

Pojam i razvoj statistike

- Skup numeričkih podataka (status – stanje)
- Prvi začetci statistike datiraju nekoliko vjekova prije nove ere (Kina 4000 g.p.n.e)
- Razvoj statistike – razvoj informatičkog društva
- Statističke tehnike se veoma često koriste u institucijama sistema, zatim, u marketingu, računovodstvu, kontroli kvaliteta, ponašanju potrošača, sportu, politici, medicini, itd...

1

Statistika – trostruki sadržaj

1. Deskriptivna statistika (u užem smislu) – prikupljanje, obrada i prezentiranje podataka
2. Statistička analiza – skup statističkih metoda kvantitativne analize pojava i njihovih odnosa
3. Statistička teorija – iznalazi stat. metode, objašnjava ih, dokazuje i usavršava.

2

Predmet i način statističkog istraživanja

- Masovne pojave (varijabilne po prirodi)
- Prirodne pojave i društvene pojave
- Statistika je naučni metod kvantitativnog istraživanja masovnih pojava.
- Statistički skup – skup svih elemenata na kojima se izvjesna pojava statistički posmatra (populacija, osnovni skup)
- Elementi ili jedinice skupa – pojedinačni elementi iz kojih se skup sastoji

3

Formiranje statističkog skupa zavisi od:

- Prirode pojave
- Cilja istraživanja
- Raspoloživih mogućnosti posmatranja

Statistički skup treba da ispunjava tri uslova:

- Homogenost
- Diferenciranost
- Cjelovitost

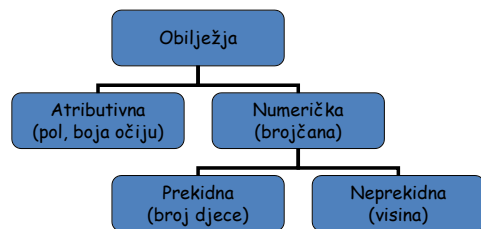
4

Skup se određuje:

- Prostorno
- Vremenski
- Pojmovno (sadržinski, stvarno)
- Statistička obilježja – osobine po kojima se jedinice određenog statističkog skupa međusobno razlikuju
- Modaliteti obilježja - različiti vidovi u kojima se obilježja mogu javiti.

5

Vrste obilježja



6

Statističko posmatranje po obimu:

- Potpuno
- Djelimično

Statistički metodi istraživanja

- Metode sređivanja, prikupljanja i prikazivanja podataka –deskriptivna statistika
- Metodi statističke analize – analitička statistika

7

Tri etape procesa statističkog istraživanja:

- Statističko posmatranje
- Sređivanje, grupisanje i obrada podataka
- Statistička analiza

8

Statističko zaključivanje

- Postupak donošenja zaključaka o karakteristikama statističkog skupa na osnovu posmatranja samo jednog dijela tog skupa – uzorka.
- Dvije osnovne oblasti:
 1. Ocjenjivanje nepoznatih parametara skupa
 2. Testiranje stat. hipoteza

9

Zašto uzorak?

- Rad sa skupom može da bude:
- Nemoguć
- Nepraktičan
- Previše skup

10

Uzorak i Populacija



Populacija (N)



Uzorak (n)

11

Mjerne skale

1. Nominalna skala: Podaci mogu biti klasifikovani samo u modalitete i ne mogu se rangirati. Između modaliteta ne postoji obavezan redoslijed, ali se oni međusobno isključuju.

PRIMJER: bračno stanje, pol, nacionalnost.

12

2. Ordinalna skala: Modaliteti se mogu rangirati, ali ne pokazuju veličinu njihovog razlikovanja.

PRIMJER: Prilikom testiranja četiri vrste sira, sir C je rangiran brojem 1, sir B brojem 2, sir A brojem 3, i sir D brojem 4.

13

3. Intervalna skala: uz redosljed modaliteta, daje i apsolutne razlike između njih. Karakteriše je određena jedinica mjere.

PRIMJER: Temperatura.

14

4. Skala odnosa: Najveći nivo mjerenja. Slična prethodnoj s tim što uključuje nulu kao nivo mjerenja

PRIMJER: Novac, visina, težina itd.

15

Grupisanje stat. podataka

- Prema obilježjima i vremenskim intervalima
- Ručno i mašinsko
- Rezultat stat. obrade – *stat. serije* – nizovi sređenih podataka po nekom obilježju ili hronologiji

16

Serije strukture

- Raspored skupa po vrijednostima obilježja: atributivnom i numeričkom obilježju
- Atributivno obilježje: pr. Geografske serije
- Numerička: **prekidno** i **neprekidno**

Br. Soba u stanovima	Broj stanova
1	90
2	125
3	26
4	7
5	2
Ukupno	250

17

Neprekidno obilježje

- Intervalne grupe
- Veličina grupnog intervala se određuje preko Sturges-ovog pravila za broj klasa, gdje je N ukupan broj podataka, X_{max} i X_{min} najveća i najmanja vrijednost obilježja:

$$k = 1 + 3,3 \log N \quad l = \frac{X_{max} - X_{min}}{k}$$

18

Primjer 1

- Izvršiti grupisanje studenata prema visini (u cm) – uzorak sadrži 33 studenta: 167, **150**, 180, 195, 200, 205, 153, 167, 194, 195, **210**, 193, 168, 183, 162, 157, 154, 190, 163, 170, 172, 184, 185, 152, 187, 167, 175, 194, 180, 156, 183, 152, 194.

19

Primjer 1

- $N=33$, $k=1+3,3\log 33=6,01=6$, $l=(210-150)/6=10$

Visina u cm	frekvencije, f
150-160	7
160,1-170	7
170,1-180	4
180,1-190	6
190,1-200	7
200,1-210	2

20

Dobro i loše grupisanje

Broj radnika	Broj proizvođača
2-6	6
7-11	8
12-16	12
17-21	8
Ukupno	34

Broj radnika	Broj proizvođača
2-7	6
7-12	8
12-17	12
17-24	8
Ukupno	34

- Podjela podataka u grupe (klase, intervale)
- Klase treba da budu:
 - Međusobno isključive
 - Da se ne preklapaju - svaka opservacija se pridružuje samo jednoj grupi
 - Potpune - iscrpne
 - Svaka opservacija je pridružena nekoj klasi
 - Jednake širine (ako je moguće)
 - Prva ili zadnja grupa mogu da budu otvoreni intervale

21

Sugestije za konstrukciju distribucije frekvencija

- Da li koristiti računski određenu širinu intervala za konstruisanje distribucije frekvencija?
VAŽNO: Preporuka je da se koristi računski ali ponekada je bolje i lakše koristiti vama prirodniju. Npr. Ako ste izračunali širinu intervala 97, možda je bolje koristiti 100.
- Izračunati broj podataka u svakoj od klasa.

Rasporedi frekvencija

- Tabela sa 2 kolone koje sadrže:
 - Svaki modalitet ili interval vrijednosti
 - Pridruženu **frekvenciju** svake klase
 - Broj opservacija pridružen svakoj klasi
 - Zbir frekvencija je jednak broju opservacija
 - N za populaciju
 - n za uzorak
- Sredina** klase je njena središnja vrijednost
- Relativna frekvencija** je udio (procenat) svake klase u ukupnim opservacijama
 - Zbir relativnih frekvencija = 1

23

Rasporedi frekvencija

- Apsolutne frekvencije – broj jedinica koji odgovara jednom atributu ili jednoj vrijednosti obilježja
- Relativne frekvencije – ako frekvenciju izvjesnog atributa stavimo u odnos prema ukupnom broju jedinica tog skupa.

$$p_i = \frac{f_i}{\sum f_i}$$

24

Primjer 1

x	f(x)	f(x)/n
Nivo zarada (\$)	Frekvencija (broj potrošača)	Relativna frekvencija
0 do manje od 100	30	0.163
100 do manje od 200	38	0.207
200 do manje od 300	50	0.272
300 do manje od 400	31	0.168
400 do manje od 500	22	0.120
500 do manje od 600	13	0.070
	184	1.000

- Primjer relativne frekvencije: $30/184 = 0.163$
- Zbir relativnih frekvencija = 1

25

Raspored kumulativnih frekvencija

x	F(x)	F(x)/n
Nivo zarada(\$)	Kumulativna frekvencija	Relativna kumulativna frekvencija
0 do manje od 100	30	0.163
100 do manje od 200	68	0.370
200 do manje od 300	118	0.641
300 do manje od 400	149	0.810
400 do manje od 500	171	0.929
500 do manje od 600	184	1.000

Kumulativna frekvencija svake klase je zbir frekvencija te i svih prethodnih klasa.

26

Vremenske serije (hronološke)

- Nizovi stat. podataka koji pokazuju varijacije pojava tokom vremena
- Momentne – nivo
- Intervalne - tok

x	F(x)
Godina	Prosječna posjećenost utakmica Premijer lige
1999	30
2000	68
2001	118
2002	149
2003	171
2004	184

27

Grafičko prikazivanje

- Kartogrami – geografske serije
- Dijagrami – tačkasti, linijski, površinski (histogrami), prostorni (stereogrami)
- Slike

28

Grafički prikaz rasporeda frekvencija

- Prekidno obilježje: pomoću tačaka, štapićastih dijagrama, linijskog dijagrama
- Neprekidno obilježje: histogram, poligon, krive frekvencije i kumulante

29

Štapićasti dijagram

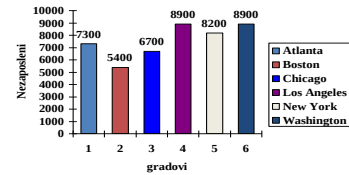
- **Štapićasti dijagram** može se koristiti za prikazivanje bilo kog nivoa mjerenja (nominalni, ordinalni, intervalni, ili racio).
- Visine štapića (pravougaonika) predstavljaju frekvencije klasa
- PRIMJER 2.: Konstruiši štapićasti dijagram za broj nezaposlenih na 100,000 za odabrane gradove.

30

Primjer 2.

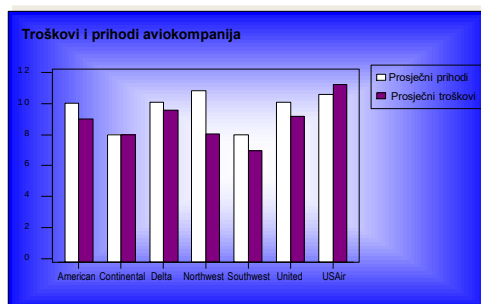
Grad	Broj nezaposlenih na 100000 stanovnika
Atlanta	7300
Boston	5400
Chicago	6700
Los Angeles	8900
New York	8200
Washington	8900

31



32

Primjer 3.



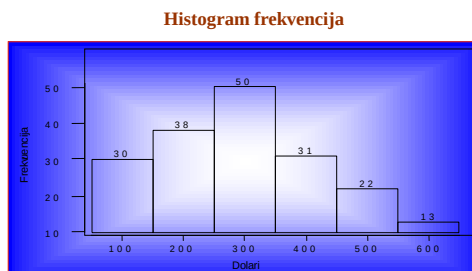
33

Histogram

- **Histogram** je karta napravljena od pravougaonika različitih visina.
 - Širine i lokacije pravougaonika odgovaraju širinama i lokacijama klasa podataka
 - Visine pravougaonika odgovaraju frekvencijama ili relativnim frekvencijama klasa podataka

34

Primjer histograma



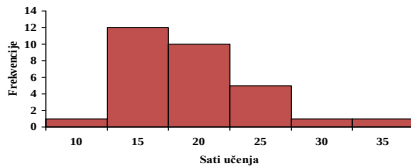
35

Primjer 4.: anketirano je 30 studenata koliko vremena mjesečno utroše na učenje. Podaci su dati u tabeli.

Sati učenja	Frekvencije f	Relativne frekvencije
9-13	1	1/30
14-18	12	12/30
19-23	10	10/30
24-28	5	5/30
29-33	1	1/30
33-38	1	1/30
TOTAL	30	30/30=1

36

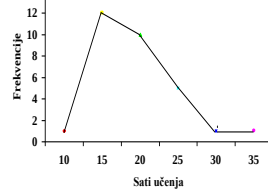
Histogram za primjer 4.



37

Poligon frekvencija

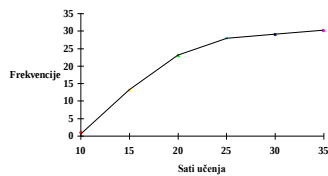
- Poligon frekvencija Sastoji se od linija koje povezuju sredine intervala



38

Raspored kumulativnih frekvencija

- Koristi se da prikaže koliko je podataka raspoređeno ispod ili iznad određene tačke u nekom od intervala.



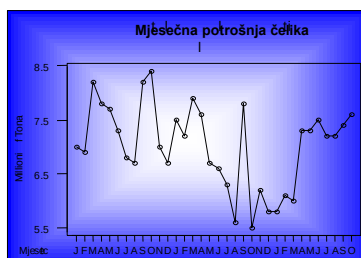
39

Grafičko prikazivanje vremenskih serija

- Aritmetički dijagram
- Polulogaritamski dijagram
- Polarni (sezonski karakter)

40

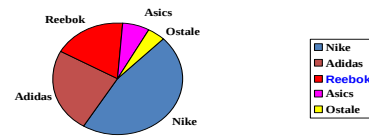
Dijagram vremenske serije



Primjer 5.

Tip patika	# trkača
Nike	92
Adidas	49
Reebok	37
Asics	13
ostale	9

43



44