostavka da je aritmetička sredina** osnovnog skupa **različita** od pretpostavljene aritmetičke sredine **(H_1)**.

Variable	Test of means against reference constant (value) (Maloprodajna cijena stolice.sta)							
	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
Cijena	160,3000	8,152028	10	2,577898	154,0000	2,443852	9	*0,037129 < 0,05*

p = 0,037129 < 0,05 = 5%, odbacuje se H_0 i prihvaća se H_1.

Zaključak:
Na razini 5% signifikantnosti ne prihvaća se pretpostavka da je prosječna maloprodajna cijena stolice 154 kn. Trgovci se ne pridržavaju prepručene cijene.

Statistica odbacivanje nulte hipoteze, dodatno označava crvenom bojom rezultata.

5.2. Testiranje hipoteze o razlici aritmetičkih sredina dvaju nezavisnih osnovnih skupova

Primjer 5.2.

Vodstvo luke «A» uspoređuje prosječan broj članova posade, po brodu, u svojoj luci i u susjednoj luci «B». Utvrđen je broj članova posade na 10 brodova u luci «A» i 8 brodova u luci «B». Podaci su dani u sljedećoj tablici:

Broj članova	posade po brodu
Luka_A	**Luka_B**
11	13
7	7
8	10
9	9
10	12
8	13
11	5
5	12
8	
7	

Izvor: Podaci luke "A"

Zadatak je odrediti razlikuje li se broj članova posade po brodu u lukama. Test izvršiti na razini 5% signifikantnosti.

Izbor brodova u jednoj luci nije ovisio o izboru brodova u drugoj luci, a to znači da su uzorci nezavisni. Potrebno je testiranti hipotezu o razlici aritmetičkih sredina dvaju nezavisnih osnovnih skupova.

$$H_0 : \overline{X}_1 - \overline{X}_2 = 0$$
$$H_0 : \overline{X}_1 - \overline{X}_2 \neq 0.$$

Postupak za rješavanje zadatka:

- U **zaglavlje** dokumenta upisati:

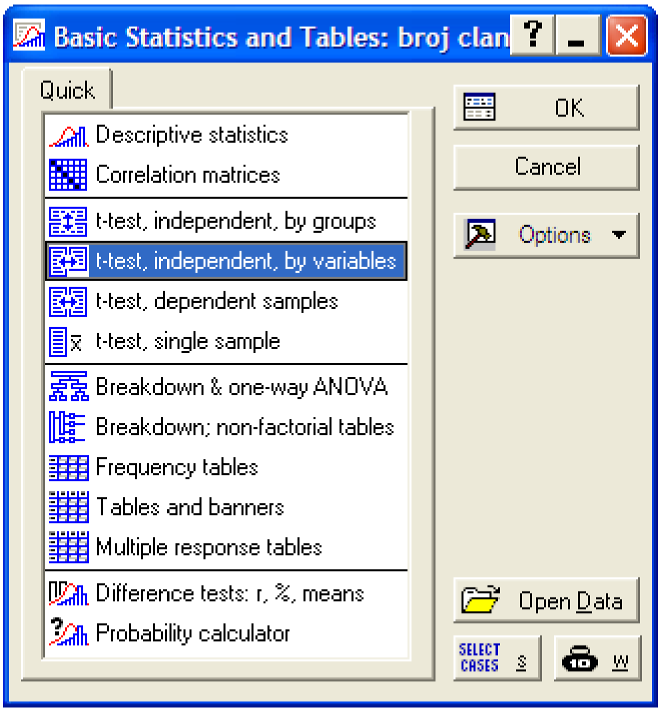
Broj članova posade po brodu u lukama «A» i «B»

- Definirati varijable:
  - Kratki naziv: ***Luka «A»***, dugi naziv:
    Broj članova posade po brodu u luci «A»
  - Kratki naziv: ***Luka «B»***, dugi naziv:
    Broj članova posade po brodu u luci «B»

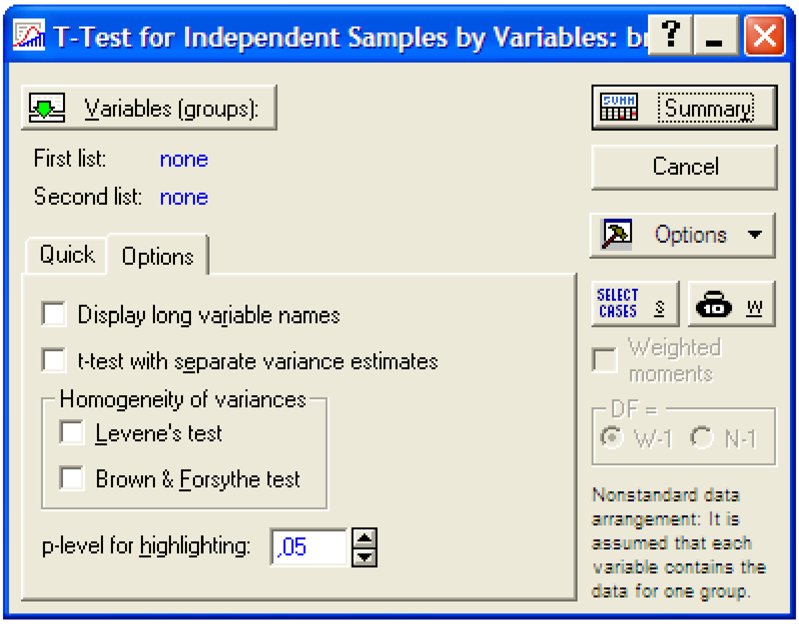
- Unijeti vrijednosti:

Luka_A	Luka_B
11	13
7	7
8	10
9	9
10	12
8	13
11	5
5	12
8	
7	

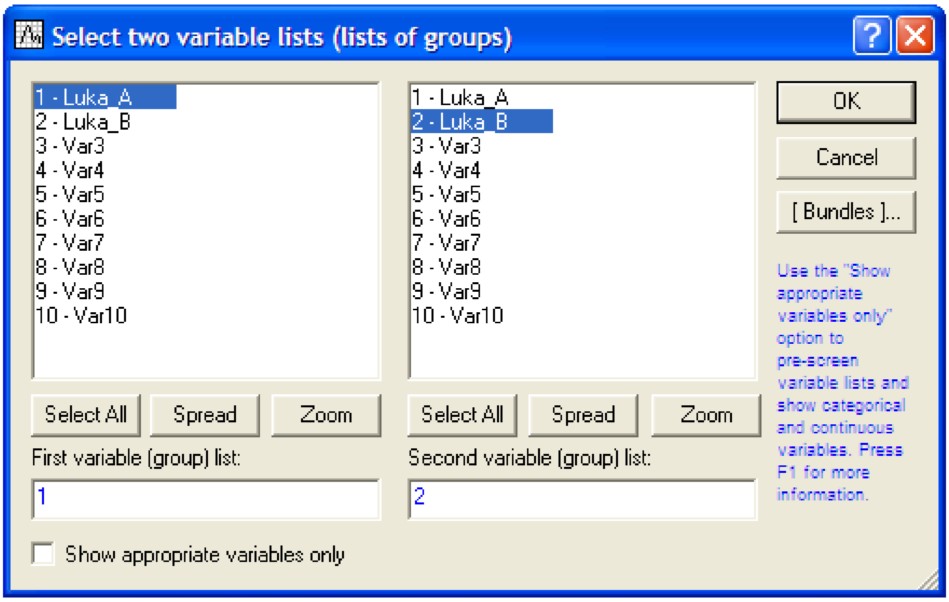
- **STATISTICS / BASIC STATATISTICS/TABLES**



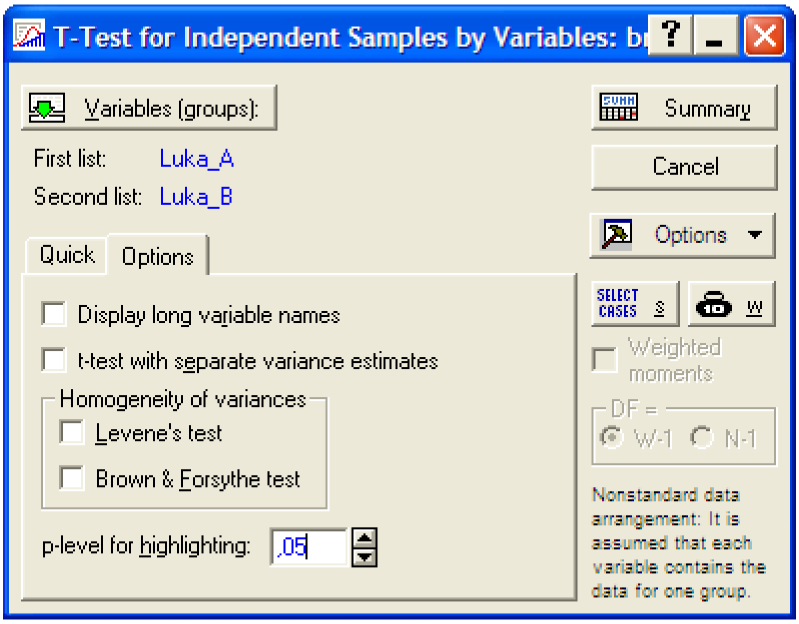
- **QUICK:** ***t-test, independent, by variables***



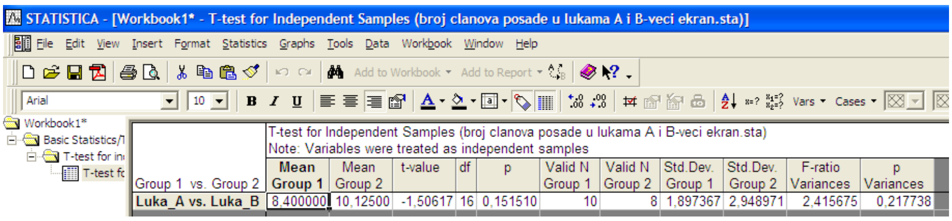
- **Variables (groups):**



  - **First variable (group) list:** ***1-Luka_A***
  - **Second variable (group) list:** ***2-Luka_B***
- **OPTIONS**
  - **P-level for highlighting:** ***0,05*** **(razina signifikantnosti)**


T-Test for Independent Samples by Variables: b
Variables (groups):
First list: Luka_A
Second list: Luka_B
Quick
Options
Display long variable names
t-test with separate variance estimates
Homogeneity of variances
Levene's test
Brown & Forsythe test
p-level for highlighting: .05
Summary
Cancel
Options
SELECT CASES
W
Weighted moments
DF =
W-1
N-1
Nonstandard data arrangement: It is assumed that each variable contains the data for one group.

Rezultati:


STATISTICA - [Workbook1* - T-test for Independent Samples (broj clanova posade u lukama A i B-veci ekran.sta)]
T-test for Independent Samples (broj clanova posade u lukama A i B-veci ekran.sta)
Note: Variables were treated as independent samples
Group 1 vs. Group 2 | Mean Group 1 | Mean Group 2 | t-value | df | p | Valid N Group 1 | Valid N Group 2 | Std.Dev. Group 1 | Std.Dev. Group 2 | F-ratio Variances | p Variances
Luka_A vs. Luka_B | 8,400000 | 10,12500 | -1,50617 | 16 | 0,151510 | 10 | 8 | 1,897367 | 2,948971 | 2,415675 | 0,217738

Nulta hipoteza testira se pomoću vrijednosti p na sljedeći način:
Ako je signifikantnost(,) ***p* > od zadane razine signifikantnosti** (5%, 1%;...), prihvaća se pretpostavka da su **aritmetičke sredine** osnovnih skupova **jednake** ***(H_0)***.

Ako je signifikantnost(,) ***p* < od zadane razine signifikantnosti** (5%, 1%;...), prihvaća se pretpostavka da su **aritmetičke sredine** osnovnih skupova **različite** ***(H_1)***.

	T-test for Independent Samples (broj clanova posade u luk Note: Variables were treated as independent samples					
Group 1 vs. Group 2	Mean Group 1	Mean Group 2	t-value	df	p	Valid N Group
Luka_A vs. Luka_B	8,400000	10,12500	-1,50617	16	***0,15150 < 0,05***	

p = 0,151510 > 0,05 = 5% prihvaća se H_0.

Zaključak:
Na razini 5% signifikantnosti prihvaća se nulta hipoteza da su aritmetičke sredine osnovnih skupova jednake, tj. da prosječan broj članova posade po brodu u lukama A i u luci B podjednak.

5.3. Testiranje hipoteze o jednakosti aritmetičkih sredina dvaju zavisnih osnovnih skupova

Primjer 5.3.

Investicijska bankarica želi provjeriti razlikuje li se povrat na investicije za slična ulaganja na domaćem i na stranom tržištu. Slučajno je izabrala 10 ulagača koji su nedavno ulagali u inozemna trgovačka drušva. Povrat na investicije za slična ulaganja u zemlji i u inozemstvu za svakog ulagača prikazan je sljedećoj tablici:

Ulagač	Povrat na investicije (%) u zemlji	u inozemstvu
1	12	9
2	14	16
3	16	11
4	11	10
5	12	12
6	14	9
7	13	14
8	19	15
9	17	14
10	10	8

Izvor: Podaci investicijske bankarice

Zadatak je odrediti razlikuje se povrat na investicije u zemlji i inozemstvu. Test izvršiti na razini 5% signifikantnosti.

Testirani su isti ulagači na različitim tržištima. Izbor ulagača na jednom tržištu odredio je izbor ulagača na drugom tržištu, a to znači da su uzorci zavisni. Testira se hipoteza o razlici aritmetičkih sredina dvaju zavisnih osnovnih skupova.

$$H_0 : \overline{X}_1 - \overline{X}_2 = 0$$
$$H_0 : \overline{X}_1 - \overline{X}_2 \neq 0.$$

Postupak za rješavanje zadatka:

- U **zaglavlje** dokumenta upisati:

Povrat na investicije u zemlji i inozemstvu (%)

Oznaka ulagača može se unijeti kao pred stupac:

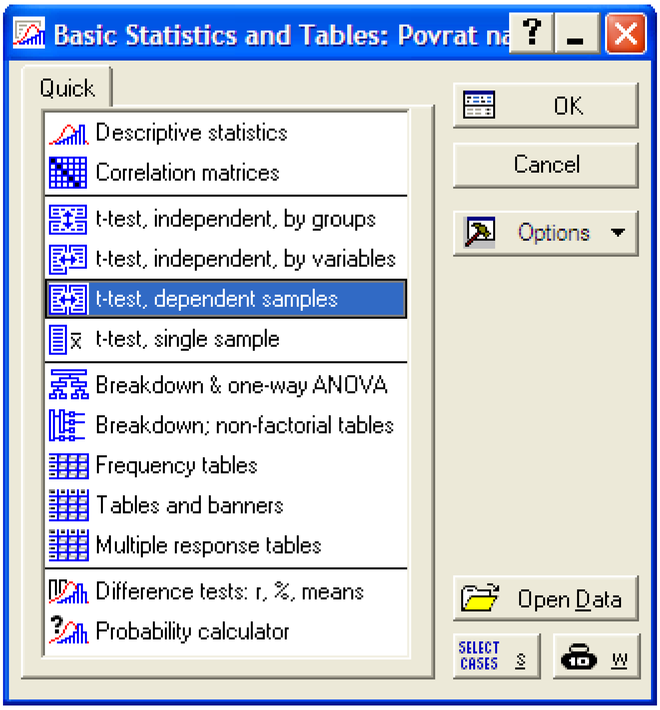
- U informativno polje upisati ***Ulagač***

- Definirati varijable:
    - Kratki naziv: ***Zemlja***, dugi naziv: ***Stopa povrata na investicije u zemlji***
    - Kratki naziv: ***Inozemstvo***, dugi naziv: ***Stopa povrata na investicije u inozemstvu***

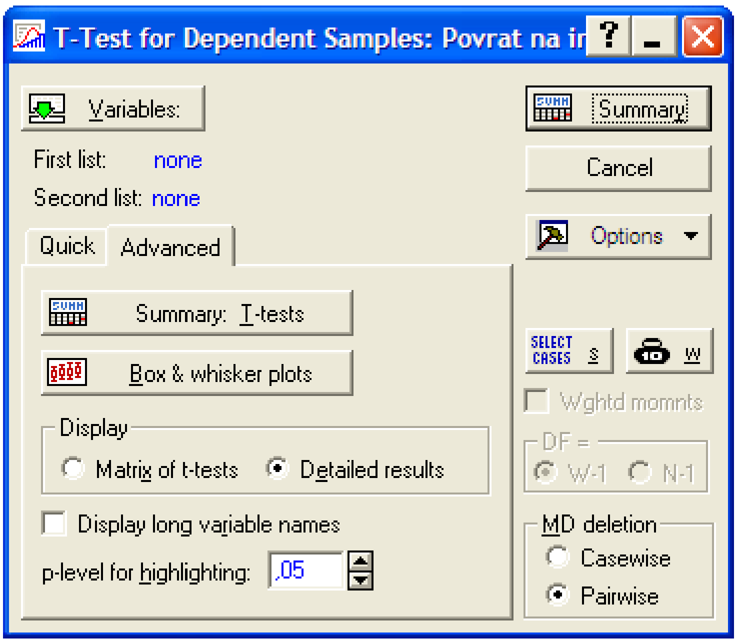
- Unijeti vrijednosti:

Ulagač	Zemlja	Inozemstvo
1	12	9
2	14	16
3	16	11
4	11	10
5	12	12
6	14	9
7	13	14
8	19	15
9	17	14
10	10	8

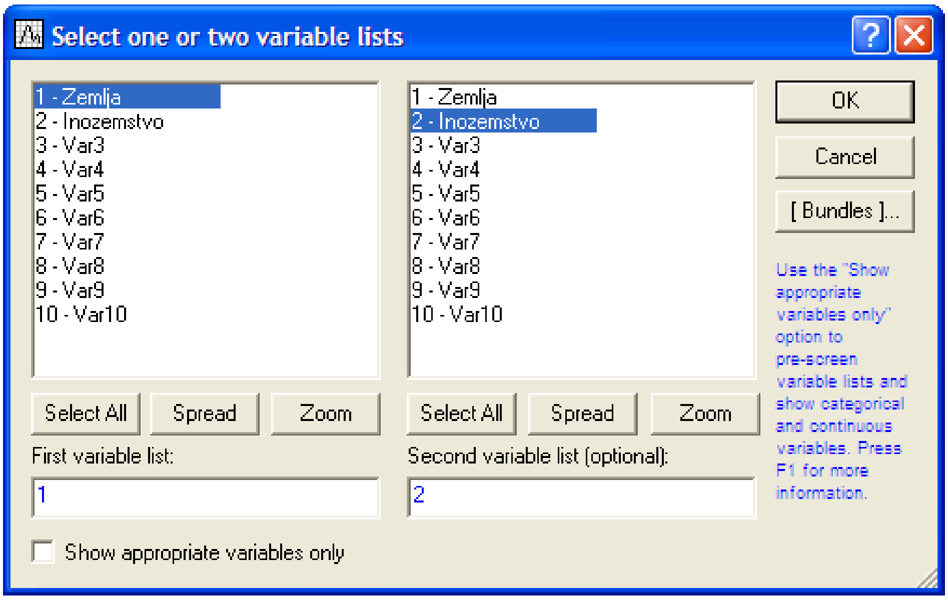
➢ **STATISTICS / BASIC STATATISTICS/TABLES**



➢ **QUICK:** ***t-test, dependent samples***

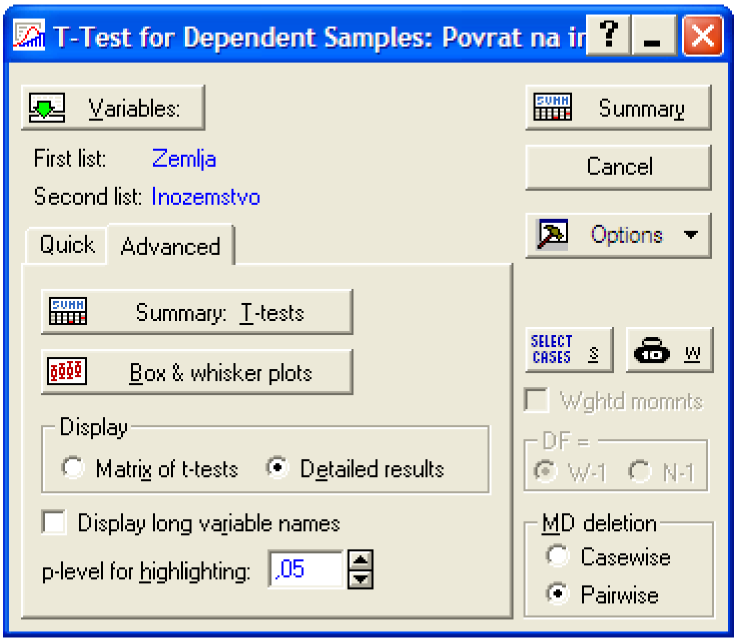


- **Variables:**

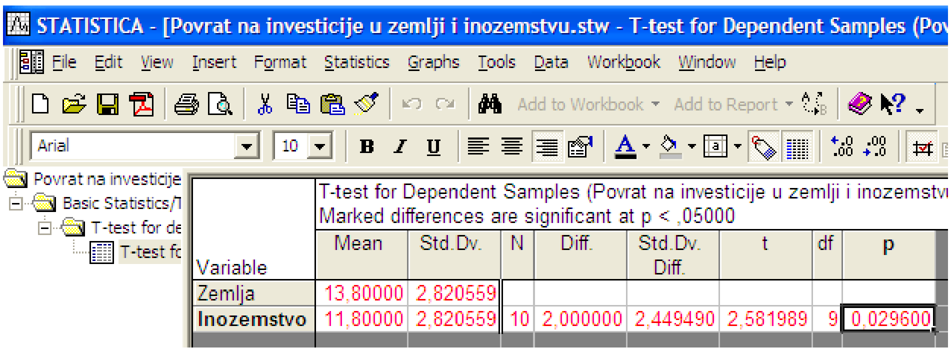


- **First variable list:** ***1 - Zemlja***
- **Second variable list:** ***2 - Inozemstvo***

- **ADVANCED**
  - **P-level for highlighting:** ***0,05*** **(razina signifikantnosti)**



Rezultati:



T-test for Dependent Samples (Povrat na investicije u zemlji i inozemstvu
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
Zemlja	13,80000	2,820559						
Inozemstvo	11,80000	2,820559	10	2,000000	2,449490	2,581989	9	0,029600

Nulta hipoteza testira se pomoću vrijednosti p na sljedeći način:
Ako je signifikantnost(,), p **> od zadane razine signifikantnosti** (5%, 1%;...), prihvaća se pretpostavka da su **aritmetičke sredine** osnovnih skupova **jednake** ***(H_0)***.

Ako je signifikantnost(,) p **< od zadane razine signifikantnosti** (5%, 1%;...), prihvaća se pretpostavka da su **aritmetičke sredine** osnovnih skupova **različite** ***(H_1)***.

Variable	T-test for Dependent Samples (Povrat na investicije u ze Marked differences are significant at p < ,05000					
	N	Diff.	Std.Dv. Diff.	t	df	p
Zemlja						
Inozemstvo	10	2,000000	2,449490	2,581989	9	***0,02960 < 0,05***

p =0,02960 < 0,05 = 5% odbacuje se H_0 i prihvaća se H_1.

Zaključak:
Na razini 5% signifikantnosti prihvaća se alternativna hipoteza da se prosječna stopa povrata na investicije u zemlji i inozemstvu razlikuje.

Statistica odbacivanje nulte hipoteze dodatno označava crvenom bojom rezultata.

6. Hi-kvadrat test

Primjer 6.1.

Odjel za ljudske resurse u trgovačkoj društvu „D“ uspoređuje je li učestalost bolovanja podjednaka cijele godine. U tu svrhu slučajno je izabrano 60 doznaka za bolovanje, a njihova raspodjela po tromjesečjima prikazana je u sljedećoj tablici:

Tromjesečje	Broj doznaka za bolovanje
I.	22
II.	14
III.	8
IV.	16

Izvor: Podaci trgovačkog društva „D“

Zadatak je odrediti je li učestalost bolovanja po tromjesečjima podjednaka ili nije. Test izvšite na razini 5% signifikantnosti.

Testira se hipoteza ima li zadana distribucija oblik jednolike distribucije:

$$H_0 : P_1 = P_2 = P_3 = P_4 = P$$

$$H_1 : \exists P_i \neq P.$$

Postupak za rješavanje zadatka:

- U **zaglavlje** dokumenta upisati:

  Raspodjela otvaranja bolovanja po tromjesečjima

Oznaka tromjesečja može se unijeti kao pred stupac:

- U informativno polje upisati ***Tromjesečje***

- Definirati varijablu:
  - Kratki naziv: ***Bolovanje***, dugi naziv:

    Broj otvorenih bolovanja u tromjesečju

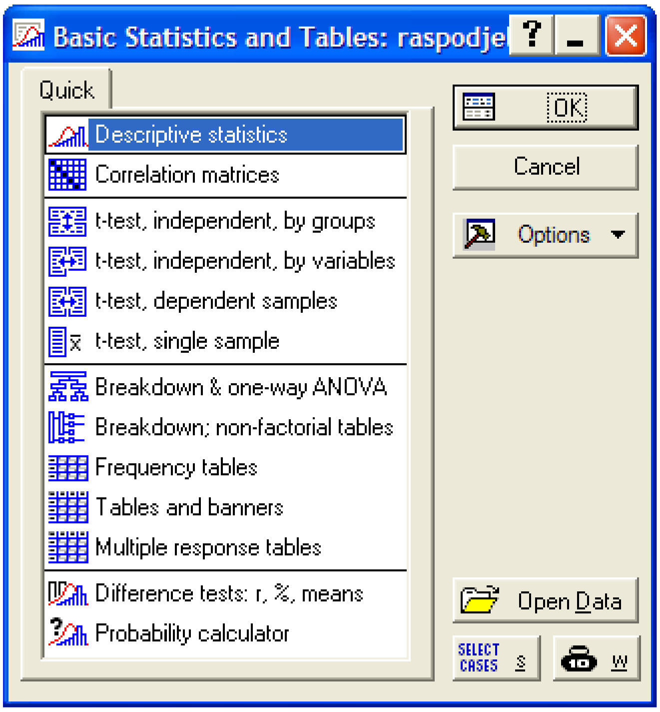
- Unijeti vrijednosti:

Tromjesečje	Bolovanje
I.	22
II.	14
III.	8
IV.	16

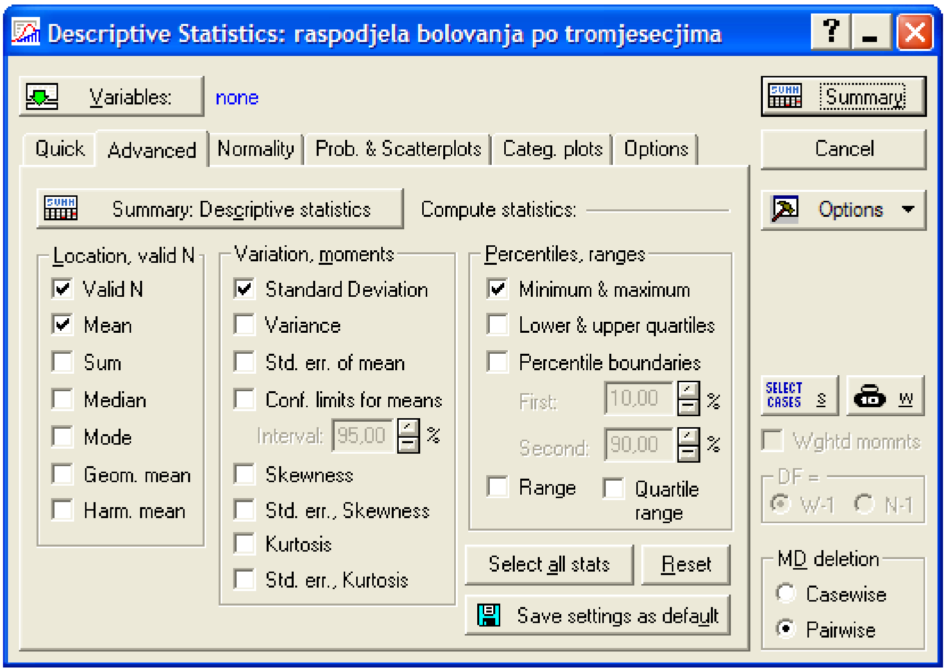
Teorijske frekvencije – e_i, računaju se kao aritmetička sredina broja doznaka po tromjesečjima (60:4 = 15)

Aritmetička sredina varijable *Bolovanje* može se izračunati na sljedeći način:

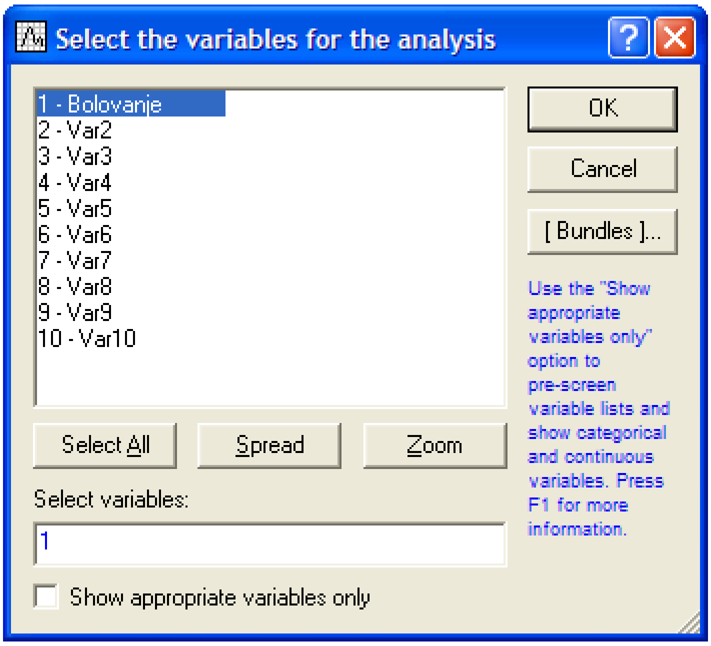
➢ **STATISTICS / BASIC STATATISTICS/TABLES**



➢ **QUICK: Descriptive statistics**

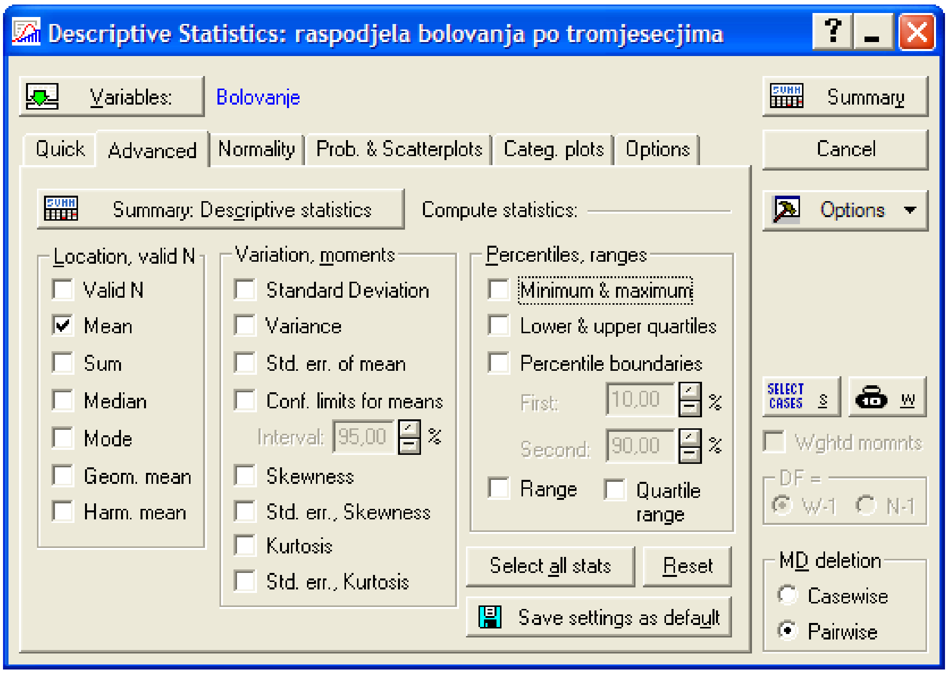


- **Variables: *1-Bolovanje***

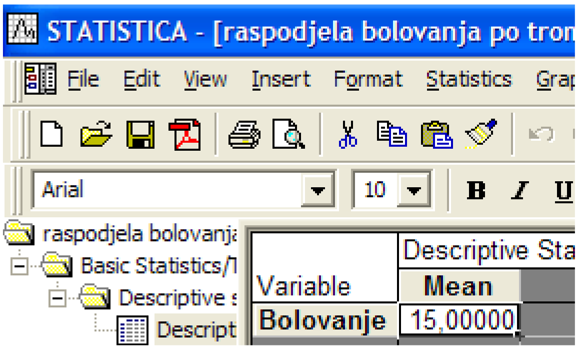


- **ADVANCED**
  - **Mean**

➢ Za ostale funkcije može se ukloniti oznaka (nije obvezatno)



Rezultat:



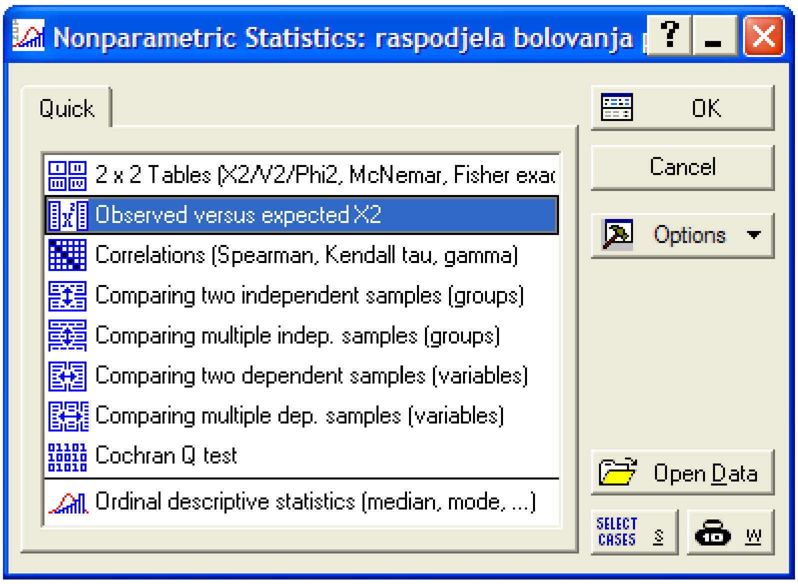
Dobiveni rezultat aritmetičke sredine varijable *Bolovanje* unosi se kao vrijednost nove varijble ei (teorijska frekvenica).

- Definirati varijablu:
    - Kratki naziv: ***ei***, dugi naziv:

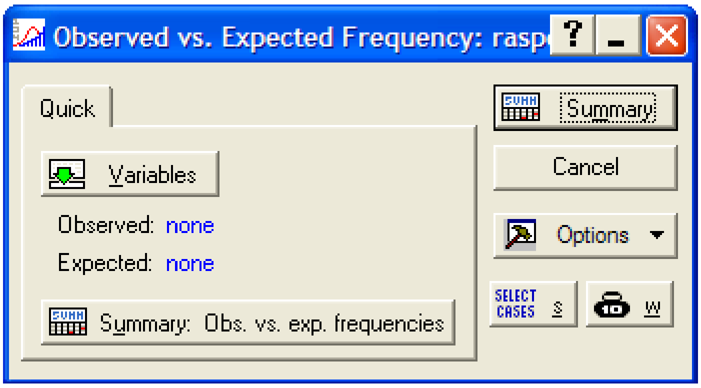
=15; teorijska frekvencija

Hi-kvadrat test provodi se na sljedeći način:

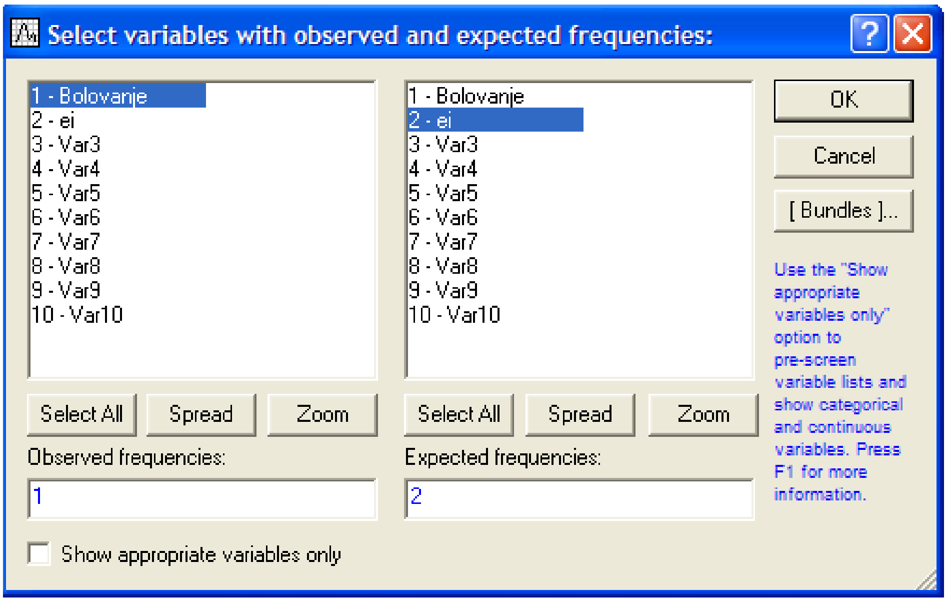
- **STATISTICS / NONPARAMETRICS**



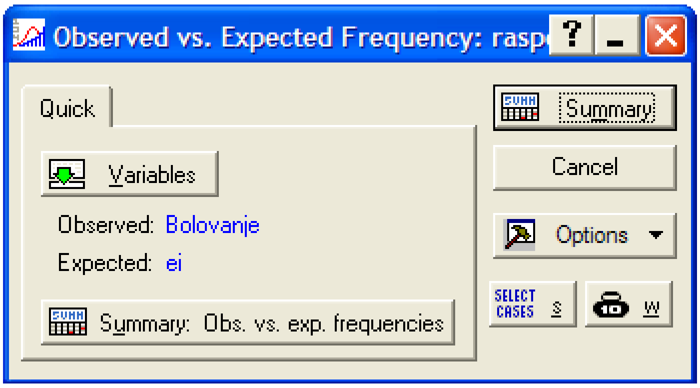
    - **QUICK: Observed versus expected $\chi 2$**



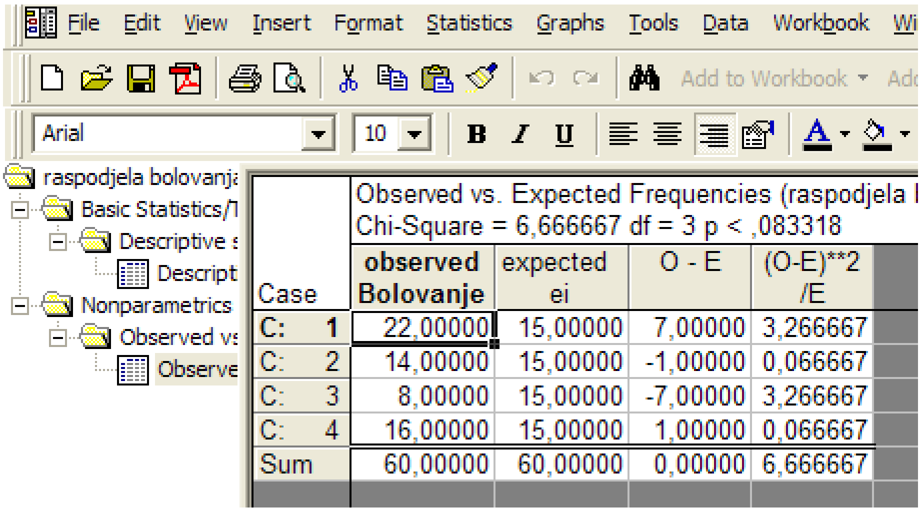
- **Variables:**



- **Observed frequencies:** ***1 -Bolovanje***
- **Expected frequences:** ***2 - ei***


Observed vs. Expected Frequency: rasp
Quick
Variables
Observed: Bolovanje
Expected: ei
Summary: Obs. vs. exp. frequencies
Summary
Cancel
Options
SELECT CASES
S
W

Rezultati:



Nulta hipoteza testira se pomoću vrijednosti *p* na sljedeći način:
Ako je signifikantnost(,) ***p* > od razine signifikantnosti 5%** prihvaća se, pretpostavka da su distribucija **ima oblik jednolike raspodjele *(H_0)***.

Nulta hipoteza testira se pomoću vrijednosti *p* na sljedeći način:
Ako je signifikantnost(,) ***p* < od razine signifikantnosti 5%** prihvaća se, pretpostavka da su distribucija **nema oblik jednolike raspodjele *(H_1)***.

Case	Observed vs. Expected Frequencies (raspodjela bolovar Chi-Square = 6,666667 df = 3 **p < ,083318 >0,05**			
	observed Bolovanje	expected ei	O - E	(O-E)**2 /E
C: I	22,00000	15,00000	7,00000	3,266667
C: II	14,00000	15,00000	-1,00000	0,066667
C: III	8,00000	15,00000	-7,00000	3,266667
C: IV	16,00000	15,00000	1,00000	0,066667
Sum	60,00000	60,00000	0,00000	6,666667

p =0,083318 > 0,05 = 5% prihvaća se H_0.

Na razini 5% signifikantnosti prihvaća se nulta hipoteza kao moguća, tj. prihvaća se pretpostavka da je učestalost bolovanja po tromjesečjima podjednaka.

Prof. dr. sc. Maja Biljan-August
Prof. dr. sc. Snježana Pivac
Doc. dr. sc. Ana Štambuk

STATISTIČKA ANALIZA U EKONOMIJI

Izdavač:
Ekonomski fakultet Sveučilišta u Rijeci

Recenzentice:
Prof. dr. sc. Jasna Horvat

Doc. dr. sc. Alemka Šegota

Doc. dr. sc. Tea Baldigara

Lektorica:
Kerol Musul-Perić, prof.

Autor naslovnice:
Luka Mičetić, dipl. oec.

Pri izradi naslovnice korišteni su materijali objavljeni na www.sxc.hu.

Objavljivanje ovog sveučilišnog udžbenika odobrilo je Povjerenstvo za izdavačku djelatnost Sveučilišta u Rijeci Odlukom – klasa: 602-09/09-01/28, ur. broj: 2170-57-05-09-3 od 25. rujna 2009.

Objavljeno na URL: http://www.efri.hr/prikaz.asp?txt_id=6327 i http://oliver.efri.hr/~statist/biljan-pivac-stambuk-statisticka.pdf.

ISBN: 978-953-6148-85-1

Rijeka, rujan, 2009.