SVEUČILIŠTE U ZAGREBU

FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Andrej Razumić

Zagreb, 2018.

SVEUČILIŠTE U ZAGREBU

FAKULTET STROJARSTVA I BRODOGRADNJE

DIPLOMSKI RAD

Mentorica: Student:

prof. dr. sc. Biserka Runje Andrej Razumić

Zagreb, 2018.

Izjavljujem da sam ovaj rad izradio samostalno koristeći stečena znanja tijekom studija i navedenu literatu.

Zahvaljujem se svojoj mentorici prof. dr. sc. Biserki Runje na stručnom vodstvu i korisnim savjetima pruženim za vrijeme izrade ovog rada.

Posebno se zahvaljujem čitavoj svojoj obitelji na razumijevanju, podršci i strpljenju iskazanom tijekom mojeg studiranja.

SVEUČILIŠTE U ZAGREBU
FAKULTET STROJARSTVA I BRODOGRADNJE
Središnje povjerenstvo za završne i diplomske ispite
Povjerenstvo za diplomske ispite studija strojarstva za smjerove:
proizvodno inženjerstvo, računalno inženjerstvo, industrijsko inženjerstvo i menadžment, inženjerstvo materijala te mehatronika i robotika

Sveučilište u Zagreb Fakultet strojarstva i brodogradnje	
Datum	Prilog
Klasa:	
Ur. broj:	

DIPLOMSKI ZADATAK

Student:	**Andrej Razumić**	Mat. br.: 0035190452
Naslov rada na hrvatskom jeziku:	**Osnovni statistički alati za analizu podataka**	
Naslov rada na engleskom jeziku:	**Basic Statistical Tools for Data Analysis**	

Opis zadatka:

Statistika je znanost o učenju iz podataka. Primjena statističkog zaključivanja neupitna je u gotovo svim znanstvenim i stručnim djelatnostima. Statističko razumijevanje, metode i alati su neophodni za racionalno kvantificiranje podataka. Primjenom statističkih metoda i alata je moguće uočiti zakonitosti i identificirati uzročno-posljedične veze između različitih pojava i procesa, te predvidjeti kretanje neke pojave.

U radu je potrebno:

- razraditi načela deskriptivne i inferencijalne statistike s posebnim naglaskom na testiranje hipoteza
- za odabrane primjere ocijeniti i opravdati adekvatnost odabranih statističkih postupaka, objasniti rezultate statističke analize i procijeniti njihov praktični značaj.

U radu je potrebno navesti korištenu literaturu i eventualno dobivenu pomoć.

Zadatak zadan:	Datum predaje rada:	Predviđeni datum obrane:
16. studenog 2017.	18. siječnja 2018.	24., 25. i 26. siječnja 2018.

Zadatak zadao:

Prof. dr. sc. Biserka Runje

Predsjednica Povjerenstva:

Prof. dr. sc. Biserka Runje

SADRŽAJ

POPIS SLIKA

POPIS TABLICA

SAŽETAK

Tema diplomskog rada je „Osnovni statistički alati za analizu podataka“. Rad je podijeljen na dva dijela. U prvom su dijelu opisane metode deskriptivne statistike, koja se bavi prikupljanjem, analizom i interpretacijom podataka. U području inferencijalne statistike, kod koje na temelju uzorka zaključujemo o populaciji, objašnjene su metode testiranja hipoteza. U drugom dijelu, odabrani su primjeri konkretne primjene opisanih statističkih alata te je procijenjen njihov praktički značaj. U svrhu lakše analize podataka upotrebljavaju se različiti računalni programi. Pri izradi ovog rada korišten je jedan takav program – Minitab 17 (probna verzija).

SUMMARY

The topic of this master thesis is “Basic Statistical Tools for Data Analysis“. The thesis is separated into two parts. The first one shows methods of descriptive statistics, which deals with data collection, analysis and interpretation. In the inferential statistics section, hypothesis testing methods are explained. In the second part, examples of the specific application of the described statistical tools are selected and their practical significance is assessed. To facilitate data analysis, various computer programs are used. Also, in preparation for this study, one such program was used – Minitab 17 (trial).

1. UVOD

Postoji mnogo definicija statistike i svaka od njih je na neki način točna. Načelno, statistika je znanost o učenju iz podataka. U svakodnevnom govoru riječ statistika koristi se i za već prikupljene i uređene podatke koji su objavljeni u obliku tabela, grafikona i slično. Sam naziv *statistika* potječe od latinskog *ratio status* – državni interes, te izvedenice *statista* – osoba koja je vješta u vođenju državnih poslova.

Područje statistike obuhvaća prikupljanje, prikazivanje, analizu i korištenje podataka za donošenje odluka, rješavanje problema i izrazu proizvoda i procesa. Budući da mnogi aspekti inženjerske prakse uključuju rad s podacima, poznavanje statistike jednako je važno za inženjera kao i druge inženjerske znanosti. Naime, statističke tehnike mogu biti moćna pomagala u oblikovanju novih proizvoda i sustava, poboljšanju postojećih dizajna i projektiranju, razvoju i poboljšanju proizvodnih procesa.

1.1. Statistika kroz povijest

Značenje pojma *statistika* mijenjalo se s vremenom. Do sredine 19. stoljeća statistikom se označuju podaci brojčane i nebrojčane prirode važni za državu. Statistika se u jednostavnijim oblicima pojavila za vrijeme babilonske, kineske i egipatske civilizacije, potom i u Rimskom Carstvu. Tada se statistika svodila na popisivanje stanovništva, poljoprivrednih prinosa te materijalnog bogatstva. U srednjem vijeku prikupljali su se podaci o činjenicama bitnima za političko-privredno stanje političkih entiteta. U 14. stoljeću nastaju zapisi *Nuova Cronica* koja sadrži niz statističkih podataka o populaciji, edukaciji i sl. o povijesti Firenze. U 17. stoljeću Herman Conring i Gottfried Achenwall utemeljili su „sveučilišnu statistiku“, koju definiraju kao znanost o stanju i političkom uređenju države. Današnje značenje, dakle prikupljanje i analiza podataka, statistika je dobila početkom 19. stoljeća. Tada su se razvile engleska, njemačka i ruska statistička škola, od kojih je svaka zaslužna za napredak pojedine grane statistike. Primjerice, engleska statistička škola zaslužna je za razvoj teorije procjenjivanja i testiranja hipoteza. Njemačka statistička škola bavila se teorijom stabilnosti statističkih redova, a ruska statistička škola pridonijela je razvoju teorije vjerojatnosti i teorije stohastičkih procesa.

Od druge polovice 20. stoljeća pa sve do danas, velike industrije pridaju ogromnu pozornost poboljšanju kvalitete i unapređenju svojih proizvoda. Primjer takve industrije dolazi iz Japana – „industrijskog čuda“ koje se počelo razvijati sredinom prošlog stoljeća. Velik dio tog uspjeha pripasan je upotrebi statističkih metoda i alata te statističkog razmišljanja i komunikaciji između menadžmenta i proizvodnje.

1.2. Podjela na deskriptivnu i inferencijalnu statistiku

Statistiku generalno možemo podijeliti u dvije skupine: deskriptivnu i inferencijalnu statistiku.

Statistika koja se bavi organiziranim prikupljanjem podataka, metodama njihove prezentacije i njihovom analizom u cilju pružanja jasne, koncizne i točne informacije o istraživanoj pojavi naziva se *deskriptivnom statistikom*.

Predmet *inferencijalne statistike* su statističke metode i tehnike koje omogućuju da se na osnovi dijela informacija koje čine podskup skupa podataka (uzorak), zaključuje o karakteristikama cijelog skupa podataka (populacije).

1.3. Minitab

Minitab je cjeloviti statistički paket koji ima sve alate potrebne za učinkovito analiziranje podataka. Deskriptivna statistika, planovi pokusa, kontrolne karte, testiranje hipoteza, procjena sposobnosti mjernog sustava samo su neka od područja koji program pokriva. Minitab 17 (probna verzija) korišten je u izradi ovog rada.

2. MJERE CENTRALNE TENDENCIJE

Suvremeni pokusi često se bave ogromnim skupovima podataka. Kako bi se stekao osjećaj za takvu količinu podataka, korisno je to sažeti nekim prikladno odabranim mjerama. U praksi se redovito pojavljuje potreba da se niz prikupljenih podataka, kojih je u pravilu mnogo, zamijeni jednom vrijednosti – *srednjom vrijednosti*. Srednja vrijednost je konstanta koja predstavlja niz varijabilnih podataka. Nju je moguće shvatiti i kao središnju vrijednost oko koje se gomilaju podaci, zbog čega se naziva još i *mjerom centralne tendencije*.

U skupinu temeljnih vrsta srednjih vrijednosti spadaju: *aritmetička*, *geometrijska* i *harmonijska sredina* te *mod* i *medijan*. Prve tri spomenute srednje vrijednosti ubrajaju se u potpune srednje vrijednosti jer se za njihovo računanje koriste svi podaci. Mod i medijan ubrajaju se u položajne vrijednosti, čija je vrijednost određena njihovim položajem unutar danog niza.

2.1. Aritmetička sredina

Aritmetička sredina je najčešće upotrebljavana i najpoznatija mjera *prosjeka*.

Aritmetička sredina $\bar{x}$ definirana je kao:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \sum_{i=1}^{n} x_i / n$$

2.1. Geometrijska sredina

Geometrijska sredina G je prema definiciji n-ti korijen iz umnožaka između *n* brojeva.

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Geometrijska se sredina ne može računati ako je vrijednost bilo kojeg člana negativna ili jednaka nuli.

Ona se pretežno koristi kao prosječna mjera brzine nekih promjera. Primjer: ako je neko mjesto 2012. godine imalo 2.000 stanovnika, 2013. godine – 9.000 stanovnika, a 2014. godine – 18.000 stanovnika. Koliko je prosječno populacija svake godine porasla?

Rješenje: $G = \sqrt{4{,}5 \cdot 2} = 3$. Na godišnjem nivou, populacija prosječno raste tri puta. Da smo računali aritmetičkom sredinom, dobili bismo netočan podatak ($\frac{4{,}5+2}{2} = 3{,}25$).

2.2. Harmonijska sredina

Harmonijska sredina H je recipročna vrijednost aritmetičke sredine recipročnih vrijednosti skupa elemenata, a računa se prema formuli:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = n / \sum_{i=1}^{n} \frac{1}{x_i}$$

U odnosu na aritmetičku i geometrijsku sredinu, harmonijska se sredina koristi znatno rjeđe. Harmonijska sredina jednaka je aritmetičkoj i geometrijskoj sredini samo kada su svi elementi skupa jednaki. U suprotnom, harmonijska sredina manja je i od aritmetičke i od geometrijske sredine ($H \leq G \leq \bar{x}$). Harmonijsku sredinu treba upotrebljavati kada želimo dobiti prosjeke nekih odnosa (pr. prosječne kilometre po satu, prosječni broj slova u minuti i sl.). Odličan primjer za primjenu harmonijske sredine nalazimo u svakodnevnom životu: Ako je automobilist udaljenost od 100 km u jednom smjeru vozio brzinom od 100 km/h, a u drugom 50 km/h, kolika je bila njegova prosječna brzina?

Rješenje: Upotrebom harmonijske sredine dolazimo do ispravnog rezultata.

$$H = \frac{2}{\frac{1}{100} + \frac{1}{50}} = 66{,}7 \text{ km/h}$$

2.3. Medijan

Medijan uzorka *Me* je vrijednost u sredini rastućeg ili padajućeg niza. Ako ne neparan broj članova u nizu, tada je medijan onaj član koji zadovoljava uvjet:

$$Me = x_{(n+1)/2}.$$

Ako je paran broj članova u nizu, medijan se računa kao aritmetička sredina dviju srednjih vrijednost:

$$Me = \frac{x_{n/2} + x_{(n+2)/2}}{2}.$$

Medijan upotrebljavamo kad moramo uzeti u obzir i neke *vrlo ekstremne* vrijednosti koje bitno mijenjaju aritmetičku sredinu.

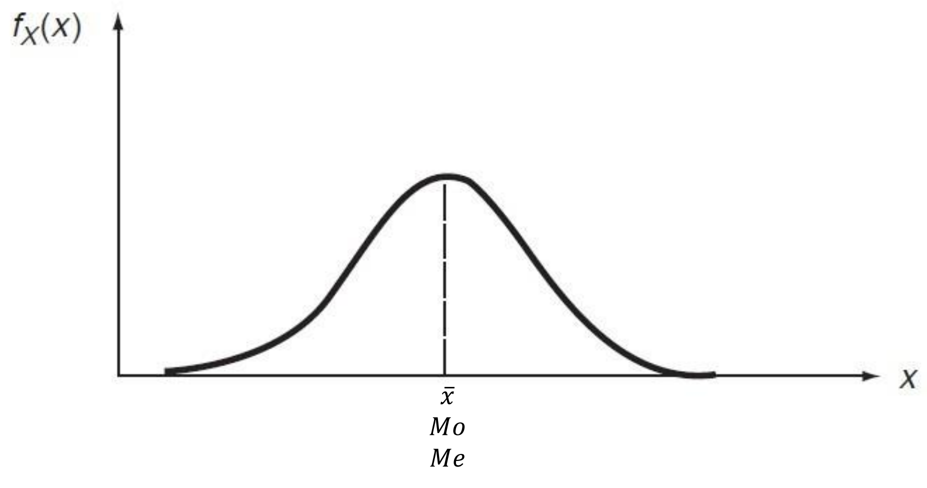
2.4. Mod

Mod uzorka Mo je podataka s najvećom frekvencijom. Ako više podataka ima istu frekvenciju pojavljivanja, svi oni predstavljaju mod.

Prednost moda pred aritmetičkom sredinom je u tome što na nju ne utječe ni broj ni vrijednost rezultata, već samo frekvencija pojedinih rezultata. Ako imamo rezultate grupirane u razrede, aproksimativna dominantna vrijednost je sredina onog razreda koji ima najveću frekvenciju.

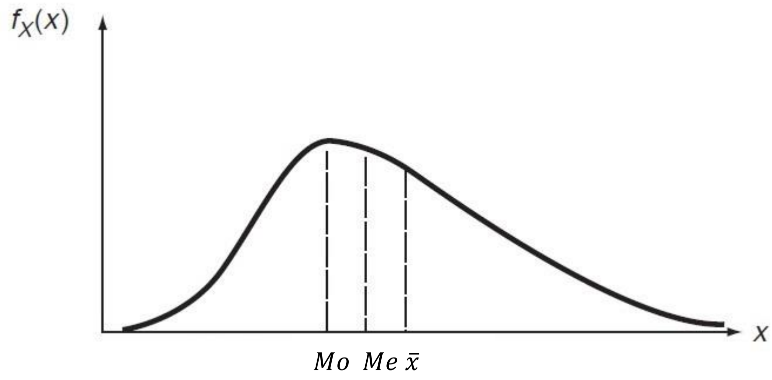
2.5. Položaj srednjih vrijednosti u distribucijama frekvencija

Položaj srednjih vrijednosti možemo ilustrirati distribucijama frekvencija.



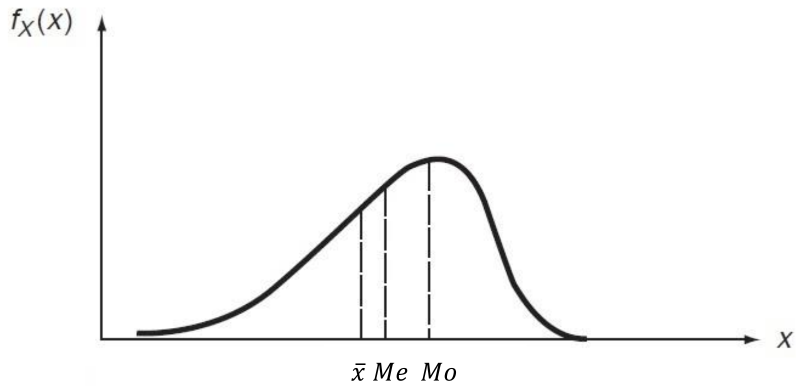
Slika 1. Simetrična distribucija

Slika 1. prikazuje simetričnu distribuciju gdje se sve tri srednje vrijednosti poklapaju, tj. vrijedi jednakost $\bar{x} = Mo = Me$. U pozitivno asimetričnoj distribuciji (Slika 2.) se ispod tjemena krivulje, kao vrijednosti s najvećom frekvencijom, smjestio mod. Aritmetička je sredina odvučena skroz u desno, dok se medijan smjestio približno trećini razmaka između aritmetičke sredine i moda. Za pozitivno asimetričnu distribuciju vrijedi $Mo < Me < \bar{x}$.



Slika 2. Pozitivno (desnostrano) asimetrična distribucija

U negativno asimetričnoj distribuciji, poredak srednjih vrijednosti je obrnut, tj. vrijedi nejednakost $\bar{x} > Me > Mo$. Medijan je i tu smješten bliže aritmetičkoj sredini nego modu.



Slika 3. Negativno (ljevostrano) asimetrična distribucija

3. MJERE RASPRŠENOSTI

Osim statističkih vrijednosti koja pokazuju središte podataka, zanimljivi su nam i oni koji opisuju širenje ili varijabilnost podataka. Među najpoznatije mjere disperzije (raspršenosti) ubrajamo raspon, standardnu devijaciju i varijancu. Mala vrijednost pokazatelja disperzije znači da je izračunata srednja vrijednost bolji predstavnik skupa podataka, i obratno.

3.1. Raspon

Raspon uzorka *R* je razlika između najvećeg i najmanjeg rezultata:

$$R = max\ (x_i) - min(x_i)$$

Raspon je vrlo nesigurna i varljiva mjera varijabilnosti rezultata jer bilo koji osamljeni *ekstremni* rezultat znatno povećava raspon, a da se grupacija rezultata oko srednje vrijednosti ipak nije bitno promijenila.

Osnovni nedostatak raspona sastoji se u tom što on obično ima veću vrijednost većim brojem mjerenja neke pojave. Sa stajališta zakona vjerojatnosti, tu pojavu je prilično lako razumjeti: uzmemo li sve rezultate nekog mjerenja u obzir, raspone je razlika između najvećeg i najmanjeg rezultata. Međutim, uzmemo li u obzir samo nekoliko rezultata, vrlo je mala vjerojatnost da će među njima biti upravo najveći i najmanji rezultat.

3.2. Varijanca

Varijanca uzorka s^2 definira se kao:

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \,/(n-1).$$

3.3. Standardna devijacija

Standardna devijacija uzorka *s* je pozitivni drugi korijen varijance uzorka:

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)}.$$

Standardna devijacija i varijanca uzorka dijele se s djeliteljem $(\mathrm{n}-1)$, a ne s 1 kao što je slučaj kod varijance cijele populacije jer želimo da uzorak daje nepristranu procjenu varijancu populacije. Pri računanju standardne devijacije uzorka računamo razlike između svakog rezultata uzorka i aritmetičke sredine uzorka. Praktički je sigurno da aritmetička sredina uzorka nije jednaka pravoj aritmetičkoj sredini populacije, već od nje manje ili više odstupa. Poznato je da je suma kvadriranih razlika između niza rezultata i njihove aritmetičke sredine manja od sume kvadriranih razlika između istih rezultata i bilo koje druge vrijednosti. Iz tog proizlazi da je suma kvadrata razlika između svakog pojedinog rezultata uzorka i aritmetičke sredine populacije veća od razlike kvadrata svakog rezultata i aritmetičke sredine uzorka. Prema tome, standardna devijacija, izračunata na osnovi razlika prema aritmetičkoj sredini uzorka, manja je od one koju bismo trebali dobiti. Da bismo tu grešku korigirali, smanjujemo vrijednost nazivnika (oduzimamo 1 u nazivniku).

3.4. Koeficijent varijabilnosti

Kad su nam poznate aritmetička sredina i standardna devijacija nekih rezultata, onda su ti rezultati definirani i možemo ih uspoređivati s nekim drugim rezultatima. Da bismo mogli međusobno uspoređivati varijabilnost različitih pojava i svojstava, služimo se koeficijentom varijabilnosti V koji nam pokazuje koliki postotak vrijednosti aritmetičke sredine iznosi vrijednost standardne devijacije.

$$V = \frac{s}{\bar{x}} \cdot 100\%$$

4. DESKRIPTIVNA STATISTIKA

Područjem opisivanja konkretnih rezultata, dobivenih prilikom nekog ispitivanja ili mjerenja bavi *deskriptivna statistika*. Njena je zadaća da opiše podatke, i to tako da ih sredi i sažme kako bi bili što pregledniji. Bez takvog sređivanja mnogi podaci bi bili nepregledni.

Numerički podaci studije, procesa ili čega drugog trebaju biti jasno i sažeto prikazani tako da osoba koja ih gleda može brzo dobiti osjećaj za bitne karakteristike podataka. Tijekom godina uočeno je da su sustavni prikaz te tablice i grafikoni vrlo korisni načini predstavljanja podataka, često otkrivajući važne značajke kao što su raspon, simetrija podataka itd.

4.1. Slučajna varijabla

Većina procesa s kojima se susrećemo ima ishode koji se mogu interpretirati kao stvari brojevi, primjerice duljina osovine, vlačna čvrstoća žice, broj prometnih nesreća na određenom pružnom prijelazu itd. Jednostavno rečeno, slučajna varijabla je numerički ishod slučajnog eksperimenta.

4.1.1. Kontinuirana slučajna varijabla

Mjerenje ponekad (kao što je struja u bakrenoj žici ili duljina strojnog dijela) može preuzeti bilo koju vrijednost u intervalu stvarnih brojeva (barem teoretski). Slučajna varijabla koja to predstavlja to mjerenje kaže se da je *kontinuirana slučajna varijabla*. Raspon slučajne varijable uključuje sve vrijednosti u intervalu realnih brojeva. Primjeri kontinuirane varijable su: struja, duljina, tlak, temperatura, vrijeme, napon, masa…

4.1.2. Diskretna slučajna varijabla

U eksperimentima u kojima bilježimo cijeli broj pr. broj prenesenih bitova u sekundi, udio neispravnih dijelova među 100 odabranih riječ je *diskretnoj slučajno varijabli*. Dakle, za slučajnu varijablu kažemo da je diskretna ako može poprimiti samo konačan broj od mogućih vrijednosti x.

4.2. Teorijske raspodjele podataka

Teorijske raspodjele su raspodjele koje se mogu očekivati u skladu s našim iskustvom. Pretpostavljamo ih u nekom statističkom modelu ili ih postavljamo kao hipotezu koju treba ispitati. Za njih su unaprijed poznate karakteristike kao što su aritmetička sredina, mod, medijan itd. Postoje još empirijske ili originalne raspodjele – to su raspodjele formirane grupiranjem opažanja ili elemenata skupa prema nekom obilježju.

Teorijske raspodjele dijele u dvije skupine:

1. Raspodjele diskretnih varijabli – binomna raspodjela, Poissonova raspodjela, hipergeometrijska raspodjela, negativna binomna raspodjela
2. Raspodjele kontinuiranih varijabli – normalna raspodjela, lognormalna, Weibullova, pravokutna, trokutasta, studentova „*t*"-razdioba, χ^2-razdioba, *F*-raspodjela

4.3. Binomna raspodjela

Binomna je raspodjela najjednostavnija teorijska raspodjela diskretnih varijabli. Slučajna varijabla x ponaša se po binomnoj raspodjeli ako:

1. Pokus sadrži fiksan broj n pokušaja.
2. Svaki pokušaj rezultira jednom od dvije moguće varijante, kao „uspjeh" ili „neuspjeh", odnosno kao pozitivan ili negativan ishod.
3. Vjerojatnost procesa p za pozitivan ishod je konstantan od pokušaja do pokušaja.
4. Pokušaji su međusobno nezavisni.
5. x označava broj uspješnih pokušaja od n pokušaja.

Mnoga istraživanja mogu se adekvatno opisati binomnom raspodjelom. Primjerice:

- Broj defektnih dijelova u uzorku veličine n na velikoj populaciji.
- Broj zaposlenih koji preferiraju određenu policu osiguranja od n ispitanih zaposlenika.
- Broj klipova u motori koji rade na ispravan način.
- Broj elektroničkih sustava koji su ovog tjedna prodani od ukupno n proizvedenih.

Binomna raspodjela definirana je dvama parametrima:

- ✓ n – broj pokušaja, veličina uzorka

- ✓ p – vjerojatnost pozitivnog ishoda, odnosno vjerojatnost da se dogodi određeni događaj A

Slučajna varijabla binomne raspodjele x označava broj uspješnih ishoda u n pokušaja. Radi lakšeg računa uvodimo još jednu oznaku – q, koja označava vjerojatnost da se ne dogodi određen događaj A. Računa se kao $q = 1 - p$.

$P(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$	$x = 0, 1, 2, \dots, n$	za $0 \leq p \leq 1$
$\mu = E(x) = np$		$\sigma^2 = V(x) = npq$

$\mu = E(x)$ – očekivana vrijednost diskretne slučajne varijable

$\sigma^2 = V(x)$ – varijanca diskretne slučajne varijable

Funkcija $P(x)$ naziva se funkcijom vjerojatnosti i za nju vrijedi:

$$P(x) \geq 0$$

$$\sum P(x) = 1$$

Općenito, ako su a i b konstante, za diskretnu slučajnu varijablu x vrijede sljedeća pravila:

$$E(ax + b) = aE(x) + b$$

$$V(ax + b) = a^2 V(x)$$

U nastavku su dani primjeri primjene binomne raspodjele, kao i rješenja i objašnjenja napisanih problema.

4.4. Primjeri

Primjer 4.1. Tvrtka proizvodi osigurače s 10% nesukladnih jedinica. Četiri osigurača iz velike pošiljke su nasumično uzeta iz pošiljke.

a) Kolika je vjerojatnost da od ta četiri nasumično odabrana osigurača samo jedan neispravan?

b) Kolika je vjerojatnost da je barem jedan nesukladan?

c) Pretpostavimo da je uzorak od četiri osigurača iz pošiljke koje smo poslali kupcu prije nego što smo ga testirali osiguran. Ako je bilo koji osigurač neispravan, dobavljač će ga popraviti bez ikakve promjene za kupca. Prema tome, trošak popravka je opisan funkcijom $C = 3x^2$, gdje x označava broj neispravnih u pošiljci od četiri komada. Koliki je očekivani trošak popravka?

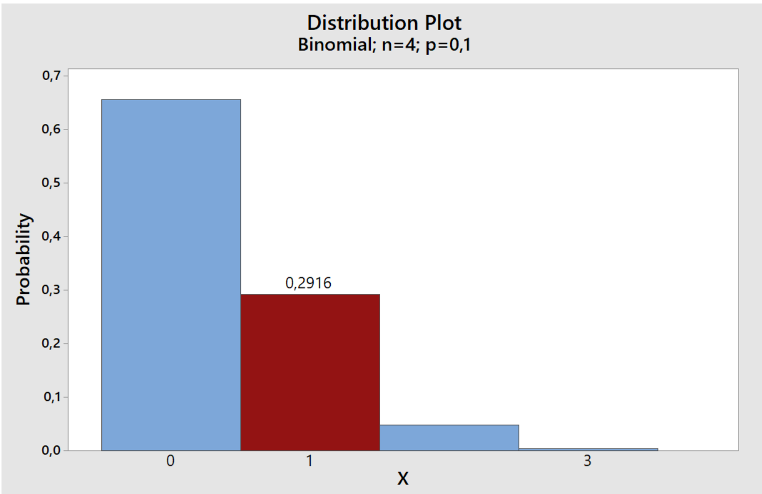
Rješenje:

Pretpostavimo da su četiri osigurača uzorkovana neovisno jedan od drugom i da je vjerojatnost jednaka (u iznosu od 0,1) za svakog od njih da bude neispravan. To je približno točno ako je pošiljka doista velika. Međutim, ako je pošiljka mala, uklanjanjem jednog od osigurača vjerojatnost drugog da bude neispravan bitno je različita.

Za velike pošiljke, binomna raspodjela pruža odgovarajući model ovog pokusa s $n = 4$ i $p = 0{,}1$. Neka je x broj nesukladnih osigurača od četiri testiranih.

a) Vjerojatnost da je samo jedan neispravan u uzorku od četiri osigurača iznosi

$$P(x = 1) = \binom{4}{1} \cdot 0{,}1^1 \cdot 0{,}9^3 = 0{,}2916$$

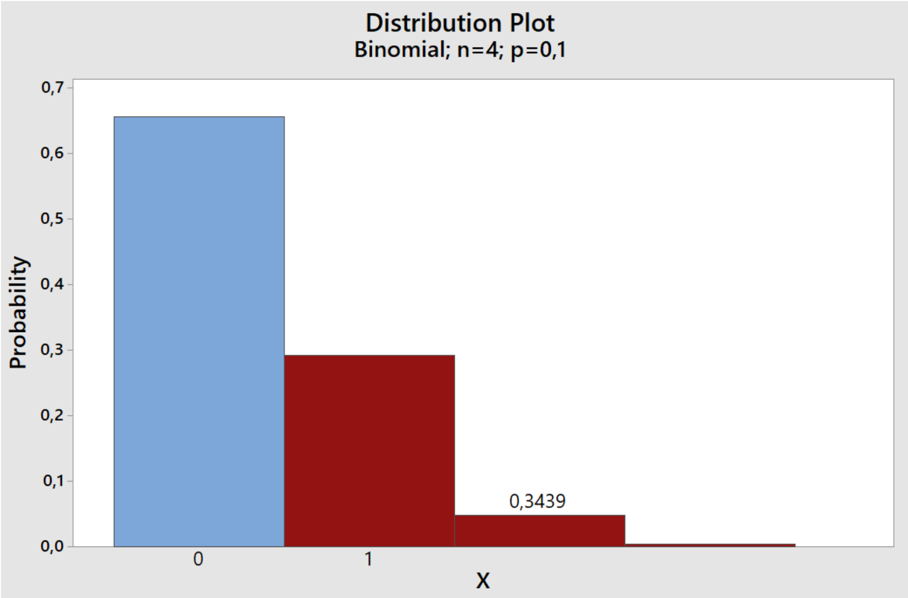


Slika 4. Minitab: Primjer 4.1. a)

Minitab: Graph > Probabilily Distribution > Plot > View probability > OK. Distribution: Binomial. Number of trials: 4. Event probability: 0,1. Shaded Area

b) Vjerojatnost da je barem jedan osigurač neispravan je

$$P(x \geq 1) = 1 - P(x = 0) = 1 - \binom{4}{0} \cdot 0{,}1^0 \cdot 0{,}9^4 = 0{,}3439$$



Slika 5. Minitab: Primjer 4.1. b)

c) Znamo da je

$$E(C) = E(3x^2) = 3E(x^2).$$

Za binomnu raspodjelu vrijedi:

$$\mu = E(x) = np \text{ i } \sigma^2 = \mathrm{V}(x) = npq.$$

Od

$$V(x) = E(x^2) - E(x)^2$$

dobijemo

$$E(x^2) = \sigma^2 + \mu^2 = npq + (np)^2.$$

Stoga,

$$E(C) = 3E(x^2) = 3[npq + (np)^2] = 3[4 \cdot 0{,}1 \cdot 0{,}9 + (4 \cdot 0{,}1)^2] = 1{,}56.$$

Ako su troškovi originalno u desecima eura, tada očekivani trošak po pošiljci od četiri osigurača iznosi 15,60 €.

Primjer 4.2. Studijom o životnom vijeku određene vrste baterija, spoznato je da vjerojatnost da baterija ima životni vijek od barem 7 sati iznosi 13,5%. Ako su tri baterije slučajnim odabirom uzete iz proizvodnje radi testiranja, kolika je vjerojatnost da:

a) Točno dvije baterije traju 7 sati ili više?

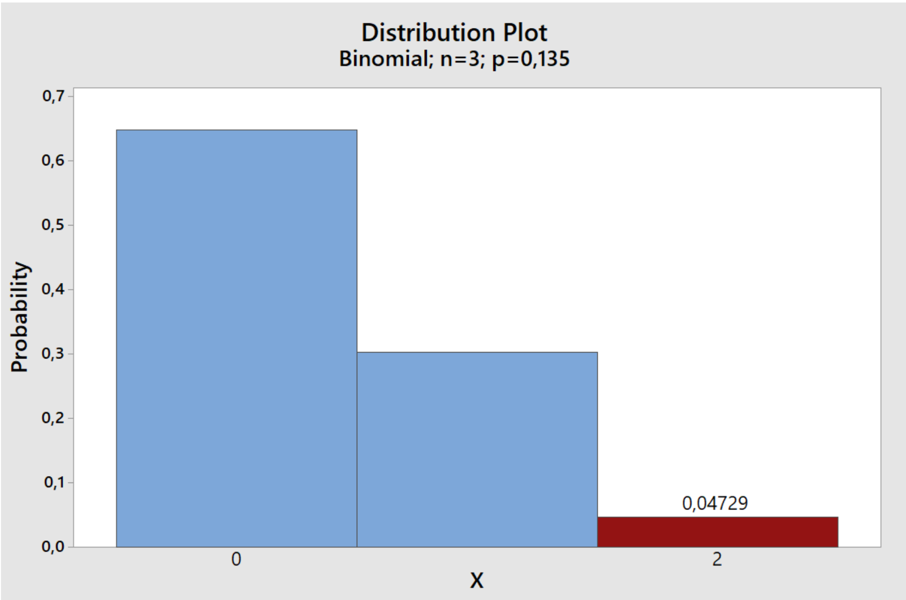
b) Nijedna baterija ne traje 7 sati ili više?

Rješenje:

Iz teksta zadatka bilježimo: $n = 3,\ p = 0{,}135$.

a) Vjerojatnost da točno dvije baterije traju barem 7 sati:

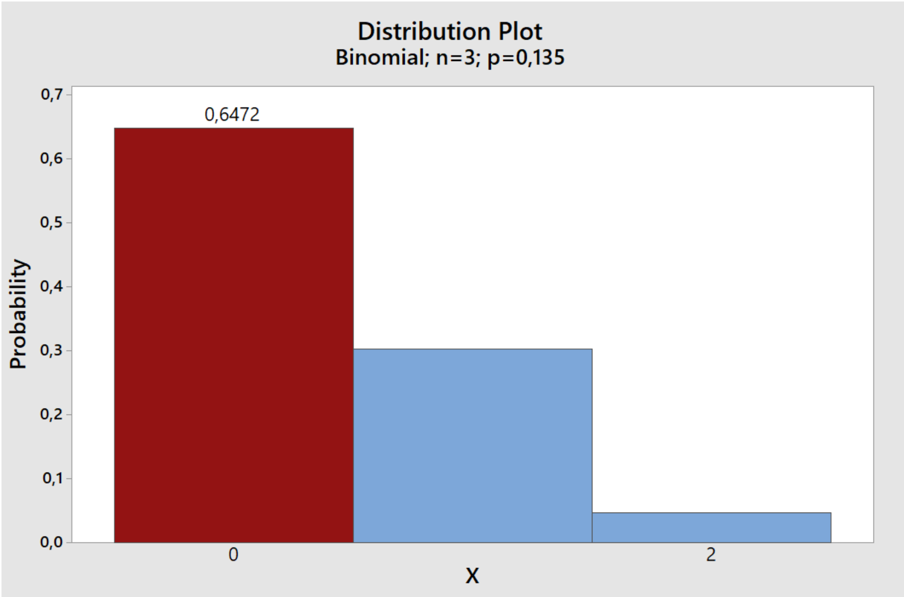
$$P(2) = \binom{3}{2} \cdot 0{,}135^2 \cdot 0{,}865^1 = 0{,}04729$$



Slika 6. Minitab: Primjer 4.2. b)

b) Vjerojatnost da nijedna baterija ne traje 7 sati ili više:

$$P(0) = \binom{3}{0} \cdot 0{,}135^0 \cdot 0{,}865^3 = 0{,}6472$$



Slika 7. Minitab: Primjer 4.2. b)

Primjer 4.3. Sustav za vođenje rakete ispravno radi s vjerojatnosti p. Nezavisni, ali identični backup sustavi instalirani su u raketu. Vjerojatnost da će barem jedan sustav ispravno proraditi kada se pozove iznosi 0,99. Neka je n broj sustava za vođenje. Koliki mora biti n da bi se postigla vjerojatnost barem jednog operativnog sustava za vođenje ako je:

a) $p = 0{,}9$
b) $p = 0{,}8$

Rješenje:

Broj sustava za vođenje označujemo s n. Ako su sustavi identični i neovisni, x podrazumijeva binomnu raspodjelu. Prema tome,

$$P(x \geq 1) = 1 - P(x = 0) = 1 - \binom{n}{0} \cdot p^0 \cdot q^{n-0} = 1 - (1-p)^n$$

$$1 - P(x \geq 1) = (1-p)^n$$

Logaritmiranjem i sređivanjem dobivamo izraz

$$n = \frac{\log[1 - P(x \geq 1)]}{\log[1 - p]}$$

a) Uvrštavanjem zadanih vrijednosti u prethodni izraz dobivamo:

$$n = \frac{\log(1-0{,}99)}{\log(1-0{,}9)} = \frac{\log(0{,}01)}{\log(0{,}1)} = 2$$

Da bi se postigla vjerojatnost barem jednog operativnog sustava, uz $p = 0{,}9$, potrebna su dva sustava za vođenje.

b) Isti slučaj, samo druge vrijednosti:

$$n = \frac{\log(1-0{,}99)}{\log(1-0{,}8)} = \frac{\log(0{,}01)}{\log(0{,}2)} = 2{,}86 \to 3$$

Znači, ako je $p = 0{,}8$, potrebna su tri sustava za vođenje da bi se postigla vjerojatnost barem jednog operativnog sustava.

Primjer 4.4. Firma koja se bavi eksploatacijom nafte namjerava napraviti 10 bušotina, svaku s vjerojatnošću od 0,1 za uspješnu proizvodnju nafte. Izrada jedne bušotine košta 10.000 €. Uspješna bušotina donosi prihod od 500.000 €. Potrebno je pronaći:

a) Očekivanu dobit od tih 10 bušotina.
b) Standardnu devijaciju dobiti firme.

Rješenje:

a) Očekivana vrijednost glasi $\mu = E(x) = np = 10 \cdot 0{,}1 = 1$. Znači, za očekivati je da će od 10 bušotina jedna biti uspješna. Dobit se računa kao razlika svih prihoda i rashoda. U našem slučaju: $\mu \cdot 500.000\text{ €} - n \cdot 10.000\text{ €} = 1 \cdot 500.000\text{ €} - 10 \cdot 10.000\text{ €} = 400.000\text{ €}$.
b) Standardna devijacija računa se kao drugi korijen varijance: $\sigma = \sqrt{npq} = \sqrt{10 \cdot 0{,}1 \cdot 0{,}9} = 0{,}9487$. Odnosno $0{,}9487 \cdot 10.000\text{ €} = 9.487\text{ €}$.

Primjer 4.5. Deset motora pripremljeno je za prodaju u određeno skladište. Motori se prodaju po cijeni od 100 €/komad, ali s garancijom o dvostrukom vraćenom iznosu za svaki neispravan motor na koji bi kupac mogao naići. Potrebno je pronaći očekivanu dobit prodavača ako je vjerojatnost da motor bude neispravan 8%. Pretpostavka je da kvaliteta jednog motora neovisna o kvaliteti ostalih.

Rješenje:

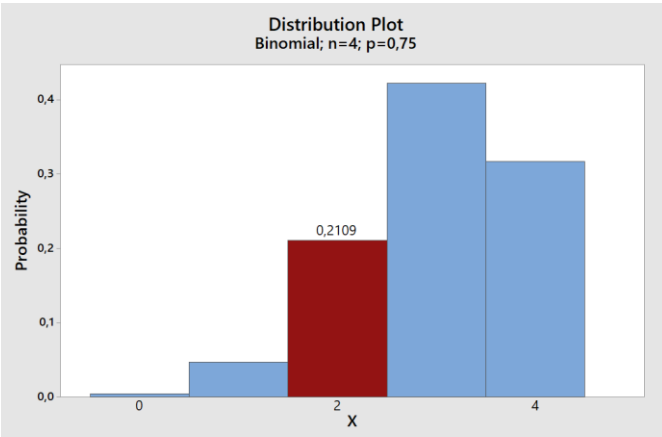
Očekivani broj nesukladnih motora je $\mu = E(x) = np = 10 \cdot 0{,}08 = 0{,}8$. Prodavač se garancijom obvezao vratiti duplu cijenu motora za svaki nesukladni, znači $2 \cdot 100\ € \cdot 0{,}8 = 160\ €$. Ukupni prihod iznosi $10 \cdot 100\ € = 1.000\ €$. Očekivana dobit prodavača je razlika prihoda i rashoda, što u ovom slučaju iznosi $1.000\ € - 160\ € = 840\ €$.

Primjena binomne raspodjele česta je u *medicini*, u području genetike. Slijedi i jedan takav primjer.

Primjer 4.6. Boja očiju kod ljudi definirana je jedinstvenim parom gena. Znamo da je gen za smeđe oči dominantniji od onoga za plave oči. To znači ako osoba ima dva gena za plave oči, imat će plave oči. Dok ako ima bar jedan gen za smeđe oči, imat će smeđe oči. Svaka osoba nasumično nasljeđuje jedan gen za boju očiju od oba roditelja. Ako najstarije dijete ima plave oči, koja je vjerojatnost da točno dvoje od još preostalo četvero djece (nijedno nisu blizanci) ima plave oči? Oba roditelja imaju smeđe oči.

Rješenje: Budući da najstarije dijete ima plave oči, a roditelji smeđe oči, to znači da svaki roditelj ima jedan gen za plave i jedan gen za smeđe oči. Vjerojatnost *p* da dijete ima plave oči iznosi $p = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0{,}25$. Stoga vjerojatnost da točno dvoje od preostalo četvero djece ima plave oči je sljedeća:

$$P(x = 2) = \binom{4}{2} \cdot 0{,}25^2 \cdot 0{,}75^2 = 0{,}2109$$



Slika 8. Minitab: Primjer 6

4.5. Poissonova raspodjela

Poissonova raspodjela predstavlja događaje koji se pojavljuju u određenom vremenskom razdoblju ili u određenoj duljini, području ili volumenu. Na primjer:

- Broj pogrešaka u kvadratnom metru tkanine.
- Broj bakterijskih kolonija u kubičnom centimetru vode.
- Broj neuspješnog paljenja stroja u tijekom radnog dana.

$$P(x) = \frac{m^x}{x!} \cdot e^{-m} \quad \text{za } x = 0, 1, 2, \ldots$$

$$\mu = E(x) = m \qquad \sigma^2 = V(x) = m$$

Poissonova raspodjela određena je jednim parametrom:

- ✓ m – očekivana vrijednost. Povezavši ju s binomnom raspodjelom, parametar m računamo i kao $m = n \cdot p$.

4.6. Primjeri

Primjer 4.7. Za određenu proizvodnju prosječno se dogode tri industrijske nesreće u tjedan dana. Potrebno je naći vjerojatnost da se:

a) Ne pojavi niti ni jedna greška u tjedan dana.

b) Pojave točno dvije greške u tjedan dana.

c) Pojave najviše četiri greške u tjedan dana.

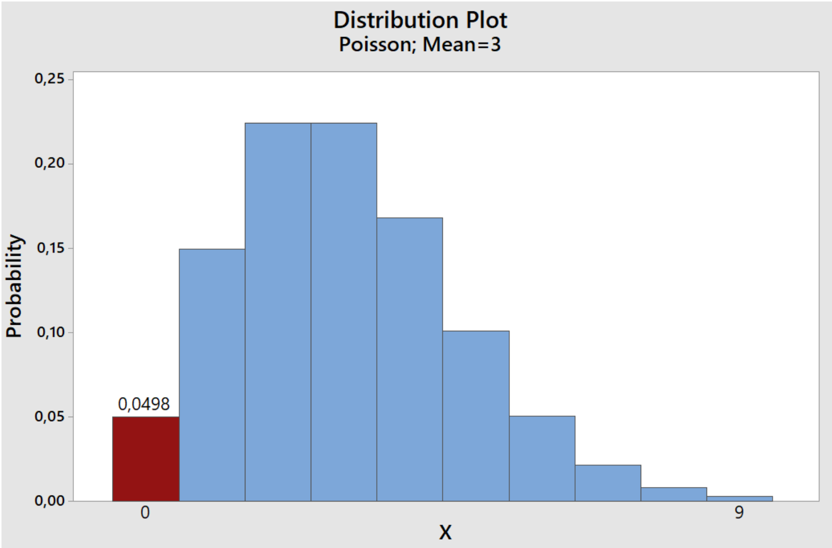
d) Pojave dvije greške u jednom danu.

Rješenje:

Iz teksta zadatka vidimo da je očekivani broj nesreće u tjedan dana $m = 3$. Stoga imamo:

a) Vjerojatnost da se ne pojavi ni jedna iznosi:

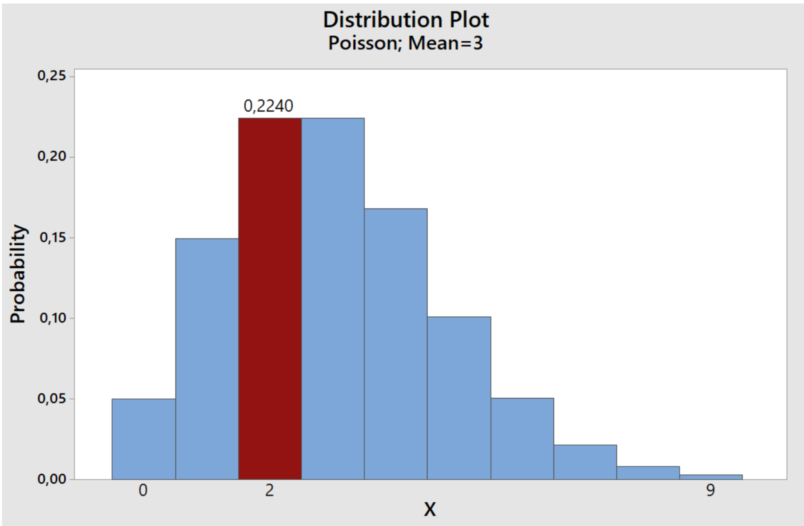
$$P(0) = \frac{3^0}{0!} \cdot e^{-3} = 0{,}0498$$



Slika 9. Minitab: Primjer 4.7. a)

b) Vjerojatnost pojave dviju grešaka:

$$P(2) = \frac{3^2}{2!} \cdot e^{-3} = 0{,}2240$$



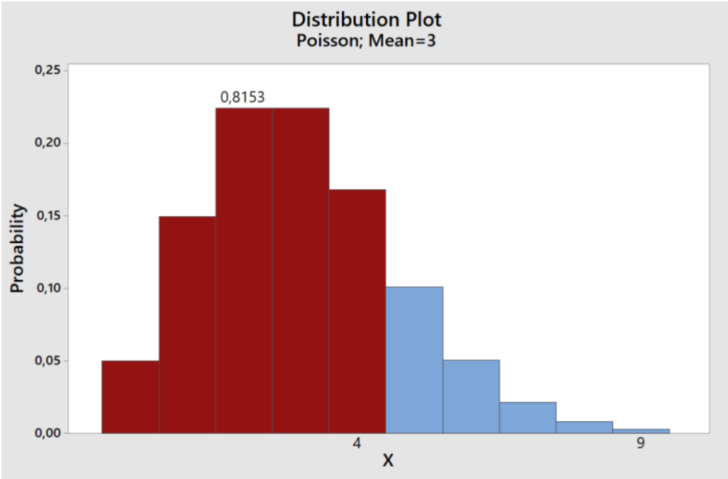
Slika 10. Minitab: Primjer 4.7. b)

c) Vjerojatnost da se ne pojavi više od četiri greške:

$$P(x \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$$

$$= \frac{3^0}{0!} \cdot e^{-3} + \frac{3^1}{1!} \cdot e^{-3} + \frac{3^2}{2!} \cdot e^{-3} + \frac{3^3}{3!} \cdot e^{-3} + \frac{3^4}{4!} \cdot e^{-3}$$

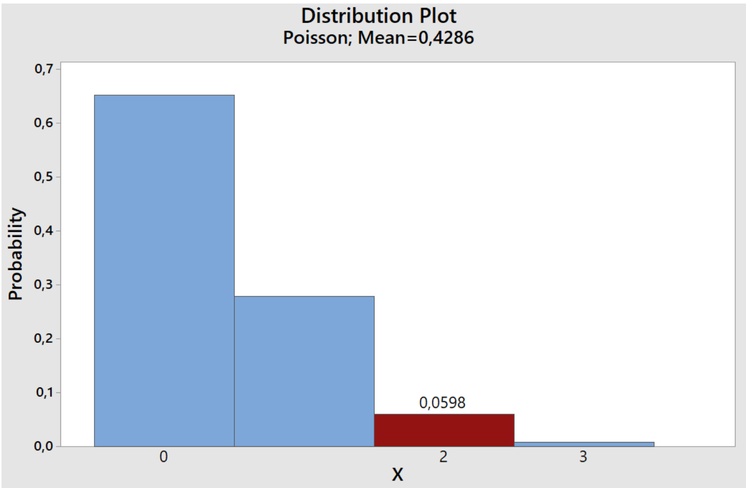
$$= 0{,}8153$$



Slika 11. Minitab: Primjer 4.7. c)

d) U ovom dijelu zadatka traži nas vjerojatnost pojave dviju grešaka u jednom danu, stoga računamo i novu očekivanu vrijednost, odnosno broj grešaka u jednom danu $m = 3/7 = 0{,}4286$. Pa je:

$$P(2) = \frac{0{,}4286^2}{2!} \cdot e^{-4286} = 0{,}0598$$



Slika 12. Minitab: Primjer 4.7. d)

Primjer 4.8. Rukovoditelj industrijskih postrojenja planira kupiti dva novi stroj – stroj tipa A ili stroj tipa B. Broj dnevnih popravka x koji su potrebni za održavanje stroja A opisan je Poissonovom varijablom s očekivanjem $0{,}10t$, gdje je t broj radnih sati stroja u danu. Broj dnevnih popravaka *y* stroja B je Poissonova varijabla s očekivanjem $0{,}12t$. Dnevni trošak stroja A opisan je funkcijom $C_{\mathrm{A}}(t) = 10t + 30x^2$, a dnevni trošak stroja B funkcijom $C_{\mathrm{B}}(t) = 8t + 30y^2$. Pretpostavimo da popravci strojeva oduzimaju neznatno vrijeme i da se strojevi čiste svaku noć tako da svaki dan rade kao novi strojevi. Koji stroj minimizira očekivani dnevni trošak svaki dan radi (a) 10 sati? (b) 20 sati?

Rješenje:

Očekivani trošak za stroj A iznosi:

$$E[C_{\mathrm{A}}(t)] = 10t + 30E(x^2)$$

$$= 10t + 30[V(x) + (E(x))^2]$$

$$= 10t + 30[0{,}10t + 0{,}01t^2]$$

$$= 13t + 0{,}3t^2$$

Analogno tome, očekivani trošak stroja B iznosi:

$$E[C_{\mathrm{B}}(t)] = 8t + 30E(y^2)$$

$$= 8t + 30[V(y) + (E(y))^2]$$

$$= 8t + 30[0{,}12t + 0{,}0144t^2]$$

$$= 11{,}6t + 0{,}432t^2$$

a) U dobivene izraze uvrstimo zadane vrijednosti, znači:

   $E[C_{\mathrm{A}}(10)] = 13 \cdot 10 + 0{,}3 \cdot 10^2 = 160$

   $E[C_{\mathrm{B}}(10)] = 11{,}6 \cdot 10 + 0{,}432 \cdot 10^2 = 159{,}2$

   Dobiveni rezultati pokazuju da povoljniji stroj B.

b) Isto tako računamo za vremenski period od 20 sati.

   $E[C_{\mathrm{A}}(20)] = 13 \cdot 20 + 0{,}3 \cdot 20^2 = 380$

   $E[C_{\mathrm{B}}(20)] = 11{,}6 \cdot 20 + 0{,}432 \cdot 20^2 = 404{,}8$

   Ovog puta povoljniji je stroj A.

Primjer 4.9. Kvaliteta tvrdog diska računala mjeri se prolaskom diska kroz uređaj koji mjeri broj preskočenih impulsa. Određeni proizvođač tvrdog diska ima prosječno 0,1 preskočeni impuls po disku. Traži se:

a) Vjerojatnost da sljedeći kontroliran disk nema ni jedan preskočeni impuls.
b) Vjerojatnost da sljedeći kontrolirani disk ima više od jednog preskočeni impulsa.

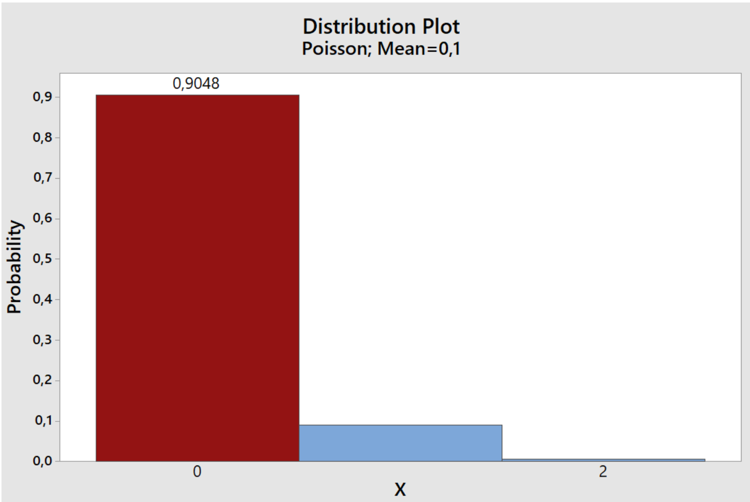
c) Vjerojatnost da ni jedan od sljedeća dva kontrolirana diska ne sadrže ni jedan preskočeni impuls.

Rješenje:

Iz tekstu zadatka je zadano da je očekivana vrijednost $m = 0{,}1$. Varijabla x označava broj preskočenih impulsa.

a) Vjerojatnost da sljedeći kontroliran disk nema ni jedan preskočeni impuls, dakle $x = 0$.

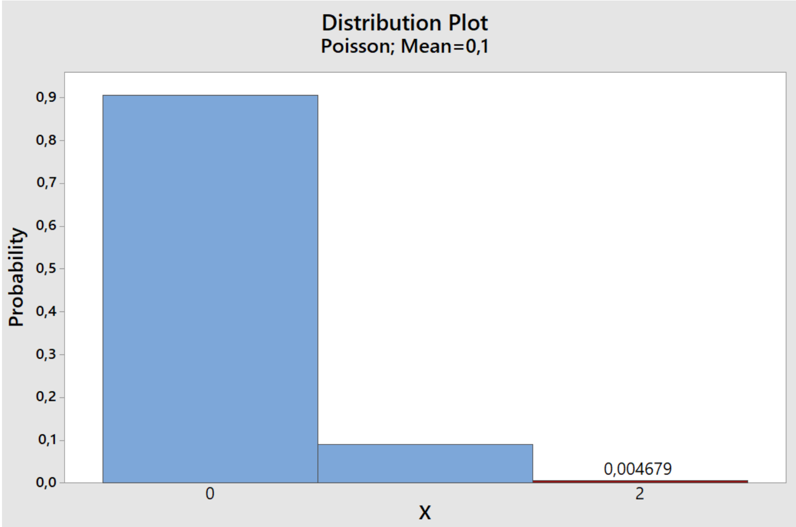
$$P(0) = \frac{0{,}1^0}{0!} \cdot e^{-0{,}1} = 0{,}9048$$



Slika 13. Minitab: Primjer 4.9. a)

b) Vjerojatnost da sljedeći kontrolirani disk ima više od jednog preskočeni impulsa.

$$P(x > 1) = 1 - P(0) - P(1) = 1 - \frac{0{,}1^0}{0!} \cdot e^{-0{,}1} - \frac{0{,}1^1}{1!} \cdot e^{-0{,}1} = 0{,}0047$$



Slika 14. Minitab: Primjer 4.9. b)

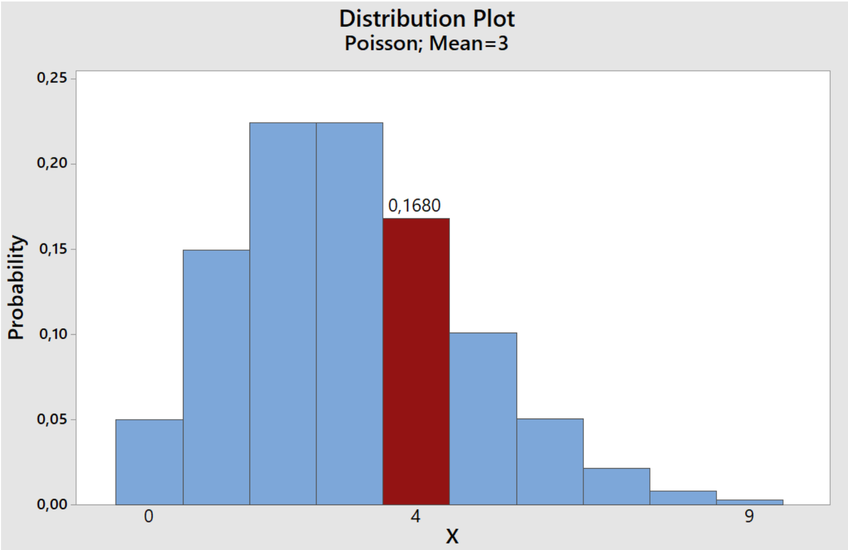
c) Vjerojatnost da sljedeći disk ne sadrži preskočeni impuls iznosi $P_1(0) = 0{,}9048$. Poissonovom raspodjelom svaki kontrolirani disk ima istu vjerojatnost za svaki ishod, stoga vjerojatnost da drugi disk nema ni jedan preskočeni impuls također iznosi $P_2(0) = 0{,}9048$. Vjerojatnost da oba diska nemaju ni jedan preskočeni disk iznosi $P_1(0) \cdot P_2(0) = 0{,}9048 \cdot 0{,}9048 = 0{,}8187$.

Primjer 4.10. U vremenskom periodu od sat vremena prosječno tri popravljena sustava paljenja zrakoplova odlaze s postrojenja za preradu zrakoplova. Odjel za sklapanje zrakoplova treba četiri sustava paljenja u sljedećih sat vremena. Kolika je vjerojatnost da će to biti izvedivo?

Rješenje:

Primjer Poissonove raspodjele s očekivanjem $m = 3$, a varijabla x označava broj popravljenih sustava paljenja zrakoplova. Vjerojatnost da četiri sustava paljenja budu gotova u sljedećih sat vremena je sljedeća:

$$P(4) = \frac{3^4}{4!} \cdot e^{-3} = 0{,}1680$$



Slika 15. Minitab: Primjer 4.10.

Primjer 4.11. Prosječno u trgovinu dođe osam kupaca u sat vremena i ti dolasci se ponašaju po Poissonovoj raspodjeli. Za vremenski period od sat vremena potrebno je pronaći vjerojatnost da:

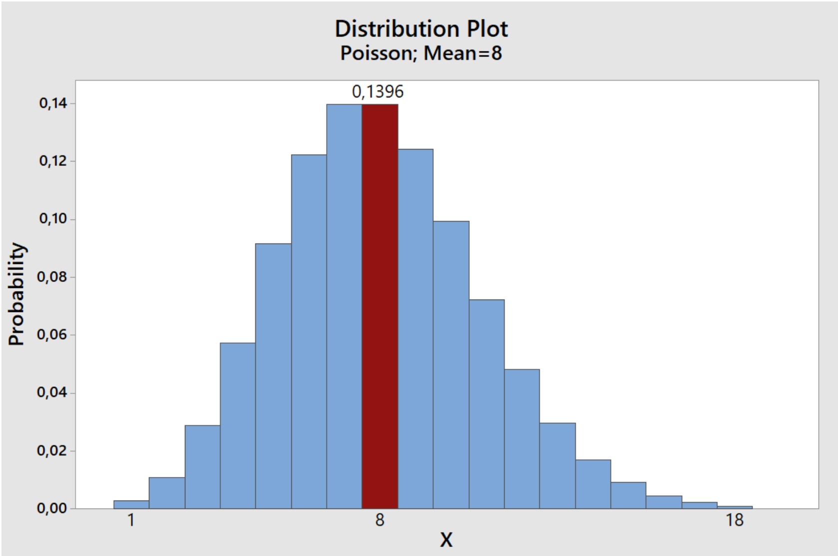
a) Točno osam kupaca dođe u trgovinu.
b) U trgovinu ne dođu više od tri kupca.
c) Najmanje tri kupca dođu u trgovinu.

Rješenje:

Očekivanje iznosi $m = 8$, varijabla x označava broj dolaska kupaca u trgovinu.

a) Vjerojatnost da osam kupaca dođe u trgovinu ($x = 8$).

$$P(8) = \frac{8^8}{8!} \cdot e^{-8} = 0{,}1396$$



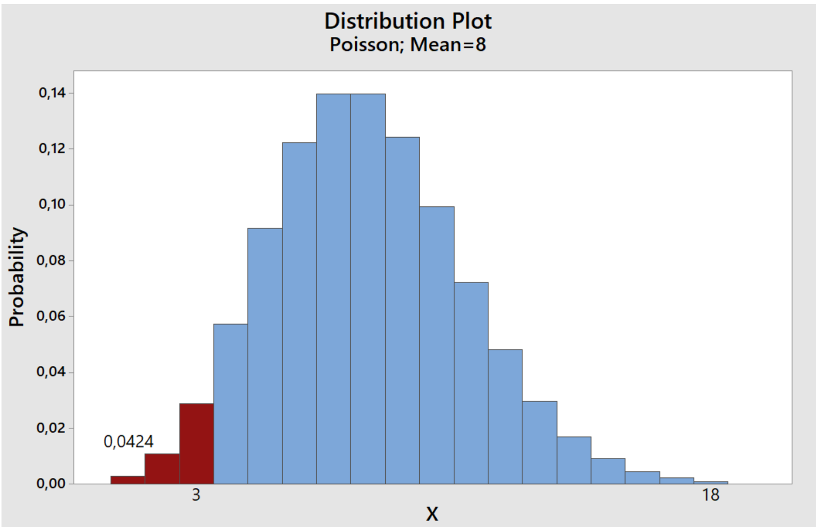
Slika 16. Minitab: Primjer 4.11. a)

b) Vjerojatnost da u trgovinu ne dođu više od tri kupca ($x \leq 3$)

$$P(x \leq\ 3) = P(0) + P(1) + P(2) + P(3)$$

$$= \frac{8^0}{0!} \cdot e^{-8} + \frac{8^1}{1!} \cdot e^{-8} + \frac{8^2}{2!} \cdot e^{-8} + \frac{8^3}{3!} \cdot e^{-8}$$

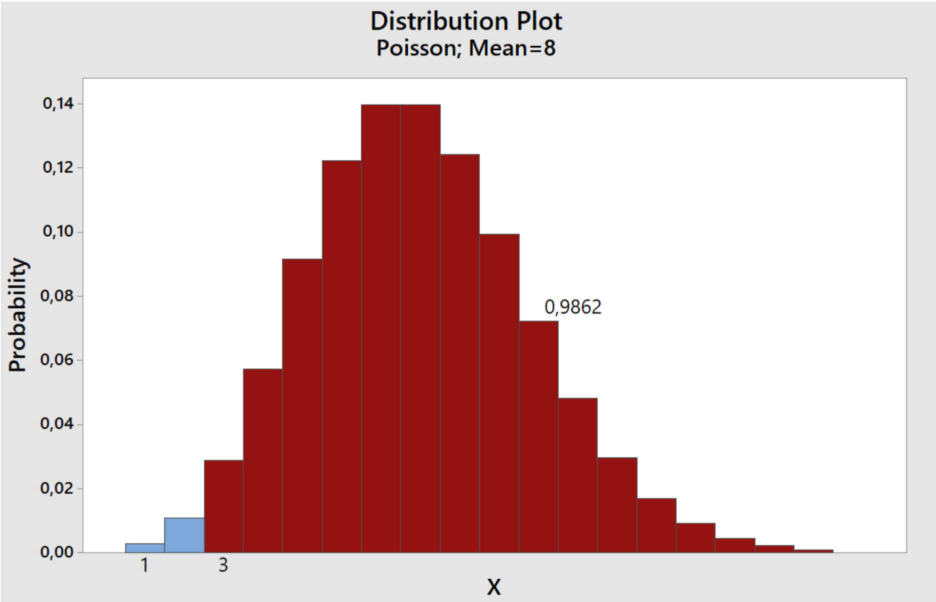
$$= \left(\frac{8^0}{0!} + \frac{8^1}{1!} + \frac{8^2}{2!} + \frac{8^3}{3!}\right) \cdot e^{-8} = 0{,}0424$$



Slika 17. Minitab: Primjer 4.11. b)

c) Vjerojatnost da dođu najmanje tri kupca u trgovinu ($x \geq 3$)

$$P(x \geq 3) = 1 - P(0) - P(1) - P(2) = 1 - \frac{8^0}{0!} \cdot e^{-8} - \frac{8^1}{1!} \cdot e^{-8} - \frac{8^2}{2!} \cdot e^{-8}$$
$$= 0{,}9862$$



Slika 18. Minitab: Primjer 4.11. c)

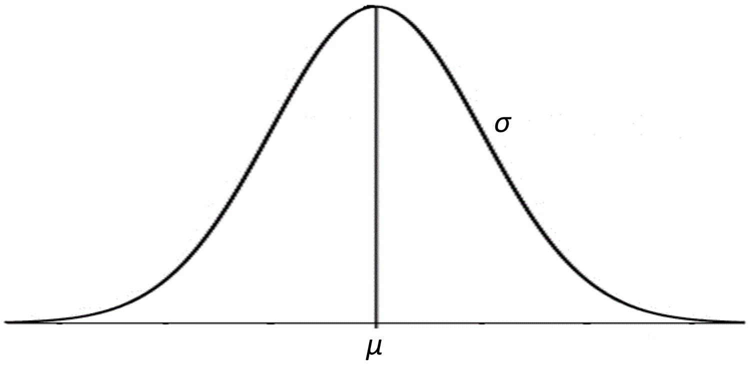
4.7. Normalna raspodjela

Najraširenija raspodjela od svih raspodjela kontinuiranih varijabli je normalna raspodjela. Normalna raspodjela naziva se još i Gaussovom raspodjelom jer ju je prvi upotrijebio Nijemac Carl Fredrich Gauss krajem 18. stoljeća.

Glavni uvjeti da se podaci nekog mjerenja ponašaju po normalnoj raspodjeli su:

1. Da se ono što mjerimo stvarno raspoređuje po normalnoj raspodjeli.
2. Da imamo veliki broj rezultata mjerenja.
3. Da su sva mjerenja provedena jednakom metodom i u što sličnijim
4. vanjskim uvjetima.

Krivulja normalne raspodjele izgleda kako prikazuje Slika 19.



Slika 19. Krivulja normalne raspodjele

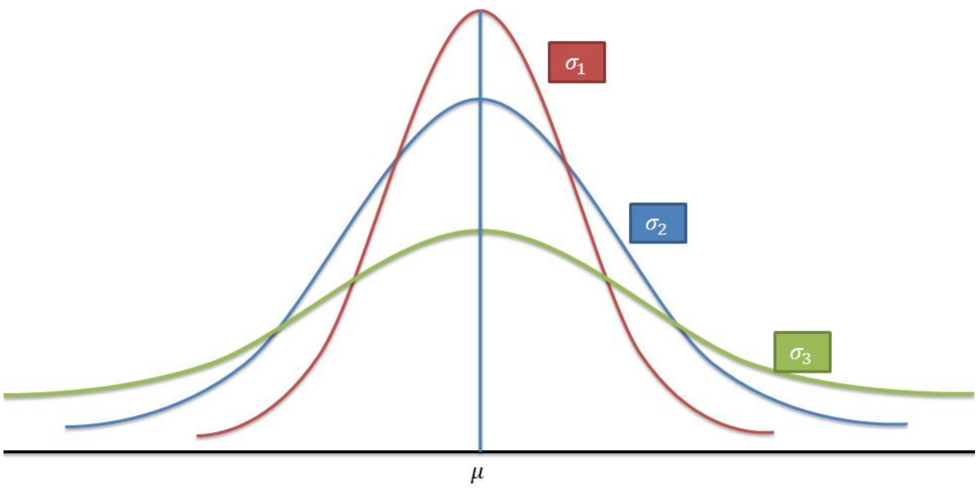
Normalna raspodjela definirana je dvama parametrima:

- ✓ μ – očekivana vrijednost
- ✓ σ – standardna devijacija

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \text{za } -\infty < x < \infty$$

$$E(x) = \mu \qquad V(x) = \sigma^2$$

Ta dva parametra u potpunosti prikazuju centar i širinu funkciju normalne raspodjele. Slika 20. prikazuje krivulje normalne raspodjele s istom aritmetičkom sredinom, ali različitim standardnim devijacijama. Za prikaz na slici vrijedi: $\sigma_1 < \sigma_2 < \sigma_3$. Što standardna devijacija ima manju vrijednost, krivulja normalne raspodjele je uža.



Slika 20. Prikaz krivulja normalnih raspodjela

4.6.1. Jedinična normalna raspodjela

Jedinična normalna raspodjela je standardizirana normalna raspodjela s parametrima $\mu = 0$ i $\sigma = 1$. U računima, sve druge normalne raspodjele *z-trasformacijom* svodimo na jediničnu normalnu raspodjelu. Simbol z označava jediničnu standardnu varijablu.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{za } -\infty < z < \infty$$

$$E(z) = 0 \qquad V(z) = 1$$

$$z = \frac{x - \mu}{\sigma}$$

U nastavku slijedi nekoliko primjera zadataka iz područja normalne raspodjele s detaljnim rješenjima i grafičkim prikazom rezultata.

4.8. Primjeri

Primjer 4.12. Prijemnom ispitu na dvaju fakulteta (fakultet A i fakultet B) svake godine pristupi više tisuća studenata. Broj bodova osvojenih na ispitu prikazanih u histogramu izgleda približno kao krivulja normalne raspodjele, stoga možemo reći da su brojevi bodova osvojenih na ispiti približno normalno distribuirani. Prošle godine fakultet A na prijemnom ispitu iz matematike imao je prosječan broj bodova 480, sa standardnom devijacijom od 100. Broj bodova prijemnog ispita iz matematike na fakultetu B imao je prosječan broj bodova 18 i standardnu devijaciju 6.

a) Fakultet A postavio je granicu prolaznosti od 550 bodova. Koji postotak studenata neće zadovoljiti uvjet?
b) Koji broj bodova na fakultetu B bi bio minimalan za prolaz, uz istu prolaznost kao na fakultetu A?
c) Koja je vjerojatnost da slučajno odabrani student ima više od 700 bodova na ispitu na fakultetu A?

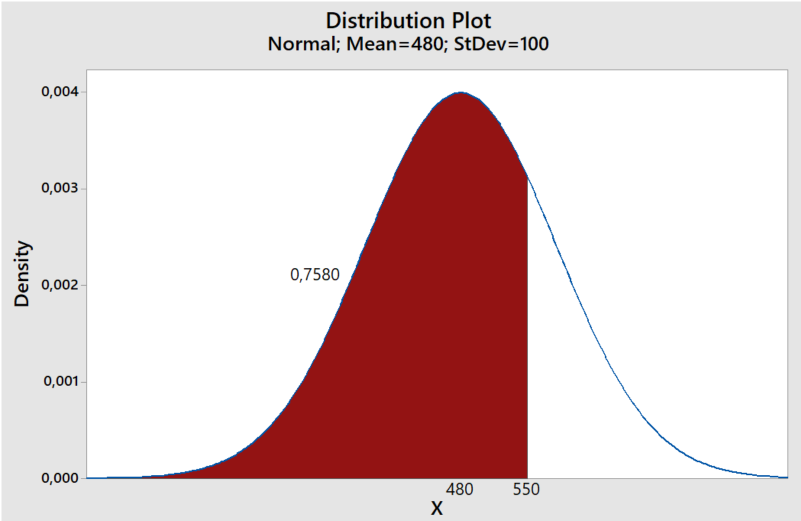
Rješenje:

a) Iz teksta zadatka očito je da je $\mu = 480$, a $\sigma = 100$. Varijablu $x = 550$ koja se ponaša po normalnoj raspodjeli transformirat ćemo u z-varijablu jedinične normalne raspodjele.

$$z = \frac{x - \mu}{\sigma} = \frac{550 - 480}{100} = 0{,}70$$

Dakle,

$$P(x < 550) = P(z < 0{,}70) = 0{,}5 + 0{,}2580 = 0{,}7580$$



Slika 21. Minitab: Primjer 4.12. a)

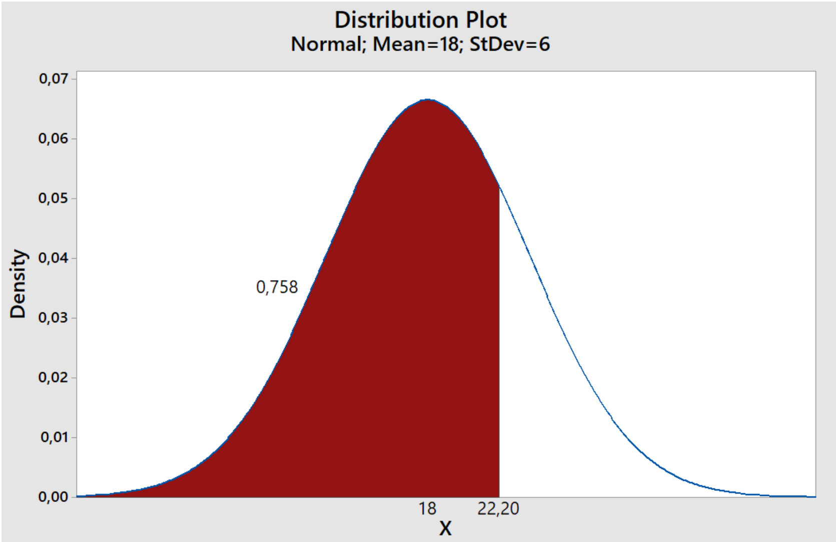
Na prijemnom ispitu fakulteta A 75,8% studenata ne bi zadovoljilo uvjet, slika 15.

b) Za slučaj fakulteta B, gdje je $\mu = 18$, a $\sigma = 6$, postavlja se pitanje „koju vrijednost ima varijabla x za vjerojatnost od 75,8%?“

Za vjerojatnost od 0,758 odgovara varijabla z jedinične normalne raspodjele u iznosu od 0,70.

$$z = \frac{x - \mu}{\sigma} \rightarrow x = \mu + z\sigma = 18 + 0{,}70 \cdot 6 = 22{,}2$$

Rezultat na prijemnom ispitu u iznosu od 22,2 bodova na fakultetu B je ekvivalentan rezultatu od 550 bodova na fakultetu A.

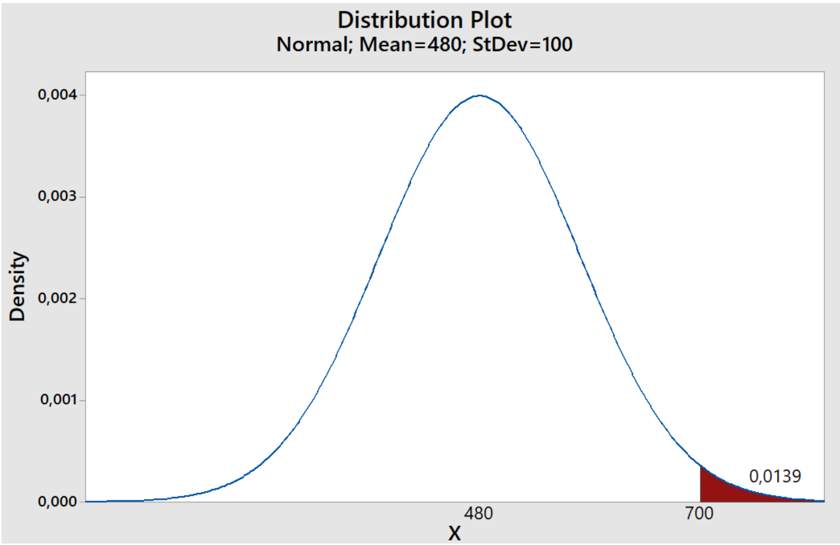


Slika 22. Minitab: Primjer 4.12. b)

c) Vjerojatnost da slučajno odabrani student ima više od 700 bodova na ispitu na fakultetu A iznosi:

$$z = \frac{x - \mu}{\sigma} = \frac{700 - 480}{100} = 2{,}2$$

$$P(x > 700) = P(z > 2{,}2) = 0{,}5 - 0{,}4861 = 0{,}0139$$



Slika 23. Minitab: Primjer 4.12. c)

Primjer 4.13. Debljina plastičnih ploča ja slučajna varijabla. Možemo pretpostaviti da je to kontinuirana slučajna varijabla koja ima normalnu distribuciju s očekivanjem 12 mm i standardnom devijacijom 0,04 mm. Kolika je vjerojatnost defektne ploče ako je kontrola dala kriterij:

a) ploča tanja od 11,96 mm?
b) ploča deblja od 12,06 mm?

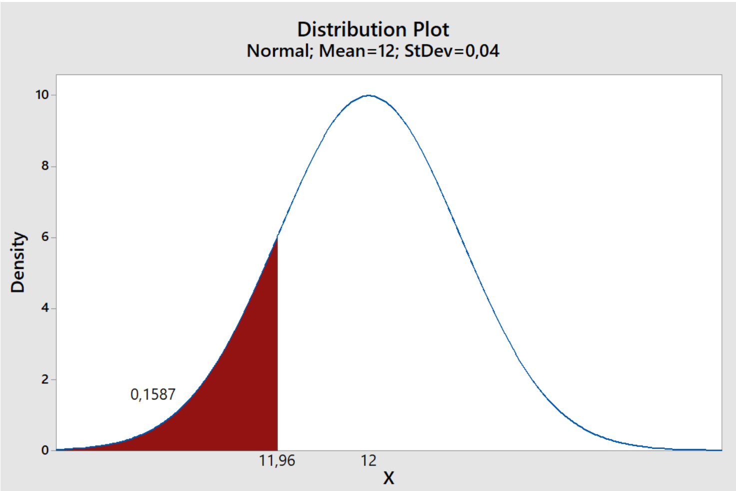
U kojim granicama treba biti debljina ploče da bi očekivani postotak defektnih ploča bio 2%?

Rješenje:

a) Vjerojatnost da je ploča tanja od 11,96 mm?

$$z = \frac{x - \mu}{\sigma} = \frac{11{,}96 - 12}{0{,}04} = -1$$

$$P(x < 11{,}96) = P(z < -1) = 0{,}5 - 0{,}3413 = 0{,}1587$$

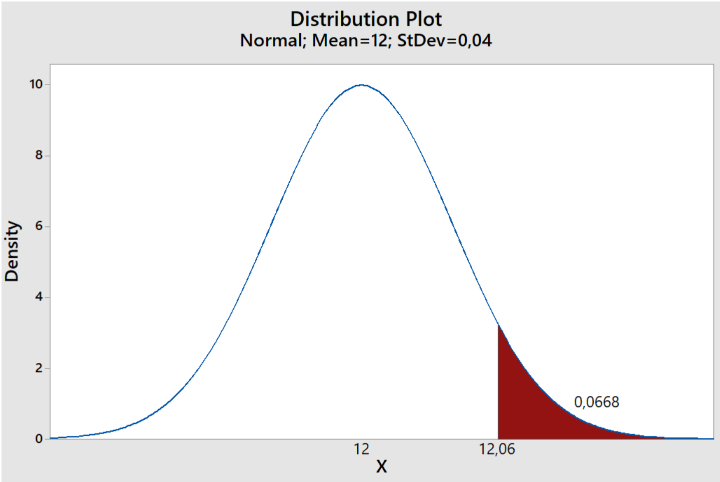


Slika 24. Minitab: Primjer 4.13. a)

b) Vjerojatnost da ploča deblja od 12,06 mm

$$z = \frac{12{,}06 - 12}{0{,}04} = \frac{12{,}06 - 12}{0{,}04} = 1{,}5$$

$$P(x > 12{,}06) = P(z > 1{,}5) = 0{,}5 - 0{,}4332 = 0{,}0668$$

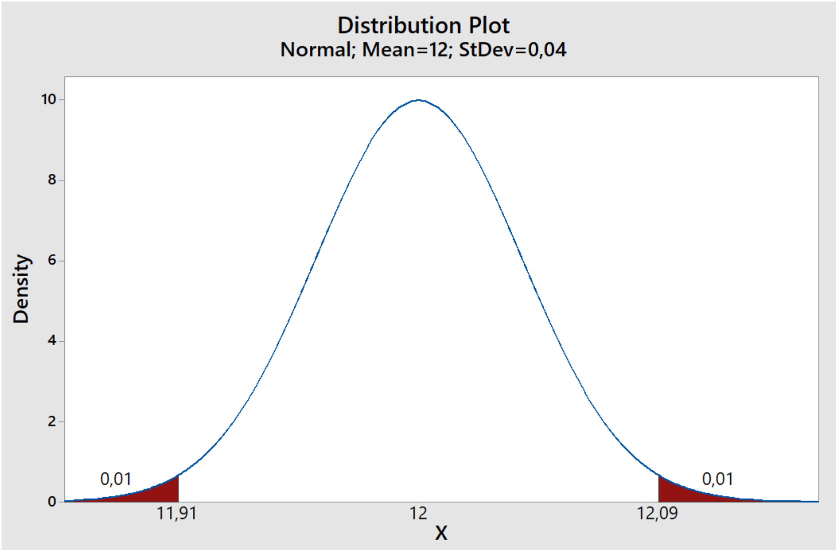


Slika 25. Minitab: Primjer 4.13. b)

Posljednje pitanje u ovom zadatku glasi: U kojim granicama treba biti debljina ploče da bi očekivani postotak defektnih ploča bio 2%? Taj postotak dijeli se na dva repa krivulje u jednakom omjeru, što znači da sa svake strane krivulje imamo 1% nesukladnih, slika 20. Taj je postotak potrebno prilagoditi korištenoj statističkoj tablici, stoga imamo: $P(x) = 0{,}5 - 0{,}01 = 0{,}49$. Iz statističke tablice potrebno je odabrati z za taj postotak i on iznosi: $z = 2{,}33$. Budući da je krivulja normalne raspodjele simetrična obzirom na očekivanu vrijednost, rezultat bilježimo kao $z = \pm 2{,}33$. Nadalje,

$$z = \frac{x - \mu}{\sigma} \rightarrow x = \mu + z\sigma = 12 + (\pm 2{,}33) \cdot 0{,}04$$

$$x_1 = 11{,}91 \text{ i } x_2 = 12{,}09$$



Slika 26. Minitab: Primjer 4.13.

Primjer 4.14. Vrijeme koje je potrebno za diobu stanice mitozom ponaša se po normalnoj raspodjeli s prosječnim vremenom od jednog sada i standardnom devijacijom od 5 minuta.

a) Koja je vjerojatnost da se dioba stanice izvrši za manje od 45 minuta?
b) Koja ja vjerojatnost da se je stanici potrebno više od 65 minuta za diobu?
c) Koliko je vremena potrebno da 99% stanica obavi mitozu?

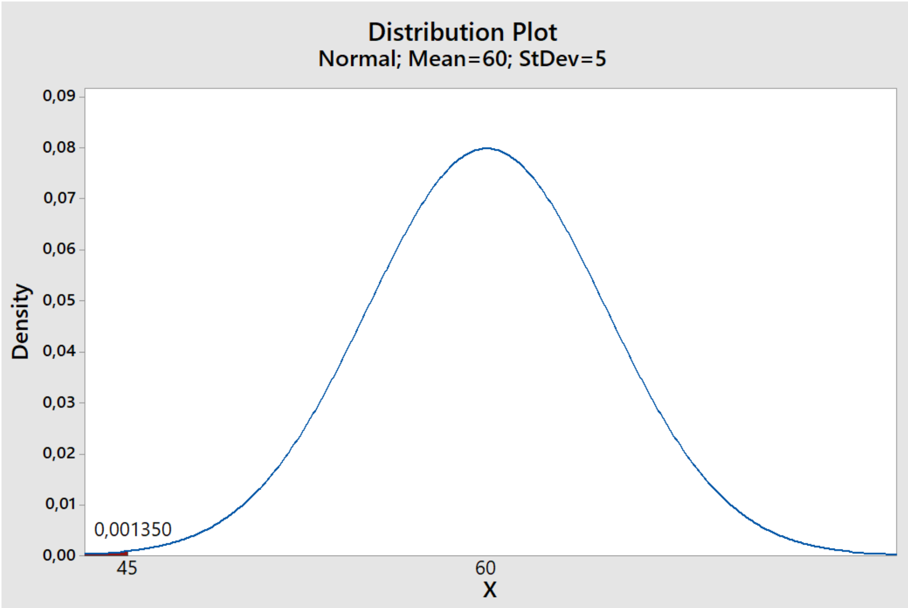
Rješenje:

Za početak, potrebno je svo vrijeme prikazati u istim mjernim jedinicama. Radi lakšeg računa, sve ćemo računati u minutama. Dakle, aritmetička sredina iznosi 60 minuta, a standardna devijacija 5 minuta.

a) Prvo slijedi z-transformacija, zatim očitavanje vjerojatnosti iz statističke tablice.

$$z = \frac{x-\mu}{\sigma} = \frac{45-60}{5} = -3$$

$$P(x < 3) = 0{,}00135$$

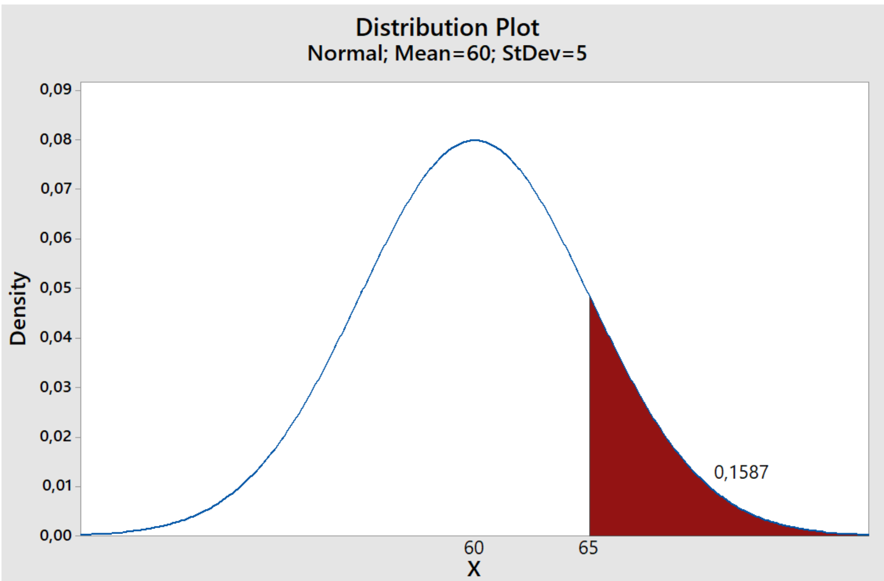


Slika 27. Minitab: Primer 4.14. a)

b) Isti postupak radimo i drugi dio zadatka:

$$z = \frac{x - \mu}{\sigma} = \frac{65 - 60}{5} = 1$$

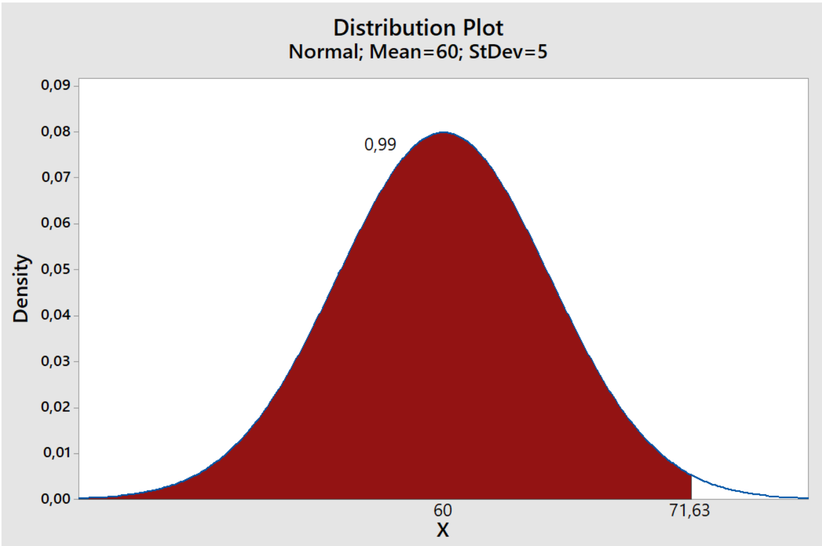
$$P(x > 1) = 0{,}1587$$



Slika 28. Minitab: Primjer 4.14. b)

c) Koliko je vremena potrebno da 99% stanica obavi mitozu? To je sada obrnuti postupak. Prvo iz statističke tablice treba očitati z-vrijednost za 99%, a zatim naći vrijednost x.

$$z(P < 0{,}99) = 2{,}33$$
$$z = \frac{x - \mu}{\sigma} \rightarrow x = \mu + z\sigma = 60 + 2{,}33 \cdot 5 = 71{,}63$$



Slika 29. Minitab: Primjer 4.14 .c)

5. INFERENCIJALNA STATISTIKA

Metode izložene u prethodnom poglavlju su metode deskriptivne statistike. Pomoću njih se opisuju i analiziraju prikupljeni podaci. Prikupljanje podataka o obilježjima svih jedinica statističkog skupa često je preskupo ili zahtijeva previše vremena, a ponekad nije ni moguće (ako je npr. beskonačan). Katkada je u svrhu kontrole kvalitete potrebno uništiti proizvode. Dio proizvoda se u tu svrhu može žrtvovati, ali cijela proizvodnja ne, a ona je ta čija nas kvaliteta zanima. U takvim slučajevima služimo se reprezentativnom promatranju kojim se obuhvaća samo dio jedinca osnovnog skupa. Tako dobiveni podaci čine dio ili podskup skupa podataka koje nazivamo *uzorak*. Do zaključka o osnovnom skupu dolazimo primjenom metoda inferencijalne statistike. Cijeli skup podataka, odnosno svi članovi neke skupine s određenom karakteristikom koju mjerimo nazivamo *populacijom*.

Da bi uzorak mogao odgovoriti zadaćama koje se na njega postavljaju, a to je prije svega da se pomoću njega dobiveni zaključci mogu koristiti na cijeli osnovni skup, on mora biti *reprezentativan*. To znači da po svojim karakteristikama mora biti nalik na osnovni skup, tj. da mora predstavljati osnovni skup u malom. To se postiže ispravno provedenim izborom jedinica u uzorak. Pouzdanu informaciju o populaciji omogućuje *slučajan izbor* elemenata u uzorak.

Pomoću uzoraka, inferencijalna statistika bavi se s dvije vrste postupaka:

1. postupci o procjenjivanju karakteristika osnovnog skupa i
2. ispitivanja istinitosti pretpostavki o nepoznatim karakteristikama populacije.

U ovom radu naglasak će biti na drugoj temi – testiranju hipoteza.

5.1. Testiranje hipoteza

U ovom poglavlju usredotočujemo se na načela testiranja hipoteza, a bit će prikazane i tehnike za rješavanje najčešćih slučajeva testiranja hipoteza koje uključuju jedan, dva ili više uzorka podataka.

Testiranje hipoteze je statistički postupak kojim se određuju da li i koliko pouzdano raspoloživi podaci podupiru postavljenu pretpostavku. Da bismo ispitali istinitost postavljene pretpostavke, postavljamo dvije hipoteze: nultu i alternativnu. Nulta hipoteza H_0 je ona koju testiramo. Definiramo ju kao hipotezu o nepostojanju razlike, bilo između jednog uzorka i neke vrijednosti, bilo između dva ili više uzorka. Alternativa hipoteza H_1 je suprotna hipoteza – hipoteza o postojanju razlike. Uvijek se testira nulta hipoteza.

Razlikujemo jednostrane i dvostrane testove.

5.2. Pogreške kod statističkih testova

U zaključivanju statističkim testovima moguće su i pojave grešaka, a to su:

1. Pogreška 1. vrste – nastaje u slučaju kada se odbaci istinita nulta hipoteza.
2. Pogreška 2. vrste – nastaje kada se ne odbaci neistinita nulta hipoteza te zaključi da nema efekta kada on stvarno postoji.

Tablica 1. Pogreške kod statističkih testova

Hipoteza H_0		Stanje	
		I S T I N I T A	N E I S T I N I T A
ODLUKA	Odbaciti	Pogreška 1. vrste α	I S P R A V N O
	Ne odbaciti	I S P R A V N O	Pogreška 2. vrste β

5.3. Razina značajnosti testa α

Razina značajnosti testa α (nivo signifikantnosti ili razina rizika) je granična vjerojatnost uz koju još uvijek valja prihvatiti eventualno istinitu nultu hipotezu. Izbor razine značajnosti statističkog testa je proizvoljan, a najčešće vrijednosti su 0,05, 0,01 ili 0,001. Prema tome, kada kažemo da je neka razlika „statistički značajna na nivou od 5%“, time zapravo smatramo da „među populacijama stvarno postoji razlika, pri čemu riskiramo oko 5% smo ipak izveli pogrešan zaključak“. Ako je neka razlika statistički značajna, onda to znači da se aritmetičke sredine populacija iz kojih su uzorci razlikuju, ali nikako ne i to da su svi individualni rezultati jedne varijable u populaciji ili uzorku veći (ili manji) od svih individualnih razlika druge varijable.

5.4. *P*-vrijednost

P-vrijednost (engl. *P*-value) je vjerojatnost opažanja podataka kakvi su na promatranom uzorku kada je nulta hipoteza istinita. Najčešća razina značajnosti testa iznosi 0,05, pa ako je *P*-vrijednost manja od 0,05 ($P\text{-vrijednost} < 0{,}05$), nultu hipotezu odbacujemo, a razlike proglašavamo statistički značajnima. Suprotno navedenom, ako je P-vrijednost $\geq 0{,}05$, nultu hipotezu ne odbacujemo, a razlike proglašavamo statistički neznačajnim. Niža *P*-vrijednost znači više dokaza protiv nulte hipoteze.

P-vrijednost se često pogrešno interpretira kao vjerojatnost da je nulta hipoteza istinita. Nulta hipoteza nije nasumična te ona nema vjerojatnost. Ona je ili istinita ili nije. Kod testiranja značajnosti razlike npr. između dvije aritmetičke sredine, ako nađemo da je razlika statistički značajna na razini manjoj od 5% (P-vrijednost $< 0{,}05$), ispravno zaključujemo sljedeće: ako među populacijama iz kojih smo uzeli uzorke nema razlike, onda se razlika kavu smo našim istraživanjem dobili mogla slučajno dogoditi samo u manje od 5% slučajeva. Stoga, uz rizik od 5% odbacujemo nultu hipotezu i zaključujemo da razlika među populacijama postoji.

5.5. Slijed radnji u testiranju hipoteza

1. Postavljanje nulte i alternativne hipoteze.
2. Izbor razine značajnosti (α).
3. Prikupljanje podataka na odgovarajućem uzorku.
4. Računanje vrijednosti rezultata statističkog testa za nultu hipotezu.
5. Usporedba rezultata statističkog testa s vrijednostima iz poznate distribucije vjerojatnosti specifične za dani test.
6. Interpretacija rezultata statističkog testa u terminima vjerojatnosti (*P*-vrijednost).

5.6. Postavljanje nulte i alternativne hipoteze

Nulta hipoteza uvijek se postavlja kao hipoteza o nepostojanju razlike. Alternativna hipoteza postavlja se ovisno o problemu. Tablica 2. prikazuje kako ispravno postaviti hipoteze za t-test, ovisno o broju uzoraka te tipu testa (jednostrani/dvostrani).

Tablica 2. Postavljanje nulte i alternativne hipoteze

Broj uzoraka	H_0	H_1	Tip testa
1 uzorak	$\mu = \mu_0$	$\mu < \mu_0$	Jednostrani test
		$\mu > \mu_0$	
		$\mu \neq \mu_0$	Dvostrani test
2 uzorka	$\mu_1 - \mu_2 = 0$ ili $\mu_1 = \mu_2$	$\mu_1 - \mu_2 < 0$ Ili $\mu_1 < \mu_2$	Jednostrani test
		$\mu_1 - \mu_2 > 0$ ili $\mu_1 > \mu_2$	
		$\mu_1 - \mu_2 \neq 0$ ili $\mu_1 \neq \mu_2$	Dvostrani test

Kod jednostranih testova, osim $\mu = \mu_0$ kao nulta hipoteza često se postavlja i $\mu \geq \mu_0$ ili $\mu \leq \mu_0$ radi bolje predodžbe testiranja podataka.

5.7. Kada odbaciti nultu hipotezu?

Postoje dva načina pri odlučivanju postoji li dovoljno dokaza iz uzorka da se odbaci H_0 ili se ne odbaci H_0, a to su:

1. u odnosu na područje prihvaćanja
   Prvo je potrebno odrediti kritičnu vrijednost statističkog testa koja odvaja područje prihvaćanja od područja odbacivanja. To se područje određuje korištenjem odgovarajuće distribucije i odlučuje se na temelju rizika kojeg smo spremni prihvatiti za odbacivanje nulte hipoteze kada je ona istinita (α-rizik). Dakle, ako statistika pada u područje odbacivanja, nultu hipotezu treba odbaciti. Sve to prikazuje Tablica 3.
2. u odnosu na *P*-vrijednost. Ako je *P*-vrijednost niža od razine značajnosti testa (α), odbacujemo nultu hipotezu. Ako je $\alpha = 0{,}05$, nultu hipotezu odbacujemo ako je *P*-vrijednost niža od 0,05 (P-vrijednost $< 0{,}05$).

Obje metode daju jednake zaključke.

Tablica 3. Područje prihvaćanja u statističkim testovima

H_0	Vrijednost statističkog testa	H_1	Odbaciti H_0 ako je
$\mu = \mu_0$	$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ σ je poznata	$\mu < \mu_0$	$z < -z_\alpha$
		$\mu > \mu_0$	$z > z_\alpha$
		$\mu \neq \mu_0$	$z < -z_{\alpha/2}$ ili $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ $df = n - 1$ σ nije poznata	$\mu < \mu_0$	$t < -t_\alpha$
		$\mu > \mu_0$	$t > t_\alpha$
		$\mu \neq \mu_0$	$t < -t_{\alpha/2}$ ili $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ σ_1 i σ_2 su poznate	$\mu_1 - \mu_2 < d_0$	$z < -z_\alpha$
		$\mu_1 - \mu_2 > d_0$	$z > z_\alpha$
		$\mu_1 - \mu_2 \neq d_0$	$z < -z_{\alpha/2}$ ili $z > z_{\alpha/2}$

$\mu_1 - \mu_2 = d_0$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $df = n_1 + n_2 - 2$ $\sigma_1 = \sigma_1$, ali nepoznate $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 < d_0$	$t < -t_\alpha$
		$\mu_1 - \mu_2 > d_0$	$t > t_\alpha$
		$\mu_1 - \mu_2 \neq d_0$	$t < -t_{\alpha/2}$ ili $t > t_{\alpha/2}$

5.8. Podjela na parametarske i neparametarske testove

Statistički testovi se dijele na parametarske i neparametarske testove. Parametarski testovi podrazumijevaju normalnu raspodjelu, a primjeri su 1-sample z-test, 1-sample t-test, 2-sample t-test, paired t-test itd. Za razliku od parametarskih testova, neparametarski testovi ne zasnivaju na pretpostavci raspodjele podataka; nema srednje vrijednosti i standardne devijacije. Podaci su po svojoj prirodi nominalni ili ordinalni, a analiza podataka se ne svodi na stvarne vrijednosti, već na rang podataka. Primjer neparametarskih testova su χ^2-test, Kruskal-Wallisov test, Friendmanov test itd.

U nastavku su navedeni primjeri problema koji se rješavaju statističkim testovima te detaljan postupak rješenja i objašnjenje svakog problema.

5.9. Primjeri

Primjer 5.1. Dvadeset studenata pristupa ispitu iz kolegija Statistika u mjeriteljstvu. Na temelju ostvarenog broja bodova studenata, možemo li zaključiti da će prosječno student imati više od 600 bodova na ispitu?

Tablica 4. Broj ostvarenih bodova studenata na ispitu

650	730	510	670	480
800	690	530	590	620
710	670	640	780	650
490	800	600	510	700

Rješenje:

Budući da znamo standardnu devijaciju cijele populacije (svi studenti koji pristupaju ispitu), ovdje se radio o z-testu. Prvo postavljamo hipoteze:

H_0: $\mu \leq 600$ i

H_1: $\mu_1 > 600$.

One-Sample Z: C1

```
Test of μ = 600 vs > 600
The assumed standard deviation = 100,1


Variable   N   Mean  StDev  SE Mean  95% Lower Bound     Z      P
C1        20  641,0  100,1     22,4            604,2  1,83  0,033
```

Slika 30. Primjer 5.1.

Ispis iz Minitaba (Slika 30.) prikazuje nam podatke o uzorku: veličina uzorka, aritmetička sredina, standardna devijacija itd., a za donošenje zaključka o postavljenim hipotezama najznačajniji parametar je P-vrijednost. U ovom slučaju P-vrijednost iznosi 0,033, a to je manje od zadane razine značajnosti testa α što znači da imamo dovoljno dokaza da odbacimo nultu hipotezu. Konkretno, studenti će na ispitu ostvariti više od 600 bodova.

Postupak za ručni izračun:

$$\bar{x} = 641 \text{ i } \sigma = 100.$$

Znamo da je

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{641 - 600}{100/\sqrt{20}} = 1{,}8336.$$

Iz statističke tablice za normalnu raspodjelu tražimo kritičnu vrijednost. Tako dolazimo do vrijednosti $z_\alpha = 1{,}645$ (za vjerojatnost od 95%). Budući da je $z > z_\alpha$, odbacujemo nultu hipotezu.

Primjer 5.2. Odjel gradskog zdravstva grada Ozlja koji se bavi analizom vode želi odrediti jesu li prosječne količine bakterija po jedinici volumena vode unutar sigurnosne razine od 200. Prikupljeno je 10 uzoraka vode i prikazana ja količina bakterija po jedinici vode:

175	190	215	198	184
207	210	193	196	180

Prema prikazanim podacima, možemo li zaključiti da je količina bakterije unutar sigurnosne razine? Razina značajnosti testa je 0,01.

Rješenje:

Prvo je potrebno napisati hipoteze.

H_0: $\mu \geq 200$ i

H_1: $\mu < 200$.

Nulta hipoteza označava slučaj kada je količina bakterija veća od dopuštene, a alternativna hipoteza predstavlja slučaj kada je količina bakterija u sigurnosnom području, odnosno ispod granične razine.

Za navedene podatke vrijedi

$$\bar{x} = 194{,}8 \text{ i } s = 13{,}14.$$

Stoga,

$$t = \frac{194{,}8 - 200}{\sqrt{13{,}14^2/10}} = -1{,}25.$$

Iz statističke tablice za slučajnu t-varijablu za $\alpha = 0{,}01$ i $\mathrm{df} = 10 - 1 = 9$ očitamo $t_\alpha = 2{,}281$. Budući da $t = -1{,}25 > -t_\alpha = -1{,}725$, nemamo dovoljno dokaza da odbacimo nultu hipotezu.

One-Sample T: C2

```
Test of μ = 200 vs < 200

Variable   N    Mean  StDev  SE Mean  99% Upper Bound      T      P
C2        10  194,80  13,14     4,15           206,52  -1,25  0,121
```

Slika 31. Primjer 5.2.

Do istog zaključka dolazimo i koristeći Minitab, Slika 31. Vidimo da P-vrijednosti iznosi 0,121 i ona je viša od razine značajnosti testa 0,01. To nam govori da nemamo dovoljno dokaza za odbaciti nultu hipotezu, odnosno količina bakterije nije unutar sigurnosne razine.

Primjer 5.3. Dvadeset i dvoje dobrovoljaca u istraživačkom institutu za virusne bolesti dobili su prehladu nakon što su bili izloženi virusu influenze (influenza je virus koji uzrokuje prehladu ili gripu). Nasumičnim odabirom, deset dobrovoljaca dobilo je tablete koje sadrže jedan gram vitamina C. Te tablete su uzimali četiri puta dnevno. Kontrolna skupina koja se sastojala od preostalih 12 dobrovoljaca dobila je *placebo* tablete koje su izgledom i okusom izgledao kao i tablete s vitaminom C. Tablete, bilo s vitaminom C, bilo *placebo*, dobrovoljci su uzimali dok liječnik koji nije znao tko spada u koju skupinu nije odlučio da više ne boluju od prehlade. U nastavku je prikazana zabilježena je duljina vremena trajanja prehlade u danima (Tablica 5.)

Tablica 5. Vrijeme trajanje prehlade

Skupina koja je uzimala tablete s vitaminom C	Skupina koja uzimala *placebo* tablete
5,5	6,5
6,0	6,0
7,0	8,5
6,0	7,0
7,5	6,5
6,0	8,0
7,5	7,5
5,5	6,5
7,0	7,5
6,5	6,0
	8,5
	7,0

Pokazuju li navedeni podaci da uzimanje 4 grama vitamina C smanjuje prosječno trajanje prehlade? Razina značajnosti testa iznosi 5%.

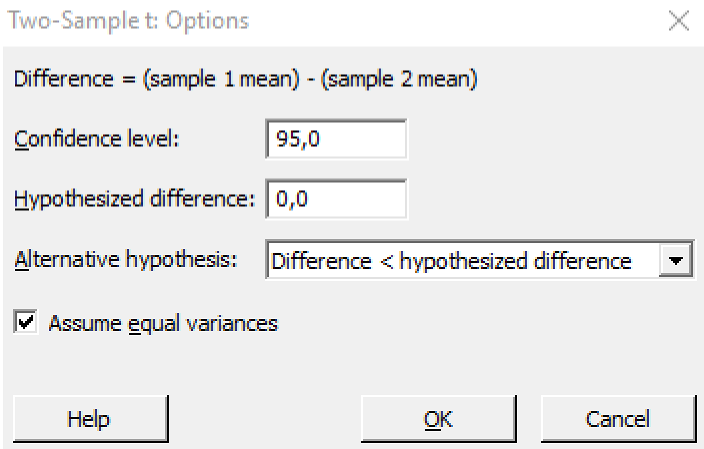
Rješenje:

Prvi korak u rješavanju navedenog problema je postavljanje nulte i alternativne hipoteze. Na ovom primjeru to glasi ovako:

H_0: $\mu_p \leq \mu_C$ i

H_1: $\mu_p > \mu_C$,

gdje je μ_p prosječno trajanje prehlade kod skupine koja je uzimala *placebo* tablete, a μ_C prosječno trajanje prehlade kod skupine koja je uzimala tablete s vitaminom C. Budući da su varijance obje skupine jednake (prije proveden F-test), možemo krenuti s T-testom. Prilikom provođenja koraka u Minitabu (Slika 32.) trebalo je paziti na odabir alternativne hipoteze. Odnosno, trebalo je odabrati da je *difference* < *hypothesized difference*, kako bi to odgovaralo unaprijed postavljenoj alternativnoj hipotezi. Slika 33. prikazuje izlist iz Minitaba, a na njemu su prikazane izračunate vrijednosti. Vidimo da je P-vrijednost u iznosu od 0,036 što znači da imamo dovoljno dokaza da odbacimo nultu hipotezu. Da odgovorimo na postavljeno pitanje: da, navedeni podaci pokazuju da uzimanje 4 grama vitamina C smanjuje prosječno trajanje prehlade.



Slika 32. Two-Sample t: Options

Two-Sample T-Test and CI: C1; C2

```
Two-sample T for C1 vs C2

     N   Mean  StDev  SE Mean
C1  10  6,450  0,762     0,24
C2  12  7,125  0,882     0,25


Difference = μ (C1) - μ (C2)
Estimate for difference:  -0,675
95% upper bound for difference:  -0,062
T-Test of difference = 0 (vs <): T-Value = -1,90  P-Value = 0,036  DF = 20
Both use Pooled StDev = 0,8303
```

Slika 33. Primjer 5.3.

Naravno, ako nemamo pristup Minitabu, isti primjer možemo riješiti ručno na sljedeći način. Prvo izračunamo aritmetičke sredine i varijance svakog uzorka:

$$\bar{x}_1 = 6{,}450,\ \bar{x}_2 = 7{,}125$$

$$s_1^2 = 0{,}581,\ s_2^2 = 0{,}778$$

Stoga,

$$s_p^2 = \frac{9}{20}s_1^2 + \frac{11}{20}s_2^2 = 0{,}689$$

I vrijednost statistike t iznosi

$$t = \frac{6{,}450 - 7{,}125}{\sqrt{0{,}689(\frac{1}{10} + \frac{1}{12})}} = -1{,}90$$

Iz statističke tablice t-raspodjele za $\alpha = 0{,}05$ i $df = 20$ očitamo t_α = 1,725. Budući da je $t\ = -1{,}90\ <\ -t_\alpha\ =\ -1{,}725$ odbacujemo nultu hipotezu.

6. PRAKTIČKI ZNAČAJ STATISTIKE NA ODABRANIM PRIMJERIMA

Primjena statističkih alata neupitna je u mnogim granama znanosti. U nastavku su prikazani primjeri različitih područja radi konkretnog prikaza primjene statistike. Opisani i objašnjeni su problemi intervala tolerancije, anketa te procjene sposobnosti mjernog sustava.

6.1. Interval tolerancije

Interval tolerancije često se koristi za otkrivanje prekomjerne varijacije uspoređujući zahtjeve klijenata s granicama tolerancije koje pokrivaju određeni dio populacije. Ako je interval tolerancije širi od zahtjeva klijenta, može doći do prevelike varijacije proizvoda. Izvedeni iz statistike uzoraka, interval tolerancije je raspon vrijednosti za određenu karakteristiku kvalitete koja vjerojatno pokriva određeni udio populacije. Alternativno, donja ili gornja granica može se postaviti tako da navedeni omjer bude veći ili manji od granice.

Da bismo generirali interval tolerancije, moramo odrediti minimalni postotak populacije koji želimo obuhvatiti, te razinu pouzdanosti. Uobičajeno, obje su vrijednosti vrlo blizu 1 (odnosno 100%). Razina pouzdanosti je vjerojatnost da interval zapravo pokriva minimalni postotak.

Minitab prikazuje vrijednosti za račun po *paramteraskoj* i po *neparametarskoj* metodi. Ako znamo po kojoj su raspodjeli podaci distribuirani, i ako njih Minitab podržava, preporučeno je koristiti rezultate parametarske metode. Inače, koristiti rezultate neparametarske metode. Rezultati parametarske metode precizniji su u odnosu na rezultate neparametarske metode. Minitab ima mogućnost utvrditi ponašaju li se podaci po određenoj raspodjeli za četrnaest raspodjela, uključujući normalnu, lognormalnu, eksponencijalnu, Weibullovu raspodjelu itd. Ako nismo sigurni ponašaju li se podaci populacije po nekoj od ponuđenih raspodjela, koristiti neparamterasku metodu. Ako koristimo neparametarsku metodu, uzorak treba sadržavati vrlo velik dio populacije. Neparametarska metoda zahtijeva samo da su podaci kontinuirani. Ako veličina uzorka nije dovoljno velika, postignuta razina pouzdanosti za naš interval tolerancije može biti znatno niža od željene razine.

Najčešća raspodjela s kojom se susrećemo u računima je normalna raspodjela. Da bismo znali ponaša li se populacija iz koje smo uzeli uzorak po normalnoj raspodjeli, potrebno je taj uzorak testirati. Testiranje se vrši pomoću testa normalnosti. Taj test spada u skupinu testova s jednim uzorkom. Nulta hipoteza testa normalnosti glasi da se populacija ponaša po normalnoj raspodjeli, dok alternativna da se ne ponaša po normalnoj raspodjeli. Da bismo to mogli odrediti, odabiremo jedan od četiri testa: Anderson-Darlingov test, Ryan-Joiner test normalnosti, Kolmogorov-Smirnov test normalnosti ili „Fat pencil“ test. Najpoznatiji je Anderson-Darlingov test – uspoređuje kumulativnu funkciju raspodjele podataka uzorka s očekivanom raspodjelom. Ako je promatrana razlika dovoljno velika, odbacuje se nulta hipoteza o normalnosti populacije.

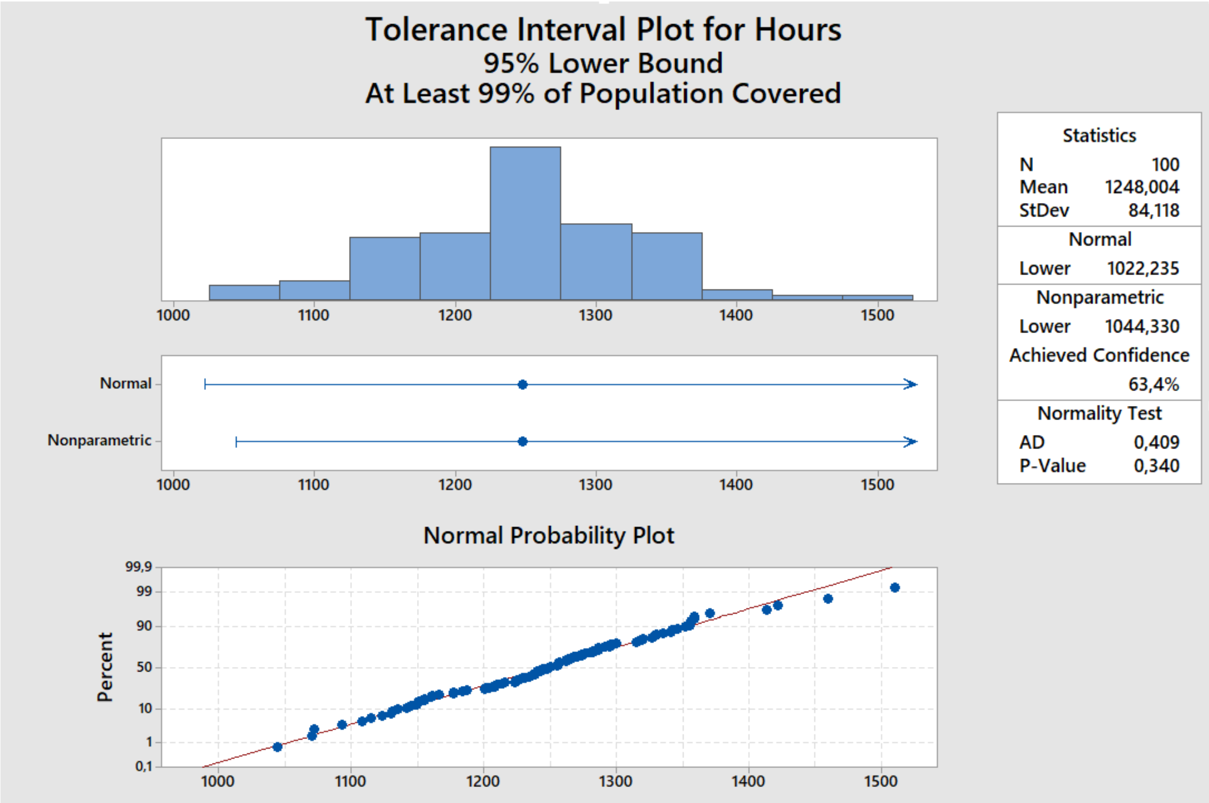
Primjer 6.1.

Proizvođač žarulja želi izračunati životni vijek žarulja (donju granicu) koju premašuje barem 99% žarulja. Proizvođač prikuplja slučajni uzorak od 100 promatranih vremena kako bi izračunao nižu toleranciju. Prikupljene podatke prikazuje Tablica 6.

Tablica 6. Vrijeme životnog vijeka žarulja, h

1215,79	1130,55	1160,95	1187,32	1210,13
1071,83	1335,57	1093,36	1342,02	1262,04
1231,25	1160,07	1135,16	1346,65	1238,68
1224,03	1257,08	1359,03	1422,32	1240,36
1145,04	1329,73	1184,25	1166,29	1358,89
1237,43	1286,74	1123,07	1356,84	1201,33
1141,86	1155,73	1319,46	1235,80	1319,96
1352,61	1070,70	1208,55	1413,94	1330,31
1044,33	1271,07	1149,76	1244,20	1200,40
1234,77	1155,32	1355,47	1286,64	1274,12
1114,90	1273,69	1240,46	1130,44	1255,55
1356,04	1317,31	1282,83	1281,16	1243,11
1256,22	1150,15	1223,81	1247,88	1267,78
1295,33	1208,50	1248,94	1291,82	1151,56
1243,79	1314,87	1267,17	1225,07	1203,85
1255,62	1282,07	1256,00	1214,48	1239,09
1263,61	1276,57	1341,08	1227,56	1263,95
1327,04	1229,51	1510,35	1295,73	1459,85
1370,65	1300,30	1256,53	1296,05	1108,45
1177,60	1177,32	1262,41	1287,07	1250,50

Nakon provedenih koraka u Minitabu, otvara se prozor koji prikazuje Slika 34.



Slika 34. Interval tolerancije

Prozor Tolerance interval plots prikazuje sljedeće:

- Histogram – prikazuje raspodjelu podataka u uzorku. Svaki stupac na histogramu predstavlja učestalost podataka unutar jednog intervala.
- Interval plot – prikazuje srednju te gornju i/ili donju granicu intervala tolerancije za svaku metodu. Okomita linija na kraju intervala predstavlja granicu, a strelica prikazuje da nema granice za tu stranu intervala.
- Normal probability plot – prikazuje koliko podaci odgovaraju normalnoj raspodjeli. Ako se podaci ponašaju po normalnoj raspodjeli, tada podaci (točke na grafu) čine ravnu liniju.
- Polje Statistics – prikazuje veličinu uzorka, aritmetičku sredinu te standardnu devijaciju.
- Polje Normal – prikazuje gornju i/ili donju granicu intervala tolerancije računate prema normalnoj metodi.
- Polje Nonparametric – prikazuje gornju i/ili donju granicu intervala tolerancije neparametarskog postupka i postignutu razinu pouzdanosti.

- Polje Normality Test – prikazuje *P*-vrijednost i vrijednost Anderson-Darlingovog testa normalnosti. Kako bismo utvrdili ponašaju li se podaci prema normalnoj raspodjeli, potrebno je usporediti *P*-vrijednost (engl. *P*-value) i usporediti je odabranom (ili zadanom) razinom značajnosti α. Ako je *P*-vrijednost ≤ α, možemo zaključiti da se podaci ne ponašaju po normalnoj raspodjeli. U tom slučaju, koristiti neparametarski interval tolerancije.

Iz prikazanih rezultata možemo zaključiti sljedeće:

- Podaci se ponašaju po normalnoj raspodjeli jer je:
  - ✓ P-vrijednost testa normalnosti $= 0{,}349 > \alpha = 0{,}05$
  - ✓ Podaci u histogramu podsjećaju na izgled funkcije normalne raspodjele
  - ✓ Normal Probability Plot – podaci privrženo prate ravnu liniju
- Donja granica intervala tolerancije iznosi $1022{,}235$ h.

Znači, s pouzdanošću od 95% možemo zaključiti da 99% žarulja će imati životni vijek duži od 1022,235 sata.

6.2. Ankete

U jednoj osnovnoj školi provedene su ankete među učenicima i njihovim roditeljima u cilju uspoređivanja navika roditelja i njihove djece.

Roditelji su odgovarali na sljedeća anketna pitanja (Tablica 7.):

Tablica 7. Anketna pitanja i odgovori roditelja

Pitanje	RODITELJI	DA	NE
1.	Smatrate li da imate dovoljno vremena za baviti se sportskim aktivnostima?	11	33
2.	Bavite li se nekom sportskom aktivnošću?	25	19
3.	Pomažete li svom djetetu pri odabiru sporta?	41	3
4.	Jeste li se u prošlosti bavili sportom?	27	17

Učenici su odgovarali na pitanja kako prikazuje Tablica 8.

Tablica 8. Anketna pitanja i odgovori učenika

Pitanje	UČENICI	DA	NE
5.	Smatrate li da imate dovoljno vremena za baviti se sportskim aktivnostima?	38	1
6.	Bavite li se nekom sportskom aktivnošću?	36	3

Ispitivače ankete zanimala se u sljedeće informacije:

a) Utječe li slobodno vrijeme roditelja (1) i na njihovu sportsku aktivnost (2)?

Postavljamo nultu i alternativnu hipotezu.

H_0: Ne postoji statistička značajna razlika između slobodnog vremena roditelja i njihova bavljenja sportskom aktivnošću.

H_1: Postoji statistička značajna razlika između slobodnog vremena roditelja i njihova bavljenja sportskom aktivnošću.

Računom kroz Minitab, dobivamo izlist kako prikazuje Slika 35.

Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: C2

Using category names in C1

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Sq
bavi se sportom	25	11	0,25	11	17,8182
ne bavi se sportom	19	33	0,75	33	5,9394

N	DF	Chi-Sq	P-Value
44	1	23,7576	0,000

Slika 35. Anketa a)

Vidimo da P-vrijednost $= 0{,}000$, stoga imamo dovoljno dokaza da odbacimo nultu hipotezu. Dakle, postoji statistička značajna razlika između slobodnog vremena roditelja i njihova bavljenja sportskom aktivnošću. Odnosno, slobodno vrijeme roditelja ne utječe na njihovu sportsku aktivnost.

b) Utječe li sportska aktivnost roditelja (2) na njihovu potporu djeci za bavljenje sportskim aktivnostima (3)?

H_0: Ne postoji statistička značajna razlika između sportske aktivnosti roditelja i njihove potpore djeci za bavljenje sportskim aktivnostima.

H_1: Postoji statistička značajna razlika između sportske aktivnosti roditelja i njihove potpore djeci za bavljenje sportskim aktivnostima.

Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: C10

Using category names in C9

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Sq
pomažu djetetu	41	25	0,568182	25	10,2400
ne pomažu djetetu	3	19	0,431818	19	13,4737

N	DF	Chi-Sq	P-Value
44	1	23,7137	0,000

Slika 36. Anketa b)

Ponovno vidimo da je P-vrijednost = 0,000 (Slika 36.). Odbacujemo nultu hipotezu i zaključujemo da postoji statistička značajna razlika između sportske aktivnosti roditelja i njihove potpore djeci za bavljenje sportskim aktivnostima. Da odgovorimo na postavljeno pitanje, sportska aktivnost roditelja ne utječe na njihovu potporu djeci za bavljenje sportom.

c) Postoji li povezanost između bavljenja sportom roditelja u mladim danima (4) i njihove potpore djeci za bavljenje sportskim aktivnostima danas (3)?

H_0: <u>Ne postoji</u> statistička značajna razlika između sportske aktivnosti roditelja u mladim danima i njihove potpore djeci za bavljenje sportskim aktivnostima danas.

H_1: <u>Postoji</u> statistička značajna razlika između sportske aktivnosti roditelja u mladim danima i njihove potpore djeci za bavljenje sportskim aktivnostima danas.

Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: C14

Using category names in C9

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Sq
pomažu djetetu	41	27	0,613636	27	7,2593
ne pomažu djetetu	3	17	0,386364	17	11,5294

N	DF	Chi-Sq	P-Value
44	1	18,7887	0,000

Slika 37. Anketa c)

Slika 37. prikazuje izlist iz Minitaba nakon provođenja koraka za provedbu ovog χ^2-testa. Zanima nas P-vrijednost koja je i u ovom primjeru jednaka nuli. Stoga odbacujemo nultu hipotezu i zaključujemo da ne postoji povezanost između bavljenja sportom roditelja u mladim danima i njihove potpore djeci za bavljenje sportom danas.

d) Utječe li sportska aktivnost roditelja (2) na sportsku aktivnost djece (6)?

H_0: Ne postoji statistička značajna razlika između sportske aktivnosti roditelja i sportske aktivnosti učenika.

H_1: Postoji statistička značajna razlika između sportske aktivnosti roditelja i sportske aktivnosti učenika.

Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: C4

Using category names in C1

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Sq
bavi se sportom	36	25	0,568182	22,1591	8,6452
ne bavi se sportom	3	19	0,431818	16,8409	11,3753

N	DF	Chi-Sq	P-Value
39	1	20,0206	0,000

Slika 38. Anketa d)

Uvidom u izračunate rezultate (Slika 38.) vidimo da *P*-vrijednost iznosi nula, stoga možemo zaključiti da ima statistički značajne razlike između sportske aktivnosti roditelja i sportske aktivnosti učenika. Drugim riječima, sportska aktivnost roditelja ne utječe na sportsku aktivnost učenika.

6.3. Procjena sposobnosti mjernog sustava

Kvalitetan mjerni sustav samo je jedan u nizu čimbenika proizvodnog procesa o kojem ovisi konačan izgleda, ali i funkcionalnost proizvedenog proizvoda. Za svaki mjerni sustav potrebno je poznavati sve njegove mogućnosti s naglaskom na pouzdanost. Svi elementi mjernog sustava (mjerni instrument, mjeritelji, računalni programi, mjerne strategije) mogu uzrokovati rasipanja rezultata mjerenja i mjernu nesigurnost. Analiza mjernog sustava razlikuje se od slučaja do slučaja, odnosno ovisi o broju mjeritelja, broju predmeta mjerenja i broju ponavljanja mjerenja.

Za procjenu mjernog sustava s više predmeta mjerenja i više mjeritelja Minitab sadrži Gage R&R studije, a to su: Crossed, Nested i Expanded. Načelno Gage R&R govori nam možemo li i koliko vjerovati rezultatima mjerenja mjernog sustava.

Izračun pojedinih komponenti može se računati na dva načina:

- prema metodi aritmetičkih sredina i raspona te
- prema ANOVA metodi.

Razlika između metode aritmetičkih sredina i raspona i ANOVA metode je ta što ANOVA razlaže obnovljivost na dva dijela: na mjeritelja (Operator) te na interakciju dijela i mjeritelja (Operator by Part). Zbog toga je metoda ANOVA preciznija u odnosu na metodu aritmetičkih sredina i raspona, ali u proračunu mnogo kompliciranija.

Primjer 6.3.

Na konkretnom primjeru iz privrede potrebno je provesti analizu mjernog sustava, komentirati dobivene rezultate te napisati način za poboljšanje mjernog sustava.

ALSTOM Croatia ltd Karlovac TMQ- 101215	Analiza mjernog sustava / MSA/ Procjena sposobnosti mjernog sustava: Ponovljivost i obnovljivost / **R & R** / Repeatabilyty /EV/ & Reproducibility /AV/	R&R ≤ 30 P= 99.73% S= ± 3

Predmet mjerenja	Naziv: Adapter za gas	Nacrt br.: HTCT 321192P0001	Objekt: GT 26- EGH Pembroke 21	IMR: 042222 RN: 40381476

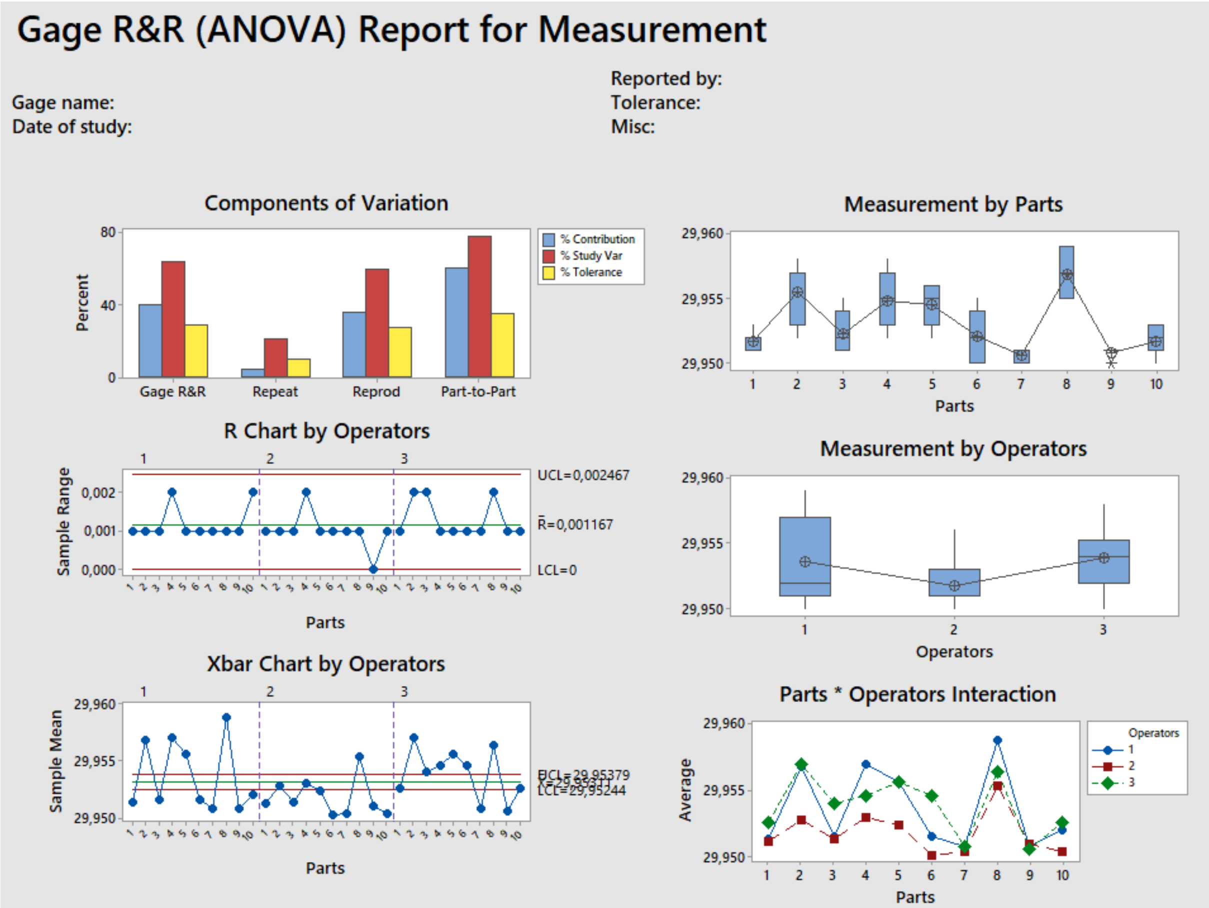
Broj operacije: 0050	Code: 131601	Stroj: Brusilica za okruglo vanjsko i unutarnje brušenje	Alat: mikrometar 25-50mm, br. 108686	Radnik: Kresojević M.

Mjerna značajka: Ø30 -0.02/ -0.05	Mjerno sredstvo: mikrometar 25-50 mm	Mjerna rezolucija: 0.01	Tolerancijsko polje T: -0.03
Datum mjerenja: 2014-10-21.	Vrijeme: 11-12 sati	Temperatura: 23°C	Vlažnost:

Mjeritelj /Ime i prezime/	Broj mjerenja	Uzorak broj									
		1 (P1-1)	2 (P1-3)	3 (P1-7)	4 (P1-8)	5 (P1-10)	6 (P1-11)	7 (P1-13)	8 (P1-14)	9 (P1-17)	10 (P1-18)
A: Alen Gajski	1	29,951	29,957	29,952	29,957	29,955	29,952	29,951	29,959	29,951	29,952
	2	29,951	29,956	29,951	29,956	29,956	29,952	29,951	29,959	29,951	29,952
	3	29,952	29,957	29,952	29,956	29,956	29,951	29,950	29,959	29,951	29,953
	4	29,951	29,957	29,952	29,957	29,956	29,952	29,951	29,958	29,951	29,951
	5	29,952	29,957	29,951	29,958	29,955	29,951	29,951	29,959	29,950	29,952
	$\overline{X}_A$=29,9536	**29,9514**	**29,9568**	**29,9516**	**29,9568**	**29,9556**	**29,9516**	**29,9508**	**29,9588**	**29,9508**	**29,9520**
	R	0,001	0,001	0,001	0,002	0,001	0,001	0,001	0,001	0,001	0,002

B: Draženko Tomac	1	29,951	29,952	29,951	29,952	29,953	29,951	29,950	29,956	29,951	29,950
	2	29,951	29,953	29,952	29,953	29,953	29,950	29,950	29,955	29,951	29,950
	3	29,951	29,953	29,951	29,953	29,952	29,950	29,951	29,956	29,951	29,951
	4	29,952	29,953	29,951	29,954	29,952	29,950	29,951	29,955	29,951	29,950
	5	29,951	29,953	29,952	29,953	29,952	29,950	29,950	29,955	29,951	29,951
	$\overline{X}_{B}$=29,9518	**29,9513**	**29,9530**	**29,9515**	**29,9533**	**29,9523**	**29,9500**	**29,9505**	**29,9553**	**29,9510**	**29,9505**
	R	0,001	0,001	0,001	0,002	0,001	0,001	0,001	0,001	0,000	0,001
C: Momir Kresojević	1	29,953	29,957	29,954	29,955	29,955	29,955	29,951	29,955	29,950	29,953
	2	29,952	29,958	29,953	29,954	29,955	29,954	29,951	29,957	29,951	29,952
	3	29,953	29,957	29,954	29,955	29,956	29,954	29,950	29,956	29,951	29,952
	4	29,953	29,957	29,955	29,955	29,956	29,955	29,951	29,957	29,951	29,953
	5	29,952	29,956	29,954	29,954	29,956	29,955	29,951	29,957	29,950	29,953
	$\overline{X}_{C}$=29,9542	**29,9526**	**29,9570**	**29,9540**	**29,9546**	**29,9556**	**29,9546**	**29,9508**	**29,9564**	**29,9506**	**29,9526**
	R	0,001	0,001	0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001
Za jedinični proizvod :	$\overline{X}_{p}$	**29,9518**	**29,9556**	**29,9524**	**29,9549**	**29,9545**	**29,9521**	**29,9507**	**29,9568**	**29,9508**	**29,9517**

Nakon provedbe koraka za analizu mjernog sustava u Minitabu, dobivamo procjene kako prikazuju Slika 39. i Slika 40.



Slika 39. Grafički prikaz procjene sposobnosti mjernog sustava

Gage R&R Study - ANOVA Method

Two-Way ANOVA Table With Interaction

Source	DF	SS	MS	F	P
Parts	9	0,0006291	0,0000699	10,0245	0,000
Operators	2	0,0001269	0,0000634	9,0994	0,002
Parts * Operators	18	0,0001255	0,0000070	22,2530	0,000
Repeatability	120	0,0000376	0,0000003		
Total	149	0,0009191			

α to remove interaction term = 0,05

Gage R&R

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0,0000028	39,81
Repeatability	0,0000003	4,50
Reproducibility	0,0000025	35,32
Operators	0,0000011	16,21
Operators*Parts	0,0000013	19,11
Part-To-Part	0,0000042	60,19
Total Variation	0,0000070	100,00

Process tolerance = 0,03

Source	StdDev (SD)	Study Var (5,15 * SD)	%Study Var (%SV)	%Tolerance (SV/Toler)
Total Gage R&R	0,0016657	0,0085785	63,10	28,60
Repeatability	0,0005598	0,0028828	21,20	9,61
Reproducibility	0,0015689	0,0080796	59,43	26,93
Operators	0,0010628	0,0054733	40,26	18,24
Operators*Parts	0,0011541	0,0059434	43,71	19,81
Part-To-Part	0,0020482	0,0105480	77,58	35,16
Total Variation	0,0026400	0,0135960	100,00	45,32

Number of Distinct Categories = 1

Slika 40. Procjene rezultata

Prikazom dobivenih procjena rezultata vidimo da mjerni sustav nije zadovoljavajući, i to iz više razloga (Tablica 9.):

Tablica 9. Usporedba zadovoljavajućeg i korištenog mjernog sustava

Kategorija	Zadovoljavajuć mjerni sustav	Korišten mjerni sustav
Broj različitih kategorija *Number of Distinct Categories*	≥ 5	1
Udio R&R *%StudyVar* za *Total Gage R&R*	< 10%, prihvatljivo i do 30%	63,10%
Doprinos *Total Gage R&R* *%Contribution (of VarComp)*	< 1%, prihvatljivo i do 9%	39,81%

Načelno, ako je mjerni sustav nezadovoljavajuć, izvore nesigurnosti i rasipanja treba potražiti u svim elementima mjernog sustava, od dijelova i mjeritelja do mjernih instrumenata i računalnih programa.

Budući da iste vrijednosti dobijemo računanjem „pješice“, računalni program kao razlog nezadovoljavajućeg mjernog sustava možemo odbaciti.

Prikazom procjena rezultata vidimo da broj različitih kategorija (*Numer of Distinct Categories*) nije dovoljno visok. Trebao bi iznositi barem 5, a procijenjen je na 1. Broj različitih kategorija računa se kao $\frac{Standardna\ devijacija\ dijelova}{Standardna\ devijacija\ mjernog\ sustava} \cdot 1{,}41$, stoga trebamo povećati standardnu devijaciju dijelova i/ili smanjiti iznos standardne devijacije mjernog sustava (ponovljivost i obnovljivost). Također, da bismo smanjili udio R&R standardne devijacije (*%StudyVar* za *Total Gage R&R*) u ukupnoj te smanjili doprinos R&R (*%Contribution*), potrebno je učiniti isto.

Ideja je prvo provjeriti utjecaj mjerila (mjernog instrumenta) u mjernom sustavu (Minitab: Type 1 Gage Study). Međutim, potreban broj mjerenja jednog predmeta od strane jednog mjeritelja je 10 (preporučeno barem 25), a u ovom je slučaj jedan mjeritelj mjerio jedan dio samo pet puta.

U prvom dijelu procjene rezultata korišten je i F-test analize varijanci. Da je *P*-vrijednost utjecaja Operatosr*Part bila veća od zadane α = 0,05, zaključili bismo da statistički značajno ne utječe na rezultat te bismo tu interakciju mogli izbaciti iz daljnjeg proračuna i pojednostavniti stvari.

7. ZAKLJUČAK

Primjena statističkog zaključivanja neupitna je u gotovo svim znanstvenim i stručnim djelatnostima. Ovim diplomskim radom razrađena su načela deskriptivne i inferencijalne statistike s posebnim naglaskom na testiranju hipoteza.

U području deskriptivne statistike zaključili smo da koristeći osnovni statističke alate, veliki skup podataka možemo jednostavnije i jezgrovitije prikazati. Opisane su i dani su primjeri najznačajnijih teorijskih raspodjela kontinuirane i diskretne varijable.

U poglavlju inferencijalne statistike razrađen je postupak zaključivanja pomoću statističkih testova. Prikazanim primjerima prikazan je značaj i mogućnost zaključivanja za cijelu populaciju na osnovnu uzorka.

Kod odabranih primjera primjene statističkih alata objašnjeni su rezultati statističke analize te je procijenjen njihov praktični značaj. Statistički alati i metode primijenjeni su na primjeru intervala tolerancije, anketa i procjene sposobnosti mjernog sustava na primjeru iz privrede. Prikazanim primjerima jasno se vidi da se poznavanje statistike i njenih alata koristi u mnogim područjima.

Na kraju vrijedi primijetiti kako izvrsno poznavanje metoda statistike također jedan od elemenata neophodnih za funkcioniranje kvalitete i produktivnosti u mnogim aktivnostima i postupcima u industriji, mnogim znanstvenim granama, ali i u svakodnevnom životu.

8. LITERATURA

[1] Richard L. Scheaffer, Madhuri S. Mulekar, James T. McClave: Probability and Statistics for Engineers, Fifth Edition, 2011.

[2] Sheldon M. Ross: Introduction to Probability and Statistics for Engineers and Scientists, Fourth Edition, 2009.

[3] Douglas C. Montgomery, George C. Runger: Applied Statistics and Probability for Engineers, Sixth Edition, 2014.

[4] Biserka Runje: Predavanja iz kolegija Statistika u mjeriteljstvu, 2017.

[5] Minitab Support: All statistics and graphs for Tolerance Intervals (Normal Distribution), dana 9.1.2018. dostupno na https://support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/quality-tools/how-to/tolerance-intervals-normal-distribution/interpret-the-results/all-statistics-and-graphs/

[6] Minitab Support: Tolerance interval basics, dana 9.1.2018. dostupno na https://support.minitab.com/en-us/minitab/18/help-and-how-to/quality-and-process-improvement/quality-tools/supporting-topics/tolerance-interval-basics/

[7] Biserka Runje, Amalija Horvatić Novak, Andrej Razumić: Measurement system analysis in production process, 2017.

[8] Zdenka Gogala: Osnove statistike, 2001.

[9] Boris Petz: Osnovne statističke metode za nematematičare, Peto izdanje, 2004.

9. PRILOG

9.1. Statističke tablice standardne normalne raspodjele

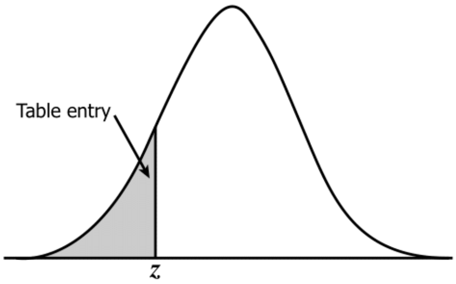


Table entry for z is the area under the standard normal curve to the left of z.

z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

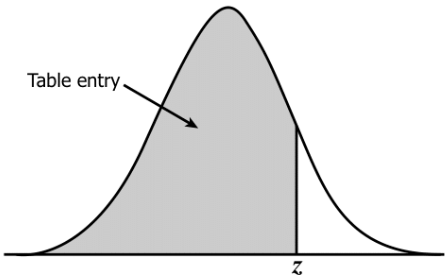


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

9.2. Statistička tablica *t*-raspodjele

_t_ Distribution: Critical Values of _t_

Degrees of freedom	Significance level — Two-tailed test: / One-tailed test:	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323
400		1.649	1.966	2.336	2.588	3.111	3.315
500		1.648	1.965	2.334	2.586	3.107	3.310
600		1.647	1.964	2.333	2.584	3.104	3.307
∞		1.645	1.960	2.326	2.576	3.090	3.291