Deskriptivna statistička analiza u programskom paketu Matlab

Rubčić, Ivan

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Preddiplomski sveučilišni studij elektrotehnike

Završni rad

DESKRIPTIVNA STATISTIČKA ANALIZA U PROGRAMSKOM PAKETU MATLAB

Rijeka, srpanj 2022.

Ivan Rubčić
0069083619

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Preddiplomski sveučilišni studij elektrotehnike

Završni rad

DESKRIPTIVNA STATISTIČKA ANALIZA U PROGRAMSKOM PAKETU MATLAB

Mentor: doc. dr. sc. Ivan Dražić

Rijeka, srpanj 2022.

Ivan Rubčić
0069083619

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
POVJERENSTVO ZA ZAVRŠNE ISPITE

Rijeka, 18. ožujka 2022.

Zavod: **Zavod za matematiku, fiziku i strane jezike**
Predmet: **Inženjerska matematika ET**
Grana: **1.01.07 primijenjena matematika i matematičko modeliranje**

ZADATAK ZA ZAVRŠNI RAD

Pristupnik: **Ivan Rubčić (0069083619)**
Studij: Preddiplomski sveučilišni studij elektrotehnike

Zadatak: **Deskriptivna statistička analiza u programskom sustavu MATLAB // Descriptive statistical analysis in the MATLAB software system**

Opis zadatka:

U radu je potrebno dati kratki uvod u programski sustav MATLAB te detaljno opisati biblioteke naredbi koje se koriste kod deskriptivne statističke analize.
Potrebno je opisati temeljne tehnike deskriptivne statističke analize u vidu formiranja tablice frekvencija, izračuna statističkih pokazatelja te izrade različitih grafičkih prikaza.
Svaku od spomenutih tehnika treba potkrijepiti primjerom iz inženjerske struke, s naglaskom na primjenu u elektrotehnici. Sve navedene primjere potrebno je riješiti unutar programskog sustava MATLAB.

Rad mora biti napisan prema Uputama za pisanje diplomskih / završnih radova koje su objavljene na mrežnim stranicama studija.

Ivan Rubčić

Zadatak uručen pristupniku: 21. ožujka 2022.

Mentor:

Doc. dr. sc. Ivan Dražić

Predsjednik povjerenstva za završni ispit:

Prof. dr. sc. Viktor Sučić

IZJAVA

Sukladno članku 8. Pravilnika o završnom radu, završnom ispitu i završetku preddiplomskih sveučilišnih studija/stručnih studija Tehničkog fakulteta Sveučilišta u Rijeci od 1. veljače 2020., izjavljujem da sam samostalno izradio završni rad prema zadatku preuzetom dana 21.03.2022.

Rijeka, 06.07.2022.

Ivan Rubčić

Ivan Rubčić

Ovim putem želim se zahvaliti svom mentoru, doc. dr. sc. Ivanu Dražiću koji mi je svojim savjetima i stalnom prisutnošću olakšao pisanje završnog rada te mi bio velika podrška i motivacija. Također, zahvaljujem se svojoj obitelji, kolegama i prijateljima koji su mi bili podrška tokom studiranja.

Sadržaj

1. Uvod

Statistika je znanstvena disciplina koja se bavi tehnikama sakupljanja, uređivanja, analize i prezentacije podataka te donošenja zaključaka o procesima i pojavama koje pokazuju ti podaci. Statistika se dijeli na deskriptivnu i inferencijalnu statistiku. Temelji se na podacima koji prikazuju oblike statističkih varijabli, tj. oznake svojstava procesa i mjera. Navedeni podaci dobivaju se promataranjem ili statističkim postupkom. Nacionalni statistički uredi, državne institucije, međunarodna politička i poslovna udruženja prikupljaju i objavljuju statističke podatke, a korisnici ih mogu vidjeti i preuzeti preko interneta i intraneta, često bez naknade ako su to javni podaci ili uz naknadu ako se radi o komercijalnim podacima.

Gledajući povijesni aspekt, statistika se pojavila još prije Krista u kineskoj, egipatskoj i babilonskoj civilizaciji te Rimskom carstvu, a glavna zadaća bila joj je popisivanje stanovništva, bogatstva te poljoprivrednih prinosa. U 16. i 17. stoljeću razvoj statistike bio je vezan za potrebe velike količine podataka koji su bili prikupljeni anketama stanovništva u rastućim europskim državama. Kasnije u 18. stoljeću statistika napreduje povezujući se s teorijom vjerojatnosti koja je inspirirana igrama na sreću, tj. kockanjem. U 19. stoljeću se razvija za potrebe proučavanja stanovništva i gospodarstva, a kasnije i kao matematički alat za analizu pripadnih numeričkih podataka. Razvojem računala i dostupnošću sve veće količine podataka, statistika se u 20. stoljeću sve brže razvija. [10]

U ovom radu obradit će se deskriptivna statistička analiza u programskom paketu Matlab. Opisat će se teorijski elementi deskriptivne statistike i navesti pripadne naredbe iz programskog paketa Matlab.

Ukratko će se opisati povijest programskog paketa Matlab, njegovo sučelje i osnovna sintaksa, usporediti Matlab s nekim njemu sličnim programskim paketima i dati njegova primjena u elektrotehnici.

Na kraju rada kroz praktične primjere će se prikazati kako se računaju svi statistički pokazatelji deskriptivne statistike te kako se prikazuju pojedini grafički prikazi pomoću programskog paketa Matlab, s naglaskom na primjere iz elektrotehnike.

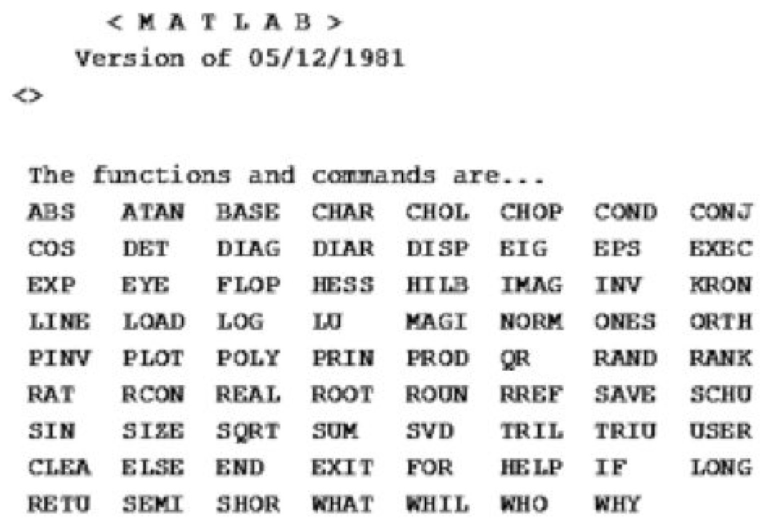
2. MATLAB

Matlab je interaktivni programski jezik visoke razine čije ime potječe od engleskog naziva MATrix LABoratory. Koristi se za rješavanje različitih matematičkih problema, vizualizaciju, proračune i simulacije koji su vezani uz obradu podataka, upravljanje i regulaciju te za programiranje. Pomoću njega se mogu analizirati i obrađivati podaci, izraditi algoritmi, modeli i aplikacije. Omogućava brzo rješavanje zadataka koji zahtijevaju veliku količinu računanja.

2.1. Povjesni razvoj MATLABA

Ovo poglavlje je obrađeno prema službenim stranicama tvrtke MathWorks [1]. Prva verzija Matlaba nastala je krajem 1970. godine na sveučilištima University of New Mexico i Stanford University kao pomoć u nastavi matrične teorije, numeričke analize te linearne algebre, a temelji se na radu matematičara i računalnog programera Clevea Molera.

Moler, tadašnji profesor matematike na sveučilištu University of Mexico, počeo je 1970.godine razvijati Matlab za svoje studente. U početku Matlab nije bio programski jezik, već jednostavni interaktivni matrični kalkulator napisan u FORTRANU pomoću potprograma LINPACK (paket za rješavanje linearnih sustava) i EISPACK (paket za rješavanje matričnih problema). U njemu je bila ugrađena samo 71 funkcija. Da bi se dodala neka nova funkcija, u izvornom kodu je bilo potrebno napisati FORTRAN podprogram, dodati njegov naziv nove riječi te ponovno kompajlirati Matlab.



Slika 2.1. Prikaz rezerviranih riječi i funkcija na prvoj verziji Matlaba. Izvor: [1]

Matlab je javnosti prvi puta predstavljen 1979. godine na Sveučilištu u Stanfordu, kao pomoć studentima u kolegiju u čijoj su osnovi bile matrice pa im je Matlab bio od velike koristi. Program je bio besplatano distribuiran na sva sveučilišta koja bi Moler posjetio te je tako sakupio velik broj korisnika. Jedan od studenata koji je kod Molera pohađao tečaj za Matlab pokazao ga je Jacku Littleu. Tako je 1983.godine Little predoložio stvaranje komercijalnog proizvoda koji bi se temeljio na Matlabu. Za pokretanje Matlaba je koristio IBM PC koji je jedva bio dovoljan za izvođenje te radnje, a izašao je svega dvije godine prije. Predvidjevši uspjeh programa, Little je napustio svoj posao, kupio Compaq PC u Searsu i preselio se u planine u zaleđu Stanforda te uz Molerovu podršku napisao novu i proširenu verziju Matlaba u programskom jeziku C. Steve Bangert, njihov prijatelj, također je u svoje slobodno vrijeme radio na razvitku Matlaba.

Matlab je 1984. godine postao dostupan kao komercijalni proizvod tvrtke MathWorks, a predstavljen je na IEEE konferenciji o odlučivanju i kontroli u Las Vegasu. Little i Bangert su napravili mnoge izmjene i razna poboljšanja te stvorili novu i proširenu verziju Matlaba sa novim funkcijama, paketima i grafikom. Tadašnji Matlab je uz matrične funkcije uključivao i brze Fourierove transformacije (FFT), da bi se narednih godina pojavili još alat za upravljački sustav, alat za obradu signala i podrška za numeričko rješavanje običnih diferencijalnih jednadžbi. Softver se sve više popularizirao razvojem različitih paketa koje su razvili stručnjaci za rješavanje različitih matematičkih problema. Do danas se program nastavio razvijati kako bi zadovoljio promjenjive potrebe inžinjera i znanstvenika te postoji više od 60 dodatnih alatna baziranih na Matlabu. [14]

2.2. Usporedba MATLABA i sličnih alata

U ovom poglavlju ukratko opisujemo programske alternative Matlabu, kao što su programski paketi GNU Octave i Python.

2.2.1. Matlab i GNU Octave

GNU Octave jezik je visoke razine namjenjen za numeričke izračune. Koristi se za rješavanje linearnih i nelinearnih problema te je kompatibilan sa Matlabom, što znači da se svi Matlab programi mogu pokrenuti u Octaveu, a isto tako i većina programa iz Octavea u Matlabu. Njegov razvoj započeo je 1988. godine kada je zamišljen kao prateći softver za udžbenik o dizajnu kemijskih reaktora. Konačan razvoj dogodio se 1992. godine, a razvio ga je John W. Eaton, da bi prva verzija izašla 1994. godine. [2]

Matlab i Octave su jedni od najtraženijih alata za matematičko izračunavanje te o njima ovisi većina znanstvenika kako bi s lakoćom izvršili neki zadatak. Matlab je kao program puno moćniji i brži od Octavea te zauzima puno više memorije na računalu. Također licenca za Matlab se plaća, što nije slučaj za Octave koji je besplatan i za koji kažemo da je besplatna aternativa za Matlab. Možemo reći da je Matlab kao program popularniji i ima širu primjenu od Octavea te će zbog svoje brzine i većeg broja alata biti bolji odabir u ovoj usporedbi, pogotovo za ozbiljnije projekte i veća poduzeća. [15]

2.2.2. Matlab i Python

Python je programski jezik visoke razine te je također interpretativni, interaktivni i objektno orijentiran progremski jezik. Jezik je otvorenog tipa, što znači da je besplatan u odnosu na Matlab. Python se sve više koristi u web i programskom dizajnu, dok se Matlab koristi za upravljanje matricama, crtanje funkcija i podataka te kreiranje korisničkih sučelja. Python je sličan Matlabu po interaktivnom sučelju koje omogućuje dinamičko unošenje i osigurava automatsko upravljanje memorijom. U Matlabu se naredbe brže izvršavaju te također ima bolje uređeniji sustav za javljanje i otklanjanje pogrešaka od Pythona. Ova dva jezika nisu kompatibilna, to jest programi iz pojedinog jezika se ne mogu pokrenuti u drugom jeziku i obrnuto. Pythonu nedostaje grafičko sučelje koje omogućuje obradu signala i modeliranje, dok Matlab ima dodatke koji to omogućavaju. [16, 17]

2.3. MATLAB u elektrotehnici

U današnje vrijeme velika većina elektrotehničkih kurikuluma koristi Matlab. Korištenje Matlaba tijekom obrazovanja ne samo da ubrzava i poboljšava učenje, već pomaže pri razvitku produktivnosti u daljnoj karijeri. Matlab ima jako široku primjenu, pa se tako primjerice koristi za projektiranje te analizu analognih i digitalnih filtara, za analizu analognih i digitalnih sustava za upravljanje, kao i za mnoge druge primjene. Pored samog uređivača koda, u svrhe projektiranja razvijene su i dodatne aplikacije za npr. automatizirano parametriranje PID regulatora, vizualizaciju sustava, itd.

Dodatni paket koji se nalazi unutar Matlaba naziva se Simulink. Simulink je grafičko programsko okruženje koje omogućuje simuliranje kontinuiranih i diskretnih sustava te se temelji na blok dijagramima, čijim povezivanjem kreiramo sustave koje simuliramo. Pokreće se unutar Matlaba naredbom "simulink" u command windowu ili pritiskom ikone na alatnoj traci. Podržava modeliranje, simulaciju, automatsko generiranje koda i analizu dinamičkih sustava s više domena koji se opisuju pomoću diferencijalnih jednadžbi. Koristi se za modeliranje sustava za upravljanje i njihovu analizu te obradu signala. Također postoji mogućnost modeliranja sustava¸za upravljanje, koji se zatim prevode u C kod za razne mikrokontrolere na kojima se regulatori izvršavaju. [9, 18]

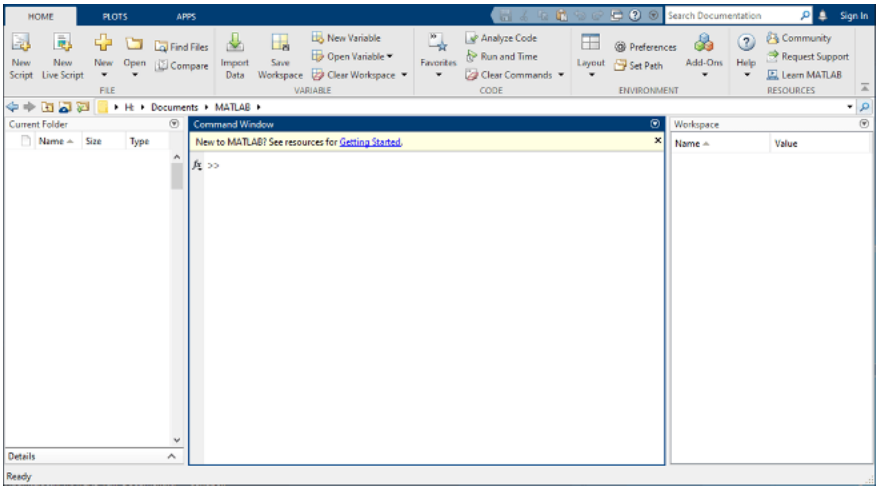
Dolazi se do zaključka da elektrotehnika kao struka voli Matlab, jer nijedan drugi program nema alternativu koja je toliko razvijena kao Matlab i njegov dodatni paket Simulink.

2.4. Sučelje MATLABA

Radni prostor je najvažniji dio korisničkog sučelja i služi za unos naredbi te ispis rezultata. Nakon pokretanja Matlaba otvara se sučelje koje se sastoji od više dijelova [19]:

- Redak s alatima (toolbar) - omogućava pristup dostupnim alatima, demo primjerima i dokumentaciji

- Command window - osnovni prozor Matlaba u koji se unose sve naredbe. Neka je vrsta terminala operacijskog sustava te u njemu vrijede i osnovne terminalske operacijske komande za manipulaciju datotekama.
- Command history - u njemu su pohranjene sve upotrebljene naredbe
- Current directory - sve datoteke koje se pokreću, učitavaju ili spremaju, nalaze se u mapi izabranoj u tom području
- Workspace – je radni prostor koji pokazuje skup varijabli koje su trenutačno dostupne



Slika 2.2. Sučelje programskog paketa Matlab. Izvor: Izrada autora

U Matlabu je moguće raditi na dva načina. Direktan način, kada se u glavnom prozoru programa (Command window) upisuje naredba i program odmah vraća rezultat. Takav način pogodan je kada radimo jednostavne operacije bez ponavljanja. Kada radimo kompliciranije naredbe koje sadrže više slijedova i ponavljanja tada korisnik koristi tekstualni editor (Matlab editor) te ga pohranjuje kao m-datoteku (.m ekstenzija). Kad u komandnom prozoru Matlaba pozovemo ime neke datoteke, pokrenut će se izvršavanje programa, tj. izvršavanje slijeda naredbi pohranjenih u pozvanoj datoteci. [9]

2.5. Osnovna sintaksa MATLABA

Ovo poglavlje izrađeno na temelju izvora [9]. Nakon pokretanja Matlaba otvara se radni prostor (Command window) u kojem se pojavljuje znak '»' koji kazuje da je program spreman za rad i pisanje naredbi, a da bi se dobio rezultat mora se pritisnuti tipka Enter. Program razlikuje velika i mala slova, pa tako ako je neka varijabla definirana malim slovom, a u određenoj funkciji se poziva sa velikim slovom, Matlab će javiti grešku. Na definirane ulazna varijable primjenjuju se matematičke operacije te se kao rezultat dobiju izlazne varijable. Ulazne varijable definiraju se na sljedeći način:

```
>> x=9.5

x =

    9.5
```

Vektor se u Matlabu prikazuje kao:

```
>> vektor=[ 1 2 5 10 ]

vektor =

    1 2 5 10
```

Niz se prikazuje na isti način kao i gore prikazani način prikaza vektora.

Korištenjem ";" razdvajamo više naredbi u jednom redu i Matlab ne ispisuje rezultat te naredbe na ekran. To je prikazano sljedećim primjerom:

```
>> x=5; y=x+3

y =

    8
```

Kada želimo opisati program ili ostaviti neki komentar koristimo znak postotka (%), što je prikazano ispod:

```
>> x=5    \% definiramo vrijednost varijable x brojem 5

x =

    5
```

Definiranje matrica vrši se na sljedeći način:

```
>> Matrica=[2 5 8; 3 5 7]   \% matrica od 2 retka i 3 stupca

Matrica =
```

```
    2    5    8
    3    5    7
```

Ako uz naredbu help upišemo i ime neke funkcije, Matlab će ispisati sve funkcije koje nam stoje na raspolaganju u tom području. Po istom pravilu će help s upisanim imenom funkcije dati informaciju o toj funkciji, kao što je prikazano u nastavku.

```
>> help sqrt

 sqrt   Square root.
    sqrt(X) is the square root of the elements of X.
    Complex results are produced if X is not positive.
```

Osnovne matematičke operacije su prikazane u sljedećoj tablici.

Tablica 2.1. Osnovne matematičke operacije

Naredba	**Opis naredbe**
+	zbrajanje
-	oduzimanje
*	množenje
/	dijeljenje
^	potenciranje

Imaginarne jedinice zapisuju se slovima i ili j, što je prikazano ispod:

```
>> x = 5+5j

x =

    0.0000+5.0000i
```

3. Uvoz podataka iz MS Excel-a

MS Excel je program koji se koristi za izradu tablica i tablično računanje. Sastoji se od stupaca i redaka, a podatke upisujemo u ćelije. Zbog takve strukture da se podaci grupiraju u tablicama, statističke podatke koji se obrađuju u Matlabu najvećim se dijelom ubacuju iz MS Excela jer su tamo podaci najpregledniji. Podatke iz MS Excela moguće je na 2 načina uvoziti u Matlab:

- Pozivom naredbe $xlsread$ ('$naziv\ spremljene\ datoteke\ iz\ MS\ Excela$') što je prikazano na slikama ispod.

	A	B	C	D	E	F	G
1	Vrijednosti otpora						
2	17,61	15,9	14,7	16,06	14,8	15,74	15,29
3	15	15,5	14	14,9	14,63	15,3	14,6
4	14,26	15,04	15	14,6	14,11	14,58	15

Slika 3.1. Vrijednosti iz MS Excela. Izvor: Izrada autora

Editor - C:\Users\Korisnik\Desktop\Prakticni primjeri.m

Prakticni_primjeri_statistika.m | Prakticni primjeri.m

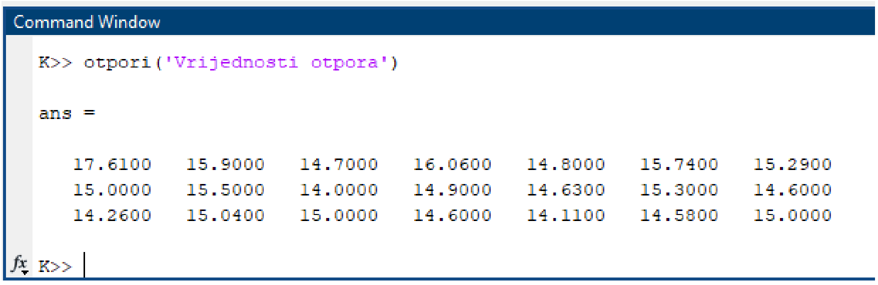
```
Vrijednosti_otpora = xlsread('Vrijednosti otpora')
```

Command Window

```
Vrijednosti_otpora =

   17.6100   15.9000   14.7000   16.0600   14.8000   15.7400   15.2900
   15.0000   15.5000   14.0000   14.9000   14.6300   15.3000   14.6000
   14.2600   15.0400   15.0000   14.6000   14.1100   14.5800   15.0000
```

Slika 3.2. Naredba iz Matlaba za uvoz iz MS Excela i rezultat. Izvor: Izrada autora

- Korištenjem ikone Import data odabiremo spremljenu datoteku iz MS Excela. Nakon toga nam se otvara Import Wizard u kojem odabiremo stupce i retke koje želimo uvesti te u padajućem izborniku Import Selection odabiremo Generate Function. Potom spremamo podatke iz editora i u command windowu pozivamo podatke iz Excela, što vidimo na sljedećoj slici. [13]



Slika 3.3. Naredba iz Matlaba za uvoz iz MS Excela i rezultat. Izvor: Izrada autora

U gore prikazanim primjerima korištene su vrijednosti otpora otpornika izmjerenih na laboratorijskim vježbama iz kolegija Mjerenja u elektrotehnici te su prikazane vrijednosti prikazane u ohmima (Ω).

4. Teorijski elementi deskriptivne statistike

Statistiku dijelimo na deskriptivnu i inferencijalnu statistiku. Deskriptivna statistika vrsta je statistike koja se bavi organizacijom sakupljenih podataka i njihovim opisom pomoću grafičkih prikaza i numeričkih pokazatelja, kao što su primjerice mjere centralne tendencije i rasapa.

Inferencijalna statistika vrsta je statistike koja se bavi metodama pomoću kojih se na osnovi prikupljenih podataka o uzorku donose zaključci o populaciji.

Statističke metode i modeli se primjenjuju u znanstvenim i stručnim djelatnostima kao što su ekonomija, demografija, sociologija, medicina, fizika, psihologija te u mnogim drugim područjima izvan znanosti. Uporabom statističkih metoda došlo je do razvoja statističke programske podrške i informacijskih sustava. [7, 11]

4.1. Mjere centralne tendencije

Mjere centralne tendencije su brojčane vrijednosti koje pokazuju oko koje vrijednosti se grupiraju rezultati. Mjere centralne tendencije obrađene su prema [7].

4.1.1. Aritmetička sredina

Aritmetička sredina jedna je od srednjih vrijednosti koje se koriste u statistici. Dobiva se zbrajanjem svih vrijednosti iz statističkog skupa i dijeljenjem s ukupnim brojem članova tog skupa, što je prikazano sljedećim izrazom:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}, \tag{4.1}$$

pri čemu $x_i, i = 0, 1, \ldots, N$ označavaju vrijednosti za koje računamo srednju vrijednost, a N označava ukupan broj tih vrijednosti u skupu. Za grupirane podatke aritmetička sredina se računa sljedećim izrazom:

$$\bar{x} = \frac{\sum\limits_{i=1}^{r} f_i \cdot x_i}{\sum\limits_{i=1}^{r} f_i}, \tag{4.2}$$

gdje je r broj razreda, f_i je frekvencija podataka, a x_i je vrijednost pojedinog podatka.

Problem koji se javlja kod računanja aritmetičke sredine u nekom skupu podataka je postojanje ekstrema koji znatko odstupa od vrijednosti ostalih podataka te nam tada aritmetička sredina nije pouzdana, što ćemo pokazati kroz sljedeća dva skupa podataka.

Tablica 4.1. Podaci bez ekstremnih odstupanja

33	32	29	35	38	27	26	34	30	29	38	40	25	31	36
26	28	33	38	37	36	34	34	29	30	40	39	31	43	45

Aritmetička sredina za podatke iz Tablice 4.1 iznosi:

$$\bar{x} = \frac{1006}{30} = 33.5333 \tag{4.3}$$

Tablica 4.2. Podaci s ekstremnim odstupanjem

33	32	29	35	38	27	**5000**	34	30	29	38	40	25	31	36
26	28	33	38	37	36	34	34	29	30	40	39	31	43	45

Aritmetička sredina za podatke iz Tablice 4.2 iznosi:

$$\bar{x} = \frac{5980}{30} = 199.3333 \tag{4.4}$$

Iz Tablice 4.1 vidimo da su sve vrijednosti skupa približno jednake i tada je rezultat aritmetičke sredine pouzdan, dok se iz Tablice 4.2 vidi da postoji jedna vrijednost koja značajno odstupa od ostalih vrijednosti skupa te tada rezultat aritmetičke sredine značajno odstupa od prijašnje vrijednosti pa se kaže da je nepouzdan, tj. aritmetička sredina tada nije reprezentativni pokazatelj centralne tendencije jer postoje ekstremna odstupanja.

4.1.2. Medijan

Medijan je srednja vrijednost po položaju koja dijeli numerički niz, u kojem su podaci poredani po veličini, na dva jednaka dijela. To znači da jedna polovina vrijednosti u poredanom nizu ima manju ili jednaku vrijednost medijanu, a druga polovina vrijednosti ima veću ili jednaku vrijednost medijanu. Ako je broj podataka u nizu neparan, tada će medijan biti središnja vrijednost tog niza.

Tablica 4.3. Medijan za neparan broj podataka u nizu

3	7	8	12	**15**	16	19	29	1500

Ako je broj podataka u nizu paran, tada medijan računamo kao aritmetičku vrijednost dva središnja podatka u nizu. Tako za podatke iz sljedeće tablice medijan računamo na sljedeći način:

Tablica 4.4. Medijan za paran broj podataka u nizu

3	7	8	12	**15**	**16**	16	19	29	1500

$$x_{med} = \frac{31}{2} = 15.5. \tag{4.5}$$

Iz navedenih primjera odmah uočavamo kako ekstremna odstupanja ne utječu na izračun medijana.

4.1.3. Mod

Mod je vrijednost statističkog skupa koja se najčešće ponavlja. Da bi odredili mod, moraju postojati barem dvije iste vrijednosti podataka u skupu. Ako nemamo nijednu istu vrijednost podataka u skupu kažemo da mod ne postoji, a ako niz ima dva moda tada taj skup nazivamo polimodalni skup. U sljedećih nekoliko tablica prikazati ćemo različite situacije koje se mogu dogoditi kod određivanja moda.

Tablica 4.5. Primjer moda

4	5	8	12	**15**	**15**	**15**	19	23	28

Tablica 4.6. Primjer polimodalnog skupa

4	**4**	8	12	**15**	**15**	19	23	25

Tablica 4.7. Primjer skupa bez moda

4	7	8	12	14	18	19	23	25

Iz Tablice 4.5 jasno se vidi da je mod 15 jer je to vrijednost koja se najviše puta ponovila. U Tablici 4.6 postoje dvije vrijednosti koje se ponavljaju jednak broj puta, broj 4 i 15 pa se za takav skup kaže da ima dva moda te se naziva polimodalnim skupom, dok se za skup u kojem se vrijednost nekog broja ne ponavlja kaže da je to skup bez moda, što je prikazano u Tablici 4.7.

Istaknimo da su podaci u gore prikazanim tablicama za prikaz pokazatelja centralne tendencije demonstracijskog karaktera, odnosno ne odgovaraju nekom stvarnom skupu podataka.

4.2. Mjere rasapa

Mjere rasapa govore koliko su podaci rasuti oko pokazatelja centralne tendencije. Tako ćemo promatrat varijancu i standardnu devijaciju kao mjeru rasapa oko aritmetičke sredine, te raspon i interkvartilni raspon kao mjeru rasapa oko medijana.

4.2.1. Varijanca i standardna devijacija

Varijancu definiramo kao srednje kvadratno odstupanje od aritmetičke sredine i opisujemo ju sljedećim izrazom: [5]

$$VAR = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2, \tag{4.6}$$

odnosno za grupirane podatke:

$$VAR = \frac{1}{N} \cdot \sum_{i=1}^{r} (f_i \cdot x_i - f_i \cdot \bar{x})^2, \tag{4.7}$$

gdje je N broj podataka, r broj razreda, f_i je frekvencija podataka, x_i je vrijednost pojedinih podataka, a $\bar{x}$ aritmetička sredina podataka.

Gore navedene formule (4.6) i (4.7) su formule koje se koriste za stvarnu varijancu skupa podataka s kojim radimo, a postoji i formula za varijancu populacije. Ona se od stvarne varijance razlikuje po nazivniku u kojem se nalazi $N-1$ umjesto N i njome procjenjujemo varijancu populacije čiji je uzorak skup podataka s kojim radimo.

Standardna devijacija pozitivna je vrijednost drugog korijena varijance te je prikazana sljedećim izrazima: [5]

$$s = \sqrt{VAR} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2}, \tag{4.8}$$

$$s = \sqrt{VAR} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (f_i \cdot x_i - f_i \cdot \bar{x})^2}, \tag{4.9}$$

uz iste oznake kao i kod izračuna varijance.

Pokažimo sada na jednom primjeru izračun aritmetičke sredine, varijance i standardne devijacije.

Primjer 4.1. *Za podatke o vrijednostima igrača preuzete s platforme Transfermarkt [20] prikazanima u sljedećoj tablici izračunajmo aritmetičku sredinu, varijancu i standardnu devijaciju.*

Primjetimo kako se radi o podacima koji nisu grupirani u razrede pa se tijekom izračuna koristimo formulama koje ne uključuju frekvenciju.

Tablica 4.8. Vrijednost igrača u milijunima eura (M€)

Igrač	**Tržišna vrijednost (M€)**
Harry Kane	100,0
Vinicius Junior	100,0
Neymar	90,0
Phil Foden	90,0
Joshua Kimmich	85,0
Raheem Sterling	85,0
Pedri	80,0

Izračunajmo najprije aritmetičku sredinu:

$$\bar{x} = \frac{100 + 100 + 90 + 90 + 85 + 85 + 80}{7} = 90. \tag{4.10}$$

Kada imamo podatak o aritmetičkoj sredini možemo izračunati i varijancu:

$$\begin{aligned} VAR = \frac{1}{7} \cdot [&(100-90)^2 + (100-90)^2 + (90-90)^2 + (90-90)^2 \\ &+(85-90)^2 + (85-90)^2 + (80-90)^2] = 50. \end{aligned} \tag{4.11}$$

Na kraju izračunavamo standardnu devijaciju.

$$s = \sqrt{VAR} = \sqrt{50} = 7.071. \tag{4.12}$$

4.2.2. Raspon i interkvartilni raspon

Raspon je gruba mjera raspršenosti podataka oko medijana, a definiran je kao razlika između maksimalne i minimalne vrijednosti skupa podataka. Izraz za raspon glasi: [5]

$$R = x_{max} - x_{min}, \tag{4.13}$$

gdje x_{max} predstavlja najveću, a x_{min} najmanju vrijednost u skupu podataka.

Primjer 4.2. *Za izmišljene podatke iz Tablice 4.9 izračunajmo raspon.*

Tablica 4.9. Izmišljeni podaci za računanje raspona

73	32	28	45	53	12	7
86	5	35	60	37	79	93

Pošto je raspon razlika maksimalne i minimalne vrijednosti, on iznosi:

$$R = x_{max} - x_{min} = 93 - 5 = 88. \tag{4.14}$$

Da bi rješili problem grube mjere raspršenosti, koristi se interkvartilni raspon koji je također mjera raspršenosti oko medijana, a definiran je kao razlika između trećeg i prvog kvartila. Bolji je pokazatelj rasapa od raspona jer ne uključuje ekstreme, dok su unutar raspona uključeni ekstremi prilikom njegovog izračuna. Prvi kvartil (Q_1) dijeli statistički niz na dva dijela u omjeru 1:3, što bi značilo da je 25% podataka statističkog skupa jednake ili manje vrijedosti od prvog kvartila, a 75% podataka je veće ili jednake vrijednosti od prvog kvartila. Treći kvartil (Q_3) također dijeli statistički niz na dva dijela, ali u omjeru 3:1, što znači da 75% podataka statističkog skupa ima jednaku ili manju vrijednost od trećeg kvartila, dok 25% podataka ima veću ili jednaku vrijednost trećem kvartilu.

Interkvartilni raspon računa se prema sljedećem izrazu: [5, 22]

$$IQR = Q_3 - Q_1, \tag{4.15}$$

pri čemu je Q_1 prvi, a Q_3 treći kvartil.

Prvi kvartil računa se na dva načina, ovisno o tome je li ukupan broj podataka (N) djeljiv s 4:

- kada N nije djeljiv sa 4

$$Q_1 = x_r, \tag{4.16}$$

gdje je

$$r = INT\left(\frac{N}{4}\right) + 1, \tag{4.17}$$

- kada je N djeljiv sa 4

$$Q_1 = \frac{x_r + x_{r+1}}{2}, \tag{4.18}$$

gdje je

$$r = INT\left(\frac{N}{4}\right). \tag{4.19}$$

Treći kvartil određuje se analogno:

- kada N nije djeljiv sa 4

$$Q_3 = x_r, \tag{4.20}$$

gdje je

$$r = INT\left(\frac{3N}{4}\right) + 1, \tag{4.21}$$

- kada je N djeljiv sa 4

$$Q_3 = \frac{x_r + x_{r+1}}{2}, \tag{4.22}$$

gdje je

$$r = INT\left(\frac{3N}{4}\right), \tag{4.23}$$

pri čemu je x_r broj na mjestu r u poredanom nizu podataka, x_{r+1} je broj na mjestu $r+1$ u poredanom nizu podataka, dok N predstavlja ukupni broj podataka u nizu. Operator INT u izrazima predstavlja cijeli dio decimalnog broja.

Pokažimo sada na dva primjera određivanje interkvartilnog raspona.

Primjer 4.3. *Za broj podataka iz Tablice 4.10 koji je djeljiv sa 4 potrebno je izračunati prvi i treći kvartil te interkvartilni raspon.*

Tablica 4.10. Podaci za računanje interkvartilnog raspona kada je N djeljiv sa 4

18	37	18	35	30	12	28	15	18	43	25	24

Prije samog računanja kvartila podatke je potrebno poredati od najmanjeg prema najvećem, što je prikazano ispod:

Tablica 4.11. Podaci poredani redoslijedom od najmanjeg prema najvećem

12	15	18	18	24	25	28	30	35	37	39	43

Prvi kvartil računa se prema sljedećim formulama:

$$r = INT\left(\frac{N}{4}\right) = INT\left(\frac{12}{4}\right) = 3, \tag{4.24}$$

$$Q_1 = \frac{x_r + x_{r+1}}{2} = \frac{18 + 18}{2} = 18 \tag{4.25}$$

Treći kvartil računa se prema sljedećim formulama:

$$r = INT\left(\frac{3 \cdot N}{4}\right) = INT\left(\frac{3 \cdot 12}{4}\right) = 9, \tag{4.26}$$

$$Q_3 = \frac{x_r + x_{r+1}}{2} = \frac{35 + 37}{2} = 36, \tag{4.27}$$

Na posljetku interkvartilni raspon za broj podataka koji je djeljiv sa 4 iznosi:

$$IQR = Q_3 - Q_1 = 36 - 18 = 18, \tag{4.28}$$

Primjer 4.4. *Za broj podataka iz Tablice 4.12 koji nije djeljiv sa 4 potrebno je izračunati prvi i treći kvartil te interkvartilni raspon.*

Tablica 4.12. Podaci za računanje interkvartilnog raspona kada N nije djeljiv sa 4

18	37	18	35	30	12	28	15	18	25	24

Prije samog računanja kvartila podatke je potrebno poredati od najmanjeg prema najvećem, što je prikazano ispod:

Tablica 4.13. Podaci poredani redoslijedom od najmanjeg prema najvećem

12	15	18	18	24	25	28	30	35	37	39

Prvi kvartil računa se prema sljedećim formulama:

$$r = INT\left(\frac{N}{4}\right) + 1 = INT\left(\frac{11}{4}\right) + 1 = 3, \tag{4.29}$$

$$Q_1 = x_r = 18, \tag{4.30}$$

Treći kvartil računa se prema sljedećim formulama:

$$r = INT\left(\frac{3N}{4}\right) + 1 = INT\left(\frac{3 \cdot 11}{4}\right) + 1 = 9, \tag{4.31}$$

$$Q_3 = x_r = 35, \tag{4.32}$$

Konačno interkvartlni raspon za broj podataka koji nije djeljiv sa 4 iznosi:

$$IQR = Q_3 - Q_1 = 35 - 18 = 17, \tag{4.33}$$

4.2.3. Relativne mjere rasapa

Koeficijent varijacije je omjer standardne devijacije i aritmetičke sredine, izražen u postocima. Koeficijentom varijacije određuje se reprezentativnost aritmetičke sredine, koja je većim rasapom sve manje reprezentativna te se određuje je li u dozvoljenim granicama da bi se mogla koristiti kao reprezentativan pokazatelj centralne tendencije. Računa se izrazom: [5]

$$KV = \frac{s}{\bar{x}} * 100, \tag{4.34}$$

gdje s predstavlja standardnu devijaciju, a $\bar{x}$ aritmetičku sredinu.

Primjer 4.5. *Mjerenjem napona baterije dobiveni su sljedeći rezultati: 1.5 V, 1.48 V, 1.45 V, 1.52 V i 1.51 V. Potrebno je izračunati varijancu, standardnu devijaciju, aritmetičku sredinu i koeficijent varijacije te na temelju dobivenuh rezultata utvrditi je li aritmetička sredina reprezentativan pokazatelj centralne tendencije.*

$$\bar{x} = \frac{1.5 + 1.48 + 1.45 + 1.52 + 1.51}{5} = 1.492, \tag{4.35}$$

$$\begin{aligned} VAR = \frac{1}{5} \cdot [&(1.5 - 1.492)^2 + (1.48 - 1.492)^2 + (1.45 - 1.492)^2 \\ &+(1.52 - 1.492)^2 + (1.51 - 1.492)^2] = 6.16 \cdot 10^{-4} \end{aligned} \tag{4.36}$$

$$s = \sqrt{VAR} = \sqrt{6.16 \cdot 10^{-4}} = 0.0248, \tag{4.37}$$

$$KV = \frac{s}{\bar{x}} * 100 = \frac{0.0248}{1.492} * 100 = 1.662, \tag{4.38}$$

Iz primjera se vidi da je aritmetička sredina reprezentativan pokazatelj centralne tendencije jer je rezultat koeficijenta varijacije manji od 30.

Koeficijent kvartilne devijacije služi za određivanje reprezetativnosti medijana i je li u dozvoljenim granicama da bi se mogao korisititi kao reprezentativan pokazatelj centralne tendencije. Rasap kod medijana čine kvartili, a koeficijent kvartilne devijacije računa se sljedećim izrazom: [5]

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100, \quad 0 \leq V_q \leq 1 \tag{4.39}$$

gdje je Q_3 treći, a Q_1 prvi kvartil.

Primjer 4.6. *Za podatke iz tablice 4.15 potrebno je odrediti koeficijent kvartilne devijacije.*

Tablica 4.14. Podaci za računanje koeficijenta kvartilne devijacije

53	35	12	22	46	8	28	19	32	8	24	18

Prije računanja prvog i trećeg kvartila, potrebno je podatke iz gornje tablice poredati od najmanjeg ka najvećem.

Tablica 4.15. Podaci za računanje koeficijenta kvartilne devijacije

8	8	12	18	19	22	24	28	32	35	46	53

Pošto se u tablici nalazi broj podataka koji je djeljiv sa 4, prvi kvartil računa se na sljedeći način:

$$r = INT\left(\frac{N}{4}\right) = INT\left(\frac{12}{4}\right) = 3, \tag{4.40}$$

$$Q_1 = \frac{x_r + x_{r+1}}{2} = \frac{12 + 18}{2} = 15 \tag{4.41}$$

Treći kvartil računa se na sljedeći način:

$$r = INT\left(\frac{3 \cdot N}{4}\right) = INT\left(\frac{3 \cdot 12}{4}\right) = 9, \tag{4.42}$$

$$Q_3 = \frac{x_r + x_{r+1}}{2} = \frac{32 + 35}{2} = 33.5 \tag{4.43}$$

Koeficijent kvartilne devijacije dobije se uvrštavanjem prvog i trećeg kvartila u sljedeću formulu:

$$V_q = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100 = \frac{33.5 - 15}{33.5 + 15} * 100 = 38.1443 \tag{4.44}$$

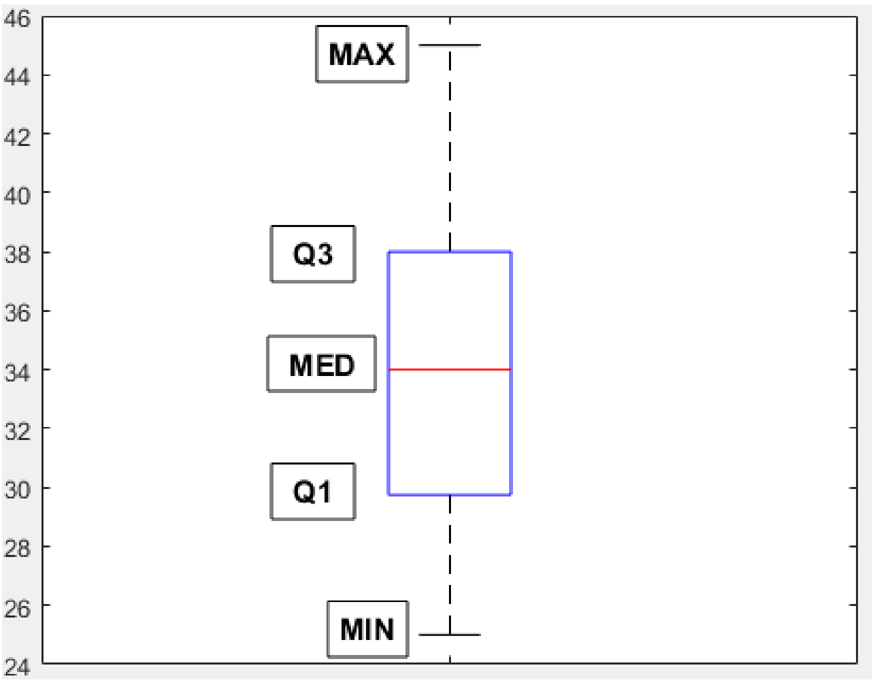
Iz primjera se vidi da medijan nije reprezentativan pokazatelj centralne tendencije jer je rezultat koeficijenta kvartilne devijacije veći od 30.

Ako su vrijednosti koeficijenta varijacije ili koeficijenta kvartilne devijacije manji od 30, kažemo da su pokazatelji centralne tendencije reprezentativni. [3]

5. Grafički prikazi deskriptivne statistike

5.1. Box plot dijagram

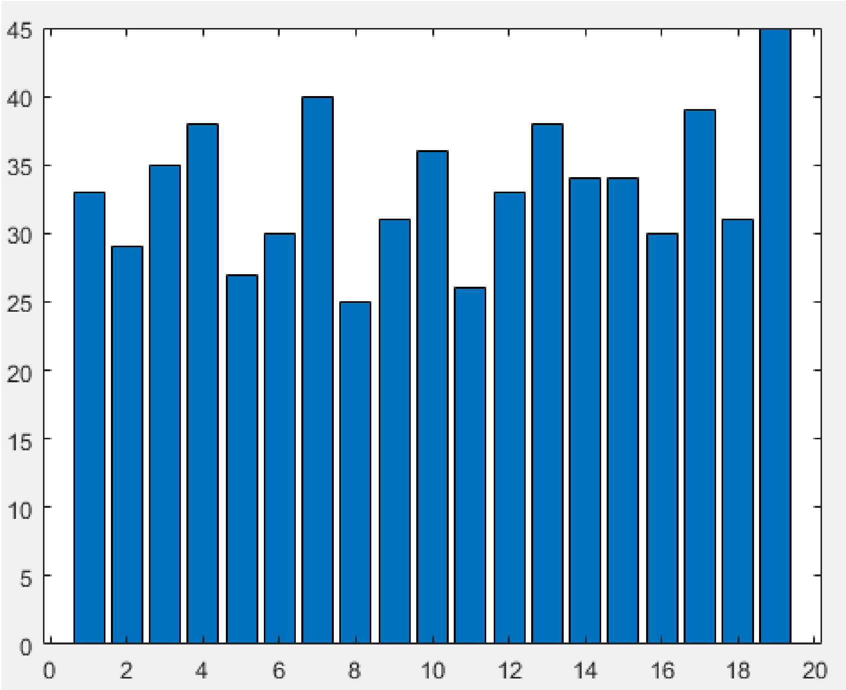
Boxplot dijagram grafički prikazuje skup izmjerenih vrijednosti. Ima oblik pravokutnika, a prikazuje odnos pet numeričkih vrijednosti: minimalna vrijednost, prvi kvartil, medijan, treći kvartil i maksimalna vrijednost. Box plot dijagram sa vrijednostima koje ga opisuju prikazan je na Slici 5.1. [23]



Slika 5.1. Box plot dijagram sa pripadajućim vrijednostima. Izvor: Izrada autora

5.2. Stupčasti dijagram

Stupčasti dijagram služi za grafički prikaz podataka u kojem je visina stupaca određena vrijednostima podataka koji se prikazuju. Stupci u stupčastom dijagramu su međusobno odvojeni i jednake su širine, a mogu biti uspravni, vodoravni ili u obliku različitih geometrijskih tijela (kvadri, valjci, piramide, . . .). Nepraktičan je za prikazivanje vrijednosti koje imaju veliki raspon vrijednosti podataka. Stupčasti dijagram prikazan je na Slici 5.2. [24]



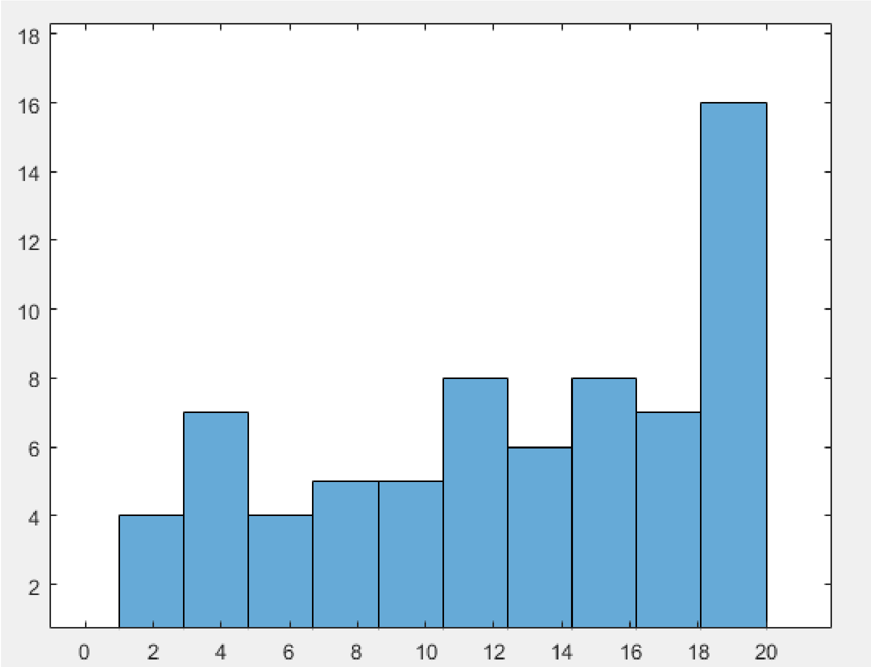
Slika 5.2. Stupčasti dijagram sa nasumičnim vrijednostima. Izvor: Izrada autora

5.3. Histogram

Histogram služi za grafički prikaz skupine statističkih podataka. Sastoji se od stupaca te je sličan stupčastom dijagramu, ali su kod histograma stupci spojeni. Prije njegove izrade, podatke je potrebno podijeliti u intervale te izraditi tablicu frekvencija grupiranih podataka. Stupci se tada postavljaju nad određenim intervalom, a visina stupca pokazuje frekvenciju nekog intervala. Histogram je prikazan na Slici 5.3. [25]

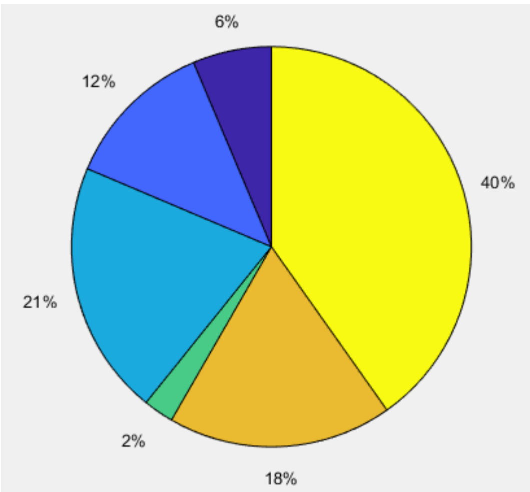
5.4. Kružni dijagram

Kružni dijagram služi za prikazivanje neke vrijednosti kao dio kruga. Podijeljen je na kružne isječke, gdje svaki isječak predstavlja određenu vrijednost, a zbroj svih isječaka predstavlja dijagram. Isječke se može prikazati različitim bojama radi lakše preglednosti, a još se mogu označavati riječima i brojevima ovisno o veličini isječka. Koriste se za prikaz raspodjela ili udjela nekih cjelina i prikazan je na Slikama 5.4a i 5.4b. [26]

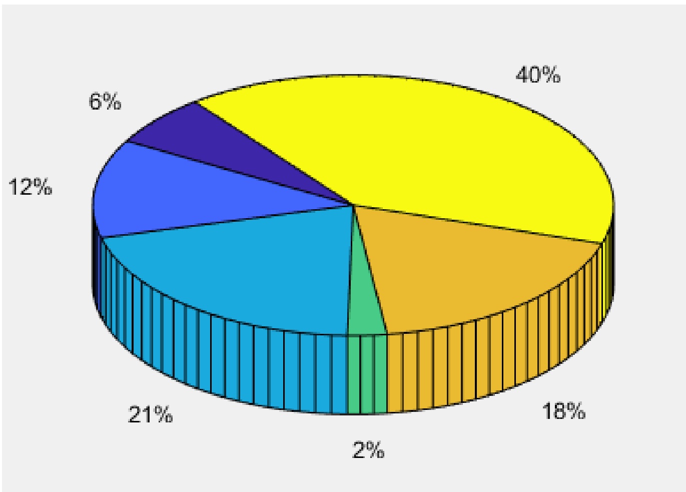


Slika 5.3. Histogram sa nasumičnim vrijednostima. Izvor: Izrada autora

(a) Kružnica



(b) Torta



Slika 5.4. Primjeri kružnog grafikona. Izvor: Izrada autora

6. Osnovne naredbe deskriptivne statistike u Matlabu

Teorijski elementi deskriptivne statistike koji su opisani u jednom od prethodnih poglavlja imaju i svoje naredbe u Matlabu pomoću kojih se dobivaju pokazatelji deskriptivne statistike. Navedene naredbe prikazane su u Tablici 6.1.

Tablica 6.1. Naredbe deskriptivne statistike u Matlabu

NAREDBA	OPIS
min()	Minimum skupa vrijednosti
max()	Maksimum skupa vrijednosti
mean()	Srednja vrijednost skupa vrijednosti
median()	Medijan skupa vrijednosti
mode()	Mod skupa vrijednosti
var()	Varijanca skupa vrijednosti
std()	Standardna devijacija skupa vrijednosti
range()	Raspon skupa vrijednosti
quantile(x,[0.25])	Prvi kvartil skupa vrijednosti
quantile(x,[0.75])	Treći kvartil skupa vrijednosti
iqr()	Interkvartilni raspon skupa vrijednosti

Koeficijent varijacije i koeficijent kvartilne devijacije nemaju svoje naredbe u Matlabu pa je za njih potrebno izvesti naredbe u Matlabu koje su prikazane ispod.

Izvedene naredbe deskriptivne statistike u Matlabu

- Koeficijent varijacije
  [std() / mean()]*100
- Koeficijent kvartilne devijacije
  [quantile(x,[0.75])-quantile(x,[0.25])] / [quantile(x,[0.75])+quantile(x,[0.25])]*100

Na sljedećim primjerima prikazat će se primjena naredbi deskriptivne statistike u Matlabu:

- Minimum skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=min(Vrijednosti)

x = 3
```

- Maksimum vrijednosti skupa

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=max(Vrijednosti)

x = 21
```

- Srednja vrijednost skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=mean(Vrijednosti)

x = 11.8571
```

- Medijan skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=median(Vrijednosti)

x = 12
```

- Mod skupa vrijednosti

```
>> Vrijednosti=[15 12 15 7 3 8 17 21];
>> x=mode(Vrijednosti)

x = 15
```

- Varijanca skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=var(Vrijednosti)          %varijanca populacije

x = 39.4762
```

- Standardna devijacija skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=std(Vrijednosti)

x = 6.283
```

- Raspon skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=range(Vrijednosti)

x = 18
```

- Prvi kvartil skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=quantile(Vrijednosti,[0.25])

x = 7.25
```

- Treći kvartil skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=quantile(Vrijednosti,[0.75])

x = 16.5
```

- Interkvartilni raspon skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=iqr(Vrijednosti)

x = 9.25
```

- Koeficijent varijacije skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x=[std(Vrijednosti)/mean(Vrijednosti)]*100

x = 52.9892
```

- Koeficijent kvartilne devijacije skupa vrijednosti

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> x={[quantile(Vrijednosti,[0.75])-quantile(Vrijednosti,[0.25])]/
[quantile(Vrijednosti,[0.75])+quantile(Vrijednosti,[0.25])]}*100

x = 38.9474
```

U Tablici 6.2 prikazane su naredbe za crtanje grafičih prikaza u Matlabu, tj.dijagrama kojima se prikazuju rezultati deskriptivne statistike.

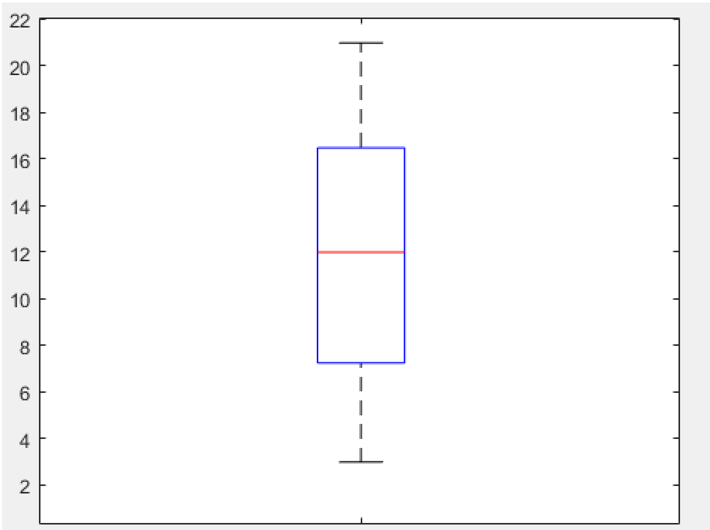
Tablica 6.2. Naredbe za crtanje dijagrama

NAREDBA	**OPIS**
boxplot()	Crtanje boxplot dijagrama
bar()	Crtanje stupčastog dijagrama
histogram()	Crtanje histograma
pie()	Crtanje kružnog dijagrama

U sljedećim primjerima prikazat će se primjena naredbi za crtanje dijagrama u Matlabu:

- Boxplot dijagram

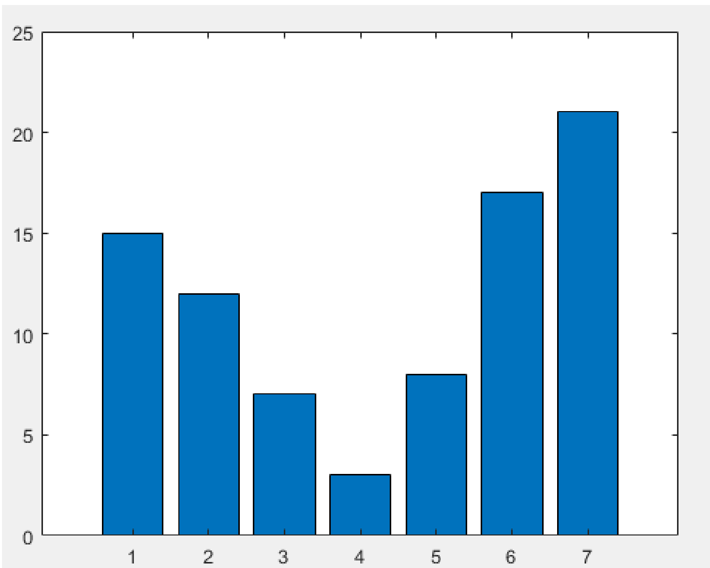
```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> boxplot(Vrijednosti)
```



Slika 6.1. Rezultat naredbe boxplot(). Izvor: Izrada autora

- Stupčasti dijagram

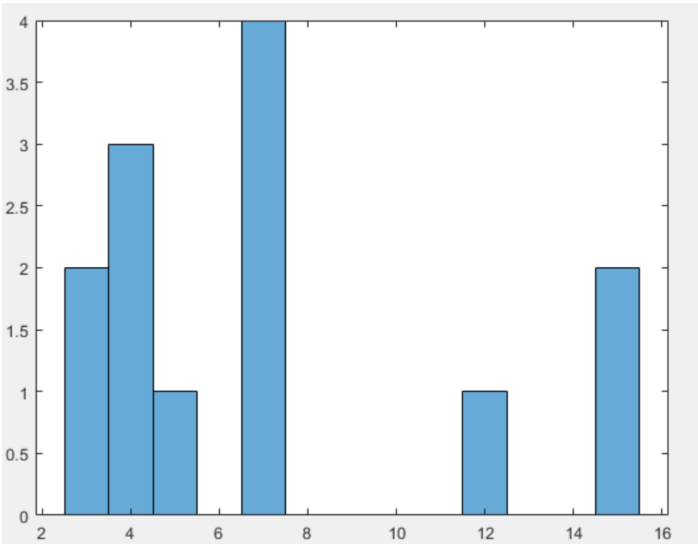
```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> bar(Vrijednosti)
```



Slika 6.2. Rezultat naredbe bar(). Izvor: Izrada autora

- Histogram

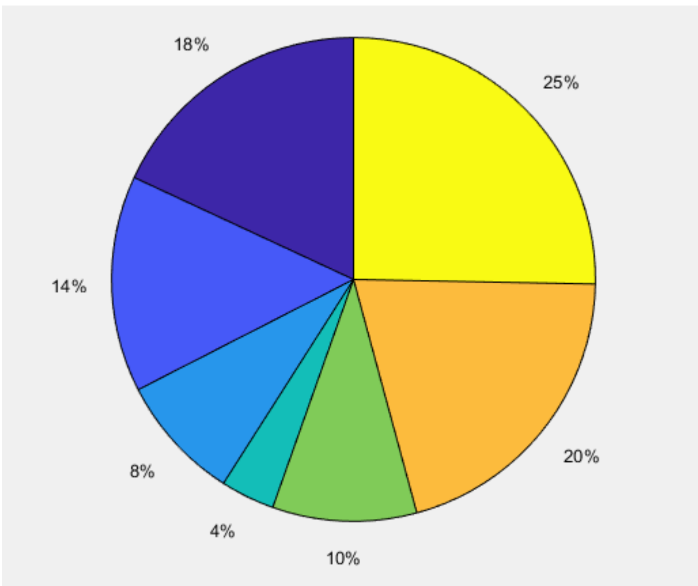
```
>> Vrijednosti=[15 3 3 4 4 4 5 7 7 7 12 15 7];
>> histogram(Vrijednosti)
```



Slika 6.3. Rezultat naredbe histogram(). Izvor: Izrada autora

- Kružni dijagram

```
>> Vrijednosti=[15 12 7 3 8 17 21];
>> pie(Vrijednosti)
```



Slika 6.4. Rezultat naredbe pie(). Izvor: Izrada autora

7. Primjeri deskriptivne statistike

7.1. Mjerenje vrijednosti otpornika ohmmetrom

Mjerenjem pojedinačnih vrijednosti otpora od 21 uzorka otpornika, na laboratorijskim vježbama iz kolegija Mjerenja u elektrotehnici, dobivene su vrijednosti iz Tablice 7.1. Potrebno je odrediti minimalnu i maksimalnu vrijednost otpora, aritmetičku sredinu, medijan i mod, zatim varijancu, standardnu devijaciju, raspon i interkvartilni raspon te odrediti jesu li aritmetička sredina i medijan reprezentativni pokazatelji centralne tendencije pomoću koeficijenta varijacije i koeficijenta kvartilne devijacije. Također je potrebno nacrtati i boxplot dijagram.

Tablica 7.1. Rezultati mjerenja otpora

17.61 kΩ	15.9 kΩ	14.7 kΩ	16.06 kΩ	14.8 kΩ	15.74 kΩ	15.29 kΩ
15 kΩ	15.5 kΩ	14 kΩ	14.9 kΩ	14.63 kΩ	15.3 kΩ	14.6 kΩ
14.26 kΩ	15.04 kΩ	15 kΩ	14.6 kΩ	14.11 kΩ	14.58 kΩ	15 kΩ

Dobivene rezultate iz Tablice 7.1 unosimo u Matlab kao ulaznu varijablu, što možemo vidjeti na Slici 7.1. Korištenjem naredbi sort(), min(), max(), mean(), median(), mode(), var(), std(), range(), iqr() dobivamo redom vrijednosti otpora poredane od najmanjeg prema najvećem (Slika 7.2), minimalnu i maksimalnu vrijednost otpora, aritmetičku sredinu, medijan, mod, varijancu, standardnu devijaciju, rapon i interkvartilni raspon. Koeficijent varijacije i koeficijent standardne devijacije dobiju se iz izvedenih formula jer za njih nema direktnih formula u Matlabu.

```
otpori=[17.61 15.9 14.7 16.06 14.8 15.74 15.29 15 15.5 ...
    14 14.9 14.63 15.3 14.6 14.26 15.04 15 14.6 14.11 14.58 15]   %definiranje ulaznih varijabli

%Izračun traženih vrijednosti
Poredak_najmanji_prema_najvecem=sort(otpori)  %slaganje vrijednosti od najmanje prema najvećoj
Minimum=min(otpori)                           %određivanje vrijednosti najmanjeg otpora
Maksimum=max(otpori)                          %određivanje vrijednosti najvećeg otpora
Aritmeticka_sredina=mean(otpori)              %određivanje vrijednosti aritmetičke sredine
Medijan=median(otpori)                        %određivanje vrijednosti medijana
Mod=mode(otpori)                              %određivanje vrijednosti moda
Varijanca=var(otpori)                         %računanje varijance
Standardna_devijacija=std(otpori)             %određivanje standardne devijacije
Raspon=range(otpori)                          %određivanje raspona
Interkvartilni_raspon=iqr(otpori)             %računanje interkvartilnog raspona

Koeficijent_varijacije=(Standardna_devijacija/Aritmeticka_sredina)*100  %određivanje koeficijenta varijacije

Q1=quantile(otpori,[0.25])                                 %određivanje prvog kvartila
Q3=quantile(otpori,[0.75])                                 %određivanje trećeg kvartila
Koeficijent_kvartilne_devijacije=((Q3-Q1)/(Q3+Q1))*100     %određivanje koeficijenta kvartilne devijacije

%Crtanje boxplot dijagrama
boxplot(otpori);
title('Mjerenje vrijednosti opornika ohmmetrom');  %naslov grafa
ylabel('Vrijednost otpora [R]kΩ')                 %definiranje y osi
```

Slika 7.1. Kod za računanje traženih vrijednosti. Izvor: Izrada autora

```
Poredak_najmanji_prema_najvecem =

  Columns 1 through 14

   14.0000   14.1100   14.2600   14.5800   14.6000   14.6000   14.6300   14.7000   14.8000   14.9000   15.0000   15.0000   15.0000   15.0400

  Columns 15 through 21

   15.2900   15.3000   15.5000   15.7400   15.9000   16.0600   17.6100
```

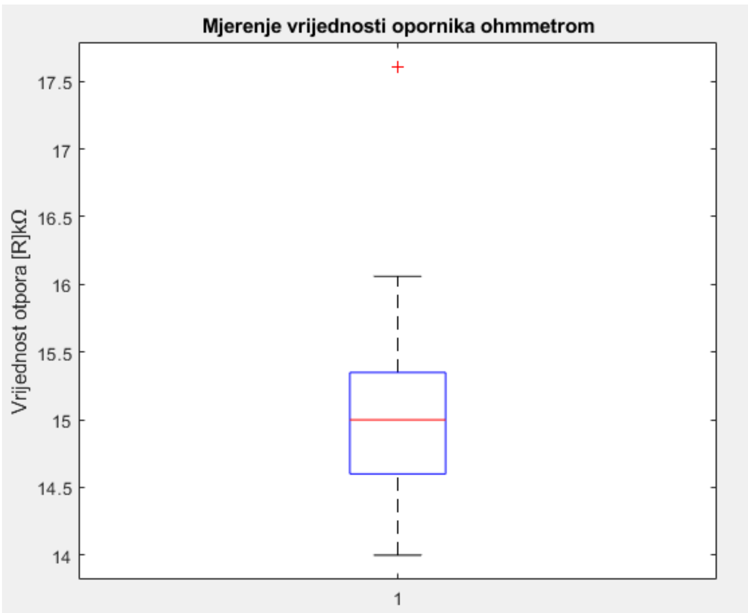
Slika 7.2. Poredak vrijednosti od najmanjeg otpora prema najvećem. Izvor: Izrada autora

Tablica 7.2. Dobivene vrijednosti iz Matlaba

Minimum	14 kΩ
Maksimum	17.61 kΩ
Aritmetička sredina	15.0771 kΩ
Medijan	15 kΩ
Mod	15 kΩ
Varijanca (populacije)	0.6354
Standardna devijacija	0.7971
Raspon	3.61 kΩ
Interkvartilni raspon	0.75
Koeficijent varijacije	5.2867
Prvi kvartil	14.6 kΩ
Treći kvaril	15.35 kΩ
Koeficijent standardne devijacije	2.5042

Iz dobivenih rezultata dolazimo do saznanja da su i aritmetička sredina i medijan reprezentativni pokazatelji centralne tendencije jer su vrijednosti koeficijenta varijacije i koeficijenta kvartilne devijacije manji od 30 (5.2867 < 30 i 2.5042 < 30).

Boxplot dijagram sa Slike 7.3 u Matlabu dobiven je pozivanjem naredbe $boxplot()$. Na slici možemo vidjeti da je minimalna vrijednost izmjerenog otpora 14 kΩ, maksimalna vrijednost 17.16 kΩ, a medijan da iznosi 15 kΩ, dok prvi kvartil iznosi 14.6 kΩ , a treći kvartil 15.35 kΩ.



Slika 7.3. Boxplot dijagram za izmjerene vrijednosti otpora. Izvor: Izrada autora

7.2. Ocjena proizvoda

Ispitivanjem kvalitete nekog proizvoda, ispitano je 100 korisnika koji su mogli za taj proizvod dati ocjenu od 1 do 10. Na temelju dobivenih rezultata potrebno je izračunati sve pokazatelje deskriptivne statistike te nacrtati histogram i kružni dijagram.

Korištenjem naredbe $randi()$ dobije se 100 nasumičnih vrijednosti od 1 do 10 koje je koriste kao ocjene za neki proizvod i te se vrijednosti koriste za računanje svih pokazatelja deskriptivne statistike.

```
Ocjene = randi([1 10],1,100) %generiranje 100 random brojeva između 1 i 10

tabulate(Ocjene)             %određivanje frekvencije ponavljanja pojedinog broja

Aritmeticka_sredina=mean(Ocjene)             %određivanje vrijednosti aritmetičke sredine
Medijan=median(Ocjene)                       %određivanje vrijednosti medijana
Mod=mode(Ocjene)                             %određivanje vrijednosti moda
Varijanca=var(Ocjene)                        %računanje varijance
Standardna_devijacija=std(Ocjene)            %određivanje standardne devijacije
Raspon=range(Ocjene)                         %određivanje raspona
Interkvartilni_raspon=iqr(Ocjene)            %računanje interkvartilnog raspona

Koeficijent_varijacije=(Standardna_devijacija/Aritmeticka_sredina)*100  %određivanje koeficijenta varijacije

Q1=quantile(Ocjene,[0.25])                                   %određivanje prvog kvartila
Q3=quantile(Ocjene,[0.75])                                   %određivanje trećeg kvartila
Koeficijent_kvartilne_devijacije=((Q3-Q1)/(Q3+Q1))*100       %određivanje koeficijenta kvartilne devijacije
```

Slika 7.4. Kod za izbacivanje nasumičnih i dobivanje traženih vrijednosti. Izvor: Izrada autora

Naredbom $tabulate()$ se prikazuje frekvencija pojedine ocjene, to jest koliko puta se koja ocjena pojavila, što je prikazano ispod.

Value	Count	Percent
1	10	10.00%
2	13	13.00%
3	7	7.00%
4	12	12.00%
5	5	5.00%
6	10	10.00%
7	10	10.00%
8	9	9.00%
9	8	8.00%
10	16	16.00%

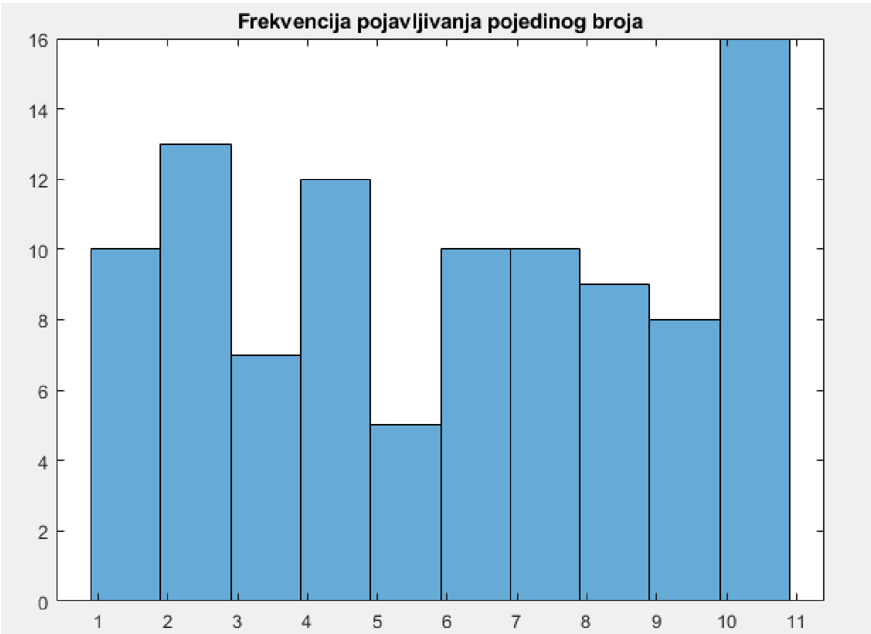
Slika 7.5. Frekvencija prikazivanja pojedine ocjene. Izvor: Izrada autora

Iz dobivenih rezultata iz Tablice 7.3 dolazimo do saznanja da ni aritmetička sredina ni medijan nisu reprezentativni pokazatelji centralne tendencije jer su vrijednosti koeficijenta varijacije i koeficijenta kvartilne devijacije veći od 30 (51.6563 > 30 i 45.4545 > 30).

Tablica 7.3. Dobivene vrijednosti iz Matlaba

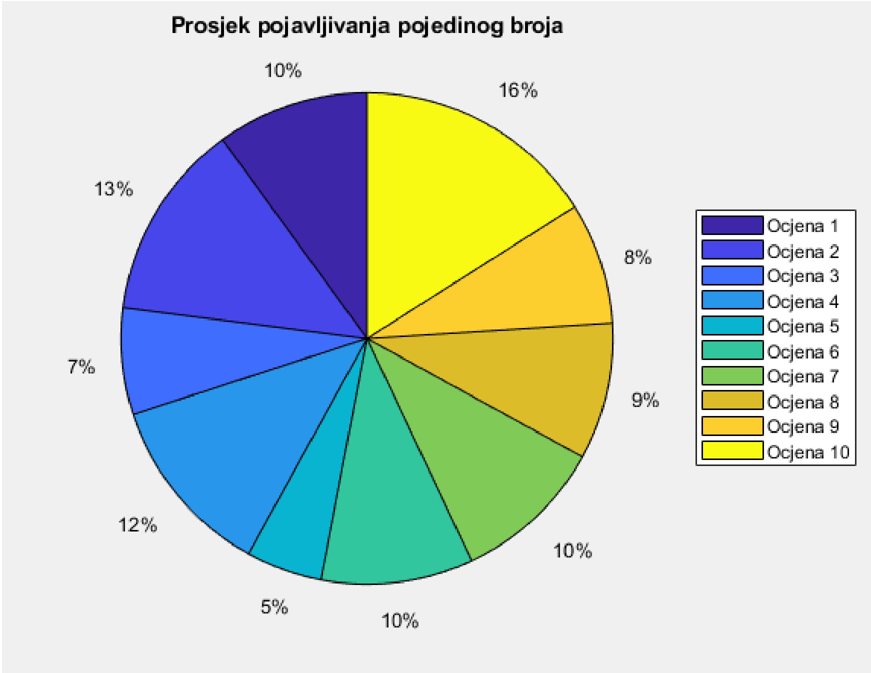
Minimum	1
Maksimum	10
Aritmetička sredina	5.64
Medijan	6
Mod	10
Varijanca (populacije)	9.4448
Standardna devijacija	3.0732
Raspon	9
Interkvartilni raspon	5
Koeficijent varijacije	54.4902
Prvi kvartil	3
Treći kvaril	8
Koeficijent standardne devijacije	45.4545

Histogram sa donje slike se dobije pomoću naredbe $histogram()$. Iz njega se može očitati frekvencija ocjena, to jest on pokazuje koliko se puta neka ocjena pojavila i prikazan je na Slici 7.6.



Slika 7.6. Histogram koji prikazuje frenvenciju prikazivanja pojedine ocjene. Izvor: Izrada autora

Kružni dijagram sa Slike 7.7 dobiven je pomoću naredbe $pie()$ i iz njega se može vidjeti postotak pojavljivanja pojedine ocjene.



Slika 7.7. Kružni dijagram sa postotkom pojavljivanja pojedine ocjene. Izvor: Izrada autora

Kod za dobivanje histograma i kružnog dijagrama prikazan je ispod:

```
Ocjene = randi([1 10],1,100) %generiranje 100 random brojeva između 1 i 10

%Crtanje hisotgrama
histogram(Ocjene,10)                                    %Crtanje hisotgrama
title('Frekvencija pojavljivanja pojedinog broja')      %naslov grafikona

Frekvencije = [ 10 13 7 12 5 10 10 9 8 16]             %frekvencija prikazivanja pojedine ocjene

%Crtanje kružnog dijagrama
labels={'Ocjena 1','Ocjena 2','Ocjena 3','Ocjena 4','Ocjena 5','Ocjena 6','Ocjena 7','Ocjena 8','Ocjena 9','Ocjena 10'}  %definiranje vrijednosti na legendi
ax1 = nexttile;                                         %crtanje kružnog grafikona
pie(ax1,Frekvencije)
title('Prosjek pojavljivanja pojedinog broja')          %naslov grafikona

lgd = legend(labels);                                   %definiranje legende-različita oja za pojedine kontinente
lgd.Layout.Tile = 'east';                               %smještaj legende
```

Slika 7.8. Kod za crtanje histograma i kružnog dijagrama. Izvor: Izrada autora

7.3. Prolaznost studenata na ispitu

Praćenjem broja studenata koji su prošli ispit na pojedinom ispitnom roku, dobiveni su sljedeći rezultati, preuzeti iz zadataka za vježbu iz kolegija Inžinjerska matematika ET:

Tablica 7.4. Prolaznost studenata na ispitnom roku

28	19	10	14	18	12	19	15	14	12
19	11	11	14	15	22	14	12	21	10

Potrebno je odrediti medijan, mod, aritmetičku sredinu, prvi i treći kvartil, koeficijent kvartilne devijacije i varijacije te utvrditi jesu li aritmetička sredina i medijan reprezentativni. Nacrtati boxplot dijagram.

Tražene vrijednosti dobit ćemo pomoću koda iz Matlaba sa Slike 7.9:

```
Studenti=[28 19 10 14 18 12 19 15 14 12 19 11 14 15 22 14 11 21 10]   %definiranje ulaznih varijabli

%Izračun traženih vrijednosti
Medijan=median(Studenti)                                              %određivanje vrijednosti medijana
Mod=mode(Studenti)                                                    %određivanje vrijednosti moda
Aritmeticka_sredina=mean(Studenti)                                    %određivanje aritmetičke sredine
Standardna_devijacija=std(Studenti)                                   %određivanje standardne devijacije
Q1=quantile(Studenti,[0.25])                                          %određivanje prvog kvartila
Q3=quantile(Studenti,[0.75])                                          %određivanje trećeg kvartila
Koeficijent_varijacije=(Standardna_devijacija/Aritmeticka_sredina)    %određivanje koeficijenta varijacije
Koeficijent_kvartilne_devijacije=(Q3-Q1)/(Q3+Q1)                      %određivanje koeficijenta kvartilne devijacije

%Crtanje boxplot dijagrama
boxplot(Studenti)
title('Prolaznost studenata na ispitnom roku');  %naslov grafa
ylabel('Broj studenata')                         %definiranje y osi
```

Slika 7.9. Kod za računanje traženih vrijednosti. Izvor: Izrada autora

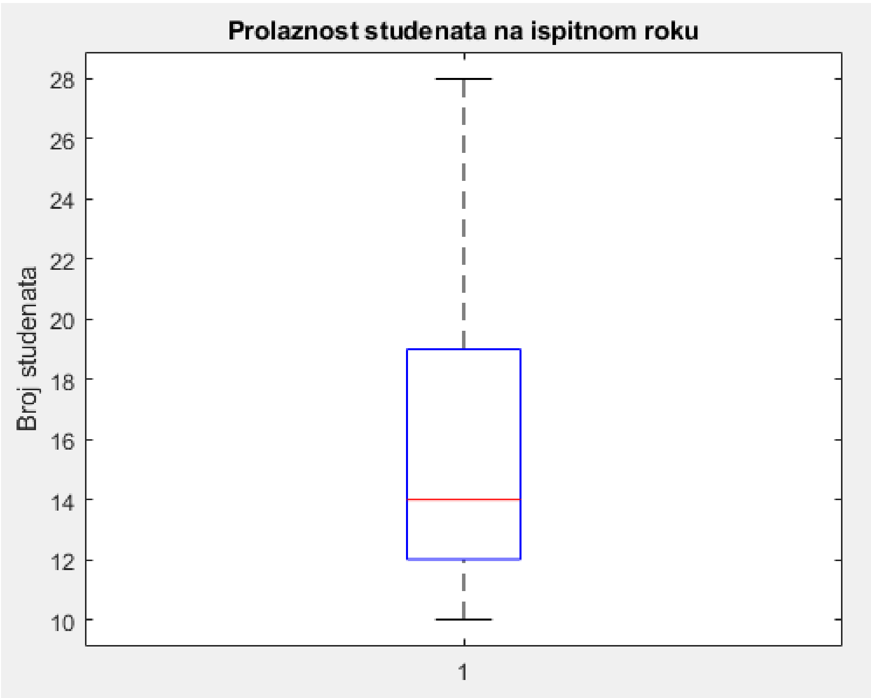
Rezultati traženih vrijednosti iznose:

Tablica 7.5. Dobivene vrijednosti iz Matlaba

Medijan	14
Mod	14
Aritmetička sredina	15.6842
Standardna devijacija	4.7499
Prvi kvartil	12
Treći kvaril	19
Koeficijent varijacije	0.3028
Koeficijent kvartilne devijacije	0.2258

Vidimo da je prosječna prolaznost studenata na pojedinom ispitnom roku 15.6842 studenta te aritmetička sredina nije reprezentativna jer je koeficijent varijacije veći od 0.3 ($0.3028 > 0.3$), dok je medijan reprezentativan jer je koeficijent kvartilne devijacije manji od 0.3 ($0.22584 < 0.3$).

Boxplot dijagram navedenih vrijednosti izleda:



Slika 7.10. Boxplot prolaznosti studenata na ispitnom roku. Izvor: Izrada autora

Iz boxplot dijagrama možemo vidjeti da je na pojedinom ispitnom roku najveći broj studenata koji je prošao ispit iznosio 28, dok je najmanja prolaznost iznosila svega 10 studenata. Također vidimo da medijan iznosi 14 studenata, dok prvi kvartil čini 12, a treći kvartil 19 studenata.

7.4. Cijena električne energije po kWh

U ovom primjeru analizirat će se cijena električne energije u pojedinim državama, članicama Europske unije, koja je izražena u eurima preme potrošenom kilovatsatu (€/kWh). Podaci o cijeni električne energije preuzeti su sa internet stranice [27] i pokazuju cijene električne energije za kućanstva u drugoj polovici 2021. godine. Prema tim podacima potrebno je izračunati sve pokazatelje deskriptivne statistike, utvrditi jesu li aritmetička sredina i medijan reprezentativni pokazatelji centralne tendencije te nacrtati boxplot dijagram i stupčasti dijagram.

Tablica 7.6. Cijena električne energije za pojedine države (€/kWh)

Država	**Cijena električne energije (€/kWh)**
Danska	0.2924
Njemačka	0.2873
Belgija	0.286
Irska	0.2546
Italija	0.2421
Švedska	0.2076
Francuska	0.1913
Finska	0.1783
Češka	0.177
Slovenija	0.1666
Slovačka	0.1585
Rumunjska	0.1421
Poljska	0.1376
Hrvatska	0.1324
Mađarska	0.1097
Bugarska	0.0958

Gore navedene podatke iz tablice o cijenama električne energije u pojedinim državama uvrštavaju se u Matlab kao ulazne varijable, a kao rezultat se dobivaju minimalna i maksimalna cijena elektična energije, aritmetička sredina, medijan, mod, varijanca i standardna devijacija, raspon i interkvartilni raspon, koeficijent varijacije te prvi i treći kvartil koji su potrebni za izračunati koeficijent kvartilne devijacije.

```
Cijene=[ 0.2924 0.2873 0.286 0.2546 0.2421 0.2076 0.1913 0.1783...
    0.177 0.1666 0.1585 0.1421 0.1376 0.1324 0.1097 0.0958 ]       %definiranje ulaznih varijabli

%Izračun traženih vrijednosti
Minimum=min(Cijene)                            %određivanje vrijednosti najmanjeg otpora
Maksimum=max(Cijene)                           %određivanje vrijednosti najvećeg otpora
Aritmeticka_sredina=mean(Cijene)               %određivanje vrijednosti aritmetičke sredine
Medijan=median(Cijene)                         %određivanje vrijednosti medijana
Mod=mode(Cijene)                               %određivanje vrijednosti moda
Varijanca=var(Cijene)                          %računanje varijance
Standardna_devijacija=std(Cijene)              %određivanje standardne devijacije
Raspon=range(Cijene)                           %odredivanje raspona
Interkvartilni_raspon=iqr(Cijene)              %računanje interkvartilnog raspona

Koeficijent_varijacije=(Standardna_devijacija/Aritmeticka_sredina)*100  %određivanje koeficijenta varijacije

Q1=quantile(Cijene,[0.25])                                 %određivanje prvog kvartila
Q3=quantile(Cijene,[0.75])                                 %određivanje trećeg kvartila
Koeficijent_kvartilne_devijacije=((Q3-Q1)/(Q3+Q1))*100     %određivanje koeficijenta kvartilne devijacije
```

Slika 7.11. Kod za računanje traženih vrijednosti. Izvor: Izrada autora

Tablica 7.7. Dobivene vrijednosti iz Matlaba

Minimum	0.0958
Maksimum	0.2924
Aritmetička sredina	0.1912
Medijan	0.1776
Mod	0.0958
Varijanca (populacije)	0.0041
Standardna devijacija	0.0642
Raspon	0.1966
Interkvartilni raspon	0.1085
Koeficijent varijacije	33.5887
Prvi kvartil	0.1399
Treći kvaril	0.2484
Koeficijent kvartilne devijacije	27.9495

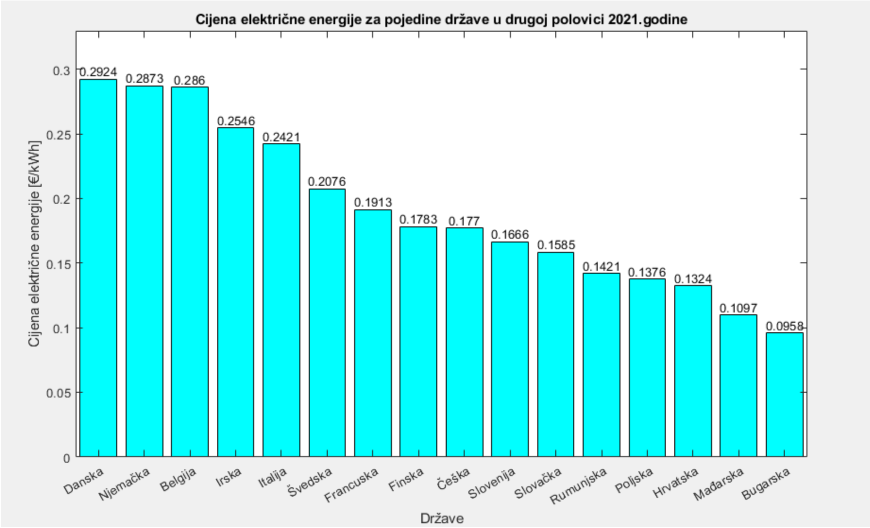
Prema dobivenim razultatima vidi se da aritmetička sredina nije reprezentativan pokazatelj centralne tendencije jer je koeficijent varijacije veći od 30 (33.5887 > 30), dok se za medijan utvrđuje da je reprezentativni pokazatelj centralne tendencije jer je koeficijent kvartilne devijacije manji od 30 (27.9495 < 30), također se vidi da Danska ima najskuplju električnu energiju, a Bugarska najjeftiniju u Europskoj uniji prema potrošenim kWh u eurima. Rezultat naredbe $mod()$ je netočan jer u skupu brojeva nema ponavljanja nekog broja pa bi to trebao biti skup bez moda, no Matlab prilikom takve situacije za mod uzima najmanju vrijednost iz tog skupa.

Stupčasti dijagram sa Slike 7.13 dobije se pomoću naredbe $bar()$ koja je prikazana ispod te se iz njega mogu očitati cijene električne energije pojedinačno za svaku državu.

```
%Crtanje stupčastog dijagrama
X = categorical({'Danska','Njemačka','Belgija','Irska','Italija','Švedska','Francuska','Finska','Češka',...
    'Slovenija','Slovačka','Rumunjska','Poljska','Hrvatska','Mađarska','Bugarska'});
X = reordercats(X,{'Danska','Njemačka','Belgija','Irska','Italija','Švedska','Francuska','Finska','Češka',...
    'Slovenija','Slovačka','Rumunjska','Poljska','Hrvatska','Mađarska','Bugarska'});     %vrijednost x osi
Y = [Cijene];                                                                           %vrijednost y osi
b=bar(X,Y,'c')                                                                          %crtanje stupčastog dijagrama
title('Cijena električne energije za pojedine države u drugoj polovici 2021.godine');    %naslov grafa
xlabel('Države');                                                                       %definiranje y osi
ylabel('Cijena električne energije [€/kWh]');                                           %definiranje x osi

xtips1 = b(1).XEndPoints;
ytips1 = b(1).YEndPoints;
labels1 = string(b(1).YData);
text(xtips1,ytips1,labels1,'HorizontalAlignment','center','VerticalAlignment','bottom')  %dodavanje vrijednosti iznad svakog stupca
```

Slika 7.12. Kod za crtanje stupčastog dijagrama. Izvor: Izrada autora

Pomoću koda dobije se sljedeći dijagram:



Slika 7.13. Stupčasti dijagram prema cijenama električne energije za pojedine države Europske unije. Izvor: Izrada autora

Iz stupčastog dijagrama vidi se da Danska ima najveću cijenu električne energije po kWh, dok Bugarska ima najnižu cijenu električne energije u Europskoj uniji po potrošenom kWh u eurima.

7.5. Ukupna mjesečna i godišnja količina oborina na otoku Hvaru u 2021. godini

Iz tablice izmjerene količine oborina kroz 12 mjeseci na otoku Hvaru, preuzete sa stranice DHMZ-a [6], potrebno je odrediti ukupnu i prosječnu količinu oborina kroz godinu, varijancu i standardnu devijaciju, interkvartilni raspon te nacrtati stupčasti dijagram prema podacima iz Tablice 7.8.

Tablica 7.8. Količina oborina na otoku Hvaru u 2021. godini

Mjesec	Količina padalina (mm)
Siječanj	108,1
Veljača	28,0
Ožujak	21,0
Travanj	54,3
Svibanj	24,7
Lipanj	2,0
Srpanj	6,3
Kolovoz	41,8
Rujan	10,1
Listopad	125,6
Studeni	328,9
Prosinac	84,3

Kod prema kojem se dobije ukupna i prosječna količina oborina kroz godinu, varijanca i standardna devijacija te interkvartilni raspon dan je slikom ispod. Navedene vrijednosti smo dobili korištenjem naredbi $sum(), mean(), var(), std()$ i $iqr()$.

```
Kolicina_oborina=[108.1 28.0 21.0 54.3 24.7 2.0 6.3 41.8 10.1 125.6 328.9 84.3];  %definiranje ulaznih varijabli

%Izračun traženih vrijednosti
Ukupna_kolicina_oborina_kroz_godinu=sum(Kolicina_oborina)              %računanje ukupne količine padalina
Prosjecna_kolicina_oborina_kroz_godinu=mean(Kolicina_oborina)          %računanje aritmetičke sredine
Varijanca=var(Kolicina_oborina)                                        %računanje varijance
Standardna_devijacija=std(Kolicina_oborina)                            %računanje standardne devijacije
Interkvartilni_raspon=iqr(Kolicina_oborina)                            %računanje interkvartilnog raspona
Koeficijent_varijacije=(Standardna_devijacija/Prosjecna_kolicina_oborina_kroz_godinu)          %računanje koeficijenta varijacije

%Crtanje stupčastog dijagrama
X = categorical({'Siječanj','Veljača','Ožujak','Travanj','Svibanj','Lipanj','Srpanj','Kolovoz','Rujan','Listopad','Studeni','Prosinac'});
X = reordercats(X,{'Siječanj','Veljača','Ožujak','Travanj','Svibanj','Lipanj','Srpanj','Kolovoz','Rujan','Listopad','Studeni','Prosinac'});%vrijednost x osi
Y = [Kolicina_oborina];                                                %vrijednost y osi
bar(X,Y,'c')                                                           %crtanje stupčastog dijagrama
title('Količina oborina na otoku Hvaru u 2021.godini');                %naslov grafa
xlabel('Mjeseci');                                                     %definiranje y osi
ylabel('Količina oborina u [mm]');                                     %definiranje x osi
text(xtips1,ytips1,labels1,'HorizontalAlignment','center','VerticalAlignment','bottom')       %dodavanje vrijednosti pojedinom stupcu na grafu
```

Slika 7.14. Kod za računanje traženih vrijednosti. Izvor: Izrada autora

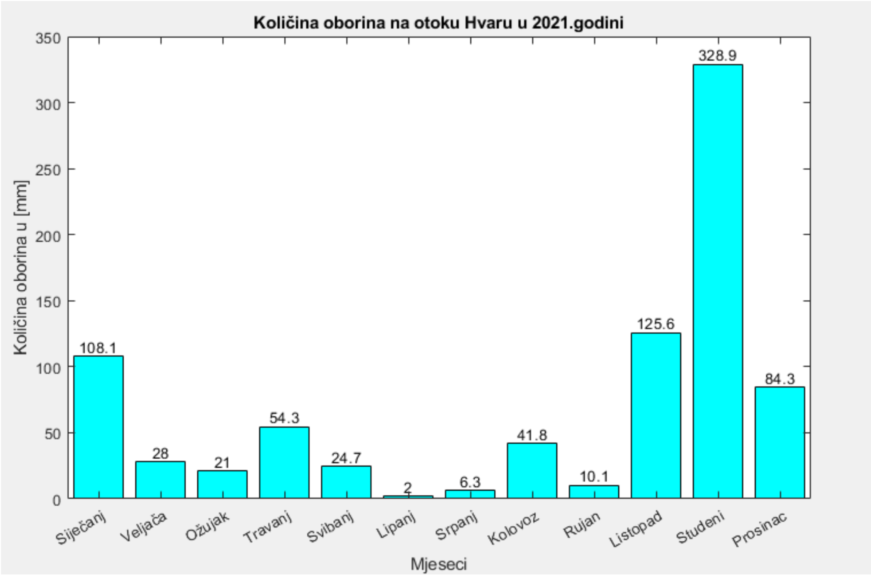
Rezultate dobivene kodom prikazat ćemo Slikom 7.15.

```
Command Window
Ukupna_kolicina_oborina_kroz_godinu =

  835.1000


Prosjecna_kolicina_oborina_kroz_godinu =

   69.5917


Varijanca =

   8.3003e+03


Standardna_devijacija =

   91.1059


Interkvartilni_raspon =

   80.6500


Koeficijent_varijacije =

    1.3091
```

Slika 7.15. Dobivene vrijednosti iz Matlaba. Izvor: Izrada autora

Stupčasti dijagram u Matlabu dobivamo pozivanjem naredbe $bar()$ i prikazan je na Slici 7.16.

Iz dijagrama možemo vidjeti količinu oborina za pojedine mjesece u godini, pa uočavamo da je studeni najkišovitiji mjesec u godini sa 328.9 mm oborina po kvadratnom metru, dok je mjesec sa najmanjom količinom oborina, mjesec lipanj, sa svega 2 mm oborina po kvadratnom metru. Prosječna količina oborina za svaki mjesec od 69.5917 mm po kvadratnom metru je nepouzdana zbog ekstrema, što možemo vidjeti i iz koeficijenta varijacije koji je veći od 0.3 (1.3091 > 0.3).



Slika 7.16. Stupčasti dijagram količine oborina na otoku Hvaru u 2021. godini. Izvor: Izrada autora

7.6. Analiza svjetske populacije stanovništva

U ovom primjeru ćemo analizirati svijetsku populaciju stanovništva s ciljem prikaza postotnog broja stanovništva pojedinog kontinenta na kružnom grafikonu. Prema podacima sa internet stranice [10], o broju stanovnika iz 2010. godine za pojedini kontinent, potrebno je odrediti ukupan broj stanovnika u svijetu, prosječan broj stanovnika po kontinentu, medijan i mod, varijancu i standardnu devijaciju, interkvartilni raspon te nacrtati kružni grafikon. Također odrediti reprezentativnost aritmetičke sredine i medijana.

Tablica 7.9. Broj stanovnika na pojedinom kontinentu

Kontinent	**Broj stanovnika**
Europa	738 849 000
Azija	4 481 757 408
Afrika	1 216 130 000
Sjeverna Amerika	579 024 000
Južna Amerika	422 535 000
Australija	38 304 000
Antartika	1106

Vrijednosti broja stanovnika unosimo kao ulazne varijable u Matlab te za njih računamo ukupan broj stanovnika u svijetu, aritmetičku sredinu, medijan i mod, varijancu i standardnu devijaciju, interkvartilni raspon, koeficijent varijacije i kvartilne devijacije te postotak broja stanovništva pojedinog kontinenta u odnosu na ukupan broj stanovnika.

```
Broj_stanovnika=[738849000 4481757408 1216130000 579024000 422535000 38304000 1106]   %definiranje ulaznih varijabli

%Izračun traženih vrijednosti
Ukupan_broj_stanovnika=sum(Broj_stanovnika)                                   %računanje ukupnog broja stanovnika u svijetu
Prosjecan_broj_stanovnika_po_kontinentu=mean(Broj_stanovnika)                 %računanje prosječnog broja stanovnika po kontinentu
Medijan=median(Broj_stanovnika)                                               %određivanje vrijednosti medijana
Mod=mode(Broj_stanovnika)                                                     %određivanje vrijednosti moda
Varijanca=var(Broj_stanovnika)                                                %računanje varijance
Standardna_devijacija=std(Broj_stanovnika)                                    %računanje standardne devijacije
Q1=quantile(Broj_stanovnika,[0.25])                                           %određivanje prvog kvartila
Q3=quantile(Broj_stanovnika,[0.75])                                           %određivanje trećeg kvartila
Interkvartilni_raspon=iqr(Broj_stanovnika)                                    %računanje interkvartilnog raspona
Koeficijent_varijacije=(Standardna_devijacija/Prosjecan_broj_stanovnika_po_kontinentu)  %određivanje koeficijenta varijacije
Koeficijent_kvartilne_devijacije=(Q3-Q1)/(Q3+Q1)                              %određivanje koeficijenta kvartilne devijacije


Prosjek_stanovnistva_po_kontinentu=(1/sum(Broj_stanovnika)*Broj_stanovnika)*100     %računanje prosječnog broja stanovnika po kontinentu

%Crtanje kružnog grafikona
x=[738849000 4481757408 1216130000 579024000 422535000 38304000 1106]                %vrijednosti na kružnom grafikonu
labels={'Europa','Azija','Afrika','Sjeverna Amerika','Južna Amerika','Australija','Antartika'}  %definiranje vrijednosti na legendi
ax1 = nexttile;
pie(ax1,x)                                                                          %crtanje kružnog grafikona
title('Prosjek stanovništva po pojedinom kontinentu')                               %naslov grafikona

lgd = legend(labels);                                                               %definiranje legende-različita oja za pojedine kontinente
lgd.Layout.Tile = 'east';                                                           %smještaj legende
```

Slika 7.17. Kod za računanje traženih vrijedosti. Izvor: Izrada autora

Rezulatate dobivene pomoću Matlaba prikazat ćemo Tablicom 7.10 i Tablicom 7.11.

Tablica 7.10. Dobivene vrijednosti iz Matlaba

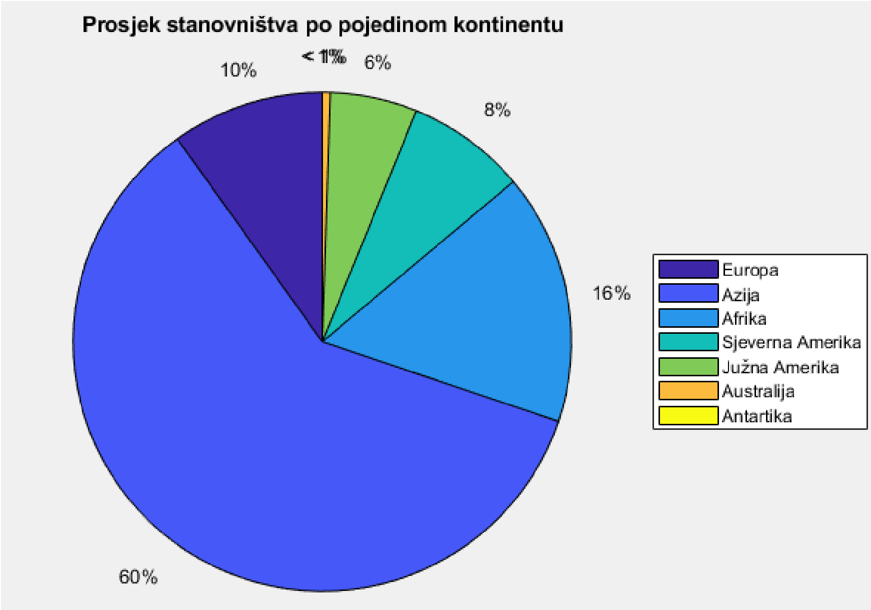
Ukupan broj stanovnika	7 476 600 514
Aritmetička sredina	1 068 100 000
Medijan	579 024 000
Mod	1106
Varijanca	$2.4401 \cdot 10^{18}$
Standardna devijacija	1 562 100 000
Prvi kvartil	134361750
Treći kvaril	$1.0968 \cdot 10^{9}$
Interkvartivni raspon	962 448 000
Koeficijent varijacije	1.4625
Koeficijent kvartilne devijacije	0.7817

Aritmetička sredina u ovom primjeru nije pouzdana jer imamo ekstreme kao npr. broj stanovnika u Aziji, pa dobivamo vrijednost od 1,0681 milijarde stanovnika po pojedinom kontinentu, što dokazuje i koeficijent varijacije (1.4625 > 0.3). Također ni medijan nije reprezentativan jer je koeficijent kvartilne devijacije veći od 0.3 (0.7817 > 0.3).

Tablica 7.11. Postotni broj stanovnika na pojedinom kontinentu

Kontinent	**Broj stanovnika**	**Postotak[%]**
Europa	738 849 000	9.88
Azija	4 481 757 408	59.94
Afrika	1 216 130 000	16.265
Sjeverna Amerika	579 024 000	7.745
Južna Amerika	422 535 000	5.65
Australija	38 304 000	0.512
Antartika	1106	<0.01
Ukupno	7 476 600 514	100

Kružni grafikon u Matlabu dobiven je pozivanjem naredbe $pie()$ i prikazan je na Slici 7.18.



Slika 7.18. Kružni grafikon svjetske populacije stanovništva

Iz kružnog grafikona možemo vidjeti da najveći postotak svjetske populacije živi u Aziji, u kojoj živi gotovo 60% svjetskog stanovništva. Zatim slijede Afrika (16%), Europa (10%), Sjeverna Amerika (8%), Južna Amerika (6%), Australija (<1%, točnije 0,5%), dok na Antartici živi toliko malo stanovnika da je nemoguće isčitati sa grafikona (<0.01%).

8. Zaključak

U ovom završnom radu obrađena je tema deskriptivne statističke analize u programskom paketu Matlab. Opisan je pojam statistike, od njezinih početaka i razvoja do danas. Opisana je podjela statistike na deskriptivnu i inferencijalnu statistiku. Također, statistika ima široku primjenu u znanstvenim i stručnim djelatnostima, kao što su ekonomija, demografija, sociologija, medicina, fizika, psihologija pa tako i u mnogim drugim područjima izvan znanosti. Opisan je i programski paket Matlab, koji nam služi za rješavanje različitih matematičkih problema, vizualizaciju, za izračunavanje i simulaciju koja je vezana uz obradu podataka, upravljanje i regulaciju te za programiranje.

Opisane su naredbe deskriptivne statističke analize, koje su podjeljene na mjere centralne tendencije, mjere rasapa i izradu različitih grafičkih prikaza na temelju dobivenih rezultata.

Kod pomoću kojeg se dobivaju rješenja na temelju određenih vrijednosti napisan je u Editoru, a rezultati su dobiveni u Command Windowu. Na kraju, svaka od naredbi deskriptivne statističke analize računana je pozivanjem određenih naredbi u programskom paketu Matlab te je potkrijepljena grafičkim prikazima.

Matlab je kao programski jezik vrlo koristan pri računanju sa naredbama deskriptivne statistike. Ima unaprijed definirane funkcije i naredbe, pa nam omogućava brže rješavanje složenijih matematičkih problema, što znači da nije potrebno pisati složene kodove, već se problem može riješiti pomoću jedne funkcije ili naredbe.

Literatura

[1] Moler, C.: "A Brief History of MATLAB", s interneta, https://ch.mathworks.com/company/newsletters/articles/a-brief-history-of-matlab.html, 2.4.2022

[2] Eaton, J.W. : "GNU Octave History", s interneta, https://www.gnu.org/software/octave/about, 15.4.2022

[3] Petković, T. : "Kratke upute za korištenje MATLAB-a", s interneta, http://nmdos.zesoi.fer.hr/laboratorij/pdf/matlab_upute.pdf, 4.4.2022

[4] Ban, Ž : "OSNOVE MATLABA", s interneta, http://download.tutoriali.org/Tutorials/Matlab/Osnove_MATLAB-a.pdf, 7.4.2022

[5] Kolegij : "Inžinjerska matematika ET - Prezentacija", Sveučilište u Rijeci, Tehnički fakultet, Rijeka, 29.4.2022

[6] Kolegij : "Inžinjerska matematika ET - Zadaci za samostalni rad", Sveučilište u Rijeci, Tehnički fakultet, Rijeka, 10.5.2022

[7] "OSNOVE STATISTIKE", s interneta, https://www.pmf.unizg.hr/_download/repository/PREDAVANJE7.pdf, 7.5.2022

[8] Državni hidrometeorološki zavod: "Ukupna mjesečna i godišnja količina oborine", s interneta, https://meteo.hr/klima.php?section=klima_podaci&param=k2_1&Godina=2021, 25.5.2022

[9] Šupuk, T: "OSNOVE MATLABA-VJEŽBA 1.", s interneta, https://docplayer.net/47889959-Osnove-matlaba-tamara-supuk-vjezba-1-fakultet-elektrotehnike-strojarstva-i-brodogradnje-la html, 3.4.2022

[10] Howard, W.E.: "A Very Brief History of Statistics", s interneta, https://www.tandfonline.com/doi/abs/10.1080/07468342.2002.11921958?journalCode=ucmj20, 22.5.2022

[11] Hrvatska enciklopedija: "Statistika", s interneta, https://www.enciklopedija.hr/natuknica.aspx?ID=57896, 7.5.2022

[12] Popis kontinenata prema broju stanovnika, s interneta, https://hr2.wiki/wiki/List_of_continents_by_population, 20.5.2022

[13] Importing and Exporting Data from MATLAB and Simulink to Excel, s internet, https://www.uml.edu/docs/Importing-Excel_tcm18-190077.pdf, 19.4.2022

[14] Moler, C.; Little, J.: "A history of MATLAB", s interneta, https://dl.acm.org/doi/10.1145/3386331, 10.4.2022

[15] Pedamkar, P.: "Matlab vs Octave", s interneta, https://www.educba.com/matlab-vs-octave/, 3.5.2022

[16] Vadapalli, P.: "MATLAB Vs Python: Difference Between Matlab & Python [2022]", s interneta, https://www.upgrad.com/blog/matlab-vs-python/, 3.5.2022

[17] Pedamkar, P.: "Python vs Matlab", s interneta, https://www.educba.com/python-vs-matlab/, 4.5.2022

[18] Tutorialspoint: "MATLAB - Simulink", s interneta, https://www.tutorialspoint.com/matlab/matlab_simulink.htm, 8.4.2022

[19] Essert, M.; Žilić, T.: "MATLAB – Matrični laboratorij", s interneta, http://titan.fsb.hr/~vmilic/RacMat/Matlab_udzbenik_print.pdf, 10.4.2022

[20] Transfermarkt: "Most valuable players", s interneta, https://www.transfermarkt.com/spieler-statistik/wertvollstespieler/marktwertetop/plus/0/galerie/0?ausrichtung=alle&spielerposition_id=alle&altersklasse=alle&jahrgang=0&land_id=0&kontinent_id=0&yt0=Show, 15.5.2022

[21] Deskriptivna statistika, s interneta, http://matematika.fkit.hr/novo/statistika_i_vjerojatnost/vjezbe/cjeline/1Deskriptivnastatistika.pdf, 30.4.2022

[22] Benšić, M.; Grahovac, D.; Šuvak, N.: "Opisivanje podataka - Kvartili i interkvartilni raspon", s interneta, https://stedy.hr/opisivanje-podataka/kvartili-i-interkvartilni-raspon, 30.4.2022

[23] Benšić, M.; Grahovac, D.; Šuvak, N.: "Opisivanje podataka - Kutijasti dijagram", s interneta, https://stedy.hr/opisivanje-podataka/kutijasti-dijagram, 2.5.2022

[24] Balat, M. i dr.: "Prikazivanje podataka", s interneta, https://edutorij.e-skole.hr/share/proxy/alfresco-noauth/edutorij/api/proxy-guest/b504e46e-b7a7-4770-bcae-f6b108769a03/html/4821_Prikazivanje_podataka.html, 2.5.2022

[25] Benšić, M.; Grahovac, D.; Šuvak, N.: "Opisivanje podataka - Stupčasti dijagram i histogram", s interneta, https://stedy.hr/opisivanje-podataka/stupicasti-dijagram-i-histogram, 2.5.2022

[26] Wikipedija: "Strukturni krug", s interneta, https://hr.wikipedia.org/wiki/Strukturni_krug, 4.5.2022

[27] Eurostat statistics explained: "Statistički podaci o cijenama električne energije", s interneta, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Statisti%C4%8Dki_podaci_o_cijenama_elektri%C4%8Dne_energije&oldid=496167, 29.5.2022

Sažetak i ključne riječi

U ovom projektnom zadatku obrađena je tema deskriptivne statističke analize u programskom paketu Matlab. Opisane su naredbe deskriptivne statističke analize, koje su podjeljene na mjere centralne tendencije (aritmetička sredina, medijan i mod), mjere rasapa (varijanca i standardna devijacija, raspon i interkvartivni raspon te relativne mjere rasapa) i izradu različitih grafičkih prikaza na temelju dobivenih rezultata (boxplot dijagram, stupčasti dijagram, histogram i kružni dijagram). Na praktičnim primjerima je prikazano kako se pomoću programskog paketa Matlab računaju pokazatelji deskriptivne statističke analize i crtaju različiti grafički prikazi.

Ključne riječi: statistika, deskriptivna statistika, matlab, mjere centralne tendencije, mjere rasapa, dijagrami

Summary and key words

This project task deals with the topic of descriptive statistical analysis in the Matlab software package. The commands of descriptive statistical analysis which are described, are divided into measures of central tendency (arithmetic mean, median and mode), measures of disintegration (variance and standard deviation, range and interquartile range and relative measures of disintegration) and making different graphs based on the obtained results (boxplot diagram, bar chart, histogram and pie chart). Practical examples show how the Matlab software package calculates descriptive statistical analysis indicators and draws various graphical representations.

Keywords: statistics, descriptive statistics, matlab, measures of central tendency, measures of disintegration, diagrams