VIOLETA ALEKSIĆ

ELEMENTI TEORIJE VEROVATNOĆE I MATEMATIČKE STATISTIKE

Materijal za pripremu ispita iz predmeta Obrada i analiza podataka

SADRŽAJ

1 DESKRIPTIVNA STATISTIČKA ANALIZA

Matematička statistika *je, kao što i naziv sugeriše, matematička disciplina. Ona predstavlja naučni metod kvantitativnog istraživanja masovnih pojava. Pojedinačni slučajevi neke pojave mogu pokazivati manja ili veća odstupanja od prosečnog ili tipičnog, pa je neophodno da se posmatraju u velikom broju, tj. u masi, da bi se otkrilo ono što je u njima opšte i zakonito. Matematička statistika daje instrukcije za pravilno sakupljanje, uređivanje, prikazivanje kao i za pravilnu analizu i interpretaciju empirijskih podataka radi dubljeg razumevanja pravilnosti neke pojave.*

1.1 POPULACIJA. OBELEŽJE

Populacija je osnovni pojam matematičke statistike. To je skup sa velikim brojem elemenata koji su istovrsni u odnosu na jednu ili više osobina. Istovrsnost ne znači jednakost. Koriste se i termini: statistička masa, statistički skup, osnovni skup, generalni skup.

Primer. Jednu populaciju čine studenti FZNR, školske 2008/9. godine. Elementi ovog skupa su istovrsni po tome što su studenti ovog fakulteta te generacije ali se oni razlikuju po starosti, visini, težini, boji očiju i kose, različiti su im brojevi košulja i cipela itd.

Osobina po kojoj se elementi populacije razlikuju, a koja je predmet statističke analize, naziva se **obeležje.** Na jednoj populaciji, tj. na svakom njenom elementu, može se posmatrati (meriti) jedna osobina. Tada se kaže da je obeležje **jednodimenzionalno**. Mogu se posmatrati dve ili više osobina istovremeno. Tada se kaže da je obeležje **višedimenzionalno.** Nadalje se razmatraju samo jednodimenzionalna obeležja.

Primer. Populaciju čini skup svih Nišlija zaposlenih u određenom mesecu, a obeležje je njihova mesečna plata.

Primer. Populaciju čine svi zaposleni u jednoj fabrici određene kalendarske godine, a obeležje je broj dana bolovanja, u toku te godine.

Primer. Populaciju čine zaposleni na radnim mestima sa povećanim rizikom neke fabrike, a obeležje je broj povreda na radu u određenom vremenskom periodu.

Primer. Treba vršiti određeno merenje, koje zbog prisustva određenih smetnji nije u potrebnoj meri ,,apsolutno" tačno. Skup svih mogućih merenja čini populaciju. Obeležje svakog merenja jeste vrednost koja se dobija u tom merenju.

Populacija može imati konačno, beskonačno prebrojivo ili beskonačno neprebrojivo elemenata. U ovoj glavi biće razmatrane samo populacije sa konačnim brojem elemenata, koji je takav da je moguće i ekonomski opravdano registrovati (izmeriti) vrednost obeležja na svakom elementu populacije. Ako je N broj elemenata populacije, onda se ona može označiti sa

$$\mathcal{E} = \{e_1, e_2, \dots e_N\}.$$

Obeležje može biti numeričko (kvantitativno) ili nenumeričko (kvalitativno, atributivno).

Numerička obeležja se izražavaju brojčano i mogu biti diskretna (prekidna) ili apsolutno neprekidna (kontinualna). **Diskretna obeležja** mogu da se broje, kao na primer, broj rođenih u jednom danu, broj članova porodice, broj studenata koji su položili ispit iz prvog puta itd. Diskretna obeležja imaju konačno ili prebrojivo beskonačno vrednosti. **Apsolutno neprekidna** obeležja uzimaju vrednosti iz nekog intervala realnih brojeva,a tih vrednosti ima neprebrojivo beskonačno. To su obeležja koja se mere kao visina, težina, potrošnja goriva, nivo šećera u krvi itd.

Atributivna obeležja se izražavaju opisno kao na primer zanimanje, boja kose, pol itd. Ona su po pravilu diskretna i mogu se na odgovarajući način predstaviti kao numerička.

Oznaka za obeležje je veliko slovo latinice $X, Y, Z \ldots$ itd. Vrednost obeležja X na $i-$tom elementu populacije e_i označava se sa $x_i = X(e_i)$.

Primer 1. Populaciju čini 40 ljudi koji su u određenom periodu kupili cipele u jednoj radnji muških cipela. Sa e_i će biti obeležen $i-$ti kupac $(i = 1, 2, \ldots, 40)$. Obeležeje elemenata populacije je veličina kupljenog para cipela. Tako je x_i veličina cipela koje je kupio kupac e_i. Za vrednosti obeležja su dobijeni sledeći rezultati (po redu kupovine):

e_i	1	2	3	4	5	6	7	8	9	10
x_i	39	40	38	43	41	43	40	38	41	42

e_i	11	12	13	14	15	16	17	18	19	20
x_i	41	42	39	41	41	38	43	41	42	38

e_i	21	22	23	24	25	26	27	28	29	30
x_i	41	40	42	41	42	44	41	40	44	42

e_i	31	32	33	34	35	36	37	38	39	40
x_i	41	41	40	42	39	40	40	41	41	44

Obeležje je ovim kompletno opisano. Ipak, prodavcima ništa ne znači to što je sedmi kupac e_7 kupio cipele broj 40, već ih interesuje koliko je kupaca kupilo cipele broj 40, zbog sledeće narudžbine. Nameće se potreba drugačijeg uređivanja ovih podataka.

1.2 UREĐIVANJE I PRIKAZIVANJE PODATAKA

Statističko proučavanje započinje tako što se definišu populacija i obeležje koje se posmatra (meri), a onda se prikupljaju podaci. Radi dalje obrade oni se predstavljaju na dva osnovna načina: tablično i grafički.

▷ **Tablični metod.** Dobijeni podaci se najpre poređaju u rastući niz, tzv. varijacioni niz i izdvoje se različite vrednosti. Jedan od načina formiranja tabela je da registrovane (izmerene) vrednosti obrazuju prvu kolonu, a da ostale kolone čine apsolutne frekvencije, relativne frekvencije, zbirne (kumulativne) frekvencije itd. po potrebi.

U prethodnom primeru varijacioni niz je: 38 38 38 38 39 39 39 40 40 40 40 40 40 40 41 41 41 41 41 41 41 41 41 41 41 41 41 42 42 42 42 42 42 42 43 43 43 44 44 44.

Može se zaključiti da ovo obeležje uzima 7 različitih vrednosti: $x_1 = 38$, $x_2 = 39$, $x_3 = 40$, $x_4 = 41$, $x_5 = 42$, $x_6 = 43$, i $x_7 = 44$.

Apsolutna frekvencija (ili samo frekvencija) f_i je broj elemenata populacije za koje je vrednost obeležja jednaka x_i, $i = 1, 2, \ldots, k$. Određenoj vrednosti obeležja x_i može se pridružiti i odgovarajuća **relativna frekvencija** $f_i^* = \dfrac{f_i}{N}$ kao i relativna frekvencija u procentima $f_i^*\% = f_i^* \cdot 100\%$.

U prethodnom primeru veličini cipela $x_3 = 40$ pridružuje se apsolutna frekvencija $f_3 = 7$ jer je sedmoro kupilo cipele veličine 40. Odgovarajuća tabela je:

vel. cipela–x_i	br. kupaca–f_i	f_i^*	$f_i^*\%$
38	4	4/40=0,1	10%
39	3	3/40=0,075	7,5%
40	7	7/40=0,175	17,5%
41	13	13/40=0,325	32,5%
42	7	7/40=0,175	17,5%
43	3	3/40=0,075	7,5%
44	3	3/40=0,075	7,5%
	$\sum f_i = 40$	$\sum f_i^* = 1$	$\sum f_i^*\% = 100\%$

Iz ove tabele sledi da je $32,5\%$ cele populacije kupilo cipele veličine 41, a 10% veličine 38. Na osnovu ovih podataka u prodavnici mogu planirati sledeću porudžbinu.

Kaže se da je na ovaj način data **raspodela obeležja** "veličina cipela", odnosno **distribucija**[1] **frekvencija obeležja.**

Osnovni problem kojim se matematička statistika bavi sastoji se u sledećem: za datu populaciju naći raspodelu vrednosti posmatranog obeležja tj. raspodelu obeležja. Ako je populacija konačna, raspodelu obeležja čine sve različite vrednosti obeležja

$$x_1, x_2, \ldots, x_k$$

sa odgovarajućim apsolutnim frekvencijama[2]

$$f_1, f_2, \ldots, fk.$$

Primer 2. Anketirana je populacija od 50 studenata o broju položenih ispita. Dobijeni su sledeći rezultati:

7 4 12 3 7 8 6 5 9 9 10 11 6 7 8 6
9 4 5 5 7 3 9 8 6 8 7 6 8 9 6 7
4 10 11 11 12 6 7 7 8 4 10 11 4 12 6
7 8 9.

[1]distribucija - raspodela, raspoređivanje, razvrstavanje, klasifikovanje

[2]frekvencija - učestalost, zastupljenost

Obeležje je broj položenih ispita, a iz podataka sledi da je prvi anketirani student položio 7 ispita, drugi 4 ispita itd.

Varijacioni niz je: 3 3 4 4 4 4 4 5 5 5 6 6 6
6 6 6 6 6 7 7 7 7 7 7 7 7 7 8 8 8
8 8 8 8 9 9 9 9 9 9 10 10 10 11 11
11 11 12 12 12.

Tabela distribucije frekvencija broja studenata prema broju položenih ispita je:

br. pol. ispita x_i	br. studenata f_i	f_i^*	$\mathcal{Z}_i$
3	2	2/50=0,04	2
4	5	5/50=0,1	7
5	3	3/50=0,06	10
6	8	8/50=0,16	18
7	9	9/50=0,18	27
8	7	7/50=0,14	34
9	6	6/50=0,12	40
10	3	3/50=0,06	43
11	4	4/50=0,08	47
12	3	3/50=0,06	50
-	$\sum f_i = 50$	$\sum f_i^* = 1$	-

Zbir svih apsolutnih frekvencija je uvek N tj. jednak broju elemenata populacije, a zbir svih relativnih frekvencija je uvek jedan. Zbirne (kumulativne) frekvencije su:
$\mathcal{Z}_1 = f_1$,
$\mathcal{Z}_2 = f_1 + f_2$,
$\mathcal{Z}_3 = f_1 + f_2 + f_3$ itd.

Prva vrsta tabele je zaglavlje, a u zadnjoj vrsti su zbirovi odgovarajućih kolona. Prethodna tabela ima još 10 vrsta jer obeležje, broj položenih ispita, ima toliko različitih vrednosti. Svaki student ove populacije je položio najmanje 3, a najviše 12 ispita. Ako obeležje ima mnogo različitih vrednosti, tabela bi imala mnogo vrsta što nije pregledno. U tom slučaju se podaci, tj. vrednosti obeležja, grupišu u itervale, najčešće jednake dužine. Osim toga, ako je obeležje apsolutno neprekidnog tipa, koristi se intervalno sređivanje podataka.

Najpre se odredi broj intervala K. Preporučuje se da taj broj bude približno $K = 1 + 3,322 \log_{10} N$, odnosno, da zadovoljava nejednačinu

$$1 + 3,322 \log_{10} N \leq K \leq 5 \log_{10} N.$$

Zatim se odredi dužina intervala Δ tako da bude približno jednaka broju koji se dobija po formuli

$$\Delta = \frac{x_{max} - x_{min}}{K},$$

gde su x_{max} i x_{min} najmanja i najveća vrednost obeležja. Intervali su oblika $[a_i, a_{i+1})$, $i = 1, 2, \ldots, K$. Na kraju se određuju frekvencije f_i, $i = 1, 2, \ldots, K$, tj. vrši se prebrojavanje vrednosti obeležja po intervalima i formira se tabela.

Primer 3. Anketirana je populacija od 50 studenata o telesnoj težini u kilogramima. Dobijeni su sledeći rezultati:

57	84	112	83	77	68	96	105	90	69	102	71	68
72	78	76	89	74	55	85	87	73	89	78	67	68
74	69	80	79	66	67	64	104	110	91	92	68	75
77	82	64	101	91	64	62	65	73	81	91		

Varijacioni niz je: 55 57 62 64 64 64 65 66 67 67
68 68 68 68 69 69 71 72 73 73 74 74 75
76 77 77 78 78 79 80 81 82 83 84 85 87
89 89 90 91 91 91 92 96 101 102 104 105
110 112.

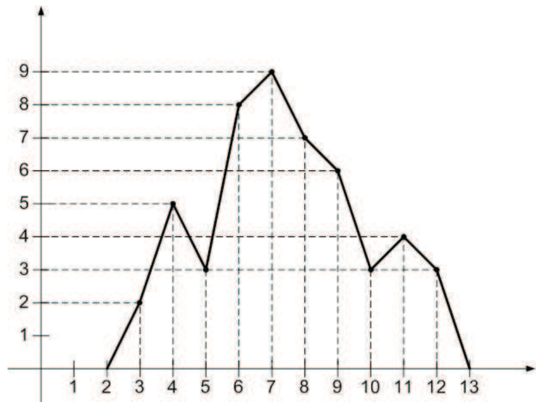
Kako je $1+3,322\log_{10}50 = 6,67$ i $5\log_{10}50 = 8,49$ odgovarajući broj intervala je $K=7$ ili $K=8$. Neka je $K=7$.
Kako je $\dfrac{x_{max}-x_{min}}{K} = \dfrac{112-55}{7} = 8,14$ dužina intervala se može korigovati. Sve vrednosti tj. težine su u intervalu $[50,120]$ pa se može uzeti $\Delta = \dfrac{120-50}{7} = 10.$

Tabela distribucije frekvencija broja studenata prema telesnoj težini je:

težina studenta x_i	br. studenata f_i	f_i^*	$\mathcal{Z}_i$
[50-60)	2	2/50=0,04	2
[60-70)	14	14/50=0,28	16
[70-80)	13	13/50=0,26	29
[80-90)	9	9/50=0,18	38
[90-100)	6	6/50=0,12	44
[100-110)	4	4/50=0,08	48
[110-120)	2	2/50=0,04	50
-	$\sum f_i = 50$	$\sum f_i^* = 1$	-

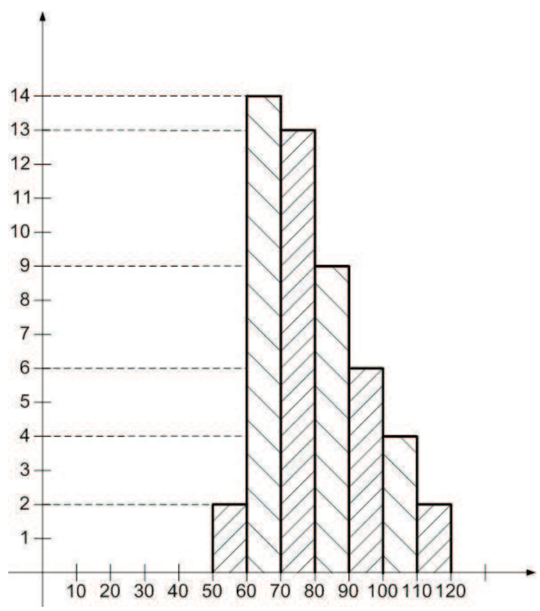
▷ **Grafički metod.** Veliki broj ljudi teško shvata količinske odnose prikazane statističkim podacima u tabelama. Primena crteža, koji se u statistici nazivaju graficima, čini podatke jasnijim i razumljivijim.

Podaci sređeni tabličnim metodom se predstavljaju u koordinatnom sistemu unošenjem tačaka čije su apscise vrednosti, a ordinate frekvencije obeležja. Za **poligon** apsolutnih frekvencija unose se tačke (x_i, f_i), $i = 1, 2, \ldots, k$ koje se spoje dužima, pri čemu se krajnje tačke spoje sa x-osom. Na isti način pomoću tačaka (x_i, f_i^*) dobija se poligon relativnih frekvencija. Ukoliko se na taj način predstavljaju zbirne frekvencije, tačke $(x_i, \mathcal{Z}_i)$, spojene dužima čine **kumulativnu krivu.** Kad su vrednosti obeležja grupisane u intervale za apscise ovih tačaka se koriste sredine intervala.



Slika 1: Poligon apsolutnih frekvencija iz primera 2.

Kod itervalno predstavljenih podataka češće se koristi **histogram** frekvencija. U tu svrhu se nad intervalima crtaju pravougaonici sa visinom jednakom frekvenciji podataka u tom intervalu.



Slika 2: Histogram apsolutnih frekvencija iz primera 3.

1.3 PARAMETRI SREDNJE VREDNOSTI

Da bi se izveli određeni zaključci o raspodeli ispitivanog obeležja u populaciji nije dovoljno podatke urediti, tablično i grafički prikazati. Neophodno je izračunati neke parametre (pokazatelje, mere) koji ukazuju na centar grupisanja vrednosti obeležja.

▷ **Aritmetička sredina.** Kao mera srednje vrednosti obeležja X najčešće se koristi aritmetička sredina ili prosek. Aritmetička sredina populacije obeležava se sa μ, a računa po nekoj od sledećih formula:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}, \qquad \mu = \frac{\sum_{i=1}^{k} f_i \cdot x_i}{N}, \qquad \mu = \frac{\sum_{i=1}^{K} f_i \cdot x_i'}{N}.$$

(k je broj različitih vrednosti obeležja; K je broj intervala; x_i' su sredine intervala.)

Primer 4. Pet novorođenih beba ima težine $4,2 \quad 3,8 \quad 4,1 \quad 5,1 \quad 4,0kg.$
Izračunati prosečnu težinu ovih novorođenih beba.

Rešenje. $\mu = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{4,2+3,8+4,1+5,1+4,0}{5} = 4,24.$
Prosečna težina ovih novorođenih beba iznosi $4,24kg$.

Primer 5. Izračunati aritmetičku sredinu podataka iz primera 2.

Rešenje. Radna tabela za izračunavanje aritmetičke sredine je:

br. položenih isp. x_i	br. studenata f_i	$x_i \cdot f_i$
3	2	6
4	5	20
5	3	15
6	8	48
7	9	63
8	7	56
9	6	54
10	3	30
11	4	44
12	3	36
-	$\sum f_i = 50$	$\sum x_i \cdot f_i = 372$

$\mu = \frac{\sum_{i=1}^{10} f_i \cdot x_i}{50} = \frac{372}{50} = 7,44,$ pa se može smatrati da je prosečan broj položenih ispita 7.

Primer 6. Izračunati aritmetičku sredinu podataka iz primera 3.

Rešenje. Radna tabela za izračunavanje aritmetičke sredine je:

težina studenta x_i	br. studenata f_i	x_i'	$f_i \cdot x_i'$
[50-60)	2	55	110
[60-70)	14	65	910
[70-80)	13	75	975
[80-90)	9	85	765
[90-100)	6	95	570
[100-110)	4	105	420
[110-120)	2	115	230
-	$\sum f_i = 50$	-	$\sum f_i \cdot x_i' = 3980$

$\mu = \dfrac{\sum_{i=1}^{7} f_i \cdot x_i'}{50} = \dfrac{3980}{50} = 79,6kg,$ tj. prosečna težina ovih studenata je $79,6kg$.

Aritmetička sredina kao numerička vrednost može biti različita od svake vrednosti obeležja populacije iz koje je izračunata. Značajna je činjenica da je zbir odstupanja vrednosti obeležja od aritmetičke sredine jednak nuli tj.

$$\sum_{i=1}^{N}(x_i - \mu) = 0 \quad \text{odnosno,} \quad \sum_{i=1}^{k} f_i \cdot (x_i - \mu) = 0.$$

Osim toga, zbir kvadrata ovih odstupanja[3]

$$\sum_{i=1}^{N}(x_i - \mu)^2$$

je minimalan. Ako se odstupanje vrednosti obeležja od proseka smatra greškom, sledi da je zbir kvadrata grešaka minimalan i da je aritmetička sredina najreprezentativnija vrednost obeležja.

Aritmetička sredina je veoma osetljiva na ekstremne vrednosti obeležja, pa je ona loš pokazatelj u distribucijama sa malim brojem ekstremnih vrednosti. To se može videti na primeru sledećih varijacionih nizova:

niz 1. 10 10 10 10 10 10 10 10 10 10,

niz 2. 1 10 10 10 10 10 10 10 10 10,

niz 3. 10 10 10 10 10 10 10 10 10 20.

Jasno je u sva tri slučaja da je broj 10 vrednost numeričkog obeležja koja dobro reprezentuje statistički skup. Međutim, u prvom slučaju je $\mu = 10$, u drugom $\mu = 9,1$, a u trećem $\mu = 11$. Zaključuje se, da pri tumačenju aritmetičke sredine treba biti oprezan.

▷ **Mod.** Mod, u oznaci M_o, je vrednost obeležja čija je apsolutna frekvencija najveća. Ukoliko su sve frekvencije jednake mod ne postoji. Mogu da postoje dva ili više moda.

[3]Ovu osobinu je F. Gaus 1795. godine nazvao metodom najmanjih kvadrata.

Tako, u primeru 1. je $M_o = 41$, jer je najveći broj ljudi (njih 13) kupilo cipele veličine 41. U primeru 2. je $M_o = 7$ ispita jer je najveći broj studenata (njih 9) položilo 7 ispita.

Ako se radi o distribuciji frekvencija sa intervalima, prvo se odredi modalni interval (onaj čija je frekvencija najveća). Onda je

$$M_o = L_{mo} + \frac{f_{mo} - f_{mo-1}}{f_{mo} - f_{mo-1} + f_{mo} - f_{mo+1}} \cdot \Delta$$

L_{mo} je početak modalnog intervala; f_{mo} je frekvencija modalnog intervala; f_{mo-1} je frekvencija intervala pre modalnog; f_{mo+1} je frekvencija intervala posle modalnog ; Δ je dužina intervala.

Primer 7. Za podatke iz primera 3. izračunati mod.

težina st. x_i	br. st. f_i
[50-60)	2
[60-70)	14
[70-80)	13
[80-90)	9
[90-100)	6
[100-110)	4
[110-120)	2
-	$\sum f_i = 50$

Rešenje. Modalni interval je $[60, 70)$ jer je njegova frekvencija najveća, iznosi 14.
Zatim, $L_{mo} = 60$, $f_{mo} = 14$, $f_{mo-1} = 2$, $f_{mo+1} = 13$, $\Delta = 10$.

$$M_o = L_{mo} + \frac{f_{mo} - f_{mo-1}}{f_{mo} - f_{mo-1} + f_{mo} - f_{mo+1}} \cdot \Delta = 60 + \frac{14-2}{14-2+14-13} \cdot 10 = 60 + 9,23 = 69,23kg$$

▷ **Napomena.** Može se desiti da je modalni interval prvi i kako nema intervala pre njega $f_{mo-1} = 0$. Ako se desi da je modalni interval poslednji, kako nema intervala posle njega $f_{mo+1} = 0$.

▷ **Medijana**[4] u oznaci M_e, je poziciona srednja vrednost i u pravom smislu središte grupisanja vrednosti obeležja. To je vrednost obeležja koja deli varijacioni niz na dva dela od kojih jedan sadrži 50% vrednosti obeležja koje su manje od medijane, a drugi 50% vrednosti obeležja koje su veće od medijane.

Kod osnovne distribucije frekvencija, ako je $\frac{N+1}{2}$ ceo broj, medijana je $M_e = x_{\frac{N+1}{2}}$.

Ako $\frac{N+1}{2}$ nije ceo broj, onda je $M_e = \dfrac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$.

Ako se radi o distribuciji frekvencija sa intervalima, onda se prvo odredi medijantni interval, a to je prvi interval za koji je zbirna učestanost veća ili jednaka $\dfrac{N}{2}$. Zatim je

[4]medijanta-središte

$$M_e = L_{me} + \frac{\frac{N}{2} - \sum\limits_{i=1}^{K_{me-1}} f_i}{f_{me}} \cdot \Delta$$

gde je L_{me} početak medijalnog intervala, a f_{me} frekvencija medijalnog intervala. Osim toga, $\sum\limits_{i=1}^{K_{me-1}} f_i$ je suma svih frekvencija do medijantnog intervala.

Primer 8. Pet novorođenih beba ima težine $4,2 \quad 3,8 \quad 4,1 \quad 5,1 \quad 4,0kg.$ Izračunati medijanu ove populacije.

Rešenje. Varijacioni niz je $3,8 \quad 4,0 \quad 4,1 \quad 4,2 \quad 5,1kg.$
Kako je $\frac{N+1}{2} = \frac{5+1}{2} = 3$ ceo broj, medijana je treća vrednost obeležja u varijacionom nizu tj. $M_e = x_3 = 4,1kg$. Dve bebe su lakše, a dve su teže od $4,1kg$.

Primer 9. Šest novorođenih beba ima težine $4,2 \quad 3,8 \quad 4,1 \quad 5,1 \quad 4,0 \quad 3,5kg.$ Izračunati medijanu ove populacije.

Rešenje. Varijacioni niz je $3,5 \quad 3,8 \quad 4,0 \quad 4,1 \quad 4,2 \quad 5,1kg.$
Kako $\frac{N+1}{2} = \frac{6+1}{2} = 3,5$ nije ceo broj, medijana je
$M_e = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2} = \frac{x_3 + x_4}{2} = \frac{4,0+4,1}{2} = 4,05kg.$

Primer 10. Izračunati medijanu za podatke iz primera 2. odnosno za sledeću distribuciju frekvencija:

br. pol. ispita x_i	br. studenata f_i	$\mathcal{Z}_i$
3	2	2
4	5	7
5	3	10
6	8	18
7	9	27
8	7	34
9	6	40
10	3	43
11	4	47
12	3	50
-	$\sum f_i = 50$	-

Rešenje. Kako $\frac{N+1}{2} = \frac{50+1}{2} = 25,5$ nije ceo broj, medijana je
$M_e = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2} = \frac{x_{25} + x_{26}}{2} = \frac{7+7}{2} = 7.$
Iz kolone za zbirne frekvencije $\mathcal{Z}_i$ sledi da je $x_{19} = x_{20} = \ldots = x_{25} = x_{26} = x_{27} = 7.$

Primer 11. Izračunati medijanu za podatke iz primera 3. odnosno za sledeću raspodelu obeležja:

težina studenta x_i	br. studenata f_i	$\mathcal{Z}_i$
[50-60)	2	2
[60-70)	14	16
[70-80)	13	29
[80-90)	9	38
[90-100)	6	44
[100-110)	4	48
[110-120)	2	50
-	$\sum f_i = 50$	-

Rešenje. Kako je $\frac{N}{2} = 25$ iz kolone za $\mathcal{Z}_i$ sledi da je medijantni interval $[70 - 80)$. Nadalje, $L_{me} = 70$, $f_{me} = 13$, $\Delta = 10$, pa je

$$M_e = L_{me} + \frac{\frac{50}{2} - \sum_{i=1}^{2} f_i}{f_{me}} \cdot \Delta = 70 + \frac{25-16}{13} \cdot 10 = 70 + 6,923 = 76,92kg$$

1.4 PARAMETRI VARIJABILITETA[5] (mere disperzije[6])

Dati su varijacioni nizovi:

niz 1.	10	10	10	10	10	10	10,
niz 2.	1	5	9	10	11	15	19,
niz 3.	1	2	3	10	11	13	30,
niz 4.	1	9	9	10	11	12	18.

U sva četiri slučaja je aritmetička sredina ista $\mu = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{70}{7} = 10.$ Pa ipak to su sasvim različite raspodele obeležja tj. različite distribucije frekvencija. Različita je raspršenost (disperzija) vrednosti numeričkog obeležja. Parametri varijabiliteta omogućavaju da se izrazi ta različita raspršenost.

▷ **Interval varijacije** u oznaci Iv je jednostavna mera disperzije. To je razlika između najveće i najmanje vrednosti obeležja

$$I_v = x_{max} - x_{min}.$$

[5] varijabilnost - promenljivost, nepostojanost

[6] disperzija - raspršenost, rasipanje, rasturanje, razbacivanje

▷**Interkvartila razlika** u oznaci I_q isključuje 25% najmanjih i 25% najvećih vrednosti obeležja i

$$I_q = Q_3 - Q_1 = P_{75} - P_{25},$$

gde su Q_i $(i = 1, 2, 3, 4)$ **kvartili**, a P_i $(i = 1, 2, \ldots, 100)$ **percentili**.
Ako se radi o osnovnoj distribuciji frekvencija, ako je broj $i \cdot \dfrac{N-1}{4} + 1 = B$ ceo broj, onda je $i-$ ti kvartil B-ta vrednost obeležja tj. $Q_i = x_B, \quad i = 1, 2, 3, 4.$
Ako $i \cdot \dfrac{N-1}{4} + 1 = B$ nije ceo broj, uzme se celi deo tog broja tj. $\left[i \cdot \dfrac{N-1}{4} + 1\right] = [B]$ pa je

$$Q_i = x_{[B]} + (B - [B]) \cdot \left(x_{[B]+1} - x_{[B]}\right), \qquad i = 1, 2, 3, 4.$$

Slično, ako je broj $i \cdot \dfrac{N-1}{100} + 1 = B$ ceo broj, onda je $i-$ ti percentil B-ta vrednost obeležja tj. $P_i = x_B$, $i = 1, 2, \ldots, 100.$
Ako $i \cdot \dfrac{N-1}{100} + 1 = B$ nije ceo broj, uzme se celi deo tog broja tj. $\left[i \cdot \dfrac{N-1}{100} + 1\right] = [B]$ pa je

$$P_i = x_{[B]} + (B - [B]) \cdot \left(x_{[B]+1} - x_{[B]}\right), \qquad i = 1, 2, \ldots, 100.$$

Primer 12. Za podatke iz primera 2. izračunati interkvartilnu razliku i trideset peti percentil.

br. pol. ispita x_i	br. studenata f_i	$\mathcal{Z}_i$
3	2	2
4	5	7
5	3	10
6	8	18
7	9	27
8	7	34
9	6	40
10	3	43
11	4	47
12	3	50
-	$\sum f_i = 50$	-

Rešenje. Za Q_1 je $B = 1 \cdot \dfrac{N-1}{4} + 1 = \dfrac{49}{4} + 1 = 13,25$ i to nije ceo broj, sledi $[B] = 13$ pa je $Q_1 = x_{13} + (13,25 - 13) \cdot (x_{14} - x_{13}) = 6 + 0,25 \cdot (6 - 6) = 6$ ispita.
(Kolona sa zbirnim frekvencijama $\mathcal{Z}_i$ pomaže da se utvrde vrednosti za x_{13} i x_{14}.)
Slično, za Q_3 je $B = 3 \cdot \dfrac{N-1}{4} + 1 = 3 \cdot \dfrac{49}{4} + 1 = 37,75$ pa je $[B] = 37$ i sledi
$Q_3 = x_{37} + (37,75 - 37) \cdot (x_{38} - x_{37}) = 9 + 0,75 \cdot (9 - 9) = 9$ ispita.
Konačno $I_q = Q_3 - Q_1 = 9 - 6 = 3$ položena ispita.
Za trideset peti percentil $B = i \cdot \dfrac{N-1}{100} + 1 = 35 \cdot \dfrac{49}{100} + 1 = 18,15$ i $[B] = 18$ pa je

$P_{35} = x_{18} + (18,15 - 18) \cdot (x_{19} - x_{18}) = 6 + 0,15 \cdot (7 - 6) = 6,15$ ispita.

Ako se radi o distribuciji frekvencija sa intervalima, broj $i \cdot \frac{N}{4}$ među kumulativnim frekvencijama određuje kvartilni interval tj. interval $i-$tog kvartila. Neka broj $K_{kv} \in \mathcal{N}$ označava koji je po redu kvartilni interval. Onda je

$$Q_i = L_{kv} + \frac{i \cdot \frac{N}{4} - \sum_{i=1}^{K_{kv}-1} f_i}{f_{kv}} \cdot \Delta$$

gde je L_{kv} početak kvartilnog intrvala, a f_{kv} frekvencija kvartilnog intrvala.

Za percentile broj $i \cdot \frac{N}{100}$ među kumulativnim frekvencijama određuje interval $i-$tog percentila. Neka broj $K_p \in \mathcal{N}$ označava koji je po redu interval $i-$tog percentila. Onda je

$$P_i = L_p + \frac{i \cdot \frac{N}{100} - \sum_{i=1}^{K_p-1} f_i}{f_p} \cdot \Delta$$

gde je L_p je početak odgovarajućeg intervala, a f_p frekvencija tog intervala.

Primer 13. Za podatke iz primera 3. izračunati interkvartilnu razliku i trideset peti percentil.

težina st. x_i	br. st. f_i	$\mathcal{Z}_i$
[50-60)	2	2
[60-70)	14	16
[70-80)	13	29
[80-90)	9	38
[90-100)	6	44
[100-110)	4	48
[110-120)	2	50
-	$\sum f_i = 50$	-

Rešenje. Kako je $1 \cdot \frac{50}{4} = 12,5$ to je interval prvog kvartila drugi po redu $[60 - 70)$.
(To se lako vidi iz kolone za zbirne frekvencije $\mathcal{Z}i$.)
Zatim, $L_{kv} = 60$, $f_{kv} = 14$, $K_{kv} - 1 = 1$, $\Delta = 10$.

$$Q_1 = L_{kv} + \frac{1 \cdot \frac{N}{4} - \sum_{i=1}^{1} f_i}{f_{kv}} \cdot \Delta = L_{kv} + \frac{1 \cdot \frac{50}{4} - f_1}{f_{kv}} \cdot \Delta = 60 + \frac{12,5 - 2}{14} \cdot 10 = 67,5kg$$

Kako je $3 \cdot \frac{N}{4} = 37,5$ to je interval trećeg kvartila četvrti po redu $[80 - 90)$.
(To se lako vidi iz kolone za zbirne frekvencije $\mathcal{Z}_i$.)
Zatim, $L_{kv} = 80$, $f_{kv} = 9$, $K_{kv} - 1 = 3$, $\Delta = 10$.

$$Q_3 = L_{kv} + \frac{3 \cdot \frac{N}{4} - \sum\limits_{i=1}^{3} f_i}{f_{kv}} \cdot \Delta = 80 + \frac{37,5 - (f_1 + f_2 + f_3)}{f_{kv}} \cdot 10 = 80 + \frac{37,5 - 29}{9} \cdot 10 = 89,44kg$$

Konačno $\quad I_q = Q_3 - Q_1 = 89,44 - 67,5 = 21,94kg$

Iz $i \cdot \frac{N}{100} = 35 \cdot \frac{50}{100} = 17,5$ sledi (iz kolone za $\mathcal{Z}_i$) da je interval koji odgovara trideset petom percentilu $[70, 80)$. Znači $L_p = 70,\ f_p = 13,\ K_p - 1 = 2$ i

$$P_{35} = L_p + \frac{i \cdot \frac{N}{100} - \sum\limits_{i=1}^{K_p-1} f_i}{f_p} \cdot \Delta = 70 + \frac{17,5 - (f_1 + f_2)}{f_p} \cdot 10 = 70 + \frac{17,5 - 16}{13} \cdot 10 = 71,15kg.$$

▷ **Srednje apsolutno odstupanje.** Srednje apsolutno odstupanje od aritmetičke sredine u oznaci $AD(\mu)$ računa se po jednoj od formula

$$AD(\mu) = \frac{\sum\limits_{i=1}^{k} f_i \cdot |x_i - \mu|}{N} \quad \text{ili} \quad AD(\mu) = \frac{\sum\limits_{i=1}^{K} f_i \cdot |x'_i - \mu|}{N}.$$

Analogno se računa srednje apsolutno odstupanje od modusa M_o ili od medijane M_e.

Upotreba ovog parametra je retka. Kako se vrednosti obeležja rasturaju oko proseka, najbolje pokazuju disperzija i standardna varijacija.

▷ **Disperzija** (ili varijansa) populacije u oznaci σ^2 je prosečno kvadratno odstupanje od aritmetičke sredine

$$\sigma^2 = \frac{\sum\limits_{i=1}^{k} f_i \cdot (x_i - \mu)^2}{N} \quad \text{odnosno,} \quad \sigma^2 = \frac{\sum\limits_{i=1}^{K} f_i \cdot (x'_i - \mu)^2}{N}.$$

Radi lakšeg računanja disperzije treba imati u vidu sledeće:

$$\sigma^2 = \frac{\sum\limits_{i=1}^{k} f_i \cdot (x_i - \mu)^2}{N} = \frac{\sum\limits_{i=1}^{k} f_i \cdot (x_i^2 - 2x_i\mu + \mu^2)}{N} = \frac{\sum\limits_{i=1}^{k} f_i \cdot x_i^2}{N} - 2\mu \cdot \frac{\sum\limits_{i=1}^{k} f_i \cdot x_i}{N} + \frac{\sum\limits_{i=1}^{k} f_i \cdot \mu^2}{N} =$$

$$= \frac{\sum\limits_{i=1}^{k} f_i \cdot x_i^2}{N} - 2\mu \cdot \mu + \mu^2 \cdot \frac{\sum\limits_{i=1}^{k} f_i}{N} = \frac{\sum\limits_{i=1}^{k} f_i \cdot x_i^2}{N} - 2\mu^2 + \mu^2 \cdot 1 = \frac{\sum\limits_{i=1}^{k} f_i \cdot x_i^2}{N} - \mu^2.$$

Znači, $\quad \sigma^2 = \frac{\sum\limits_{i=1}^{k} f_i \cdot (x_i - \mu)^2}{N} = \frac{\sum\limits_{i=1}^{k} f_i \cdot x_i^2}{N} - \mu^2.$

Analogno se pokazuje da je $\quad \sigma^2 = \frac{\sum\limits_{i=1}^{K} f_i \cdot (x'_i - \mu)^2}{N} = \frac{\sum\limits_{i=1}^{K} f_i \cdot (x'_i)^2}{N} - \mu^2.$

▷ **Standardna devijacija** (ili standardno odstupanje). Disperzija je izražena u istim jedinicama u kojima je dato numeričko obeležje , ali na drugom stepenu. Zato se računa drugi koren iz disperzije i dobija se novi parametar poznat kao standardna devijacija (u oznaci σ). Znači

$$\sigma = \sqrt{\sigma^2}.$$

Standardna devijacija je izražena u istim jedinicama u kojima je dato numeričko obeležje. Aritmetička sredina i standardna devijacija pokazuju kakav je raspored vrednosti obeležja u populaciji. Ukoliko je vrednost standardne devijacije manja, to je sabijenost vrednosti obeležja oko aritmetičke sredine veća pa je i njena reprezentativnost veća i obrnuto.

Primer 14. Za podatke iz primera 2. izračunati disperziju i standardnu devijaciju.

Rešenje. Radna tabela za izračunavanje disperzije i standardne devijacije:

x_i	f_i	x_i^2	$f_i \cdot x_i^2$
3	2	9	18
4	5	16	80
5	3	25	75
6	8	36	288
7	9	49	441
8	7	64	448
9	6	81	486
10	3	100	300
11	4	121	484
12	3	144	432
-	$\sum f_i = 50$	-	$\sum f_i \cdot x_i^2 = 3052$

Disperzija je $\sigma^2 = \dfrac{\sum_{i=1}^{K} f_i \cdot (x_i)^2}{N} - \mu^2 = \dfrac{3052}{50} - (7,44)^2 = 5,69.$

Standardna devijacija je $\sigma = \sqrt{\sigma^2} = \sqrt{5,69} = 2,385$ položenih ispita.

Primer 15. Za podatke iz primera 3. izračunati disperziju i standardnu devijaciju.

Rešenje. Radna tabela za izračunavanje disperzije i standardne devijacije:

x_i	f_i	x_i'	$(x_i')^2$	$f_i \cdot (x_i')^2$
[50-60)	2	55	3025	6050
[60-70)	14	65	4225	59150
[70-80)	13	75	5625	73125
[80-90)	9	85	7225	65025
[90-100)	6	95	9025	54150
[100-110)	4	105	11025	44100
[110-120)	2	115	13225	26450
-	$\sum f_i = 50$	-	-	$\sum f_i \cdot (x_i')^2 = 328050$

Disperzija je $\sigma^2 = \dfrac{\sum\limits_{i=1}^{7} f_i \cdot (x_i')^2}{N} - \mu^2 = \dfrac{328050}{50} - (79,6)^2 = 224,84kg^2.$

Standardna devijacija je $\sigma = \sqrt{\sigma^2} = \sqrt{224,84} = 14,99kg$

▷ **Napomena.** Kada se mnogo vrednosti obeležja iz jednog intervala zamenjuju sredinom intervala disperzija se može korigovati. Korigovana disperzija je $\sigma^2 - \dfrac{\Delta^2}{12}.$

Zadatak. Anketirane su dve grupe od po 10 studenata o broju položenih ispita. Dobijeni su sledeći podaci:

$$A: \quad 1,2,3,3,5,5,7,7,8,9 \qquad B: \quad 2,4,4,4,5,5,6,6,6,8.$$

Koja grupa studenata ima bolji uspeh?

Rezultat. U obe grupe je prosečan broj položenih ispita 5. Grupa B ima bolji uspeh jer je za tu grupu standardna devijacija manja.

1.5 KOEFICIJENT VARIJACIJE

Za razliku od aritmetičke sredine i standardne devijacije koje se izražavaju u istim jedinicama kao i obeležje, koeficijent varijacije je neimenovani broj. Omogućava da se poredi rasturanje vrednosti dva različita obeležja u različitim populacijama. Označava se sa C_v i iznosi

$$C_v = \frac{\sigma}{\mu}.$$

Može da se računa i u procentima i onda je

$$C_v\% = C_v \cdot 100\%.$$

Ako je $C_v\% < 30\%$, smatra se da je statistički skup homogen i aritmetička sredina je reprezentativna centralna vrednost.

Primer 16. Jedna grupa studenata ima prosečnu težinu $80kg$ sa disperzijom $49kg^2$, a prosečnu visinu $185cm$ sa standardnom devijacijom $12cm$. Da li je veće variranje težine ili visine studenata?

Rešenje.

Za težinu je $C_v = \dfrac{\sigma}{\mu} = \dfrac{\sqrt{49}}{80} = \dfrac{7}{80} = 0,0875$ odnosno $C_v\% = 8,75\%.$

Za visinu je $C_v = \dfrac{\sigma}{\mu} = \dfrac{12}{185} = 0,0649$ odnosno $C_v\% = 6,49\%.$

Više varira težina studenata nego visina.

Primer 17. Mesečni prosek broja zaposlenih u jednom preduzeću iznosio je 2002. godine 15000 zaposlenih sa standardnom devijacijom $\sigma = 420$. U 2003. godini prosek je iznosio 12000 zaposlenih sa standardnom devijacijom $\sigma = 350$. Da li je stabilnost zaposlenih bila veća u 2002. ili u 2003. godini?

Rešenje.
Za 2002 godinu je $\quad C_v\% = \dfrac{\sigma}{\mu} \cdot 100\% = \dfrac{420}{15000} \cdot 100\% = 2,8\%.$
Za 2003 godinu je $\quad C_v\% = \dfrac{\sigma}{\mu} \cdot 100\% = \dfrac{350}{12000} \cdot 100\% = 2,92\%.$
Stabilnost broja zaposlenih je bila veća u 2002. godini (variranje je bilo manje).

1.6 NORMALIZOVANO STANDARDNO ODSTUPANJE

Normalizovano standardno odstupanje, u oznaci Z_i, ukazuje na odstupanje $i-$te vrednosti obeležja ($i-$tog člana populacije) od aritmetičke sredine. Ima se u vidu stepen stabilnosti (homogenosti) vrednosti obeležja u toj populaciji i

$$Z_i = \frac{x_i - \mu}{\sigma}.$$

Primer 18. Jedan student iz populacije opisane u primeru 3. je težak $75kg$, a drugi $84kg$. Koji student bolje reprezentuje ovu populaciju svojom težinom?

Rešenje. Iz primera 7. je prosečna težina studenata $\mu = 79,6kg$ sa standardnom devijacijom $\sigma = 14,99kg$ (iz primera 15). Sledi
$Z_1 = \dfrac{x_1 - \mu}{\sigma} = \dfrac{75 - 79,6}{14,99} = -0,307$
$Z_2 = \dfrac{x_2 - \mu}{\sigma} = \dfrac{84 - 79,6}{14,99} = 0,2935.$
Drugi student bolje reprezentuje ovu populaciju. On je malo iznad prosečne težine, a drugi student je (nešto više) ispod prosečne težine.

Primer 19. Jedan student iz populacije opisane u primerima 2. i 3. je položio 7 ispita i težak je $75kg$. Da li ovaj student bolje reprezentuje tu populaciju svojom težinom ili brojem položenih ispita?

Rešenje. Iz prethodnog primera $Z_1 = -0,307$.
Prosečan broj položenih ispita $\mu = 7,44$ (primer 5.), sa standardnom devijacijom $\sigma = 2,38$ (primer 15). Sledi
$Z_1' = \dfrac{x_1 - \mu}{\sigma} = \dfrac{7 - 7,44}{2,38} = -0,185.$
Ovaj student bolje predstavlja ovu populaciju brojem položenih ispita odnosno, po broju položenih ispita je bliži proseku.

Zadatak 1. Pri sistematskom pregledu jednog odeljenja srednje škole dobijeni su sledeći rezultati o telesnoj masi učenika:

73,4	63,7	74,3	69,2	77,0	76,3	75,0	75,0	76,3	80,6	79,2
78,6	84,0	84,0	86,2	86,2	80,2	81,3	82,8	82,3	81,3	81,0
82,3	80,2	82,3	85,4	83,4	87,3	83,0	87,3	83,1	83,0.	

Izračunati aritmetičku sredinu, mod, medijanu, interkvartilnu razliku, disperziju i standardnu devijaciju. Zadatak uraditi na dva načina, sa i bez grupisanja vrednosti obeležja u intervale.

Sve navedeno u ovoj glavi odnosi se na konačne populacije. Ako je populacija mnogo velika sakupljanje podataka može biti skupo i neekonomično. Ako je populacija beskonačna (prebrojivo ili neprebrojivo) tada se ne može dati tačna raspodela obeležja. Onda se (u opštem slučaju) ne može govoriti o broju elemenata populacije za koji obeležje ima neku vrednost, već se govori o delovima ili procentima populacije sa tom vrednošću ili sa vrednošću iz nekog skupa (intervala). Tada se raspodela obeležja procenjuje na osnovu raznih merenja (uzorka) ili se izvodi iz teorijskih pretpostavki. **TEORIJA VEROVATNOĆE** *je druga matematička disciplina koja je teorijska podloga Matematičke statistike.*

2 OSNOVNI ELEMENTI TEORIJE VEROVATNOĆE

Eksperiment koji se izvodi u praksi može biti **deterministički** i **slučajni**. Kod determinističkog eksperimenta u svakom ponavljanju eksperimenta pri istim uslovima dobija se uvek isti rezultat. Rezultat eksperimenta se naziva i **ishodom**. Dakle, kod determinističkog eksperimenta se unapred zna ishod. Kod slučajnog eksperimenta se ne može predvideti ishod. Primer slučajnog eksperimenta je bacanje kockice za igru. Ona na svakoj strani ima od jedne do šest tačkica. Registruje se broj tačkica na gornjoj strani kocke. Iako se zna da je ishod jedan od brojeva $1, 2, 3, 4, 5, 6$, ne može se tvrditi koji će se broj registrovati u sledećem bacanju. Eksperiment koji se sastoji u bacanju jednog novčića, pri čemu se registruje gornja strana novčića, ima samo dva moguća ishoda: ,,pismo" i ,,glava". Ipak, ishod svakog pojedinog bacanja je nepredvidiv. Ovo su tipični primeri statističkih eksperimenata.

Statistički eksperiment je onaj koji zadovoljava sledeće uslove:
1) može se ponavljati proizvoljan broj puta pod istim uslovima,
2) unapred je definisano šta se registruje u eksperimentu i poznati su svi mogući ishodi,
3) ishod svakog pojedinačnog eksperimenta nije unapred poznat.

Statistički eksperiment nije samo ,,laboratorijski", tj. namerno izazvan eksperiment. Mnoga posmatranja prirodnih i društvenih pojava imaju osobine statističkog eksperimenta. Neka je Ω skup svih logički mogućih ishoda posmatrane pojave, odnosno skup svih rezultata eksperimenta. Ovaj skup može biti konačan, prebrojivo ili neprebrojivo beskonačan. Ako je eksperiment deterministički, Ω ima samo jedan element, a ako je slučajni (nedeterministički), onda Ω ima bar dva elementa.

Element ovog skupa tj. $\omega \in \Omega$ je **elementarni događaj** ili **ishod**, a svaki podskup od Ω je **slučajni događaj.** Slučajni događaji se obeležavaju velikim slovima latinice $A, B, C, \ldots$ Kako je $\{\omega\} \subset \Omega$, to je i svaki ishod slučajni događaj.

Kaže se da se slučajni događaj $A \subseteq \Omega$ **realizovao** (**ostvario**) ako se realizovao bilo koji od elementarnih događaja (ishoda) $\omega \in A$. Zato je $\omega \in A$ **povoljan ishod** za događaj A.

Kako se u teoriji verovatnoće razmatraju samo slučajni događaji, to se reč ,,slučajni" najčešće izostavlja.

Primer 1. Eksperiment je bacanje kocke za igru i registruje se broj (broj tačkica) na gornjoj strani.

Skup svih ishoda je $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Kako je $\{4, 5, 6\} \subset \Omega$, to je $A = \{4, 5, 6\}$ jedan slučajni događaj.
Ako posle bacanja kocke na gornjoj strani bude pet tačkica, tj. ako se registruje broj 5, događaj A se realizovao, a ako se registruje broj 2, događaj A se nije realizovao. Povoljni ishodi za događaj A su $\{4\}$, $\{5\}$, i $\{6\}$. Događaj A se može opisati kao događaj da se registruje broj veći od tri. Kraće se piše A: ,,da je broj veći od tri".

Kako je $\{2, 4, 6\} \subset \Omega$, to je $B = \{2, 4, 6\}$ jedan slučajni događaj. Događaj B se može opisati sa B: ,,da je broj paran"

Događaj C : ,, da je broj neparan" je skup $\{1,3,5\} \subset \Omega$.

Događaj D : ,, da je broj deljiv sa tri" je $D = \{3,6\}$.

Događaj E : ,, da je broj deljiv sa pet" je $E = \{5\}$.
Može se primetiti da je slučajni događaj E zapravo elementarni događaj ili ishod.

Primer 2. Eksperiment se sastoji u bacanju jednog novčića četiri puta. Registruje se koliko je ukupno puta palo ,,pismo". Skup svih ishoda je $\Omega = \{0,1,2,3,4\}$.

Događaj A: ,, da padne pismo više od tri puta" je $A = \{4\}$.

Događaj B: ,, da padne pismo bar dva puta" je $B = \{2,3,4\}$.

Primer 3. Eksperiment se sastoji u bacanju jednog novčića četiri puta. Registruje se niz ,,pisama" (p) i ,,glava" (g). Skup svih ishoda je

$$\Omega = \{\, pppp,\, pppg,\, ppgp,\, \ldots,\, gggg\}.$$

To su svi nizovi dužine četiri od slova p i g. Ima ih $2^4 = 16$.

Događaj A: ,, da padne glava više nego pisama" je
$A = \{gggg,\ gggp,\ ggpg,\ gpgg,\ pggg\}$.

Primer 4. Eksperiment je partija šaha. Ako se registruje rezultat, onda je

$$\Omega = \{\text{pobeda belog}, \quad \text{pobeda crnog}, \quad \text{remi}\}.$$

Ako se registruje broj poteza, onda je $\Omega = \{1,2,3,\ldots\}$. Ovo je prebrojivo beskonačan skup ishoda.

Primer 5. Eksperiment je merenje vlažnosti vazduha na određenom mestu u određeno vreme. Registruje se vlažnost vazduha u procentima pa je
$\Omega = \{v \in \mathcal{R} \,|\, 0 \leq v \leq 100\}$, tj. $\Omega = [0,100]$. Skup ishoda Ω je neprebrojiv.

2.1 ALGEBRA DOGAĐAJA

Pošto je $\Omega \subseteq \Omega$, sledi da je Ω slučajan događaj i naziva se **siguran događaj** jer se mora realizovati pri vršenju eksperimenta. Iz $\emptyset \subseteq \Omega$ sledi da je i prazan skup slučajni događaj. To je **nemoguć događaj** i on se nikad ne ostvaruje.

Za događaj A se kaže da **povlači** (implicira) događaj B ako je $A \subseteq B$ odnosno, kad god se realizuje događaj A, realizuje se i događaj B.

Za događaje A i B skup $A \cap B$ je **presek događaja** A i B. To je novi događaj koji se realizuje ako i samo ako se realizuje događaj A i realizuje događaj B. Odnosno, događaj $A \cap B$ se realizuje ako i samo ako se realizuju oba događaja A i B istovremeno. Češće se umesto termina presek događaja koristi termin **proizvod događaja** i tada se koristi oznaka AB.

Ako je presek događaja prazan skup, tj. ako je $AB = \emptyset$, kaže se da su događaji A i B disjunktni ili da se uzajamno isključuju.

Tri događaja A, B i C se uzajamno isključuju ako je $AB = \emptyset$, $AC = \emptyset$ i $BC = \emptyset$.

Događaji $A_1, A_2, \ldots, A_n \ldots \quad n \in N$ se uzajamno isključuju (međusobno su disjunktni) ako je $A_iA_j = \emptyset$ za $i \neq j$, $\quad i, j = 1, 2, 3, \ldots$

Za događaje A i B skup $A \cup B$ je **unija događaja** A i B. To je događaj koji se realizuje ako i samo ako se realizuje događaj A ili se realizuje događaj B. Odnosno, događaj $A \cup B$ se realizuje ako i samo ako se realizuje bar jedan od događaja A i B (mogu i oba). U slučaju da je $AB = \emptyset$, unija događaja se obeležava sa $A + B$ i koristi se termin **zbir događaja**.

Napomena. Uobičajen je sledeći zapis

$$A_1 + A_2 + \ldots + A_n = \sum_{i=1}^{n} A_i \quad \text{odnosno,} \quad A_1 + A_2 + \ldots + A_n + \ldots = \sum_{i=1}^{+\infty} A_i \quad \text{za}$$

uniju konačno mnogo ili prebrojivo beskonačno događaja koji se uzajamno isključuju.

Za događaje A i B skup $A \backslash B$ je **razlika događaja** A i B. To je događaj koji se realizuje ako i samo ako se realizuje događaj A i ne realizuje događaj B.

Ako je A događaj, onda se komplement skupa A, tj. $\Omega \backslash A$, naziva **suprotan događaj**, a obeležava se sa $\overline{A}$. Događaj $\overline{A}$ se realizuje ako i samo ako se događaj A ne realizuje.

Kako su operacije unije, preseka, komplementa itd. uvedene nad skupom događaja na potpuno isti način kao što je to učinjeno u teoriji skupova, očigledno da sva pravila, odnosi i operacije koje važe u teoriji skupova važe i nad skupom događaja.

Primer 6. Neka su eksperiment i događaji A, B, C, D i E kao u primeru 1.

Događaj S: „broj je manji od 20" je siguran događaj jer je $S = \{1, 2, 3, 4, 5, 6\} = \Omega$.

Događaj N: „broj je deljiv sa 7" je nemoguć događaj, jer među brojevima od 1 do 6 nema broja koji je deljiv sa 7, te je $N = \emptyset$.

Događaj $E = \{5\}$ povlači događaj $A = \{4, 5, 6\}$ tj. $E \subseteq A$.

Događaj $C = \{1, 3, 5\}$ je suprotan događaju $B = \{2, 4, 6\}$ tj. $C = \overline{B}$.

Presek događaja A i B je $A \cap B = \{4, 5, 6\} \cap \{2, 4, 6\} = \{4, 6\}$.
Presek događaja A i D je $A \cap D = \{4, 5, 6\} \cap \{3, 6\} = \{6\}$,
a presek događaja D i E je događaj $D \cap E = \{3, 6\} \cap \{5\} = \emptyset$. Zaključuje se da se događaji D i E se uzajamno isključuju.

Unija događaja A i B je $A \cup B = \{4, 5, 6\} \cup \{2, 4, 6\} = \{2, 4, 5, 6\}$. Zbir događaja B i E je $B + E = \{2, 4, 6\} \cup \{5\} = \{2, 4, 5, 6\}$, dok je zbir događaja D i E događaj $D + E = \{3, 6\} \cup \{5\} = \{3, 5, 6\}$.

Razlike događaja A i D, odnosno A i B su redom $A \backslash D = \{4, 5, 6\} \backslash \{3, 6\} = \{4, 5\}$ i $A \backslash B = \{4, 5, 6\} \backslash \{2, 4, 6\} = \{5\}$.

2.2 VEROVATNOĆA I OSNOVNE OSOBINE

Statistička definicija verovatnoće. Jedna od osobina statističkog eksperimenta je da se on može ponavljati neograničen broj puta pod istim uslovima. Pretpostavimo da se eksperiment u kome se može realizovati događaj A ponavlja n puta.

Neka je f_A broj realizacija događaja A u jednoj seriji od n ponavljanja eksperimenta. Broj f_A je apsolutna frekvencija događaja A, a broj $\dfrac{f_A}{n}$ je relativna frekvencija događaja A u n ponavljanja eksperimenta. U drugoj seriji od n ponavljanja eksperimenta relativna frekvencija $\dfrac{f_A}{n}$ imaće drugu vrednost. Međutim, ako se u velikim serijama ponavljanja eksperimenta relativne frekvencije $\dfrac{f_A}{n}$ grupišu oko nekog broja (označićemo ga sa $P(A)$), onda se taj broj uzima za verovatnoću događaja A.

Aksiomatska (savremena) definicija verovatnoće.[1] Verovatnoća je funkcija P koja svakom događaju $A \subseteq \Omega$ pridružuje realan broj $P(A)$, sa sledećim osobinama:

1. verovatnoća bilo kog događaja je nenegativna, tj. $P(A) \geq 0$ za svako $A \subseteq \Omega$,
2. verovatnoća sigurnog događaja je 1, odnosno $P(\Omega) = 1$, (verovatnoća je normirana funkcija)
3. za međusobno disjunktne događaje $A_1, A_2, A_3 \dots$ važi $\sigma-$aditivnost

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Konkretno, treća osobina verovatnoće znači da za dva <u>disjunktna</u> događaja ($AB = \emptyset$), važi

$$P(A+B) = P(A) + P(B),$$

a za tri <u>međusobno disjunktna</u> događaja, ($AB = \emptyset$, $AC = \emptyset$, $BC = \emptyset$), važi

$$P(A+B+C) = P(A) + P(B) + P(C).$$

Iz ove definicije proizilaze i sledeće osobine verovatnoće:

1. verovatnoća nemogućeg događaja je 0, odnosno $P(\emptyset) = 0$,
2. verovatnoća suprotnog događaja je $P(\overline{A}) = 1 - P(A)$,
3. ako događaj A implicira događaj B $(A \subseteq B)$ tada je $P(A) \leq P(B)$,
4. za svaki događaj A važi, $0 \leq P(A) \leq 1$,
5. verovatnoća unije dva događaja je $P(A \cup B) = P(A) + P(B) - P(AB)$.

U praksi često eksperiment ima konačno mnogo ishoda koji su jednako verovatni, tj. skup svih ishoda je na primer $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$ i važi $P(\omega_1) = P(\omega_2) = \cdots = P(\omega_k) = \dfrac{1}{k}$ (kao što je , recimo, pri bacanju kocke za igru). Onda je od značaja sledeća definicija.

[1]definicija je prilagođena čitaocima kojima je namenjen udžbenik

Klasična definicija verovatnoće. Neka je Ω konačan skup ishoda koji su jednako verovatni. Verovatnoća događaja A jednaka je odnosu broja povoljnih ishoda m prema broju svih mogućih ishoda k, tj.

$$P(A) = \frac{m}{k}.$$

Primer 7. Neka su događaji A, B, C, D i E kao u primeru 1. Odrediti verovatnoće ovih događaja.

Rešenje. Ukupan broj ishoda je 6. Svi su ishodi jednako verovatni. Kako je $A = \{4, 5, 6\}$, broj povoljnih ishoda za događaj A je 3. Znači, $P(A) = \frac{3}{6} = \frac{1}{2}$.
Slično, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{2}$, $P(D) = \frac{1}{3}$, $P(E) = \frac{1}{6}$.

Primer 8. Jedan novčić se baca dva puta. Registruje se gornja strana novčića. Izračunati verovatnoće događaja A : ,,da tačno jednom padne pismo”, događaja B : ,,da bar jednom padne pismo” i događaja C : ,,da je u prvom bacanju pismo, a u drugom glava”.

Rešenje. Ukupan broj ishoda je 4, jer je $\Omega = \{pp, pg, gp, gg\}$. Kako je $A = \{pg, gp\}$, sledi $P(A) = \frac{2}{4} = \frac{1}{2}$. Iz $B = \{pp, pg, gp\}$, sledi $P(B) = \frac{3}{4}$. Pošto je $C = \{pg\}$, zaključuje se $P(C) = \frac{1}{4}$.

2.3 USLOVNA VEROVATNOĆA. NEZAVISNOST

Često je prilika da se traži verovatnoća nekog događaja B, a da se ima informacija da se određeni događaj A realizovao. Neka su eksperiment i događaj C kao u primeru 8. Recimo da se ima informacija da je u oba bacanja novčić pao na istu stranu. Posle te informacije skup svih ishoda je $\{pp, gg\}$ i verovatnoća događaja C je $P(C) = 0$. Pre dodatne informacije ta verovatnoća je bila $\frac{1}{4}$.
Uslovna verovatnoća događaja B, pri uslovu da se realizovao događaj A sa $P(A) > 0$, u oznaci $P(B|A)$, jednaka je

$$P(B|A) = \frac{P(AB)}{P(A)}.$$

Odavde sledi da je verovatnoća proizvoda dva događaja jednaka

$$P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B).$$

Ako verovatnoća realizacije događaja B ne zavisi od realizacije događaja A sa $P(A) > 0$, tada mora biti

$$P(B|A) = P(B).$$

Zato se za dva događaja A i B kaže da su **nezavisni** (stohastički nezavisni) ako je

$$P(AB) = P(A) \cdot P(B).$$

Ukoliko ovaj uslov nije zadovoljen, kaže se da su događaji A i B **zavisni**.

Primer 9. Iz špila (52 karte) se slučajno izvlači jedna karta. Da li su događaji A :,, izvučena je dama" i B :,, izvučen je pik" nezavisni?

Rešenje. Događaj A ima četiri povoljna ishoda: ,, izvučena je dama pik", ,, izvučena je dama herc", ,, izvučena je dama tref" i ,, izvučena je dama karo". Sledi,

$$P(A) = \frac{4}{52} = \frac{1}{13}.$$

Događaj B ima 13 povoljnih ishoda: ,, izvučena je jedinica pik", ,, izvučena je dvojka pik", . . . ,, izvučen je kralj pik". Zato je

$$P(B) = \frac{13}{52} = \frac{1}{4}.$$

Događaj AB ima samo jedan povoljan ishod ,,izvučena je dama pik" pa je

$$P(AB) = \frac{1}{52}.$$

Iz $\quad P(A) \cdot P(B) = \frac{1}{13} \cdot \frac{1}{4} = \frac{1}{52} = P(AB) \quad$ sledi da su ovi događaji nezavisni.

Zadatak 1. U jednom preduzeću je 200 zaposlenih svrstano po polu i godinama starosti sledećom tabelom

starost	muški	ženski	ukupno
[0, 30)	20	30	50
[30, 50)	70	60	130
[50, 65]	12	8	20
			200

Kolika je verovatnoća da je slučajno izabrani zaposleni

a) od 30 do 50 godina starosti

b) muškog pola

c) ženskog pola ili ispod 30 godina starosti

d) ženskog pola ako se zna da je ispod 30 godina starosti.

Rezultat. a) 0,65 b) 0,51 c) 0,59 d) 0,6.

Zadatak 2. Iz skupa $S = \{1, 2, \ldots, 20\}$ je slučajno izabran jedan broj. Ako je poznato da je izabran broj deljiv sa 3, kolika je verovatnoća da je izabran paran broj?

Rezultat. 0,5.

Zadatak 3. Broj prodatih jakni, u jednoj prodavnici, svrstan je po boji i proizvođaču

proizvođač	boja	br. jakni
najk	teget	34
najk	crna	10
najk	braon	4
adidas	teget	26
adidas	crna	4
adidas	braon	2
		80

Naći verovatnoću da će sledeća kupljena jakna biti

a) marke najk,
b) braon, marke adidas,
c) teget ili crna,
d) crna ili marke najk.

Rezultat. a) 0,6 b) 0,025 c) 0,925 d) 0,65.

Zadatak 4. Kolika je verovatnoća da se pri bacanju kockice dobije neparan broj ako se zna da je pao broj manji od pet.

Rezultat. 0,5.

Zadatak 5. Ako je poznato da je iz špila (52 karte) izvučena karta crvene boje, naći verovatnoću da je izvučena desetka.

Rezultat. $\frac{1}{13}$.

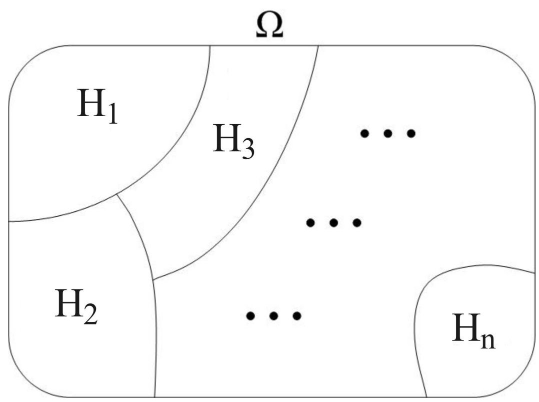
Zadatak 6. U jednoj prodavnici je 95 ispravnih i 5 neispravnih sijalica. U želji da kupi dve sijalice, kupac isprobava jednu po jednu. Ako isproba samo dve sijalice, naći verovatnoću da su obe ispravne. (Jasno da se jednom isprobana sijalica ne vraća bilo da je ispravna ili neispravna.)

Rezultat. 0,902.

2.4 FORMULA TOTALNE VEROVATNOĆE. BAJESOVA FORMULA

Formula totalne verovatnoće i Bajesova formula u pojedinim slučajevima olakšavaju nalaženje ,,običnih" i uslovnih verovatnoća. One važe za potpun sistem događaja.

Za međusobno disjunktne događaje $H_1, H_2, \ldots, H_n$ kaže se da čine **potpun sistem događaja** ako je $H_1 + H_2 + \ldots + H_n = \Omega$ i $P(H_i) > 0$ za svako $i = 1, 2, \ldots, n$.
Događaji $H_1, H_2, \ldots, H_n$ se nazivaju **hipotezama** ili **uzrocima**.



Slika 3: Potpun sistem događaja H_i.

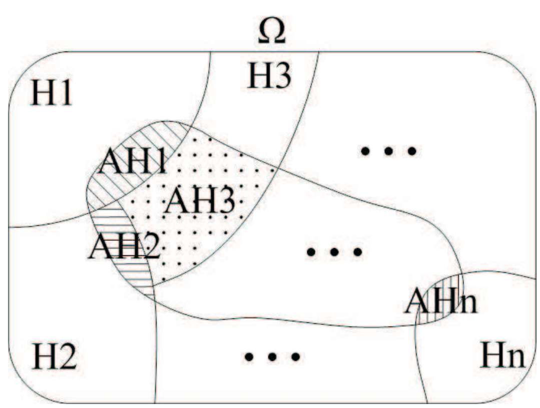
Ako se događaj A realizuje kao posledica raznih uzroka H_i i ako se znaju uslovne verovatnoće $P(A|H_i)$, može se naći verovatnoća događaja A.

Formula totalne verovatnoće. Ako događaji $H_1, H_2, \ldots, H_n$ čine potpun sistem događaja, tada za bilo koji događaj $A \subseteq \Omega$ važi

$$P(A) = \sum_{i=1}^{n} P(A|H_i) \cdot P(H_i).$$

Iz činjenice da za svaki događaj $A \subseteq \Omega$ važi (slika 4.) da je $A = \sum\limits_{i=1}^{n} AH_i,$ sledi

$$P(A) = P\left(\sum_{i=1}^{n} AH_i\right) = \sum_{i=1}^{n} P(AH_i) = \sum_{i=1}^{n} P(A|H_i) \cdot P(H_i).$$



Slika 4: $A = \sum\limits_{i=1}^{n} AH_i,$

Primer 10. Jedna serija od 200 proizvoda ima 5% neispravnih, a druga serija od 150 proizvoda ima 6% neispravnih. Iz prve serije se slučajno bira 60, a iz druge 40 proizvoda koji se nose u prodavnicu. Kolika je verovatnoća da kupac kupi jedan ispravan proizvod?

Rešenje.

Događaj A : ,,kupljeni proizvod je ispravan”.

Događaj H_1 : ,,kupljeni proizvod je iz prve serije” $\Longrightarrow P(H_1) = \dfrac{60}{60+40} = 0,6.$

Događaj H_2 : ,,kupljeni proizvod je iz druge serije” $\Longrightarrow P(H_2) = \dfrac{40}{60+40} = 0,4.$

Verovatnoća da je kupljen ispravan proizvod iz prve serije je $P(A|H_1) = 1 - 0,05 = 0,95.$

Verovatnoća da je kupljen ispravan proizvod druge serije je $P(A|H_1) = 1 - 0,06 = 0,94.$

$P(A) = P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) = 0,95 \cdot 0,6 + 0,94 \cdot 0,4 = 0,946.$

Neka se događaj A realizuje kao posledica raznih uzroka H_i $(i = 1, 2, \ldots, n)$. Ako se zna da se A realizovao, može se naći kolika je verovatnoća da se on realizovao pod određenim uzrokom H_k tj. $P(H_k|A)$.

Bajesova formula ili formula verovatnoće hipoteza (uzroka).

Ako događaji $H_1, H_2, \ldots, H_n$ čine potpun sistem događaja, tada za bilo koji događaj $A \subseteq \Omega$, sa $P(A) > 0$, važi

$$P(H_k|A) = \frac{P(A|H_k) \cdot P(H_k)}{P(A)} = \frac{P(A|H_k) \cdot P(H_k)}{\sum\limits_{i=1}^{n} P(A|H_i) \cdot P(H_i)} \qquad k = 1, 2, 3, \ldots, n.$$

Primer 11. Akumulator može pripadati jednoj od tri serije sa verovatnoćama $0,25$; $0,5$; i $0,25$ redom. Verovatnoća da će akumulator prve serije raditi tri zime je $0,1$. Verovatnoća da će akumulator druge serije raditi tri zime je $0,2$ i treće serije je $0,4$.

a) Naći verovatnoću da kupljen akumulator radi tri zime.

b) Ako je akumulator radio tri zime, naći verovatnoću da je iz treće serije.

Rešenje. Neka je događaj A : ,,akumulator radi tri zime”.

Za događaj H_1 : ,,akumulator je iz prve serije” $\Longrightarrow P(H_1) = 0,25$ i za događaj

$A|H_1$: ,,akumulator prve serije radi tri zime” $\Longrightarrow P(A|H_1) = 0,1.$

Za događaj H_2 : ,,akumulator je iz druge serije” $\Longrightarrow P(H_2) = 0,5$ i za događaj

$A|H_2$: ,,akumulator druge serije radi tri zime” $\Longrightarrow P(A|H_2) = 0,2.$

Za događaj H_3 : ,,akumulator je iz treće serije” $\Longrightarrow P(H_3) = 0,25$ i za događaj

$A|H_3$: ,,akumulator treće serije radi tri zime” $\Longrightarrow P(A|H_3) = 0,4.$

a) $P(A) = P(A|H_1) \cdot P(H_1) + P(A|H_2) \cdot P(H_2) + P(A|H_3) \cdot P(H_3) =$
$= 0,1 \cdot 0,25 + 0,2 \cdot 0,5 + 0,25 \cdot 0,4 = 0,225.$

b) $P(H_3|A) = \dfrac{P(A|H_3) \cdot P(H_3)}{P(A)} = \dfrac{0,4 \cdot 0,25}{0,225} = \dfrac{4}{9}.$

Zadatak 7. U jednoj kutiji se nalaze dve bele i tri crne kuglice, a u drugoj četiri bele i tri crne. Ako se slučajno bira jedna kutija i iz nje izvlači jedna kuglica, naći verovatnoću da se izvuče crna kuglica. Smatra se da su izbori kutija jednako verovatni događaji.

Rezultat. $\dfrac{18}{35}.$

Zadatak 8. U jedan objekat se ugrađuju cevi dveju fabrika I i II. Fabrika I dostavlja 65%, a II fabrika 35% potrebnih cevi. Propisanom standardu odgovara 95% cevi I fabrike i 90% cevi II fabrike.

a) Odrediti verovatnoću da slučajno izabrana cev bude standardna,

b) odrediti verovatnoću da je ugrađena cev iz I fabrike ako se zna da je ugrađena cev standardna.

Rezultat. a) 0,93 b) 0,66.

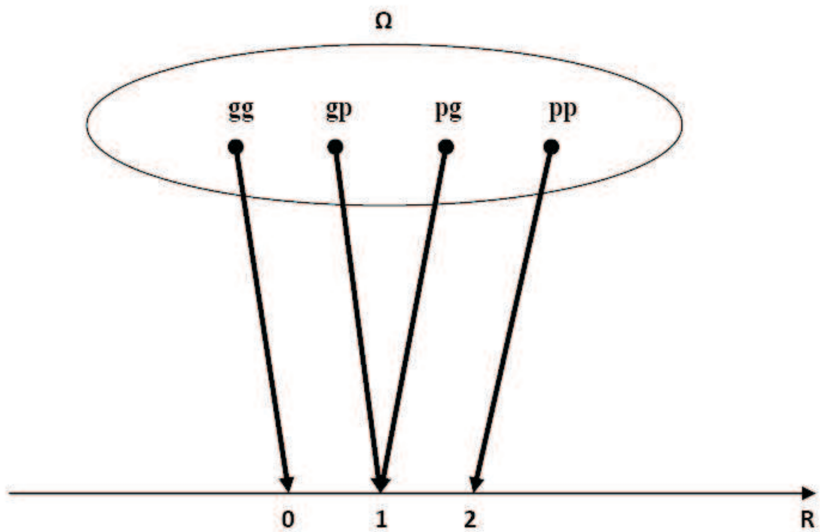
3 SLUČAJNA PROMENLJIVA

Da bi se prevazišao problem beskonačne populacije, kada se ne može (u opštem slučaju) govoriti o tome za koliko elemenata populacije obeležje ima određenu vrednost, pojam obeležja se povezuje sa pojmom slučajne promenljive u Teoriji verovatnoće.

Definicija. Funkcija koja svakom ishodu $\omega \in \Omega$, dodeljuje realan broj $X(\omega)$ je **slučajna promenljiva**, ako zadovoljava uslov da je za svako $x \in \mathcal{R}$ skup $\{\omega \,|\, X(\omega) \leq x\}$ događaj, tj. podskup od Ω.

Slučajne promenljive se označavaju velikim slovima X, Y, Z, X_1 itd. Vrednosti slučajne promenljive su **realizacije slučajne promenljive** i označavaju se malim slovima x, y, z, x_1 itd.

Primer 1. Eksperiment je bacanje novčića dva puta. Registruje se niz ,,glava" (g) i ,,pisama" (p) . Skup svih ishoda je $\Omega = \{gg,\, gp,\, pg,\, pp\}$. Neka funkcija X svakom ishodu dodeljuje broj registrovanih ,,pisama" (slika 5.).



Slika 5: Slučajna promenljiva X

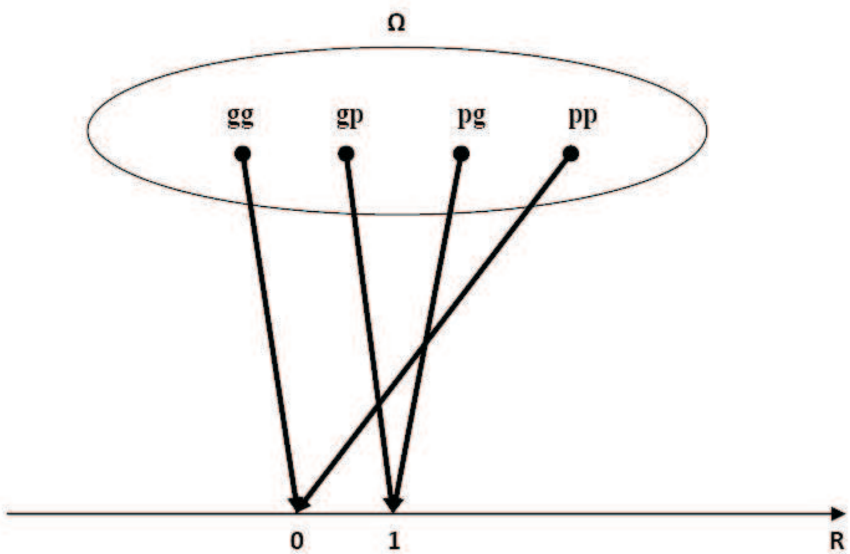
Kako je

$$\{\omega \,|\, X(\omega) \leq x\} = \begin{cases} \emptyset & \text{za} \quad x < 0 \\ \{gg\} & \text{za} \quad 0 \leq x < 1 \\ \{gg, gp, pg\} & \text{za} \quad 1 \leq x < 2 \\ \Omega & \text{za} \quad x \geq 2, \end{cases}$$

ispunjen je uslov da je za svako $x \in \mathcal{R}$ skup $\{\omega \,|\, X(\omega) \leq x\}$ događaj, pa je X slučajna promenljiva.

Slučajni događaj $\{gp, pg\} \subseteq \Omega$ je sada događaj da slučajna promenljiva X uzme vrednost jedan i može se kraće zapisati, $\{X = 1\}$. U tom smislu, događaj $\{\omega \,|\, X(\omega) = x_i\}$ se najčešće zapisuje u obliku $\{X = x_i\}$.
Slučajni događaj $\{gg, gp, pg\} \subseteq \Omega$ je sada događaj da slučajna promenljiva X uzme vrednost nula ili jedan. Kako su događaj $\{X = 0\} = \{gg\}$ i događaj $\{X = 1\} = \{gp, pg\}$ disjunktni, sledi
$\{pg, gp, pg\} = \{X = 0 \vee X = 1\} = \{X = 0\} + \{X = 1\} = \{1 \leq X \leq 2\}$.
Skup realizacija slučajne promenljive X je $\{0, 1, 2\}$.

Nad istim skupom ishoda može se definisati i druga slučajna promenljiva Y koja ishodu dodeljuje 0 ako u oba bacanja padne ista strana novčića, a 1 u suprotnom (slika 6).



Slika 6: Slučajna promenljiva Y

Kako je

$$\{\omega \,|\, Y(\omega) \leq x\} = \begin{cases} \emptyset & \text{za} \quad x < 0 \\ \{gg, pp\} & \text{za} \quad 0 \leq x < 1 \\ \Omega & \text{za} \quad x \geq 1, \end{cases}$$

ispunjen je uslov da je za svako $x \in \mathcal{R}$ skup $\{\omega \,|\, Y(\omega) \leq x\}$ događaj, pa je Y slučajna promenljiva.
Skup realizacija slučajne promenljive Y je $\{0, 1\}$. Primeri nekih događaja su:
$\{Y = 1\} = \{gp, pg\}$, $\{Y = 3\} = \emptyset$, $\{Y > -1\} = \Omega$, $\{Y < -2\} = \emptyset$, $\{Y < 5\} = \Omega$,
$\{Y = 1/2\} = \emptyset$, $\{Y > 1/2\} = \{Y = 1\} = \{gp, pg\}$, $\{0 \leq Y < 1\} = \{Y = 0\} = \{gg, pp\}$ itd.

Nadalje se razmatraju dva tipa slučajnih promenljivih: **diskretne** i **apsolutno neprekidne**. Diskretne slučajne promenljive imaju konačno mnogo vrednosti $\{x_1, x_2, \ldots, x_n\}$ ili prebrojivo beskonačno mnogo vrednosti $\{x_1, x_2, x_3, \ldots\}$. Apsolutno neprekidne slučajne promenljive imaju neprebrojivo mnogo vrednosti. Slučajne promenljive X i Y iz primera 1. su diskretne slučajne promenljive.

Postoje slučajne promenljive koje ne odgovaraju ni jednom od ova dva tipa.

3.1 DISKRETNA SLUČAJNA PROMENLJIVA

Slučajna promenljiva diskretnog tipa je određena ako su poznate sve njene realizacije x_i i sve verovatnoće realizacija događaja $\{X = x_i\}$, tj. verovatnoće $p_i = P\{X = x_i\}$. Sve realizacije x_i i sve verovatnoće $p_i = P\{X = x_i\}$ čine **zakon raspodele verovatnoća** diskretne slučajne promenljive (kaže se i zakon raspodele). Ako je diskretna slučajna promenljiva sa konačno mnogo vrednosti, tada se njen zakon raspodele zapisuje u obliku

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}.$$

Sve verovatnoće p_i su pozitivni brojevi i njihov zbir mora biti 1, tj. $p_1 + p_2 + \cdots + p_n = 1$.

Ukoliko je slučajna promenljiva diskretna sa prebrojivo beskonačno mnogo vrednosti, tada se njen zakon raspodele zapisuje u obliku

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}.$$

I ovde su verovatnoće p_i pozitivni brojevi i zbir svih verovatnoća mora biti 1.

Primer 2. Neka su slučajne promenljive kao u primeru 1. Zakon raspodele slučajne promenljive X je

$$\begin{pmatrix} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

jer su verovatnoće

$$p_1 = P\{X = 0\} = P\{gg\} = \frac{1}{4}, \qquad p_2 = P\{X = 1\} = P\{gp, pg\} = \frac{2}{4} = \frac{1}{2} \quad \text{i}$$

$$p_3 = P\{X = 2\} = P\{pp\} = \frac{1}{4}.$$

Sada, kada se zna raspodela slučajne promenljive X, mogu se odrediti verovatnoće svih događaja oblika $\{X \in S\}$, gde je S neki skup realnih brojeva. Ako je $S = [-5, 3/2)$, onda je

$$P\{X \in S\} = P\{-5 \le X < 3/2\} = P\{X = 0 \vee X = 1\} = P\{X = 0\} + P\{X = 1\} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}.$$

Zakon raspodele slučajne promenljive Y je

$$\begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}.$$

Primer 3. Novčić se baca sve dok ne padne ,,glava". Odrediti raspodelu slučajne promenljive Z koja predstavlja broj bacanja.

Rešenje. Očigledno je Z diskretna slučajna promenljiva sa prebrojivo beskonačno mnogo vrednosti. Ona uzima vrednosti $1, 2, 3,$ itd.

$p_1 = P\{Z = 1\} = P\{g\} = \dfrac{1}{2}$ (jedan je povoljan ishod od dva moguća)

$p_2 = P\{Z = 2\} = P\{pg\} = \dfrac{1}{4}$ (jedan je povoljan ishod od četiri moguća)

$p_2 = P\{Z = 2\} = P\{ppg\} = \dfrac{1}{8}$ (jedan je povoljan ishod od mogućih $2^3 = 8$)

$\dots$

$\dots$

$\dots$

$p_i = P\{Z = i\} = P\{\underbrace{pp\dots p}_{i-1}g\} = \dfrac{1}{2^i}$ (jedan je povoljan ishod od mogućih 2^i)

$\dots$

$\dots$

$\dots$

Sledi da je zakon raspodele slučajne promenljive Z

$$\begin{pmatrix} 1 & 2 & 3 & \dots & i & \dots \\ 1/2 & 1/4 & 1/8 & \dots & 1/2^i & \dots \end{pmatrix}.$$

Napomena. Suma $\dfrac{1}{2} + \dfrac{1}{4} + \ldots + \dfrac{1}{2^i} + \ldots = 1$. To je geometrijski red sa količnikom $q = \dfrac{1}{2} < 1$.

U primeru 1. prve glave za populaciju od 40 kupaca i za obeležje ,,veličina cipela" razmatrana je sledeća raspodela obeležja:

vel. cipela-x_i	38	39	40	41	42	43	44	$\sum$
br. kupaca-f_i	4	3	7	13	7	3	3	40
f_i^*	0,1	0,075	0,175	0,325	0,175	0,075	0,075	1

Neka je eksperiment, slučajan izbor jednog kupca. Skup svih ishoda je 40 kupaca tj. cela populacija. Obeležje ,,veličina cipela" postaje slučajna promenljiva X koja svakom kupcu (ishodu) dodeljuje broj i to baš veličinu kupljenog para cipela. Skup realizacija slučajne promenljive X je $\{38, 39, 40, 41, 42, 43, 44\}$. Odgovarajuće verovatnoće su:

$p_1 = P\{X = 38\} = \dfrac{4}{40} = 0,1$

$p_2 = P\{X = 39\} = \dfrac{3}{40} = 0,075$

$\dots$

$\dots$

$\dots$

$p_7 = P\{X = 44\} = \dfrac{3}{40} = 0,075$

zakon raspodele verovatnoća slučajne promenljive X je

$$\begin{pmatrix} 38 & 39 & 40 & 41 & 42 & 43 & 44 \\ 0,1 & 0,075 & 0,175 & 0,325 & 0,175 & 0,075 & 0,075 \end{pmatrix}.$$

Dakle, raspodela slučajne promenljive X je ista kao raspodela posmatranog obeležja .

Kad je populacija konačna, uvek se može definisati eksperiment tako da se raspodela obeležja poklapa sa zakonom raspodele odgovarajuće slučajne promenljive. Kod beskonačne populacije se zaključci o obeležju uvek izvode preko nekog zadatog eksperimenta tako da se to obeležje može poistovetiti sa odgovarajućom slučajnom promenljivom. Na taj način se problem raspodele obeležja svodi na određivanje zakona raspodele verovatnoća odgovarajuće slučajne promenljive.

Svakoj slučajnoj promenljivoj pridružuje se i **funkcija raspodele verovatnoća** (kraće, funkcija raspodele).

Definicija. Realna funkcija F definisana sa

$$F(x) = P\{X \leq x\}, \qquad x \in \mathcal{R},$$

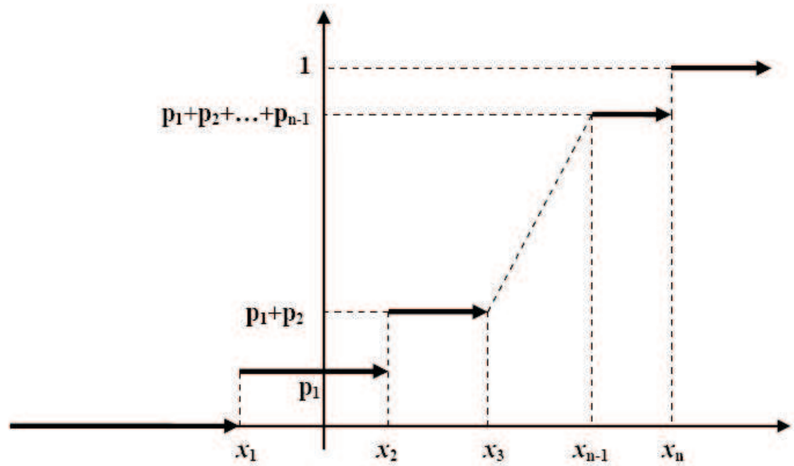
je funkcija raspodele verovatnoća slučajne promenljive X.

Funkcija raspodele $F(x)$ slučajne promenljive X je neopadajuća funkcija , neprekidna sa desne strane i za nju važi: $F(-\infty) = 0$ i $F(\infty) = 1$.

Ako je X diskretna slučajna promenljiva sa konačno mnogo vrednosti $\{x_1, x_2, \ldots, x_n\}$, tada je njena funkcija raspodele zadata sa

$$F(x) = \begin{cases} 0, & x < x_1, \\ p_1, & x_1 \leq x < x_2, \\ p_1 + p_2, & x_2 \leq x < x_3, \\ \vdots & \vdots \\ p_1 + p_2 + \cdots + P_{n-1}, & x_{n-1} \leq x < x_n, \\ 1, & x \geq x_n. \end{cases}.$$

Na sledećoj slici se vidi da je grafik funkcije raspodele diskretne slučajne promenljive stepenastog oblika.



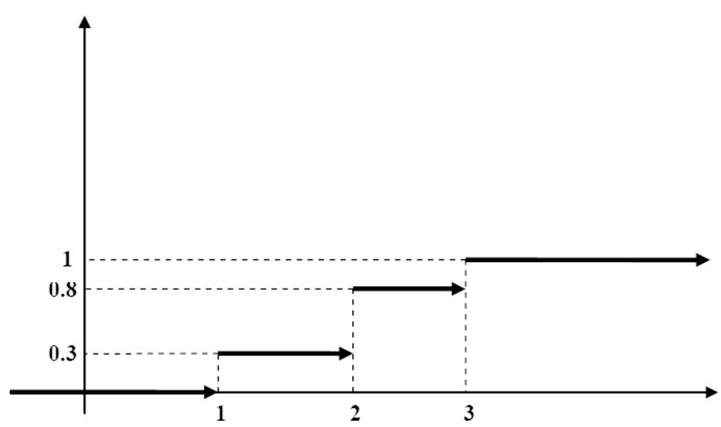
Slika 7: Grafik funkcije raspodele diskretne slučajne promenljive.

Primer 4. Ako slučajna promenljiva X ima sledeći zakon raspodele

$$\begin{pmatrix} 1 & 2 & 3 \\ 0,3 & 0,5 & 0,2 \end{pmatrix}$$

njena funkcija raspodele je

$$F(x) = \begin{cases} 0 & \text{za} \quad x < 1, \\ 0,3 & \text{za} \quad 1 \leq x < 2, \\ 0,8 & \text{za} \quad 2 \leq x < 3, \\ 1 & \text{za} \quad x \geq 3. \end{cases}$$



Slika 8: Grafik funkcije raspodele iz primera 4.

3.2 APSOLUTNO NEPREKIDNA SLUČAJNA PROMENLJIVA

Skup vrednosti slučajne promenljive X apsolutno neprekidnog tipa je neki interval, podskup skupa realnih brojeva ili ceo skup realnih brojeva. Pošto je skup vrednosti neprebrojiv, intuitivno je jasno da je verovatnoća da slučajna promenljiva X uzme tačno određenu vrednost, recimo i, jednaka nuli, tj. $P\{X = i\} = 0$. Zato su od značaja verovatnoće $P\{a \leq X < b\}$.

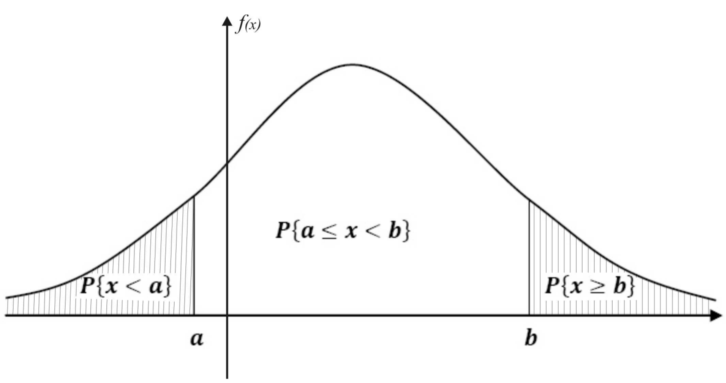
Definicija. Slučajna promenljiva X je apsolutno neprekidnog tipa ako postoji nenegativna funkcija $f : \mathcal{R} \longrightarrow \mathcal{R}$, tako da je za bilo koji interval $[a, b) \subset (-\infty, \infty)$,

$$P\{a \leq X < b\} = \int_a^b f(x)dx.$$

Funkcija $f(x)$ mora zadovoljavati uslov $\quad P\{-\infty < X < \infty\} = P\{\Omega\} = \int_{-\infty}^{\infty} f(x)dx = 1.$
Funkcija $f(x)$ je **gustina raspodele verovatnoća** (kraće, gustina raspodele) slučajne promenljive X. Ovo ime je zato što je verovatnoća da X uzme vrednost iz nekog intervala proporcionalna dužini tog intervala, a koeficijent proporcionalnosti je baš $f(x)$.

Diskretna slučajna promenljiva nema gustinu raspodele.

Kako je $\left|\int_a^b f(x)dx\right|$ merni broj površine zatvorene oblasti ograničene krivom (gustinom) $y = f(x)$, pravama $x = a$, $x = b$ i osom Ox, to je i verovatnoća da slučajna promenljiva uzme vrednost iz intervala $[a, b) \subset (-\infty, \infty)$, tj. $P\{a \leq X < b\}$, jednaka mernom broju pomenute površine (slika 9.). Sledi i da je površina ispod grafika gustine raspodele $f(x)$, a iznad x-ose, jednaka jedan.



Slika 9: Grafik gustine $f(x)$ neke apsolutno neprekidne slučajne promenljive X.

Za svaku slučajnu promenljivu funkcija raspodele $F(x)$ je definisana za sve realne vrednosti x. Ako je X slučajna promenljiva apsolutno neprekidnog tipa sa gustinom $f(x)$, onda je

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(t)dt.$$

Dakle, $F(x)$ je neopadajuća, neprekidna funkcija od x i

$$F'(x) = f(x)$$

u svim tačkama u kojima je $f(x)$ neprekidna funkcija. Takođe, $0 \leq F(x) \leq 1$ za svako x. Iz definicije funkcije raspodele $F(x)$, zbog $P\{X=a\} = P\{X=b\} = 0$, sledi

$$P\{a < X < b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = P\{a \leq X \leq b\} = \int_a^b f(x)dx = F(b) - F(a),$$

$$P\{X > a\} = P\{X \geq a\} = 1 - P\{X < a\} = 1 - P\{X \leq a\} = 1 - F(a).$$

Primer 5. Odrediti konstantu c tako da funkcija:

$$f(x) = \begin{cases} c \cdot x^2, & 0 \leq x \leq 2 \\ 0, & x < 0 \lor x > 2 \end{cases}$$

bude gustina raspodele verovatnoća. Zatim naći funkciju raspodele.

Rešenje. Da bi ova funkcija bila gustina raspodele, mora da zadovoljava uslov $\int\limits_{-\infty}^{\infty} f(x)dx = 1$, tj. $\int\limits_0^2 c \cdot x^2 dx = 1$. Sledi, $c \cdot \left.\dfrac{x^3}{3}\right|_0^2 = 1$ odakle je $c = \dfrac{3}{8}$.
Znači, gustina raspodele je

$$f(x) = \begin{cases} \frac{3}{8} \cdot x^2, & 0 \leq x \leq 2 \\ 0, & x < 0 \lor x > 2 \end{cases},$$

a funkcija raspodele je

$$F(x) = \begin{cases} 0, & x < 0 \\ \int\limits_0^x \frac{3}{8} t^2 dt = \dfrac{x^3}{8}, & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases}.$$

Primer 6. Ako slučajna promenljiva X ima gustinu raspodele

$$f(x) = \begin{cases} \dfrac{1}{2} \cdot \sin x, & 0 \leq x < \pi, \\ 0, & \text{inače}, \end{cases}$$

naći verovatnoće
a) $P\left\{0 < X \leq \dfrac{\pi}{4}\right\}$,
b) $P\left\{\dfrac{\pi}{2} \leq X \leq \dfrac{3\pi}{2}\right\}$.

Rešenje.
a) $P\left\{0 < X \leq \dfrac{\pi}{4}\right\} = \int\limits_0^{\frac{\pi}{4}} f(x) \cdot dx = \int\limits_0^{\frac{\pi}{4}} \dfrac{1}{2} \cdot \sin x \cdot dx = -\dfrac{1}{2} \cos x \Big|_0^{\frac{\pi}{4}} = \dfrac{2 - \sqrt{2}}{4}.$

b) $P\left\{\frac{\pi}{2} < X \leq \frac{3\pi}{2}\right\} = \int\limits_{\frac{\pi}{2}}^{\frac{3\pi}{2}} f(x)\cdot dx = \int\limits_{\frac{\pi}{2}}^{\pi} f(x)\cdot dx + \int\limits_{\pi}^{\frac{3\pi}{2}} f(x)\cdot dx = \int\limits_{\frac{\pi}{2}}^{\pi} \frac{1}{2}\sin x\cdot dx + \int\limits_{\pi}^{\frac{3\pi}{2}} 0\cdot dx = \frac{1}{2}.$

Primer 7. Ako slučajna promenljiva X ima funkciju raspodele

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \leq x < 1, \\ 1 & x \geq 1, \end{cases}$$

naći verovatnoće

a) $P\left\{X \leq \frac{1}{2}\right\}$,

b) $P\left\{\{-1 \leq X < \frac{1}{3}\right\}$.

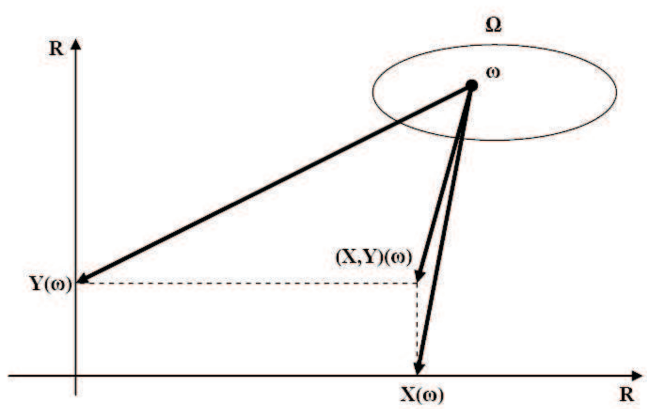
Rešenje. a) $P\left\{X \leq \frac{1}{2}\right\} = F\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$

b) $P\left\{-1 \leq X < \frac{1}{3}\right\} = F\left(\frac{1}{3}\right) - F(-1) = \left(\frac{1}{3}\right)^2 - 0 = \frac{1}{9}.$

3.3 VIŠEDIMENZIONALNA SLUČAJNA PROMENLJIVA

Do sada su predmet razmatranja bile samo jednodimenzionalne slučajne promenljive. Međutim, istovremeno se može posmatrati i više slučajnih promenljivih na istoj populaciji. Na primer, temperatura, vlažnost vazduha i nivo buke na nekom radnom mestu.

Ako su $X : \Omega \to \mathcal{R}$, $Y : \Omega \to \mathcal{R}$ slučajne promenljive, tada se uređeni par (X, Y) naziva **dvodimenzionalna slučajna promenljiva**. Uređeni par (X, Y) svakom ishodu $\omega \in \Omega$ dodeljuje uređeni par brojeva $(X(\omega), Y(\omega)) = (x, y) \in \mathcal{R} \times \mathcal{R} = \mathcal{R}^2$.



Slika 10: Slučajna promenljiva (X, Y).

Funkcija raspodele dvodimenzionalne slučajne promenljive $F_{XY} : \mathcal{R}^2 \to [0,1]$ se definiše kao verovatnoća realizacije događaja $\{X \le x, Y \le y\}$ tj.

$$F_{XY}(x,y) = P\{X \le x, Y \le y\} \qquad x, y \in \mathcal{R}.$$

Nadalje se razmatraju samo dvodimenzionalne slučajne promenljive diskretnog tipa.

Dvodimenzionalna slučajna promenljiva je diskretnog tipa ako su X i Y diskretnog tipa. Ako je $\{x_1, x_2, \ldots, x_m\} \subset \mathcal{R}$ skup vrednosti slučajne promenljive X, a $\{y_1, y_2, \ldots, y_n\} \subset \mathcal{R}$ skup vrednosti slučajne promenljive Y, onda je

$$R_{XY} = \{(x_1,y_1),(x_1,y_2),(x_1,y_3),\ldots,(x_2,y_1),(x_2,y_2),\ldots,(x_m,y_n)\} \subset \mathcal{R}^2$$

skup vrednosti diskretne dvodimenzionalne slučajne promenljive (X,Y). Vrednostima (x_i, y_j) ovog skupa odgovaraju verovatnoće

$$p_{ij} = P\{(X,Y) = (x_i,y_j)\} = P\{X = x_i, Y = y_j\}.$$

Skup vrednosti R_{XY} zajedno sa odgovarajućim verovatnoćama p_{ij} predstavlja zakon raspodele diskretne dvodimenzionalne slučajne promenljive (X,Y). On se obično zapisuje analitički ili u obliku tabele

$X \backslash Y$	y_1	y_2	$\ldots$	y_n
x_1	p_{11}	p_{12}	$\ldots$	p_{1n}
x_2	p_{21}	p_{22}	$\ldots$	p_{2n}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\ldots$	p_{mn}

Zbir svih verovatnoća p_{ij} mora biti jedan.

Primer 8. Novčić se baca tri puta. Posmatraju se dve slučajne promenljive

X: broj grbova u tri bacanja i Y: broj grbova u poslednja dva bacanja.

Odrediti zakon raspodele dvodimenzionalne slučajne promenljive (X,Y).

Rešenje. Skup vrednosti slučajne promenljive X je $R_X = \{0,1,2,3\}$, a skup vrednosti slučajne promenljive Y je $R_Y = \{0,1,2\}$.

$$\begin{array}{rcccccccc}
\Omega = & \{ppp, & ppg, & pgp, & gpp, & pgg, & gpg, & ggp, & ggg\} \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
R_{XY} = & \{(0,0) & (1,1) & (1,1) & (1,0) & (2,2) & (2,1) & (2,1) & (3,2)\}
\end{array}$$

Verovatnoće su redom:

$p_{11} = P\{(X,Y) = (0,0)\} = P\{X = 0, Y = 0\} = P\{ppp\} = \dfrac{1}{8}$

$p_{21} = P\{(X,Y) = (1,0)\} = P\{X = 1, Y = 0\} = P\{gpp\} = \dfrac{1}{8}$

$p_{22} = P\{(X,Y) = (1,1)\} = P\{ppg, pgp\} = \dfrac{2}{8} = \dfrac{1}{4}$

$p_{32} = P\{(X,Y) = (2,1)\} = P\{gpg, ggp\} = \dfrac{2}{4} = \dfrac{1}{4}$

$p_{33} = P\{(X,Y) = (2,2)\} = P\{pgg\} = \dfrac{1}{8}$

$p_{43} = P\{(X,Y) = (3,2)\} = P\{ggg\} = \dfrac{1}{8}.$

Ostale verovatnoće $p_{12}, p_{13}, p_{23}, p_{31}, p_{41}, p_{42}$ su jednake nuli. Na primer $p_{12} = P\{(X,Y) = (0,1)\} = P\{X = 0, Y = 1\} = 0$ jer je događaj $\{(X,Y) = (0,1)\}$ nemoguć. Sledi da je zakon raspodele

$X\backslash Y$	0	1	2
0	1/8	0	0
1	1/8	1/4	0
2	0	1/4	1/8
3	0	0	1/8

.

Iz zajedničke raspodele diskretnih slučajnih promenljivih X i Y mogu se odrediti pojedinačni zakoni raspodela jednodimenzionalnih diskretnih slučajnih promenljivih X i Y. Ako je zajednička raspodela diskretnih slučajnih promenljivih X i Y data tabelom tada je zakon raspodele slučajne promenljive X oblika

$$\begin{pmatrix} x_1 & x_2 & \dots & x_m \\ p_{1\bullet} & p_{2\bullet} & \dots & p_{m\bullet} \end{pmatrix},$$

gde je $p_{i\bullet} = p_{i1} + p_{i2} + \ldots + p_{in}$, zbir verovatnoća u i-toj vrsti ($i = 1, 2, \ldots, m$). Zakon raspodele slučajne promenljive Y je oblika

$$\begin{pmatrix} y_1 & 2_2 & \dots & x_n \\ p_{\bullet 1} & p_{\bullet 2} & \dots & p_{\bullet n} \end{pmatrix},$$

gde je $p_{\bullet j} = p_{1j} + p_{2j} + \ldots + p_{mj}$, zbir verovatnoća u j-toj koloni ($j = 1, 2, \ldots, n$).

Raspodele slučajnih promenljivih X i Y koje proizlaze iz zajedničke raspodele slučajnih promenljivih X i Y nazivaju se **marginalne raspodele** dok se verovatnoće $p_{i\bullet}$ i $p_{\bullet j}$ nazivaju **marginalne verovatnoće**.

Primer 9. Odrediti marginalne raspodele slučajnih promenljivih X i Y iz prethodnog primera.

Rešenje. Slučajna promenljiva X uzima vrednosti iz skupa $\{0, 1, 2, 3\}$ sa verovatnoćama

$$P\{X = 0\} = p_{1\bullet} = \frac{1}{8} + 0 + 0 = \frac{1}{8},$$

$$P\{X = 1\} = p_{2\bullet} = \frac{1}{8} + \frac{1}{4} + 0 = \frac{3}{8},$$

$$P\{X = 2\} = p_{3\bullet} = 0 + \frac{1}{4} + \frac{1}{8} = \frac{3}{8},$$

$$P\{X = 3\} = p_{4\bullet} = 0 + 0 + \frac{1}{8} = \frac{1}{8}.$$

Prema tome, marginalna raspodela slučajne promenljive X je

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{pmatrix}.$$

Skup realizacija slučajne promenljive Y je $\{0,1,2\}$, a verovatnoće su

$$P\{Y=0\} = p_{\bullet 1} = \frac{1}{8} + \frac{1}{8} + 0 + 0 = \frac{1}{4},$$

$$P\{Y=1\} = p_{\bullet 2} = 0 + \frac{1}{4} + \frac{1}{4} + 0 = \frac{1}{2},$$

$$P\{Y=2\} = p_{\bullet 3} = 0 + \frac{1}{4} + \frac{1}{8} = \frac{1}{4}.$$

Prema tome, marginalna raspodela slučajne promenljive Y je

$$\begin{pmatrix} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}.$$

3.4 NEZAVISNE SLUČAJNE PROMENLJIVE

Pojam nezavisnih slučajnih događaja može se proširiti na slučajne promenljive.

Neka su X i Y slučajne promenljive diskretnog tipa. One su **nezavisne** ako i samo ako je za svako i, j

$$p_{ij} = p_i \cdot p_j \quad \text{tj.} \quad P\{(X,Y) = (x_i, y_j)\} = P\{X = x_i, Y = y_j\} = P\{X = x_i\} \cdot \{Y = y_j\}.$$

Diskretne slučajne promenljive X i Y iz prethodnog primera nisu nezavisne.

$$P\{(X,Y) = (1,2)\} = P\{X = 1, Y = 2\} = 0, \qquad P\{X = 1\} = \frac{3}{8}, \qquad P\{Y = 2\} = \frac{1}{4}.$$

Može se, neformalno, reći: ako su slučajne promenljive X i Y nezavisne, umesto jedne dvodimenzionalne slučajne promenljive (X,Y) mogu se posmatrati dve (jednodimenzionalne) slučajne promenljive X i Y.

U praksi nezavisnost slučajnih promenljivih X i Y obično sledi iz fizičkih uslova eksperimenta za koji su te slučajne promenljive vezane. Uvek kada se kaze da se vrše nezavisna merenja dveju (ili više) veličina, to znači da su odgovarajuće slučajne promenljive nezavisne u smislu prethodne definicije.

NUMERIČKE KARAKTERISTIKE SLUČAJNIH PROMENLJIVIH

Gustina raspodele za apsolutno neprekidnu i zakon raspodele za diskretnu slučajnu promenljivu predstavljaju potpune karakteristike ovih slučajnih promenljivih. U mnogim praktičnim problemima nije neophodno poznavati sve karakteristike slučajne promenljive. Ponekad je dovoljno poznavati bitne numeričke karakteristike. Najveći praktični značaj imaju matematičko očekivanje i disperzija.

3.5 CENTRI GRUPISANJA VREDNOSTI SLUČAJNIH PROMENLJIVIH

Matematičko očekivanje, medijana i mod su numeričke karakteristike koje ukazuju na centar grupisanja vrednosti slučajne promenljive.

Matematičko očekivanje slučajne promenljive X u oznaci $E(X)$ je osnovna numerička karakteristika slučajne promenljive.

Neka je slučajna promenljiva diskretnog tipa sa konačno mnogo vrednosti:

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \qquad \sum_{i=1}^{n} p_i = 1.$$

Tada je matematičko očekivanje broj

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i.$$

Ako slučajna promenljiva ima prebrojivo beskonačno vrednosti, matematičko očekivanje broj

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot p_i, \qquad \text{ukoliko postoji.}$$

Matematičko očekivanje slučajne promenljive apsolutno neprekidnog tipa sa gustinom $f(x)$ je broj

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx, \qquad \text{ukoliko postoji.}$$

Termin matematičko očekivanje često se skraćuje pa se koristi naziv očekivanje ili očekivana vrednost. Osim toga, koristi se i termin srednja vrednost slučajne promenljive.

Primer 10. Slučajna promenljiva X

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,05 & 0,2 & 0,4 & 0,25 & 0,1 \end{pmatrix}$$

ima matematičko očekivanje

$$E(X) = 1 \cdot 0,05 + 2 \cdot 0,2 + 3 \cdot 0,4 + 4 \cdot 0,25 + 5 \cdot 0,1 = 3,15.$$

Primer 11. Slučajna promenljiva X sa gustinom

$$f(x) = \begin{cases} \frac{3}{8} \cdot x^2, & 0 \le x \le 2 \\ 0, & x < 0 \lor x > 2 \end{cases}$$

ima matematičko očekivanje

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{0}^{2} x \cdot \frac{3}{8} \cdot x^2 dx = \frac{3}{8} \int_{0}^{2} x^3 dx = \frac{3}{8} \cdot \frac{x^4}{4} \Bigg|_0^2 = \frac{3}{32}(2^4 - 0) = \frac{3}{2}.$$

Osobine matematičkog očekivanja slučajne promenljive X su:

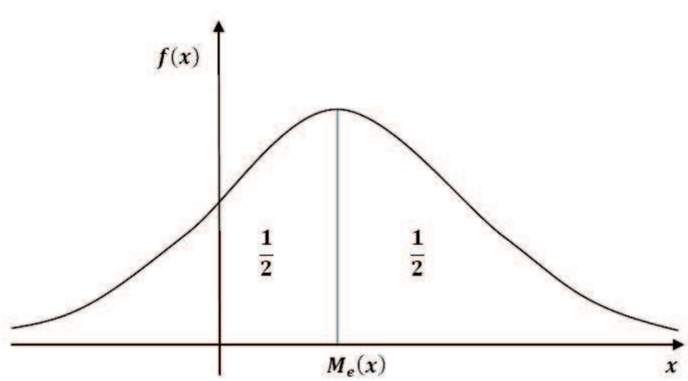
1. Ako je $X = c$, gde je c konstanta, tada je $E(c) = c$,
2. $E(a \cdot X + b) = a \cdot E(X) + b$, gde su a i b konstante,
3. $E(X + Y) = E(X) + E(Y)$ za bilo koje dve slučajne promenljive X i Y,
4. $E(X \cdot Y) = E(X) \cdot E(Y)$, ukoliko su slučajne promenljive X i Y nezavisne.

Medijana slučajne promenljive X, se označava sa $M_e(X)$.
Kod diskretne slučajne promenljive X medijana se definiše kao rešenje sistema nejednačina

$$P\{X < M_e(X)\} \le \frac{1}{2} \le P\{X \le M_e(X)\}.$$

Moguće je da ne postoji takva jedinstvena vrednost kod diskretne slučajne promenljive.
Kod apsolutno neprekidne slučajne promenljive medijana je ona vrednost slučajne promenljive za koju je

$$P\{X \le M_e(X)\} = \frac{1}{2}.$$



Slika 11: Medijana apsolutno neprekidne slučajne promenljive sa gustinom $f(x)$.

Kod apsolutno neprekidne slučajne promenljive medijana je uvek jedinstvena i odgovara ordinati koja deli površinu ispod krive gustine $f(x)$ na dva jednaka dela. Kako je ukupna površina ispod krive gustine jednaka jedan, to svaki ovaj deo ima površinu $\frac{1}{2}$ (slika 10).

Mod (modus) slučajne promenljive se označava sa $M_o(X)$. Kod diskretne slučajne promenljive to je ona vrednost slučajne promenljive koja ima najveću verovatnoću u svojoj okolini.

Kod apsolutno neprekidne slučajne promenljive, mod je ona vrednost u kojoj gustina $f(x)$ ima maksimum (lokalni).

Primer 12. Slučajna promenljiva X sa raspodelom

$$X : \begin{pmatrix} 2 & 4 & 8 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

ima mod $M_o(X) = 8$, jer vrednost 8 slučajna promenljiva uzima sa najvećom verovatnoćom, a koja iznosi $1/2$.

Moment reda k ($k \in \mathcal{N}$) slučajne promenljive X, u oznaci m_k, je matematičko očekivanje slučajne promenljive X^k (ukoliko ovo postoji).
U diskretnom slučaju, tj. za

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \qquad m_k = E(X^k) = \sum_{i=1}^{n} x_i^k \cdot p_i.$$

Za slučajnu promenljivu apsolutno neprekidnog tipa sa gustinom $f(x)$

$$m_k = E(X^k) = \int_{-\infty}^{\infty} x^k \cdot f(x)dx.$$

Matematičko očekivanje je moment prvog reda.

3.6 MERE RASTURANJA OKO CENTARA GRUPISANJA

Mere odstupanja vrednosti slučajne promenjive od njenog matematičkog očekivanja su disperzija i standardna devijacija.

Disperzija (varijansa) je numerička karakteristika vezana za rasturanje ili rasejavanje slučajne promenljive oko njenog očekivanja. Disperzija slučajne promenljive u oznaci $D(X)$ je broj

$$D(X) = E\left(X - E(X)\right)^2,$$

ukoliko očekivanje $E(X)$ postoji. Disperzija se može efektivno izračunati na sledeći način

$$D(X) = E\left(X^2\right) - \left(E(X)\right)^2.$$

Ako je X diskretna slučajna promenljiva sa raspodelom

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

onda slučajna promenljiva X^2 ima raspodelu

$$X : \begin{pmatrix} x_1^2 & x_2^2 & \dots & x_n^2 \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

i njeno matematičko očekivanje je $E(X^2) = \sum\limits_{i=1}^{n} x_i^2 \cdot p_i$.
Za $E(X) = \mu$, sledi da je matematičko očekivanje diskretne slučajne promenljive

$$D(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i = \sum_{i=1}^{n} x_i^2 \cdot p_i - \mu^2.$$

Ako je X apsolutno neprekidna slučajna promenljiva sa gustinom $f(x)$, čije je matematičko očekivanje $E(X) = \int\limits_{-\infty}^{+\infty} x \cdot f(x) \cdot dx = \mu$ disperzija je

$$D(X) = E(X^2) - (E(X))^2 = \int_{-\infty}^{+\infty} x^2 \cdot f(x) \cdot dx - \mu^2.$$

Primer 13. Slučajna promenljiva X

$$X : \begin{pmatrix} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

ima matematičko očekivanje $E(X) = 0 \cdot \dfrac{1}{4} + 1 \cdot \dfrac{1}{2} + 2 \cdot \dfrac{1}{4} = 1.$ Kako je $E(X^2) = 0^2 \cdot \dfrac{1}{4} + 1^2 \cdot \dfrac{1}{2} + 2^2 \cdot \dfrac{1}{4} = \dfrac{3}{2},$ sledi $D(X) = E(X^2) - (E(X))^2 = \dfrac{3}{2} - 1 = \dfrac{1}{2}.$

Primer 14. Slučajna promenljiva X sa gustinom

$$f(x) = \begin{cases} \frac{3}{8} \cdot x^2, & 0 \leq x \leq 2 \\ 0, & x < 0 \vee x > 2 \end{cases}$$

ima matematičko očekivanje $E(X) = \dfrac{3}{2}$ (primer 11.).

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \int_0^2 x^2 \cdot \frac{3}{8} x^2 dx = \frac{3}{8} \int_0^2 x^4 dx = \frac{3}{8} \cdot \frac{x^5}{5} \Bigg|_0^2 = \frac{3}{40} \cdot (2^5 - 0) = \frac{12}{5}.$$

Disperzija ove slučajne promenljive je $D(X) = E(X^2) - (E(X))^2 = \dfrac{12}{5} - \left(\dfrac{3}{2}\right)^2 = \dfrac{3}{20}.$

Osobine disperzije su:

1. $D(X) \geq 0$ za svaku slučajnu promenljivu,
2. $D(X) = 0$ ako i samo ako je X konstanta,
3. $D(a \cdot X + b) = a^2 \cdot D(X)$, gde su a i b konstante, a X slučajna promenljiva,
4. $D(X + Y) = D(X) + D(Y)$ ukoliko su X i Y nezavisne slučajne promenljive.

Standardno odstupanje ili **standardna devijacija** je pozitivan broj

$$\sigma(X) = \sqrt{D(X)}.$$

Centralni moment reda k ($k \in \mathcal{N}$) slučajne promenljive X, u oznaci M_k, postoji ukoliko postoji moment reda k tj. $m_k = E\left(X^k\right)$. Tada je

$$M_k = E\left(X - E(X)\right)^k.$$

Disperzija je centralni moment drugog reda.

Koeficijent varijacije omogućava da se porede rasturanja, oko matematičkog očekivanja, vrednosti slučajnih promenljivih sa različitim raspodelama. Ako matematičko očekivanje i disperzija postoje i ako je $E(X) \neq 0$ to je broj

$$C_v = \frac{\sqrt{D(X)}}{E(X)} \quad \text{ili u procentima} \quad C_v\% = \frac{\sqrt{D(X)}}{E(X)} \cdot 100\%.$$

3.7 NUMERIČKE KARAKTERISTIKE DVODIMENZIONALNE SLUČAJNE PROMENLJIVE

Centar grupisanja vrednosti dvodimenzionalne slučajne promenljive najčešće karakteriše matematičko očekivanje

$$E(X, Y) = (E(X), E(Y))$$

gde su $E(X)$ i $E(Y)$ matematička očekivanja slučajnih promenljivih X i Y.

Disperzija dvodimenzionalne slučajne promenljive (X, Y) je uređen par $(D(X), D(Y))$ gde su $D(X)$ i $D(Y)$ disperzije slučajnih promenljivih X i Y.

Koeficijent korelacije meri stepen linearne zavisnosti slučajnih promenljivih X i Y. Koeficijent korelacije slučajnih promenljivih X i Y se definiše kao broj

$$\rho_{X,Y} = \frac{E\left((X - E(X))\left(Y - E(Y)\right)\right)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} = \frac{E(X \cdot Y) - E(X) \cdot E(Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}}.$$

Koeficijent korelacije ima sledeće osobine:

1. Uzima vrednosti iz segmenta $[-1, 1]$.
2. Ako između slučajnih promenljivih X i Y postoji linearna veza $Y = a \cdot X + b$ pri čemu je $a > 0$, onda je $\rho_{X,Y} = 1$.
3. Ako između slučajnih promenljivih X i Y postoji linearna veza $Y = a \cdot X + b$ pri čemu je $a < 0$, onda je $\rho_{X,Y} = -1$.
4. Ako su slučajne promenljive X i Y nezavisne, tada je koeficijent korelacije $\rho_{X,Y} = 0$. Obrnuto u opštem slučaju ne važi.

Primer 15. Dvodimenzionalna slučajna promenljiva (X, Y) je zadata zakonom raspodele

$X\backslash Y$	0	1	2
-1	0,05	0,4	0,15
1	0,2	0,1	0,1

Odrediti koeficijent korelacije slučajne promenljive (X, Y).

Rešenje. Iz zajedničke raspodele slučajnih promenljivih X i Y, sledi da je marginalna raspodela slučajne promenljive X

$$\begin{pmatrix} -1 & 1 \\ 0,6 & 0,4 \end{pmatrix}.$$

Zato su njeno očekivanje i disperzija jednaki

$E(X) = (-1) \cdot 0,6 + 1 \cdot 0,4 = -0,6 + 0,4 = -0,2$

$D(X) = (-1)^2 \cdot 0,6 + 1^2 \cdot 0,4 - (-0,2)^2 = 0,6 + 0,4 - 0,04 = 0,96.$

Marginalna raspodela slučajne promenljive Y je

$$\begin{pmatrix} 0 & 1 & 2 \\ 0,25 & 0,5 & 0,25 \end{pmatrix},$$

pa su njeno očekivanje i disperzija jednaki

$E(Y) = 0 \cdot 0,25 + 1 \cdot 0,5 + 2 \cdot 0,25 = 0 + 0,5 + 0,5 = 1$

$D(Y) = 0^2 \cdot 0,25 + 1^2 \cdot 0,5 + 2^2 \cdot 0,25 - 1^2 = 0 + 0,5 + 1 - 1 = 0,5.$

Očekivanje $E(XY)$ se računa po formuli

$$E(XY) = \sum_{i,j} x_i \cdot y_j \cdot p_{ij}$$

gde je $p_{ij} = P\{X = x_i, Y = y_j\}$. Sledi

$E(XY) = (-1) \cdot 0 \cdot 0,05 + (-1) \cdot 1 \cdot 0,4 + (-1) \cdot 2 \cdot 0,15 + 1 \cdot 0 \cdot 0,2 + 1 \cdot 1 \cdot 0,1 + 1 \cdot 2 \cdot 0,1 =$
$= -0,4 - 0,3 + 0,1 + 0,2 = -0,4.$

Konačno, koeficijent korelacije između slučajnih promenljivih X i Y je

$$\rho_{XY} = \frac{-0,4 - (-0,2) \cdot 1}{\sqrt{0,96} \cdot \sqrt{0,5}} = -0,29.$$

VAŽNIJE RASPODELE SLUČAJNIH PROMENLJIVIH

3.8 RASPODELE DISKRETNIH SLUČAJNIH PROMENLJIVIH

Zakon raspodele verovatnoća diskretne slučajne promenljive najčešće se naziva diskretnom raspodelom.

Binomna raspodela. U nekom eksperimentu registruje se samo da li se događaj A realizovao ili ne. Neka je $P(A) = p$, odnosno $P(\overline{A}) = 1 - p = q$. Neka se taj eksperiment ponavlja n puta pod istim uslovima tako da su ta ponavljanja nezavisna. Slučajna promenljiva S_n-broj realizacija događaja A u n ponavljanja eksperimenta je diskretna slučajna promenljiva sa vrednostima $0, 1, 2, \ldots, n$. Slučajna promenljiva S_n se naziva binomna slučajna promenljiva.

Verovatnoća da slučajna promenljiva S_n uzme određenu vrednost $i \in \{0, 1, 2, \ldots, n\}$ je

$$p_i = P\{S_n = i\} = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} = \binom{n}{i} \cdot p^i \cdot q^{n-i} \qquad i = 0, 1, 2, \cdots, n,$$

gde je

$$\binom{n}{i} = \frac{n!}{i! \cdot (n-i)!} = \frac{n \cdot (n-1) \cdots (n-i+1)}{i!}, \qquad n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1.$$

Raspodela verovatnoća

$$\begin{pmatrix} 0 & 1 & 2 & \ldots & i & \ldots & n \\ q^n & npq^{n-1} & \binom{n}{2} p^2 q^{n-2} & \ldots & \binom{n}{i} p^i q^{n-i} & \ldots & p^n \end{pmatrix}$$

naziva se **binomna raspodela** sa parametrima n i p i kraće se piše $S_n : B(n, p)$.

Matematičko očekivanje i disperzija ove slučajne promenljive su:

$$E(S_n) = np \quad \text{i} \quad D(S_n) = np(1-p).$$

Zadatak 1. Kolika je verovatnoća da se pri osam bacanja kocke , broj jedan pojavi tri puta?

Rezultat. $X : B(8, 1/6)$; tražena verovatnoća je $P\{X = 3\} = \dfrac{56 \cdot 5^5}{6^8}$.

Zadatak 2. Kocka se baca deset puta. Kolika je verovatnoća da se broj dva pojavi bar jednom.

Rezultat. $X : B(10, 1/6)$; tražena verovatnoća je $P\{X \geq 1\} = 1 - \left(\dfrac{5}{6}\right)^{10}$.

Zadatak 3. Strelac gađa u metu deset puta. Verovatnoća da pogodi metu je u svakom pokušaju ista i iznosi $0,7$. Naći verovatnoću da pogodi osam puta.

Rezultat. $X : B(10; 0,7);$ tražena verovatnoća je $P\{X = 8\} = 0,234.$

Zadatak 4. Student polaže kolokvijum koji se sastoji od pet pitanja. Pod pretpostavkom da on na svako pitanje odgovara tačno sa verovatnoćom $1/2$, odrediti verovatnoće

a) da je tačno odgovorio na svih pet pitanja,
b) da je tačno odgovorio na tri pitanja,
c) da je tačno odgovorio najviše na dva pitanja
d) da je tačno odgovorio na više od tri pitanja.

Rezultat. $X : B(5; 0,5);$ tražene verovatnoće su

a) $P\{X = 5\} = \dfrac{1}{32}$,

b) $P\{X = 3\} = \dfrac{5}{16}$,

c) $P\{X \leq 2\} = P\{X = 0\} + P\{X = 1\} + P\{X = 2\} = \dfrac{1}{2^5} + \dfrac{5}{2^5} + \dfrac{10}{2^5} = \dfrac{1}{2}.$

d) $P\{X > 3\} = P\{X = 4\} + P\{X = 5\} = \dfrac{5}{2^5} + \dfrac{1}{2^5} = \dfrac{3}{16}.$

Puasonova raspodela. Ako je u binomnoj raspodeli $B(n, p)$, p malo i n veliko $(n \geq 50)$ onda za $\lambda = n \cdot p < 10$ važi da su verovatnoće $p_i = P\{S_n = i\}$ približno jednake $e^{-\lambda} \cdot \dfrac{\lambda^i}{i!}$ tj. važi

$$p_i = P\{S_n = i\} = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} \quad \approx \quad \frac{\lambda^i}{i!} \cdot e^{-\lambda} \qquad i = 0, 1, 2, \ldots.$$

Diskretna raspodela verovatnoća

$$\begin{pmatrix} 0 & 1 & 2 & \ldots & i & \ldots \\ e^{-\lambda} & \lambda \cdot e^{-\lambda} & \dfrac{\lambda^2}{2} \cdot e^{-\lambda} & \ldots & \dfrac{\lambda^i}{i!} \cdot e^{-\lambda} & \ldots \end{pmatrix}$$

se naziva Puasonova raspodela.

Slučajna promenljiva sa ovom raspodelom (u oznaci S_∞) naziva se Puasonova slučajna promenljiva. Kraće se zapisuje $S_\infty : \mathcal{P}(\lambda)$. Matematičko očekivanje i disperzija ove slučajne promenljive su:

$$E(S_\infty) = \lambda \quad \text{i} \quad D(S_\infty) = \lambda.$$

Puasonova rasodela ne služi samo za aproksimaciju binomne raspodele već i za određivanje verovatnoće broja javljanja (pojavljivanja) nekog događaja u prostoru i vremenu. Slučajna promenljiva X_t−broj pojavljivanja nekog događaja u intervalu $[0, t)$ ima Puasonovu raspodelu sa parametrom λ, tj. $X_t : P(\lambda)$ gde je λ očekivani broj pojavljivanja posmatranog događaja A odnosno, gde λ na neki način karakteriše intenzitet protoka događaja A.

Kod Puasonove raspodele uočava se sledeće:

$$p_{k+1} = P\{X = k+1\} = e^{-\lambda} \cdot \frac{\lambda^k \cdot \lambda}{(k+1) \cdot k!} = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \cdot \frac{\lambda}{k+1} = P\{X = k\} \cdot \frac{\lambda}{k+1} = p_k \cdot \frac{\lambda}{k+1}.$$

Napomena. Za $np \geq 10$ koristi se druga aproksimacija binomne raspodele tj. koristi se druga formula za približno izračunavanje ovih verovatnoća i o tome će biti reči kasnije.

Zadatak 5. Verovatnoća da je jedan proizvod neispravan iznosi $0,01$. Iz velikog skladišta uzima se 100 proizvoda. Kolika je verovatnoća da među njima bude 5 neispravnih.

Rezultat. $X : \mathcal{P}(1)$; tražena verovatnoća je $P\{X = 5\} = 0,003$.

Zadatak 6. Strelac gađa u metu i verovatnoća pogodka je $0,002$. Naći verovatnoću najmanje dva pogodka u 1000 gađanja.

Rezultat. $X : \mathcal{P}(2)$; tražena verovatnoća je $P\{X \geq 2\} = 1 - P\{X < 2\} = 0,588$.

Primer 16. U telefonskoj centrali, u toku jednog sata bilo je 180 poziva.Izračunati verovatnoću da u toku dva minuta
a) nije bilo poziva,
b) je bilo više od tri poziva,
c) je bilo tačno šest poziva.

Rešenje. Slučajna promenljiva X je broj telefonskih poziva u toku dva minuta. Slučajna promenljiva $X : \mathcal{P}(\lambda)$. Parametar λ, je očekivani broj poziva u dva minuta tj. $\lambda = \frac{180}{30} = 6$ jer jedan sat tj. 60 minuta ima 30 puta po dva minuta.
Tražene verovatnoće su

a) $p_0 = P\{X = 0\} = e^{-6} \cdot \frac{6^0}{0!} = e^{-6} = 0,0025.$

b) $P\{X > 3\} = 1 - P\{X \leq 3\} = 1 - (P\{X = 0\} + P\{X = 1\} + P\{X = 2\} + P\{X = 3\}) =$
$= 1 - (p_0 + p_1 + p_2 + p_3)$

Kako je

$p_1 = p_0 \cdot \frac{\lambda}{1} = 0,015$ i $p_2 = p_1 \cdot \frac{\lambda}{2} = 0,045$ i $p_3 = p_2 \cdot \frac{\lambda}{3} = 0,09$ sledi $P\{X > 3\} = 0,85.$

c) $P\{X = 6\} = e^{-6} \cdot \frac{6^6}{6!} = e^{-6} \cdot \frac{6^4}{20} = 0,17.$

Diskretno uniformna (ravnomerna) raspodela. Neka slučajna promenljiva X uzima vrednosti iz skupa $\{x_1, x_2, \ldots, x_n\}$. Kaže se da slučajna promenljiva X ima uniformnu diskretnu raspodelu na skupu $\{x_1, x_2, \ldots, x_n\}$, ako se sve vrednosti realizuju sa jednakim verovatnoćama

$$P\{X = x_1\} = P\{X = x_2\} = \ldots = P\{X = x_n\} = \frac{1}{n}.$$

U tom slučaju zakon raspodele verovatnoća slučajne promenljive X je

$$\begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_n \\ 1/n & 1/n & 1/n & \ldots & 1/n \end{pmatrix}.$$

3.9 RASPODELE APSOLUTNO NEPREKIDNIH SLUČAJNIH PROMENLJIVIH

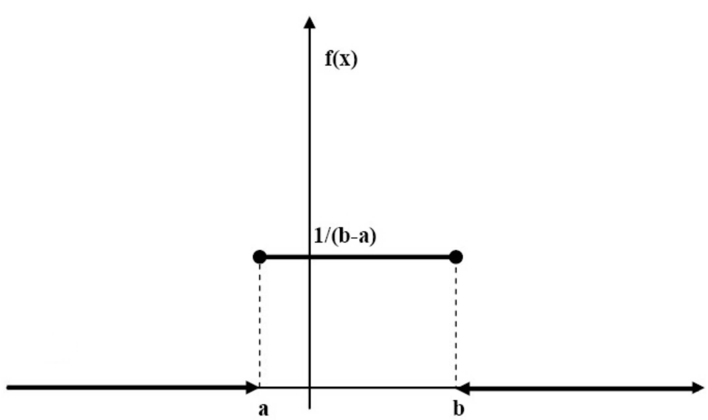
Uniformna raspodela. Ova raspodela ima dva parametra $a, b \in \mathcal{R}$ pri čemu je $a < b$. Gustina ove raspodele je

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b]. \end{cases}$$

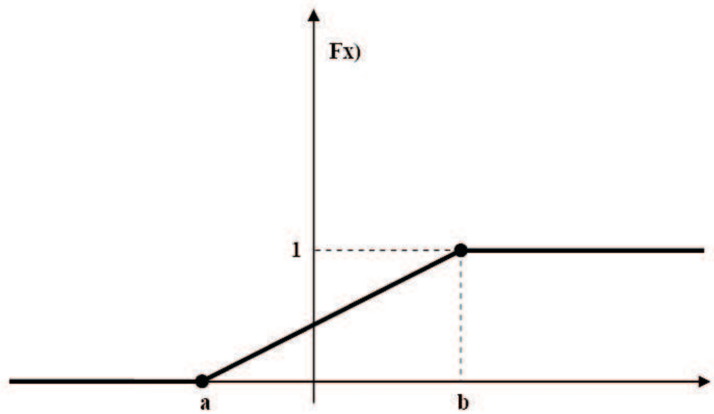
Odgovarajuća funkcija raspodele je

$$F(x) = \begin{cases} 0, & x \leq a, \\ \frac{1}{b-a} \cdot x - \frac{a}{b-a}, & a < x \leq b, \\ 1, & x \geq b. \end{cases}$$

Da slučajna promenljiva X ima uniformnu raspodelu piše se $X : \mathcal{U}(a, b)$.



Slika 12: Gustina $f(x)$ uniformne raspodele $\mathcal{U}(a, b)$.



Slika 13: Funkcija raspodele $F(x)$ uniformne raspodele $\mathcal{U}(a,b)$.

Primer 17. Slučajna promenljiva X ima uniformnu raspodelu na intervalu $[2,4]$.
a) Izračunati verovatnoće $P\{X \leq 3\}$, $P\{1 < X \leq 2.5\}$ $P\{2,5 < X \leq 3,5\}$.
b) Odrediti funkciju raspodele slučajne promenljive X, a zatim izračunati ove verovatnoće.

Rešenje. a) Kako je $f(x) = \begin{cases} \frac{1}{2}, & x \in [2,4] \\ 0, & x \notin [2,4]. \end{cases}$, sledi

$$P\{X \leq 3\} = \int_2^3 \frac{1}{2} \cdot dx = \frac{1}{2} \cdot x \Big|_2^3 = \frac{1}{2}(3-2) = \frac{1}{2}.$$

$$P\{1 < X \leq 2.5\} = \int_2^{2,5} \frac{1}{2} \cdot dx = \frac{1}{2} \cdot x \Big|_2^{2,5} = \frac{1}{4}.$$

$$P\{2,5 < X \leq 3,5\} = \int_{2,5}^{3,5} \frac{1}{2} \cdot dx = \frac{1}{2} \cdot x \Big|_{2,5}^{3,5} = \frac{1}{2}.$$

b) Funkcija raspodele verovatnoća je

$$F(x) = \begin{cases} 0 & \text{za} \quad x < 2, \\ \frac{1}{2} \cdot x - 1 & \text{za} \quad 2 \leq x < 4, \\ 1 & \text{za} \quad x \geq 4. \end{cases}$$

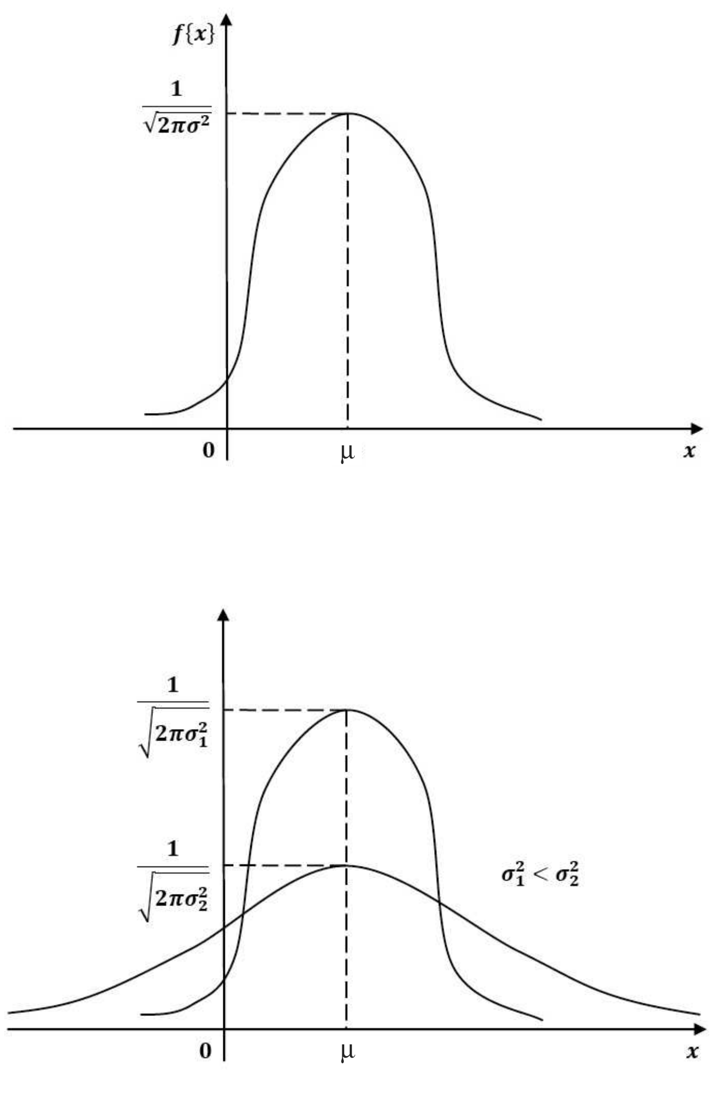
Normalna (Gausova)raspodela. Ovo je najznačajnija i najčešće korišćena apsolutno neprekidna raspodela.

Za slučajnu promenljivu X se kaže da ima normalnu raspodelu sa parametrima $\mu \in \mathcal{R}$ i $\sigma^2 > 0$ ako je njena gustina raspodele

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathcal{R},$$

a piše se $X : N(\mu, \sigma^2)$.

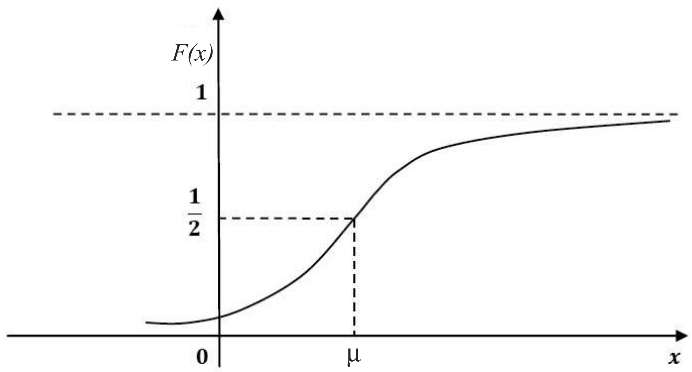
Pokazuje se da su matematičko očekivanje i disperzija slučajne promenljive sa normalnom raspodelom jednaki parametrima μ i σ^2, redom.



Slika 14: Gustina $f(x)$ normalne raspodele $N(\mu, \sigma^2)$.

Ako slučajna promenljiva ima normalnu raspodelu, $N(\mu, \sigma^2)$ njena funkcija raspodele je

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \, dt, \qquad x \in \mathcal{R}.$$



Slika 15: Funkcija raspodele $F(x)$ normalne raspodele $N(\mu, \sigma^2)$.

Za slučajnu promenljivu sa normalnom raspodelom važi (pravilo tri sigme):

$$\begin{aligned} P\{\mu - \sigma < X < \mu + \sigma\} &\approx 0,68 \\ P\{\mu - 2\sigma < X < \mu + 2\sigma\} &\approx 0,95 \\ P\{\mu - 3\sigma < X < \mu + 3\sigma\} &\approx 0,997. \end{aligned}$$

Normalnu raspodelu imaju visina ljudi, količina padavina u toku godine, pritisak itd.

Napomena. Mnoge diskretne raspodele se mogu aproksimirati normalnom raspodelom ako je broj različitih vrednosti slučajne promrnljive veliki i ako su one bliske jedna drugoj. Tako, za $n \geq 50$ i $n \cdot p \geq 10$, kada se binomna raspodela ne može aproksimirati Puasonovom, ona se aproksimira normalnom raspodelom, tj.

$$B(n,p) \sim N(np, np(1-p)).$$

Ako su parametri normalne raspodele $\mu = 0$ i $\sigma^2 = 1$, kaže se da je to **standardna normalna raspodela**. Ako slučajna promenljiva X ima normalnu raspodelu $N(\mu, \sigma^2)$, tada slučajna promenljiva $\dfrac{X-\mu}{\sigma}$ ima standardnu normalnu raspodelu $N(0,1)$. Zato se slučajna promenljiva $\dfrac{X-\mu}{\sigma}$ naziva standardizovanim oblikom slučajne promenljive X i ona se označava sa X^*, tj. $X^* = \dfrac{X-\mu}{\sigma}$. U tom slučaju gustina i funkcija raspodele su

$$f^*(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \quad \text{i} \quad F^*(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, dt, \qquad x \in \mathcal{R}.$$

Da bi se lako izračunale verovatnoće $P\{a < X^* \leq b\}$ definiše se nova funkcija $\Phi(x)$ čije su vrednosti tabelirane,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}} dt \qquad \text{za svako} \quad x \in \mathcal{R}.$$

Pokazuje se da važi

$$1.\ \Phi(0) = 0, \qquad 2.\ \Phi(+\infty) = 0.5, \qquad 3.\ \Phi(-x) = -\Phi(x).$$

Nadalje, za $x > 0$ je

$$F^*(x) = P\{-\infty < X^* < x\} = P\{-\infty < X^* < 0\} + P\{0 < X^* < x\} = 0.5 + \Phi(x).$$

Za $x < 0$ tj. za $x = -u$ $(u > 0)$ je

$F^*(x) = P\{-\infty < X^* < x\} = P\{-\infty < X^* < -u\} = P\{u < X^* < +\infty\} =$

$= F^*(+\infty) - F^*(u) = 1 - (0,5 + \Phi(u)) = 1 - 0,5 - \Phi(u) = 0,5 + \Phi(-u) = 0,5 + \Phi(x).$

Sledi, da za svako $x \in \mathcal{R}$ važi

$$F^*(x) = 0,5 + \Phi(x).$$

Tako za proizvoljne $a, b \in \mathcal{R}$ je

$$P\{a < X^* < b\} = F^*(b) - F^*(a) = 0,5 + \Phi(b) - (0,5 + \Phi(a)) = 0,5 + \Phi(b) - 0,5 - \Phi(a), \text{ tj.}$$

$$P\{a < X^* < b\} = \Phi(b) - \Phi(a).$$

Znači da bi izračunali verovatnoću $P\{a \leq X^* \leq b\}$ dovoljno je da iz tablice pročitamo vrednosti $\Phi(b)$ i $\Phi(a)$ i da ih oduzmemo.

Primer 18. Za slučajnu promenljivu $Z : N(0,1)$, naći $P\{0 < Z < 1,42\}$.

Rešenje. $P\{0 < Z < 1,42\} = \Phi(1,42) - \Phi(0) = 0,4222 - 0 = 0,4222.$

Primer 19. Za slučajnu promenljivu $Z : N(0,1)$, naći $P\{-0,73 < Z < 0\}$.

Rešenje. $P\{-0,73 < Z < 0\} = \Phi(0) - \Phi(-0,73) = 0 + \Phi(0,73) = 0,2673.$

Primer 20. Za slučajnu promenljivu $Z : N(0,1)$, naći $P\{-1,37 < Z < 2,01\}$.

Rešenje. $P\{-1,37 < Z < 2,01\} = \Phi(2,01) - \Phi(-1,37) = \Phi(2,01) + \Phi(1,37) =$
$= 0,4778 + 0,4147 = 0,8925.$

Primer 21. Za slučajnu promenljivu $Z : N(0,1)$, naći $P\{0,65 < Z < 1,26\}$.

Rešenje. $P\{0,65 < Z < 1,26\} = \Phi(1,26) - \Phi(0,65) = 0,3962 - 0,2422 = 0,1540.$

Primer 22. Za slučajnu promenljivu $Z : N(0,1)$, naći $P\{-1,79 < Z < -0,54\}$.

Rešenje. $P\{-1,79 < Z < -0,54\} = \Phi(-0,54) - \Phi(-1,79) = -\Phi(0,54) + \Phi(1,79) =$
$= -0,2054 + 0,4633 = 0,2579.$

Primer 23. Za slučajnu promenljivu $Z : N(0,1)$, naći $P\{Z > 1,13\}$.

Rešenje. $P\{Z > 1,13\} = P\{1,13 < Z < +\infty\} = \Phi(+\infty) - \Phi(1,13) = 0,5 - 0,3708 = 0,1292$

Primer 24. Ako je $Z : N(0,1)$, naći t tako da je $P\{Z < t\} = 0,7967$.

Rešenje. $P\{Z < t\} = P\{-\infty < Z < t\} = \Phi(t) - \Phi(-\infty) = \Phi(t) + \Phi(+\infty) = \Phi(t) + 0,5$
Iz $\Phi(t) + 0,5 = 0,7967$ sledi $\Phi(t) = 0,2967$. Iz tablice se čita $t = 0,83$.

Primer 25. Ako je $Z : N(0,1)$, naći t tako da je $P\{t < Z < 2\} = 0,3772$.

Rešenje. Iz $P\{t < Z < 2\} = 0,3772$ i $P\{t < Z < 2\} = \Phi(2) - \Phi(t)$ sledi
$\Phi(t) = \Phi(2) - 0,3772 = 0,4772 - 0,3772 = 0,1000.$
Broja $0,1000$ nema u tablici ali najpribližniji broj u tablici je $0,0987$. Njemu odgovara $t = 0,25$.

Zadatak 7. Koeficijent inteligencije ljudi je slučajna promenljiva X sa normalnom raspodelom $X : N(110,100)$. Posmatra se jedna osoba i meri se njen koeficijent inteligencije. Odrediti verovatnoće da posmatrana osoba ima koeficijent inteligencije

a) veći od 120,
b) manji od 100,
c) izmedju 110 i 130.

Rezultat.

a) $P\{X > 120\} = P\{1 < X^* < +\infty\} = 0,1587.$

b) $P\{X < 100\} = P\{-\infty < X^* < -1\} == 0,1587.$

c) $P\{110 < X < 130\} = P\{0 < X^* < 2\} = 0,4772.$

Zadatak 8. Pretpostavimo da telesne mase 800 studenata imaju normalnu raspodelu sa srednjom težinom $m = 66kg$ i standardnim odstupanjem $\sigma = 5kg$. Naći broj studenata čija je težina

a) između $65kg$ i $75kg$.
b) veća od $72kg$.

Rezultat.

a) $P\{65 < X < 75\} = P\{-0,2 < X^* < 1,8\} = 0,5434.$

Znači da $54,34\%$ studenata ima težinu od $65kg$ do $75kg$ odnosno $800 \cdot 0,5434 = 434,72 \approx 435$ studenata.

b) $P\{X > 72\} = P\{1,2 < X^* < +\infty\} = 0,1151.$

Znači da $11,51\%$ studenata ima težinu veću od $72kg$ odnosno $800 \cdot 0,1151 = 92,08 \approx 92$ studenta.

Primer 26. Ako slučajna promenljiva X ima normalnu raspodelu $X : N(3, 4^2)$ naći verovatnoće

a) $P\{X > 9\}$,

b) $P\{X > 9 | X > 5\}$.

Rešenje.

a) $P\{X > 9\} = P\left\{\dfrac{X-3}{4} > \dfrac{9-3}{4}\right\} = P\{X^* > 1,5\} = P\{1,5 < X^* < +\infty\} =$

$= \Phi(+\infty) - \Phi(1,5) = 0,5 - 0,4332 = 0,0668.$

b) $P\{X > 9 | X > 5\} = \dfrac{P\{X > 9 \wedge X > 5\}}{P\{X > 5\}} = \dfrac{P\{X > 9\}}{P\{X > 5\}}$

Kako je $P\{X > 5\} = P\left\{\dfrac{X-3}{4} > \dfrac{5-3}{4}\right\} == P\{X^* > 0,5\} = P\{0,5 < X^* < +\infty\} =$

$= \Phi(+\infty) - \Phi(0,5) = 0,5 - 0,1915 = 0,3085,$ sledi

$P\{X > 9 | X > 5\} = \dfrac{0,0668}{0,3085} = 0,217.$

Primer 27. Pretpostavimi da debljina metalnih ploča koje se proizvode u izvesnoj fabrici ima normalnu raspodelu sa parametrima $m = 0,25cm$ i $\sigma^2 = (0,028)^2 cm^2$. Naći procenat nestandardnih ploča ako ploča ne odgovara standardu kad je njena debljina manja od $0,20cm$ ili veća od $0,28cm$.

Rešenje. $X-$ debljina proizvedene ploče $\Longrightarrow$ $X : N(0,25\,;(0,028)^2)$.

$P\{X < 0,20 \vee X > 0,28\} = P\{X < 0,20\} + P\{X > 0,28\}.$

$P\{X < 0,2\} = P\left\{\dfrac{X-0,25}{0,028} < \dfrac{0,2-025}{0,028}\right\} = P\{-\infty < X^* < -1,79\} =$

$= \Phi(-1,79) - \Phi(-\infty) = -\Phi(1,79) + \Phi(+\infty) = -0,4633 + 0,5 = 0,0367$

$P\{X > 0,28\} = P\left\{\dfrac{X-0,25}{0,028} > \dfrac{0,28-025}{0,028}\right\} = P\{1,07 < X^* < +\infty\} =$

$= \Phi(+\infty) - \Phi(1,07) = 0,5 - 0,3577 = 0,1423.$

$P\{X < 0,20 \vee X > 0,28\} = P\{X < 0,20\} + P\{X > 0,28\} = 0,04 + 0,14 = 0,18.$

Primer 28. Pretpostavimo da je poznato da se u 8 od 10 dana prosečna frekvencija električne struje u gradskoj mreži nalazi u dozvoljenim granicama. Kolika je verovatnoća da će u 50 slučajno izabranih dana, broj dana u kojima je prosečna frekvencija u dozvoljenim granicama biti veća od 45.

Rešenje.

A : događaj daje prosečna frekvencija u dozvoljenim granicama, $p = P(A) = \dfrac{8}{10} = 0,8$

X −broj dana sa dozvoljenim frekvencijama u 50 dana $\Longrightarrow X : B(50; 0,8)$ Kako je $n \geq 50$, binomna raspodela se može aproksimirati normalnom raspodelom sa parametrima

$$\mu = n \cdot p = 50 \cdot 0,8 = 40 \quad \text{i} \quad \sigma^2 = n \cdot p \cdot (1-p) = 50 \cdot 0,8 \cdot 0,2 = 8.$$

Znači $\quad X \sim N(40; 8)$.

$$P\{45 < X \leq 50\} = P\left\{\frac{45-40}{\sqrt{8}} < \frac{X-40}{\sqrt{8}} \leq \frac{50-40}{\sqrt{8}}\right\} = P\{1,77 < X^* \leq 3,54\} =$$
$$= \Phi(3,54) - \Phi(1,77) = 0,4998 - 0,4616 = 0,0382.$$

4 STATISTIČKO ZAKLJUČIVANJE

4.1 UZORAK

Osnovni problem Matematičke statistike je da se nađe raspodela obeležja na elementima date populacije. Često je taj problem u praksi gotovo nemoguće pouzdano rešiti. Razlog može biti brojnost populacije (npr. neprebrojivost) ali može biti problem i u velikim troškovima merenja vrednosti obeležja na jednom elementu populacije. Osim toga, može doći do razaranja elementa na kome se vrši merenje, npr. kad se meri radni vek sijalice.

Onda se rešava jednostavniji problem, na primer, ocenjuju se vrednosti nekog parametra koji tu populaciju, odnosno raspodelu posmatranog obeležja na njoj, dovoljno dobro opisuje. U tu svrhu se iz populacije izdvaja jedan njen deo koji se naziva **uzorak**. Zatim se proučava taj deo i ako se mogu doneti zaključci koji se odnose na celu populaciju, kaže se da je uzorak **reprezentativan**.

Izbor, na slučajan način, jednog elementa iz populacije se može smatrati eksperimentom. Skup svih ishoda ovog eksperimenta je cela populacija. Kako se svakom ishodu, tj. svakom elementu populacije, pridružuje broj (vrednost na njemu izmerenog obeležja), obeležje možemo smatrati slučajnom promenljivom. Kvalitativna obeležja su nenumerička ali se i njihove vrednosti mogu registrovati kao brojevi u postupku koji se zove **kodiranje**. Tako se i kvalitativna obeležja poistovećuju sa slučajnom promenljivom. Sledi da je svako obeležje okarakterisano svojom funkcijom raspodele $F(x)$.

Broj elemenata u uzorku n je uvek konačan i predstavlja **obim uzorka**. Neka se na populaciji Ω posmatra (meri) obeležje X ili ekvivalentno tome, neka je na skupu svih ishoda Ω zadata slučajna promenljiva X. Izbor prvog elementa za uzorak i registrovanje vrednosti obeležja X postaje nova slučajna promenljiva X_1 definisana nad istim skupom ishoda Ω (svi elementi populacije ravnopravno učestvuju u izboru). Preme tome, za $n-$dimenzionalni uzorak raspolažemo sa n slučajnih promenljivih $X_1, X_2, \dots, X_n$, odnosno raspolažemo sa $n-$dimenzionalnom slučajnom promenljivom $(X_1, X_2, \dots, X_n)$. Posle formiranja uzorka ima se niz konkretnih podataka $(x_1, x_2, \dots, x_n)$ koji se naziva **realizovani uzorak**.

Definicija. Neka se na populaciji Ω posmatra obeležje X. **Prost slučajni uzorak** obima n za obeležje X je niz nezavisnih slučajnih promenljivih $X_1, X_2, \dots, X_n$ od kojih svaka ima istu raspodelu kao i obeležje X. Odnosno, prost slučajni uzorak obima n je $n-$dimenzionalna slučajna promenljiva $(X_1, X_2, \dots, X_n)$.

Reč "slučajan" označava da su elementi uzorka na slučajan način birani iz populacije, a reč "prost" označava da su sve slučajne promenljive iz uzorka uzajamno nezavisne i da imaju istu raspodelu kao obeležje X.

Odmah se postavlja pitanje reprezentativnosti uzorka, tj. da li prost slučajni uzorak može da pruži kompletnu informaciju o raspodeli proizvoljnog obeležja X. Centralna teorema Matematičke statistike daje potvrdan odgovor na ovo pitanje. Pri tome se zahteva da obim uzorka n teži beskonačno. Pošto u praksi može da se radi samo sa konačnim obimom uzorka, znači da raspodela za obeležje X može da se odredi samo približno tačno, a tačnije ukoliko je obim uzorka n veći.

Nadalje se radi samo sa prostim slučajnim uzorkom pa se kraće naziva samo uzorak.

4.2 ETAPE STATISTIČKOG PROUČAVANJA

Statističko proučavanje čine tri etape.

1. Prikupljanje podataka. Ovo je prva etapa, najpre se definiše populacija koja se proučava kao i obeležje koje će se razmatrati. U ovoj etapi se zatim odredi obim uzorka i način izbora uzorka. Odredi se i tehnička procedura prikupljanja podataka. Podaci se mogu prikupiti raznim anketama, intervjuima, popunjavanjem raznih obrazaca, merenjima itd.

2. Sređivanje i grafičko prikazivanje podataka. Prikupljeni podaci se sređuju tabelarno i prikazuju grafički na način kako je to objašnjeno u prvoj glavi.

3. Statistička obrada podataka. Unapred planiranim statističkim metodama se obrađuju podaci. Cilj je da se dobiju statističke ocene karakteristika populacije.

Kao prvo, na osnovu uzorka se vrši ocena (procena) parametara raspodele obeležja, kao što su parametri srednje vrednosti i parametri varijabiliteta. Znači, rešavaju se problemi ocena.

Kao drugo, donose se odluke da li prihvatiti ili odbaciti određenu pretpostavku (hipotezu) koja se odnosi na osobine populacije. Znači, rešavaju se problemi testiranja hipoteza.

4.3 STRATEGIJA IZBORA UZORKA

Kao što je rečeno, u prvoj etapi statističkog proučavanja određuje se obim uzorka kao i način njegovog izbora. Uzorak se može izdvojiti iz populacije na više načina. Strategija izbora uzorka predstavlja način izdvajanja uzorka iz populacije. Navodimo neke češće primenjivane strategije.

1) Uzorak formiran po tablici slučajnih brojeva, tj. slučajni uzorak. Da bi mogla da se koristi tablica slučajnih brojeva za formiranje uzorka, populacija mora biti konačna i mora da se zna broj elemenata populacije. Zatim se svakom elementu populacije pridruži tačno jedan prirodan broj.

Tablicu slučajnih brojeva čine dekadne cifre poređane u niz i grupisane po četiri ili pet radi lakšeg čitanja. Cifre se generišu raznim tehnikama (npr. ruletom) koje garantuju slučajnost.

Primer 1. Neka se populacija Ω sastoji od 100 elemenata, tj. $\Omega = \{e_1, e_2, e_3, \ldots, e_{99}, e_{100}\}$. Indeks i elementa e_i predstavlja prirodan broj pridružen svakom elementu populacije. Formirati slučajni uzorak obima 10.

Rešenje. Jedan deo tablice slučajnih brojeva je:

4247	2149	0753	0965	0095	1826
2910	1727	8461	4465	7209	6594
4731	2591	7894	3148	7462	4435
6454	5232	0503	7255	4742	5819
5146	3267	9615	9488	6043	7436
8196	0946	4332	3956	1508	4586
0649	8169	5031	5236	6617	9369
8704	9814	3726	7721	3814	4953
2560	6405	7827	3187	7465	3278
3476	7047	9251	2851	1559	5795

Dovoljno je čitati po dve cifre iz tablice. Ako se pročita 00 uzeće se 100−ti element e_{100}. Ako se pročita 01 uzeće se prvi element e_1. Ako se pročita 07 uzeće se sedmi element e_7. Ako se pročita 37 uzeće se tridesetsedmi element e_{37}.

Slučajno se bira i cifra od koje se počinje, a da bi čitaocu bilo jasnije sada se čitaju redom. Skup brojeva je

$$42,\quad 47,\quad 21,\quad 49,\quad 07,\quad 53,\quad 09,\quad 65,\quad 00,\quad 95.$$

Dobijen je uzorak

$$\{e_{42},\quad e_{47},\quad e_{21},\quad e_{49},\quad e_7,\quad e_{53},\quad e_9,\quad e_{65},\quad e_{100},\quad e_{95}\}.$$

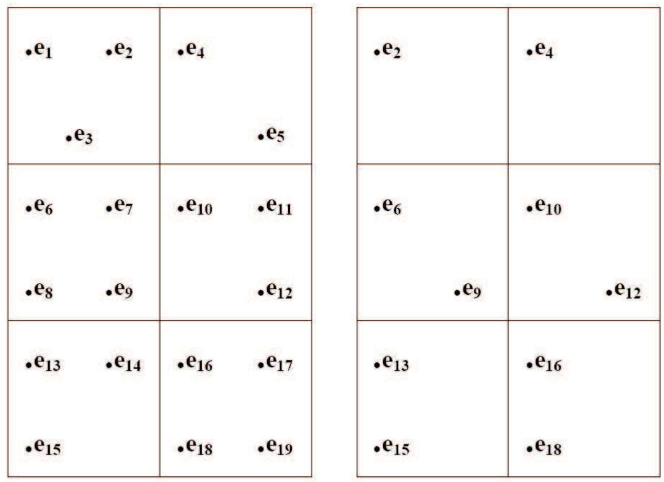
U postupku formiranja slučajnog izorka razlikuju se dva načina.

1) Kada se iz populacije izabere jedan element, na njemu se izmeri vrednost obeležja X, a zatim se on vraća u populaciju i ravnopravno učestvuje u sledećem izvlačenju. Pri tom je bitan redosled izvlačenja. To je **slučajni uzorak sa vraćanjem (sa ponavljanjem)**. U ovom uzorku se jedan isti element može pojaviti više puta.

2) Kada se iz populacije izabere jedan element, na njemu se izmeri vrednost obeležja X ali se on ne vraća u populaciju. On ne učestvuje u sledećem izvlačenju. To je **slučajni uzorak bez vraćanja (bez ponavljanja)**. Karakteristično za ovaj uzorak obima n je da je izvlačenje elemenata jedan za drugim ekvivalentno izvlačenju svih n odjednom.

Periodični uzorak. Periodični uzorak zahteva uređenost elemenata populacije, a da bi bio reprezentativan ta uređenost ne sme biti u vezi sa posmatranim obeležjem. Uređenost znači da elementi populacije formiraju niz. Zatim se iz tog niza biraju elementi na istom razmaku, npr. svaki sedmi, svaki deseti itd. Ovaj uzorak se vrlo jednostavno formira i ravnomerno je raspoređen po populaciji.

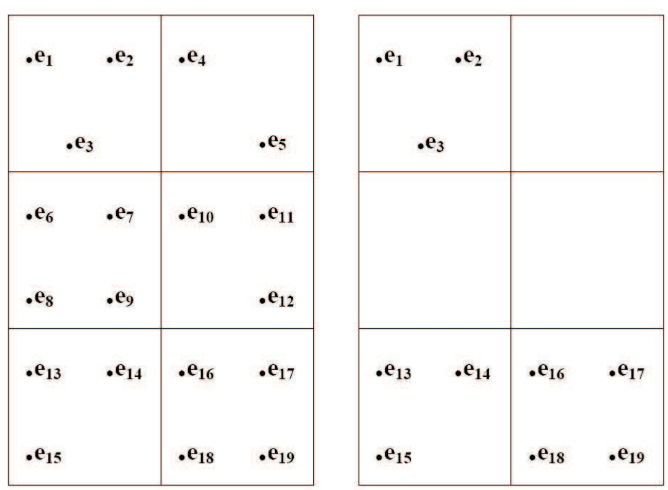
Stratifikovani (slojevit) uzorak Kod ovog uzorka se najpre populacija deli po nekom pravilu na disjunktne delove (stratume ili slojeve). To pravilo treba da obezbedi da svaki stratum bude homogen u odnosu na neko svojstvo. Zatim se iz svakog stratuma na slučajan način bira unapred određeni broj elemenata. Kod **ravnomernog stratifikovanog uzorka** se iz svakog stratuma uzima, ukoliko je moguće, isti broj elemenata. Ukoliko to nije moguće uzima se približno jednak broj elemenata. Kod **proporcionalnog stratifikovanog uzorka** se iz svakog stratuma uzima broj elemenata koji je proporcionalan njegovoj veličini.



Slika 16: Stratifikovani (slojevit) uzorak.

Grupni uzorak. Populacija se deli na disjunktne delove (grupe) koje ne moraju biti homogene po bilo kom svojstvu. Zatim se na slučajan način bira određeni broj grupa i uzimaju se svi elementi iz izabranih grupa.

Grupni uzorak se dobija jednostavnije od stratifikovanog uzorka ali stratifikovani uzorak bolje reprezentuje populaciju.



Slika 17: Grupni uzorak.

Višeetapni uzorak. Za stratifikovani uzorak se još kaže da je jednoetapni. Iz svakog stratuma se na slučajan način izvlači izvestan broj elemenata.

Međutim, ako se ne uzimaju u uzorak svi stratumi, već se na slučajan način izvuče jedan podskup stratuma pa se zatim iz svakog stratuma izvlači deo elemenata, za tako formirani uzorak se kaže da je dvoetapni stratifikovani uzorak. Mogu se formirati i višeetapni stratifikovani uzorci.

4.4 STATISTIKA

U trećoj etapi statističkog proučavanja se, između ostalog, na osnovu realizovanog uzorka ocenjuju numeričke karakteristike obeležja kao što su matematičko očekivanje, mod, medijana, disperzija itd. U tu svrhu koriste se različite funkcije uzorka, tj. funkcije koje zavise od slučajnih promenljivih $X_1, X_2, \ldots, X_n$, a koje su i same slučajne promenljive.

Definicija. **Statistika**[1] je funkcija uzorka $(X_1, X_2, \ldots, X_n)$ čiji analitički izraz ne zavisi od nepoznatih parametara raspodele obeležja populacije iz koje je uzorak uzet.

Važnije statistike su:

X_{min} – minimum uzorka

X_{max} – maksimum uzorka

$R = X_{max} - X_{min}$ – raspon uzorka

$\overline{X}_n = \dfrac{1}{n}\sum_{k=1}^{n} X_k$ – uzoračka sredina

$\overline{S}_n^2 = \dfrac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X}_n\right)^2 = \dfrac{1}{n}\sum_{k=1}^{n} X_k^2 - (\overline{X}_n)^2$ – uzoračka disperzija $(n \geq 30)$

$\widetilde{S}_n^2 = \dfrac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}_n\right)^2,$ odnosno

$\widetilde{S}_n^2 = \dfrac{1}{n-1}\sum_{k=1}^{n} X_k^2 - \dfrac{n}{n-1}(\overline{X}_n)^2$ –popravljena uzoračka disperzija $(n < 30)$

$\overline{S}_n = \sqrt{\overline{S}_n^2}$ – uzoračka standardna devijacija $(n \geq 30)$

$\widetilde{S}_n = \sqrt{\widetilde{S}_n^2}$ – popravljena uzoračka standardna devijacija $(n < 30)$

[1] Reč "statistika" se upotrebljava u dva potpuno različita značenja, da označi određenu naučnu disciplinu i da označi funkciju uzorka. Iz konteksta je uvek jasno o kom značenju se radi.

Za dvodimenzionalnu slučajnu promenljivu (X,Y) ističe se uzorački koeficijent korelacije:

$$R_{X,Y} = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(X_i-\overline{X}_n)\cdot(Y_i-\overline{Y}_n)}{\overline{S}_X\cdot\overline{S}_Y} = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}X_iY_i \ - \overline{X}_n\cdot\overline{Y}_n}{\overline{S}_X\cdot\overline{S}_Y} \qquad (n\geq 30),$$

gde je $\overline{S}_X$ standardna devijacija $\overline{S}_n$ obeležja X, a $\overline{S}_Y$ standardna devijacija $\overline{S}_n$ obeležja Y.

$$R_{X,Y} = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(X_i-\overline{X}_n)\cdot(Y_i-\overline{Y}_n)}{\widetilde{S}_X\cdot\widetilde{S}_Y} = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}X_iY_i \ - \overline{X}_n\cdot\overline{Y}_n}{\widetilde{S}_X\cdot\widetilde{S}_Y} \qquad (n<30),$$

gde je $\widetilde{S}_X$ popravljena standardna devijacija $\widetilde{S}_n$ obeležja X, a $\widetilde{S}_Y$ popravljena standardna devijacija $\widetilde{S}_n$ obeležja Y.

Statistike su slučajne promenljive. Međutim, ako se raspolaže sa realizovanim uzorkom $(x_1,x_2,\ldots,x_n)$, što je niz od n određenih brojeva, statistike postaju određeni brojevi. To su realizovane vrednosti statistika i one se obeležavaju malim slovima, na primer, $\overline{x}_n$, $\overline{s}_n^2$, $\widetilde{s}_n^2$, $r_{X,Y}$. Odnosno, to su vrednosti slučajnih promenljivih $\overline{X}_n$, $\overline{S}_n^2$, $\widetilde{S}_n^2$, $R_{X,Y}$ redom.

4.5 TAČKASTE OCENE PARAMETARA OBELEŽJA[2]

Osnovni problem Matematičke statistike je da se na osnovu uzorka $(X_1,X_2,\ldots,X_n)$ zaključi kakva je raspodela obeležja X. Često se na osnovu nekih razmatranja zna da obeležje X ima određeni tip raspodele, ali se ne znaju parametri te raspodele. Potrebno je oceniti nepoznate parametre na osnovu uzorka.

Normalna raspodela $X:N(\mu,\sigma^2)$ zavisi od dva parametra μ i σ^2. Zna se da je parametar μ matematičko očekivanje, tj. $\mu=E(X)$. Takođe se zna da je disperzija $D(X)$ jednaka parametru σ^2. Puasonova raspodela $X:\mathcal{P}(\lambda)$ zavisi od jednog parametra λ za koji važi $\lambda=E(X)=D(X)$. U slučaju da slučajna promenljiva ima binomnu raspodelu, tj. $X:B(n,p)$, može se pokazati da je $p=1-\dfrac{D(X)}{E(X)}$. Znači, ocena nepoznatih parametara najčešće se svodi na ocenu matematičkog očekivanja i disperzije obeležja.

Ako je ocena nepoznatog parametra raspodele obeležja konkretan broj kaže se da je to **tačkasta ocena nepoznatog parametra**.

Uzoračka sredina. Neka je $(X_1,X_2,\ldots,X_n)$ uzorak obima n. Uzoračka sredina, u oznaci $\overline{X}_n$, je statistika definisana sa

$$\overline{X}_n=\frac{1}{n}\cdot(X_1+X_2+\cdots+X_n)=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

Za realizovan uzorak $(x_1,x_2,\ldots,x_n)$, dobija se **realizovana uzoračka sredina**

$$\overline{x}_n=\frac{1}{n}(x_1+x_2+\cdots+x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

[2]Parametri obeležja je skraćeno od parametri raspodele obeležja (tj. slučajne promenljive).

Realizovan uzorak je niz konkretnih brojeva, to su podaci koji se mogu tabelarno i grafički urediti na načine koji su predstavljeni u prvoj glavi ove knjige. Tako se realizovana uzoračka sredina može računati i po nekoj od sledećih formula

$$\overline{x}_n = \frac{\sum\limits_{i=1}^{n} x_i}{n}, \qquad \overline{x}_n = \frac{\sum\limits_{i=1}^{k} f_i \cdot x_i}{n}, \qquad \overline{x}_n = \frac{\sum\limits_{i=1}^{K} f_i \cdot x_i'}{n}.$$

(f_i su apsolutne frekvencije; k je broj različitih vrednosti obeležja; K je broj intervala u koje su grupisane vrednosti obeležja; x_i' su sredine intervala.)

Uzoračka sredina je slučajna promenljiva, a realizovana uzoračka sredina je konkretan broj. Zato je uzoračka sredina $\overline{X}_n$ tačkasta ocena. To je najčešće korišćena ocena, pre svega kao ocena za nepoznato matematičko očekivanje.

Primer 2. Iz partije jednotipnih visokoomskih otpornika izabrano je radi kontrole 10 komada. Merenja su dala sledeća odstupanja (u kiloomima):

odstupanje- 1, 3, -2, 2, 4, -1, 5, 3, -2, 4.

Naći uzoračku sredinu odstupanja.

Rešenje. Realizovana uzoračka sredina je

$$\overline{x}_{10} = \frac{\sum\limits_{i=1}^{n} x_i}{n} = \frac{1+3-2+2+4-1+5+3-2+4}{10} = \frac{17}{10} = 1,7.$$

Primer 3. U prošloj školskoj godini ispit iz Matematičke statistike položilo je 276 studenata. Na slučajan način su izabrana 32 studenta i njihove ocene su bile:
6 6, 7, 8, 9, 6, 6, 6, 6, 10, 8, 9, 7, 7, 10, 7, 8, 8, 7, 10, 9,
9, 6, 7, 8, 8, 6, 6, 7, 9, 10, 8.
Naći uzoračku sredinu ocena.

Rešenje. Najpre se načine varijacioni niz[3] i radna tabela

x_i	f_i	$x_i \cdot f_i$
6	9	54
7	7	49
8	7	56
9	5	45
10	4	40
-	$\sum = 32$	$\sum = 244$

Realizovana uzoračka sredina je $\overline{x}_{32} = \frac{\sum\limits_{i=1}^{5} x_i \cdot f_i}{32} = \frac{244}{32} = 7,625.$

[3] Videti u prvoj glavi ove knjige.

Primer 4. Prinos pšenice u tonama po hektaru na 32 slučajno izabrane parcele iznosio je:
$12,2$ $15,2$ $13,1$ $11,2$ $13,4$ $15,4$ $16,8$ $15,2$ $12,3$ $11,2$ $11,2$ $12,6$
$13,1$ $16,5$ $15,5$ $14,2$ $15,4$ $15,6$ $15,2$ $13,2$ $16,4$ $16,2$ $13,3$ $15,6$
$15,2$ $11,9$ $16,3$ $16,2$ $11,4$ $16,5$ $15,2$ $14,0.$
Koliki je očekivani prinos pšenice?

Rešenje. Ove vrednosti obeležja je najbolje grupisati u intervale. Broj intervala K mora da zadovoljava nejednakosti

$$1 + 3,322 \log 32 \leq K \leq 5 \log 32 \quad \Longleftrightarrow \quad 6,00 \leq K \leq 7,53$$

Neka je broj intervala $K = 6$. Dužina intervala biće

$$\Delta = \frac{16,8 - 11,2}{6} \approx \frac{17 - 11}{6} = 1.$$

Sada se formira pomoćna tabela sa 6 intervala dužine 1.

x_i	x'_i	f_i	$x'_i \cdot f_i$
$[11, 12)$	11,5	4	46
$[12, 13)$	12,5	4	50
$[13, 14)$	13,5	5	67,5
$[14, 15)$	14,5	2	29
$[15, 16)$	15,5	10	155
$[16, 17)$	16,5	7	115,5
-	-	$\sum = 32$	$\sum = 463$

Očekivani prinos pšenice je

$$\overline{x}_{32} = \frac{\sum\limits_{i=1}^{6} x'_i \cdot f_i}{32} = \frac{463}{32} = 14,47.$$

Uzoračka disperzija i uzoračka standardna devijacija. Posle uzoračke sredine najčešće korišćena ocena je uzoračka disperzija. Neka je $(X_1, X_2, \ldots, X_n)$ uzorak obima $n \geq 30$. Uzoračka disperzija, u oznaci $\overline{S}_n^2$ je statistika definisana sa

$$\overline{S}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

Ako se raspolaže sa realizovanim uzorkom, onda u zavisnosti od toga kako su ti podaci uređeni, za izračunavanje realizovane uzoračke disperzije može se koristiti neka od sledećih formula:

$$\overline{s}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - (\overline{x}_n)^2$$

ili

$$\overline{s}_n^2 = \frac{1}{n} \sum_{i=1}^{k} f_i \cdot (x_i - \overline{x}_n)^2 = \frac{1}{n} \sum_{i=1}^{k} f_i \cdot x_i^2 - (\overline{x}_n)^2$$

ili

$$\overline{s}_n^2 = \frac{1}{n}\sum_{i=1}^{K} f_i \cdot (x_i' - \overline{x}_n)^2 = \frac{1}{n}\sum_{i=1}^{K} f_i \cdot (x_i')^2 - (\overline{x}_n)^2.$$

(k je broj različitih vrednosti obeležja; K je broj intervala; x_i' su sredine intervala.)
Vidi se da je za izračunavanje uzoračke disperzije potrebno prvo odrediti uzoračku sredinu. Naravno, realizovana uzoračka disperzija je konkretan broj i on predstavlja meru rasturanja vrednosti obeležja oko srednje vrednosti obeležja. Ovo su ocene za nepoznatu disperziju obeležja.

Ukoliko se raspolaže sa uzorkom obima $n < 30,$ tada se umesto uzoračke disperzije izračunava **popravljena uzoračka disperzija**. To je nova statistika, u oznaci $\widetilde{S}_n^2$, definisana sa

$$\widetilde{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Realizovana popravljena uzoračka disperzija se može računati po nekoj od formula:

$$\widetilde{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}(\overline{x}_n)^2$$

ili

$$\widetilde{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i \cdot (x_i - \overline{x}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i \cdot x_i^2 - \frac{n}{n-1}(\overline{x}_n)^2$$

ili

$$\widetilde{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{K} f_i \cdot (x_i' - \overline{x}_n)^2 = \frac{1}{n-1}\sum_{i=1}^{K} f_i \cdot (x_i')^2 - \frac{n}{n-1}(\overline{x}_n)^2$$

(k je broj različitih vrednosti obeležja; K je broj intervala; x_i' su sredine intervala.)

Uzoračka standardna devijacija se označava sa $\overline{S}_n$ i definiše se sa

$$\overline{S}_n = \sqrt{\overline{S}_n^2}.$$

Popravljena uzoračka standardna devijacija se označava sa $\widetilde{S}_n$ i definiše se sa

$$\widetilde{S}_n = \sqrt{\widetilde{S}_n^2}.$$

Primer 5. Minimalne mesečne temperature na Dunavu kod Zemuna 1998. godine su bile:
3; 2; 6; 7; 13; 19,5; 20,5; 15; 11; 4; i 1 stepen.
(Izvor: Stat. god. Jug.,1999., str. 19(izvod iz tabele)). Odrediti uzoračku sredinu i disperziju.

Rešenje. Suma vrednosti obeležja-minimalna mesečna temperatura je
$\sum_{i=1}^{n} x_i = 3 + 2 + 6 + 7 + 13 + 19 + 19,5 + 20,5 + 15 + 11 + 4 + 1 = 121.$ Sledi da je uzoračka sredina

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{121}{12} = 10,08 \text{ stepeni.}$$

Kako je obim uzorka $n = 12 < 30$ računa se realizovana popravljena uzoračka disperzija.
$\sum\limits_{i=1}^{n} x_i^2 = 3^2 + 2^2 + 6^2 + 7^2 + 13^2 + 19^2 + 19,5^2 + 20,5^2 + 15^2 + 11^2 + 4^2 + 1^2 = 1791,5$

$$\tilde{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}(\overline{x}_n)^2 = \frac{1791,5}{11} - \frac{12}{11}\cdot(10,08)^2 = 52,02.$$

Primer 6. Izračunati uzoračku disperziju u primeru 4.

Rešenje. Pomoćna tabela je:

x_i	x_i'	f_i	$x_i'\cdot f_i$	$(x_i')^2\cdot f_i$
$[11,12)$	11,5	4	46	529
$[12,13)$	12,5	4	50	625
$[13,14)$	13,5	5	67,5	911,25
$[14,15)$	14,5	2	29	420,5
$[15,16)$	15,5	10	155	2402,5
$[16,17)$	16,5	7	115,5	1905,75
-	-	$\sum = 32$	$\sum = 463$	$\sum = 6794$

$$\overline{s}_n^2 = \frac{1}{n}\sum_{i=1}^{K} f_i\cdot(x_i')^2 - (\overline{x}_n)^2 = \frac{6794}{32} - (14,47)^2 = 2,93.$$

Primer 7. U jednom mestu slučajno je izabrano 50 stanovnika koji su u radnom odnosu. Registrovane su njihove zarade u hiljadama dinara. Dobijeni su podaci

zarada x_i	[20, 30)	[30, 40)	[40, 50)	[50, 60)	[60, 70)	[70, 80)	[80, 90)
br. radnika f_i	13	14	12	8	1	1	1

.

Oceniti prosečnu zaradu zaposlenih. Izračunati uzoračku disperziju.

Rešenje. Radna tabela je

x_i	x_i'	f_i	$x_i'\cdot f_i$	$(x_i')^2\cdot f_i$
[20, 30)	25	13	325	8125
[30, 40)	35	14	490	17150
[40, 50)	45	12	540	24300
[50, 60)	55	8	440	24200
[60, 70)	65	1	65	4225
[70, 80)	75	1	75	5625
[80, 90)	85	1	85	7225
-	-	$\sum = 50$	$\sum = 2020$	$\sum = 90850$

.

$$\overline{x}_{50} = \frac{\sum\limits_{i=1}^{7} x_i' \cdot f_i}{50} = \frac{2020}{50} = 40,4 \quad \text{i} \quad \overline{s}_n^2 = \frac{1}{n}\sum_{i=1}^{7} f_i \cdot (x_i')^2 - (\overline{x}_n)^2 = \frac{90850}{50} - (40,4)^2 = 184,84.$$

Transformaciju uzorka. Može se desiti da su realizovane vrednosti velike. Da bi se lakše računalo, moguće je izvršiti transformaciju uzorka. Za pogodno izabrane vrednosti $a, b \in \mathcal{R}$ neka je $y_i = \frac{x_i - a}{b}$. Onda važi

$$\overline{x}_n = a + b \cdot \overline{y}_n \qquad \text{i} \qquad \overline{s}_X^2 = b^2 \overline{s}_Y^2$$

gde je $\overline{s}_X^2$ uzoračka disperzija pre transformacije, a $\overline{s}_Y^2$ uzoračka disperzija posle transformacije.

Primer 8. Transformacijom uzorka izračunati uzoračku sredinu i disperziju u prethodnom primeru.

Rešenje. Neka je $a = 55$, a $b = 10$, tj. $y_i = \frac{x_i' - 55}{10}$. Sada je radna tabela oblika:

x_i	x_i'	f_i	$y_i = \frac{x_i' - 55}{10}$	$y_i \cdot f_i$	$(y_i)^2 \cdot f_i$
[20, 30)	25	13	-3	-39	117
[30, 40)	35	14	-2	-28	56
[40, 50)	45	12	-1	-12	12
[50, 60)	55	8	0	0	0
[60, 70)	65	1	1	1	1
[70, 80)	75	1	2	2	4
[80, 90)	85	1	3	3	9
-	-	$\Sigma = 50$	-	$\Sigma = -73$	$\Sigma = 199$

$\overline{y} = \frac{\sum\limits_{i=1}^{7} y_i \cdot f_i}{50} = \frac{-73}{50} = -1,46$ pa je tražena uzoračka sredina
$\overline{x} = a + b \cdot \overline{y} = 55 + 10 \cdot (-1,46) = 40,4$
$\overline{s}_Y^2 = \frac{1}{n}\sum\limits_{i=1}^{7} f_i \cdot (y_i)^2 - (\overline{y})^2 = \frac{199}{50} - (-1,46)^2 = 1,8484$ sledi da je tražena uzoračka disperzija
$\overline{s}_X^2 = b^2 \cdot \overline{s}_Y^2 = 10^2 \cdot 1,8484 = 184,84.$

Uzorački koeficijent korelacije. Nekad je potrebno utvrditi, da li postoji linearna povezanost i koliko jaka, između dva obeležja X i Y. Neka je $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ uzorak obima $n \geq 30$. Uzorački koeficijent korelacije jednak je

$$R_{X,Y} = \frac{\frac{1}{n} \cdot \sum\limits_{i=1}^{n} (X_i - \overline{X}_n) \cdot (Y_i - \overline{Y}_n)}{\overline{S}_X \cdot \overline{S}_Y},$$

gde su $\overline{X}_n$ i $\overline{Y}_n$ uzoračke sredine obeležja X i Y, a $\overline{S}_X$ i $\overline{S}_Y$ uzoračke standardne devijacije ovih obeležja. Uzorački koeficijent korelacije se može izračunati jednostavnije na sledeči način:

$$R_{X,Y} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} X_i Y_i - \overline{X}_n \cdot \overline{Y}_n}{\overline{S}_X \cdot \overline{S}_Y}.$$

ukoliko je uzorak mali, $n < 30$, onda se umesto uzoračkih standardnih devijacija $\overline{S}_X$ i $\overline{S}_Y$ koriste popravljene uzoračke standardne devijacije $\widetilde{S}_X$ i $\widetilde{S}_Y$, tj.

$$R_{X,Y} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X}_n) \cdot (Y_i - \overline{Y}_n)}{\widetilde{S}_X \cdot \widetilde{S}_Y} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} X_i Y_i - \overline{X}_n \cdot \overline{Y}_n}{\widetilde{S}_X \cdot \widetilde{S}_Y}.$$

Uzorački koeficijent korelacije uzima vrednosti iz intervala $[-1, 1]$. Još važi:
$r_{X,Y} = 1$ ako između obeležja X i Y, postoji linearna veza $Y = aX + b$ pri čemu je $a > 0$,
$r_{X,Y} = -1$ ako između obeležja X i Y, postoji linearna veza $Y = aX + b$ pri čemu je $a < 0$.

Jedno tumačenje vrednosti uzoračkog koeficijenta korelacije je sledeće:
1) ako je $|r_{X,Y}| < 0,25$ smatra se da između obeležja X i Y, **ne** postoji linearna povezanost.
2) ako je $0,25 \leq |r_{X,Y}| < 0,5$ smatra se da između obeležja X i Y, postoji sasvim **neznatna** linearna povezanost.
3) ako je $0,5 \leq |r_{X,Y}| < 0,7$ smatra se da između obeležja X i Y, postoji **značajna** linearna povezanost.
4) ako je $0,7 \leq |r_{X,Y}| < 0,9$ smatra se da između obeležja X i Y, postoji **visoko značajna** linearna povezanost.
5) ako je $0,9 \leq |r_{X,Y}| < 1$ smatra se da je povezanost obeležja X i Y, praktično **linearna**.

Ako su X i Y, nezavisna obeležja onda je $r_{X,Y} = 0$, ali obrnuto ne mora da važi.

Napomena. Za velike realizovane vrednosti vrši se transformacija uzorka, npr. $Z_i = aX_i + b$ i $W_i = cY_i + d$. Ako su a i c istog znaka, onda je $R_{X,Y} = R_{Z,W}$, a ako su a i c suprotnog znaka $R_{X,Y} = -R_{Z,W}$.

Primer 9. Na uzorku od sedam porodica registrovana su primanja po članu domaćinstva i potrošnja voća.

primanja u hiljadama dinara	25	26	28	26	29	31	31
potrošnja voća u kilogramoma	6	5	8	7	8	9	8

Da li postoji linearna zavisnost između primanja i potrošnje voća u domaćinstvima?

Rešenje. Izračunaćemo uzorački koeficijent korelacije pomoću tabele:

x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
25	6	150	625	36
26	5	130	676	25
28	8	224	784	64
26	7	182	676	49
29	8	232	841	64
31	9	279	961	81
31	8	248	961	64
$\Sigma = 196$	$\Sigma = 51$	$\Sigma = 1445$	$\Sigma = 5524$	$\Sigma = 383$

Uzoračke sredine su

$$\overline{x}_7 = \frac{196}{7} = 28 \quad \text{i} \quad \overline{y}_7 = \frac{51}{7} = 7,29.$$

Pošto je uzorak mali, obima $n = 7$, računaju se popravljene uzoračke standardne devijacije.

$$\widetilde{s}_X^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} \cdot (\overline{x}_n)^2 = \frac{1}{6} \cdot 5524 - \frac{7}{6} \cdot (28)^2 = 6 \quad \Longrightarrow \quad \widetilde{s}_X = \sqrt{6} = 2,45$$

$$\widetilde{s}_Y^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} y_i^2 - \frac{n}{n-1} \cdot (\overline{y}_n)^2 = \frac{1}{6} \cdot 383 - \frac{7}{6} \cdot (7,29)^2 = 1,83 \quad \Longrightarrow \quad \widetilde{s}_Y = \sqrt{1,83} = 1,35$$

$$r_{X,Y} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i \cdot y_i - \overline{x}_n \cdot \overline{y}_n}{\widetilde{s}_X \cdot \widetilde{s}_Y} = \frac{\frac{1}{7} \cdot 1445 - 28 \cdot 7,29}{2,45 \cdot 1,35} = 0,70$$

Postoji značajna linearna povezanost između primanja i potrošnje voća u domaćinstvima.

Primer 10. Raspodela porodičnog prihoda i broja dece 100 slučajno izabranih porodica data je sledećom tabelom

	prihod u hiljadama dinara			
broj dece	[0, 10)	[10, 20)	[20, 30)	[30, 40)
0	9	14	8	6
1	11	20	9	4
2	4	7	3	1
3	1	2	1	0

Odrediti uzorački koeficijent korelacije između porodičnog prihoda i broja dece u porodici.

Rešenje. Neka su obeležja $X-$broj dece i $Y-$porodični prihod. Obeležje X ima $r = 4$ različite vrednosti: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ i $x_4 = 3$. Obeležje Y ima vrednosti grupisane u intervale koji se predstavljaju svojim sredinama: $y_1' = 5$, $y_2' = 15$, $y_3' = 25$ i $y_4' = 35$. Uslovno rečeno, obeležje Y ima $s = 4$ različite vrednosti. Sa f_{ij} se obeležava apsolutna frekvencija para (x_i, y_j'). Označimo sa $f_{i\bullet}$ i $f_{\bullet j}$ sume

$$f_{i\bullet} = f_{i1} + f_{i2} + \cdots + f_{is} \qquad \text{i} \qquad f_{\bullet j} = f_{1j} + f_{2j} + \cdots + f_{rj}.$$

Uzorački koeficijent korelacije biće

$$r_{X,Y} = \frac{\frac{1}{n}\sum\limits_{j=1}^{s} y'_j \sum\limits_{i=1}^{r} x_i f_{ij} - \overline{x}_n\overline{y}_n}{\overline{s}_X\,\overline{s}_Y}.$$

Prva pomoćna tabela je:

$X\backslash Y$	5	15	25	35	$f_{i\bullet}$	$f_{i\bullet}\cdot x_i$	$f_{i\bullet}\cdot x_i^2$
0	9	14	8	6	37	0	0
1	11	20	9	4	44	44	44
2	4	7	3	1	15	30	60
3	1	2	1	0	4	12	36
$f_{\bullet j}$	25	43	21	11	$n = 100$	$\Sigma = 86$	$\Sigma = 140$

Ova tabela pomaže da se odrede $\overline{x}_{100} = \overline{x}$ i $\overline{s}_X^2$.

$$\overline{x} = \frac{1}{100}\sum_{i=1}^{4} f_{i\bullet}\, x_i = \frac{86}{100} = 0,86$$

$$\overline{s}_X^2 = \frac{1}{100}\sum_{i=1}^{4} f_{i\bullet}\, x_i^2 - (\overline{x})^2 = \frac{140}{100} - (0,86)^2 = 0,66$$

Druga pomoćna tabela je:

y'_j	$f_{\bullet j}$	$f_{\bullet j}\cdot y'_j$	$f_{\bullet j}\cdot (y'_j)^2$
5	25	125	625
15	43	645	9675
25	21	525	13125
35	11	385	13475
		$\Sigma = 1680$	$\Sigma = 36900$

Sada je

$$\overline{y}_{100} = \overline{y} = \frac{1}{100}\sum_{j=1}^{4} f_{\bullet j}\, y'_j = \frac{1680}{100} = 16,8$$

$$\overline{s}_Y^2 = \frac{1}{100}\sum_{j=1}^{4} f_{\bullet j}\, (y'_j)^2 - (\overline{y})^2 = \frac{36900}{100} - (16,8)^2 = 86,76$$

Potrebno je još naći dvostruku sumu $\sum\limits_{j=1}^{s} y'_j \sum\limits_{i=1}^{r} f_{ij}\, x_i$ za $r = 4$ i $s = 4$. Opet će od pomoći biti tabela

x_i	f_{ij}							
0	9	14	8	6	$0\cdot 9 = 0$	$0\cdot 14 = 0$	$0\cdot 8 = 0$	$0\cdot 6 = 0$
1	11	20	9	4	$1\cdot 11 = 11$	$1\cdot 20 = 20$	$1\cdot 9 = 9$	$1\cdot 4 = 4$
2	4	7	3	1	$2\cdot 4 = 8$	$2\cdot 7 = 14$	$2\cdot 3 = 6$	$2\cdot 1 = 2$
3	1	2	1	0	$3\cdot 1 = 3$	$3\cdot 2 = 6$	$3\cdot 1 = 3$	$3\cdot 0 = 0$
					$\sum_1 = 22$	$\sum_2 = 40$	$\sum_3 = 18$	$\sum_4 = 6$

$y'_j \cdot \sum f_{ij}\, x_i = y'_1 \cdot \sum_1 + y'_2 \cdot \sum_2 + y'_3 \cdot \sum_3 + y'_4 \cdot \sum_4 = 5 \cdot 22 + 15 \cdot 40 + 25 \cdot 18 + 35 \cdot 6 = 1370$

Konačno

$$r_{X,Y} = \frac{\frac{1}{100} \sum_{j=1}^{4} y'_j \sum_{i=1}^{4} x_i f_{ij} - \overline{x} \cdot \overline{y}}{\overline{s}_X\, \overline{s}_Y} = \frac{\frac{1}{100} \cdot 1370 - 0,86 \cdot 16,8}{\sqrt{0,66} \cdot \sqrt{86,76}} = -0,10$$

Pošto je vrednost uzoračkog koeficijenta korelacije veoma mala, zaključuje se da između porodičnih prihoda i broja dece u porodici ne postoji linearna veza.

4.6 INTERVALNO OCENJIVANJE PARAMETARA OBELEŽJA

Pored tačkastih ocena nepoznatih parametara raspodele obeležja, postoje i intervalne ocene ovih parametara. Nekada je dovoljna informacija u kojim se granicama kreće prava vrednost nepoznatog parametra.

Neka obeležje X populacije Ω ima raspodelu koja zavisi od nepoznatog parametra θ. Iz populacije Ω se, na slučajan način, izvlači uzorak obima n. Taj uzorak $(x_1, x_2, \ldots, x_n)$ je realizacija $n-$dimenzionalne slučajne promenljive $(X_1, X_2, \ldots, X_n)$. Za unapred izabranu verovatnoću $1-\alpha$, koja je bliska jedinici, potrebno je naći dve statistike $\varphi_1(X_1, X_2, \ldots, X_n)$ i $\varphi_2(X_1, X_2, \ldots, X_n)$ takve da je

$$P\{\varphi_1(X_1, X_2, \ldots, X_n) \le \theta \le \varphi_2(X_1, X_2, \ldots, X_n)\} = 1-\alpha.$$

Interval $[\varphi_1, \varphi_2]$ se naziva **slučajni interval parametra θ nivoa poverenja** $(1-\alpha)\cdot 100\%$ ili kraće, **interval poverenja**.

Ako su obe granice intervala poverenja slučajne promenljive, tada se kaže da je to dvostrani interval poverenja. Ako je samo jedna granica intervala poverenja slučajna promenljiva, tada se kaže da je to jednostrani interval poverenja.

Statistike $\varphi_1(X_1, X_2, \ldots, X_n)$ i $\varphi_2(X_1, X_2, \ldots, X_n)$ su slučajne promenljive, ali ove statistike od realizovanog uzorka su brojne vrednosti. Tako, za realizovani uzorak je interval poverenja brojni interval (deo realne ose). Za jedan realizovani uzorak dobije se, na primer, jedan iterval poverenja nivoa poverenja 95%. Za drugi realizovani uzorak dobije se drugi iterval poverenja nivoa poverenja 95%, itd. Nivo poverenja 95%, znači da 95% tih intervala (ne svi) zahvata pravu vrednost nepoznatog parametra θ.

Intervalne ocene parametara normalne raspodele

Neka obeležje X ima normalnu raspodelu $N(\mu, \sigma^2)$ i neka je $(X_1, X_2, \ldots, X_n)$ uzorak obima n. Intervalna ocena za nepoznato matematičko očekivanje μ se razlikuje uzavisnosti od toga da li je disperzija σ^2 poznata ili ne.

Ako je disperzija σ^2 **poznata**, **interval poverenja za matematičko očekivanje** μ **sa nivoom poverenja** $1-\alpha$ je oblika

$$\mathbf{I}_\mu = \left[\overline{X}_n - z_{\frac{1-\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\ ,\quad \overline{X}_n + z_{\frac{1-\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right],$$

pri čemu je $\overline{X}_n$ uzoračka sredina, a broj $z_{\frac{1-\alpha}{2}}$ se čita iz tablice normalne raspodele i zadovoljava uslov $\Phi\left(z_{\frac{1-\alpha}{2}}\right) = \dfrac{1-\alpha}{2}$.

Primer 11. Izabran je uzorak od 100 porodica da se proceni količina dnevnog otpada. Dobijeni podaci su:

količina otpada x_i	[1,5; 2,5)	[2,5; 3,5)	[3,5; 4,5)	[4,5; 5,5)
br. porodica f_i	15	30	35	20

Ako se zna da je količina dnevnog otpada po porodici obeležje sa normalnom raspodelom $N(\mu;\, 2)$, odrediti interval poverenja za srednju vrednost μ sa nivoom poverenja $0,99$.

Rešenje. Disperzija je poznata $\sigma^2 = 2$. Za interval poverenja parametra μ potrebno je izračunati realizovanu uzoračku sredinu $\overline{x}_{100}$ i naći broj $z_{\frac{1-\alpha}{2}}$. Iz pomoćne tabele sledi,

x_i	x'_i	f_i	$x'_i \cdot f_i$
$[1,5;\,2,5)$	2	15	30
$[2,5;\,3,5)$	3	30	90
$[3,5;\,4,5)$	4	35	140
$[4,5;\,5,5)$	5	20	100
-	-	$\sum = 100$	$\sum = 360$

realizovana uzoračka sredina iznosi $\overline{x}_{100} = \dfrac{360}{100} = 3,6$.

$\Phi\left(z_{\frac{1-\alpha}{2}}\right) = \dfrac{1-\alpha}{2} \iff \Phi(z_{0,495}) = 0,495$, a nalaženjem broja $0,495$ u tablici normalne raspodele dobija se da je $z_{0,495} = 2,575$. Sledi da je

$\mathbf{I}_\mu = \left[\overline{x}_n - z_{\frac{1-\alpha}{2}} \cdot \dfrac{\sigma}{\sqrt{n}}\ ,\quad \overline{x}_n + z_{\frac{1-\alpha}{2}} \cdot \dfrac{\sigma}{\sqrt{n}}\right] = \left[3,6 - 2,575 \cdot \dfrac{1,41}{\sqrt{100}};\ 3,6 + 2,575 \cdot \dfrac{1,41}{\sqrt{100}}\right].$

$\mathbf{I}_\mu = [3,24;\ 3,96]$.

Ukoliko je parametar σ^2 **nepoznat**, a obim uzorka $n \geq 30$, **interval poverenja za matematičko očekivanje** μ sa nivoom poverenja $1-\alpha$ je oblika

$$\mathbf{I}_\mu = \left[\overline{X}_n - t_{n-1;\,\frac{1-\alpha}{2}} \cdot \frac{\overline{S}_n}{\sqrt{n-1}}\ ,\quad \overline{X}_n + t_{n-1;\,\frac{1-\alpha}{2}} \cdot \frac{\overline{S}_n}{\sqrt{n-1}}\right]$$

gde je $\overline{S}_n$ uzoračka standardna devijacija, a $t_{n-1;\,\frac{1-\alpha}{2}}$ je vrednost koja se čita iz tablice Studentove raspodele.

Kod malog uzorka $n < 30$, **interval poverenja za matematičko očekivanje** μ, kad je parametar σ^2 nepoznat, je

$$\mathbf{I}_\mu = \left[\overline{X}_n - t_{n-1;\,\frac{1-\alpha}{2}} \cdot \frac{\widetilde{S}_n}{\sqrt{n}}\,, \quad \overline{X}_n + t_{n-1;\,\frac{1-\alpha}{2}} \cdot \frac{\widetilde{S}_n}{\sqrt{n}}\right],$$

$\widetilde{S}_n$ je popravljena uzoračka standardna devijacija, a $t_{n-1;\,\frac{1-\alpha}{2}}$ je vrednost koja se čita iz tablice Studentove raspodele.

Primer 12. Na osnovu 9 nezavisnih merenja nekog objekta izračunata je prosečna masa $\overline{x}_9 = 42,319g$ i popravljena uzoračka disperzija $\widetilde{s}_9^2 = 5^2$. Treba oceniti stvarnu vrednost izmerene veličine sa nivoom poverenja 0,95.

Rešenje. Treba naći interval poverenja za nepoznat parametar μ, a pri tome i disperzija σ^2 je nepoznata. Kako je
$t_{n-1;\,\frac{1-\alpha}{2}} = t_{8;\,0,475} = 2,306$, sledi
$\mathbf{I}_\mu = \left[\overline{x}_n - t_{n-1;\,\frac{1-\alpha}{2}} \cdot \frac{\widetilde{s}_n}{\sqrt{n}}\,, \quad \overline{x}_n + t_{n-1;\,\frac{1-\alpha}{2}} \cdot \frac{\widetilde{s}_n}{\sqrt{n}}\right] = \left[42,319 - 2,306 \cdot \frac{5}{3};\ 42,319 + 2,306 \cdot \frac{5}{3}\right]$
$\mathbf{I}_\mu = [38,476;\, 46,162]$.

Intervalna ocena za nepoznatu disperziju σ^2 razlikuje se u slučaju da je matematičko očekivanje μ poznato ili ne.

Ako je **poznato** matematičko očekivanje μ, dvostrani **interval poverenja za nepoznatu disperziju** σ^2 sa nivoom poverenja $1-\alpha$ je

$$\mathbf{I}_{\sigma^2} = \left[\frac{\sum\limits_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{n;\,1-\frac{\alpha}{2}}}\,, \quad \frac{\sum\limits_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{n;\,\frac{\alpha}{2}}}\right],$$

pri čemu su $\chi^2_{n;\,1-\frac{\alpha}{2}}$ i $\chi^2_{n;\,\frac{\alpha}{2}}$ vrednosti koje se čitaju iz tablice χ^2-raspodele.

Odgovarajući gornji jednostrani interval poverenja je

$$\mathbf{I}_{\sigma^2} = \left(0\,, \quad \frac{\sum\limits_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{n;\,\alpha}}\right].$$

Ako je **nepoznato** matematičko očekivanje μ, dvostrani **interval poverenja za nepoznatu disperziju** σ^2 sa nivoom poverenja $1-\alpha$ je

$$\mathbf{I}_{\sigma^2} = \left[\frac{n \cdot \overline{S}_n^2}{\chi^2_{n-1;\,1-\frac{\alpha}{2}}}\,, \quad \frac{n \cdot \overline{S}_n^2}{\chi^2_{n-1;\,\frac{\alpha}{2}}}\right].$$

Vrednosti $\chi^2_{n-1;\,1-\frac{\alpha}{2}}$ i $\chi^2_{n-1;\,\frac{\alpha}{2}}$ se čitaju iz tablice χ^2-raspodele.

Odgovarajući gornji jednostrani interval poverenja je

$$\mathbf{I}_{\sigma^2} = \left(0\ ,\quad \frac{n\cdot \overline{S}_n^2}{\chi^2_{n-1;\,\alpha}}\right].$$

Primer 13. Vršeno je 30 merenja neke fizičke veličine jednim aparatom. Dobijena je uzoračka disperzija $\overline{s}_{12}^2 = 0,36$. Odrediti tačnost aparata sa nivoom poverenja 99%.

Rešenje. Tačnost aparata karakteriše disperzija merenja. Treba naći interval poverenja za nepoznatu disperziju σ^2, a parametar μ je nepoznat.
Iz $1-\alpha = 0,99 \implies \alpha = 0,01 \implies \dfrac{\alpha}{2} = 0,005$, pa je
$\chi^2_{n-1;\,1-\frac{\alpha}{2}} = \chi^2_{29;\,0,995} = 52,3$
$\chi^2_{n-1;\,\frac{\alpha}{2}} = \chi^2_{29;\,0,005} = 13,1$

$$\mathbf{I}_{\sigma^2} = \left[\frac{n\cdot \overline{s}_n^2}{\chi^2_{n-1;\,1-\frac{\alpha}{2}}}\ ,\quad \frac{n\cdot \overline{s}_n^2}{\chi^2_{n-1;\,\frac{\alpha}{2}}}\right] = \left[\frac{30\cdot 0,36}{52,3};\ \frac{30\cdot 0,36}{13,1}\right] = [0,21;\ 0,82].$$

4.7 TESTIRANJE STATISTIČKIH HIPOTEZA

Osnovni problem kojim se bavi Matematička statistika je da se utvrdi raspodela obeležja. Nekada je moguće pretpostaviti o kojoj se raspodeli radi ili je moguće učiniti pretpostavke o pojedinim karakteristikama obeležja kao što su aritmetička sredina , medijana, disperzija i slično.

Svaka pretpostavka koja se odnosi na raspodelu obeležja je **statistička hipoteza.** Postupak njenog verifikovanja pomoću uzorka naziva se **statistički test**.

Svaki statistički test istovremeno razmatra dve hipoteze koje su jedna drugoj suprostavljene. Jedna se uzima za polaznu ili **nultu hipotezu** i označava se sa H_0. Druga hipoteza se naziva **alternativna hipoteza** i označava se se sa H_1 (ili H_a). Po pravilu se za nultu hipotezu uzima ona koja se lakše verifikuje.

Za testiranje hipoteze raspolaže se sa realizovanim uzorkom $(x_1, x_2, \ldots, x_n)$ (niz brojeva) i najčešće se koristi neka statistika. Statistika koju u tom postupku koristi statistički test naziva se **test statistika**. Test statistika od realizovanog uzorka je realizovana vrednost test statistike.

Svaki statistički test zatim definiše skup C koji se naziva **kritična oblast**. Ako realizovana vrednost test statistike pripada oblasti C nulta hipoteza se odbacuje, u suprotnom se ona prihvata.

U ovom postupku moguće je načiniti dve greške: grešku prve i grešku druge vrste. **Greška prve vrste** se čini kada se odbaci nulta hipoteza H_0, a zapravo je tačna. Verovatnoća da se napravi greška prve vrste se označava sa α i naziva se **prag značajnosti testa**,

$$\alpha = P_{H_0}\{(X_1, X_2, \ldots, X_n) \in C\}.$$

Za prag značajnosti testa α najčešće se uzimaju vrednosti $0,1;\ 0,01;\ 0,05.$

Greška druge vrste se čini kada se prihvati nulta hipoteza H_0, a ona faktički nije tačna. Verovatnoća da se napravi greška druge vrste se označava sa β,

$$\beta = P_{H_1}\{(X_1, X_2, \dots, X_n) \notin C\}.$$

Statistički testovi mogu biti parametarski i neparametarski. Parametarski testovi su oni kod kojih raspodela test statistike zavisi od raspodele razmatranog obeležja. Kod neparametarskih testova raspodela test statistike ne zavisi od raspodele posmatranog obeležja.

Testiranje statističkih hipoteza ima veliku praktičnu primenu. Savremena biologija, medicina, ekonomija, agronomija, neke oblasti tehnike i druge naučne oblasti, koriste testiranje statističkih hipoteza pri obradi rezultata istraživanja.

4.8 TESTOVI HIPOTEZA O PARAMETRIMA NORMALNE RASPODELE

Testovi koji se odnose na testiranje parametara normalne raspodele spadaju u grupu parametarskih testova. Statistika koju koriste zavisi od parametara normalne raspodele.

Znači, ovi testovi mogu da se koriste kada se zna da obeležje ima normalnu raspodelu. Posle izvesnog vremena može se postaviti pitanje da li su parametri normalne raspodele nepromenjeni? Isto pitanje se postavlja i ako su se na populaciji sprovodile neke aktivnosti.

Testiranje hipoteza o parametru μ kada je σ^2 poznato

Nulta hipoteza je $H_0\,(\mu = \mu_0)$, a alternativna hipoteza je jedna od sledećih:

$$H_1\,(\mu \neq \mu_0), \qquad H_1\,(\mu > \mu_0) \quad \text{ili} \quad H_1\,(\mu < \mu_0)$$

(μ_0 je konkretan broj). Test statistika je

$$Z_0 = \frac{\overline{X}_n - \mu_0}{\sigma} \cdot \sqrt{n}$$

i ona ima standardnu normalnu raspodelu ako je nulta hipoteza tačna. Obim uzorka je n, a $\overline{X}_n$ je uzoračka sredina.

U zavisnosti od toga kako glasi alternativna hipoteza određuje se kritična oblast C, a što je prikazano u tabeli

H_0	H_1	kritična oblast C
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert z_0 \rvert \geq z_{\frac{1-\alpha}{2}}$
$\mu = \mu_0$	$\mu > \mu_0$	$z_0 \geq z_{\frac{1}{2}-\alpha}$
$\mu = \mu_0$	$\mu < \mu_0$	$z_0 \leq -z_{\frac{1}{2}-\alpha}$

.

Broj $z_{\frac{1-\alpha}{2}}$ je rešenje jednačine $\Phi(z_{\frac{1-\alpha}{2}}) = \frac{1-\alpha}{2}$ i određuje se iz tablice normalne raspodele. Takođe broj $z_{\frac{1}{2}-\alpha}$ je rešenje jednačine $\Phi(z_{\frac{1}{2}-\alpha}) = z_{\frac{1}{2}-\alpha}$ i određuje se iz tablice normalne raspodele.

Na kraju, ako realizovana vrednost test statistike z_0 upada u kritičnu oblast, nulta hipoteza se odbacuje, a alternativna prihvata. Odnosno, sa pragom značajnosti α, zaključuje se da je došlo do promene parametra μ. U suprotnom se, sa pragom značajnosti α, prihvata nulta hipoteza i može se smatrati da nije došlo do promene parametra μ.

Primer 14. Iz uzorka od 25 vekni hleba nađeno je da je srednja vrednost težine vekne $594g$. Ako je standardna devijacija
a) $\sigma = 30g$,
b) $\sigma = 10g$,
sa pragom značajnosti $\alpha = 0,05$, testirati hipotezu $H_0(\mu = 600g)$ protiv alternativne hipoteze $H_1(\mu \neq 600g)$. Pretpostavlja se normalna raspodela.

Rešenje.
a) Realizovana vrednost test statistike je $z_0 = \dfrac{\overline{x}_n - \mu_0}{\sigma} \cdot \sqrt{n} = \dfrac{594 - 600}{30} \cdot \sqrt{25} = -1.$
Kritična oblast (iz tabele) je $C: \ |z_0| \geq z_{\frac{1-\alpha}{2}}$.
Iz $\Phi(z_{\frac{1-\alpha}{2}}) = z_{\frac{1-\alpha}{2}} \iff \Phi(z_{0,475}) = 0,475$ (kad se nađe broj $0,475$ u tablici normalne raspodele) sledi da je $z_{\frac{1-\alpha}{2}} = z_{0,475} = 1,96$. Znači kritična oblast je $C: \ |-1| \geq 1,96$ ili interval $(-\infty; -1,96] \cup (1,96; +\infty)$. Realizovana vrednost test statistike ne pripada kritičnoj oblasti i nulta hipoteza se ne odbacuje. Pri standardnoj devijaciji od $30g$ težina vekni ne odstupa značajno.

b) Realizovana vrednost test statistike je $z_0 = \dfrac{\overline{x}_n - \mu_0}{\sigma} \cdot \sqrt{n} = \dfrac{594 - 600}{10} \cdot \sqrt{25} = -3.$
Kritična oblast (iz tabele) je $C: \ |z_0| \geq z_{\frac{1-\alpha}{2}}$.
Kao pod a) kritična oblast je $C: \ |-3| \geq 1,96$ ili interval $(-\infty; -1,96] \cup (1,96; +\infty)$. Realizovana vrednost test statistike pripada kritičnoj oblasti i nulta hipoteza se odbacuje. Pri standardnoj devijaciji od $10g$ težina vekni odstupa značajno.

Testiranje hipoteza o parametru μ kada je σ^2 nepoznato

Ispituje se da li je došlo do promene očekivanja μ ali disperzija nije poznata. Testira se nulta hipoteza $H_0\,(\mu = \mu_0)$, a alternativna hipoteza je jedna od sledećih:

$$H_1\,(\mu \neq \mu_0), \qquad H_1\,(\mu > \mu_0) \quad \text{ili} \quad H_1\,(\mu < \mu_0)$$

(μ_0 je konkretan broj).

Za velike uzorke (obima $n \geq 30$) test statistika je $\quad t_{n-1} = \dfrac{\overline{X}_n - \mu_0}{\overline{S}_n} \cdot \sqrt{n-1}$,

a za male uzorke (obima $n < 30$) test statistika je $\quad t_{n-1} = \dfrac{\overline{X}_n - \mu_0}{\widetilde{S}_n} \cdot \sqrt{n}.$

Obe ove statistike, pod uslovom da je nulta hipoteza tačna, imaju Studentovu raspodelu sa $n-1$ stepeni slobode i obe zavise od uzoračke sredine $\overline{X}_n$. Ako je uzorak veliki, test statistika zavisi od uzoračke standardne devijacije $\overline{S}_n$, a ako je uzorak mali od popravljene uzoračke standardne devijacije $\widetilde{S}_n$.

Kritična oblast zavisi od alternativne hipoteze i data je u tabeli

H_0	H_1	kritična oblast C
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert t_{n-1}\rvert \geq t_{n-1\,;\,\frac{1-\alpha}{2}}$
$\mu = \mu_0$	$\mu > \mu_0$	$t_{n-1} \geq t_{n-1\,;\,\frac{1}{2}-\alpha}$
$\mu = \mu_0$	$\mu < \mu_0$	$t_{n-1} \leq -t_{n-1\,;\,\frac{1}{2}-\alpha}$

.

Brojevi $t_{n-1\,;\,\frac{1-\alpha}{2}}$ i $t_{n-1\,;\,\frac{1}{2}-\alpha}$ se određuju iz tablice Studentove raspodele.

Zaključak o prihvatanju ili odbacivanju nulte hipoteze izvodi se kao u prethodnom testiranju.

Primer 15. Izvršeno je 30 merenja procenta metala u rudi na jednom geološkom lokalitetu. Rezultati su sledeći:

procenat metala x_i%	[0,3; 0,8)	[0,8; 1,3)	[1,3; 1,8)	[1,8; 2,3)	[2,3; 2,8)	[2,8; 3,3)
f_i	5	5	4	9	5	2

.

Pod pretpostavkom da procenat metala ima normalnu raspodelu, testirati hipotezu da je očekivani procenat $1,75$ protiv alternativne da je različit od $1,75$, sa pragom značajnosti $0,05$.

Rešenje. Za uzorak $n = 30$ test statistika je $t_{n-1} = \dfrac{\overline{X}_n - \mu_0}{\overline{S}_n} \cdot \sqrt{n-1}$. Najpre treba naći realizovanu uzoračku sredinu i realizovanu uzoračku standardnu devijaciju.

x_i	x_i'	f_i	$x_i' \cdot f_i$	$(x_i')^2 \cdot f_i$
[0,3; 0,8)	0,55	5	2,75	1,51
[0,8; 1,3)	1,05	5	5,25	5,51
[1,3; 1,8)	1,55	4	6,20	9,61
[1,8; 2,3)	2,05	9	18,45	37,82
[2,3; 2,8)	2,55	5	12,75	32,51
[2,8; 3,3)	3,05	2	6,10	18,60
-	-	$\Sigma = 30$	$\Sigma = 51,50$	$\Sigma = 105,56$

$$\overline{x}_{30} = \frac{51,50}{30} = 1,72$$

$$\overline{s}_{30}^2 = \frac{105,56}{30} - (1,72)^2 = 0,56 \quad \Longrightarrow \quad \overline{s}_{30} = \sqrt{0,56} = 0,75$$

Realizovana vrednost test statistike je $t_{29} = \dfrac{1,72 - 1,75}{0,75} \cdot \sqrt{29} = -0,22$.

Iz tablice Studentove raspodele se čita da je broj $t_{n-1\,;\,\frac{1-\alpha}{2}} = t_{29;\,0,475} = 2,045$. Kako je $\lvert -0,22\rvert \geq 2,045$ netačno, zaključuje se da realizovana vrednost test statistike ne pripada kritičnoj oblasti pa se nulta hipoteza prihvata. Sa pragom značajnosti $0,05$ može se smatrati da je očekivani procenat metala u rudi $1,75$.

Testiranje hipoteze o parametru σ^2 kada je μ poznato

Nulta hipoteza je $H_0\,(\sigma^2 = \sigma_0^2)$, a alternativna hipoteza je jedna od sledećih:

$$H_1\,(\sigma^2 \neq \sigma_0^2), \qquad H_1\,(\sigma^2 > \sigma_0^2) \quad \text{ili} \quad H_1\,(\sigma^2 < \sigma_0^2)$$

(σ_0^2 je konkretan broj). Test statistika je

$$\chi_0^2 = \frac{\sum\limits_{i=1}^{n} (X_i - \mu)^2}{\sigma_0^2}$$

i ona ima χ^2 raspodelu sa n stepeni slobode ako je nulta hipoteza tačna.
Kritična oblast zavisi od alternativne hipoteze i data je u tabeli

H_0	H_1	kritična oblast C
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi_0^2 \leq \chi^2_{n\,;\,\frac{\alpha}{2}} \vee \chi_0^2 \geq \chi^2_{n\,;\,1-\frac{\alpha}{2}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi_0^2 \geq \chi^2_{n\,;\,1-\alpha}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi_0^2 \leq \chi^2_{n\,;\,\alpha}$

Brojevi: $\quad \chi^2_{n\,;\,\frac{\alpha}{2}} \quad \chi^2_{n\,;\,1-\frac{\alpha}{2}} \quad \chi^2_{n\,;\,1-\alpha} \quad \chi^2_{n\,;\,\alpha} \quad$ se određuje iz tablice χ^2 raspodele.
Zaključak o prihvatanju ili odbacivanju nulte hipoteze izvodi se kao u prethodnim testiranjima.

Primer 16. U slučajno uzetim uzorcima vode iz jednog jezera, ispitan je sadržaj B_2O_3 (bor trioksida) i dobijeni su, u 1 : 100000, rezultati: 45, 43, 37, 41, 43, 60, 58, 61, 60, 58, 60, 60, 61, i 58. Pretpostavljajući da sadržaj bor trioksida ima normalnu raspodelu $N(50, \sigma^2)$, sa pragom značajnosti $\alpha = 0,01$ testirati hipotezu da je disperzija jednaka 100.

Rešenje. Testira se hipoteza $H_0(\sigma^2 = 100)$ protiv alternativne $H_1(\sigma^2 \neq 100)$. Test statistika je $\chi_0^2 = \dfrac{\sum\limits_{i=1}^{n} (X_i - \mu)^2}{\sigma_0^2}$, a njena realizovana vrednost $\dfrac{\sum\limits_{i=1}^{n} (x_i - \mu)^2 \cdot f_i}{\sigma_0^2}$ se računa pomoću tabele:

x_i	f_i	$(x_i - 50)^2$	$(x_i - 50)^2 \cdot f_i$
37	1	169	169
41	1	81	81
43	2	49	98
45	1	25	25
58	3	64	192
60	4	100	400
61	2	121	242
-	-	-	$\Sigma = 1207$

.

Realizovana vrednost test statistike je $\chi_0^2 = \dfrac{1207}{100} = 12,07$. Iz tablice se čita
$\chi^2_{n\,;\,\frac{\alpha}{2}} = \chi^2_{14;\,0,005} = 4,07$ i $\chi^2_{n\,;\,1-\frac{\alpha}{2}} = \chi^2_{14;\,0,995} = 31,3$. Kritična oblast je
$12,07 \leq 4,07$ ili $12,07 \geq 31,3$. Kako je ovo netačno, sledi da realizovana vrednost test statistike ne upada u kritičnu oblast. Nulta hipoteza da je $\sigma^2 = 100$ se prihvata.

Testiranje hipoteze o parametru σ^2 kada je μ nepoznato

Nulta hipoteza je $H_0\,(\sigma^2 = \sigma_0^2)$, a alternativna hipoteza je jedna od sledećih:

$$H_1\,(\sigma^2 \neq \sigma_0^2), \qquad H_1\,(\sigma^2 > \sigma_0^2) \quad \text{ili} \quad H_1\,(\sigma^2 < \sigma_0^2)$$

(σ_0^2 je konkretan broj).

Za velike uzorke (obima $n \geq 30$,) test statistika je $\chi_0^2 = \dfrac{n \cdot \overline{S}_n^2}{\sigma_0^2}$,

a za male uzorke ($n < 30$,) test statistika je $\chi_0^2 = \dfrac{(n-1)\widetilde{S}_n^2}{\sigma_0^2}$.

Obe ove statistike, pod uslovom da je nulta hipoteza tačna, imaju χ^2 raspodelu sa $n-1$ stepeni slobode. Za veliki uzorak test statistika koristi uzoračku disperziju $\overline{S}_n^2$, a za mali uzorak, uzoračku popravljenu disperziju $\widetilde{S}_n^2$.
Kritične oblasti C zavise od alternativnih hipoteza i prikazane su u tabeli:

H_0	H_1	kritična oblast C
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\chi_0^2 \leq \chi^2_{n-1\,;\,\frac{\alpha}{2}} \vee \chi_0^2 \geq \chi^2_{n-1\,;\,1-\frac{\alpha}{2}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi_0^2 \geq \chi^2_{n-1\,;\,1-\alpha}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi_0^2 \leq \chi^2_{n-1\,;\,\alpha}$

.

Brojevi: $\chi^2_{n-1\,;\,\frac{\alpha}{2}}$ $\chi^2_{n-1\,;\,1-\frac{\alpha}{2}}$ $\chi^2_{n-1\,;\,1-\alpha}$ $\chi^2_{n-1\,;\,\alpha}$ se određuje iz tablice χ^2 raspodele.

Ako realizovana vrednost test statistike χ_0^2 upada u kritičnu oblast, tada se sa pragom značajnosti α zaključuje da je došlo do promene vrednosti parametra σ^2. U suprotnom se smatra da nije došlo do promene vrednosti parametra σ^2.

Primer 17. Mašina pakuje prašak sa disperzijom $\sigma_0^2 = 0,8g$. Posle pet časova rada na uzorku obima $n = 25$ konstatovano je da je realizovana uzoračka disperzija $\widetilde{s}^2 = 1g$. Da li se sa pragom značajnosti $\alpha = 0,05$ može zaključiti da je proces rada u okviru standardnog.

Rešenje. Testira se hipoteza $H_0(\sigma^2 = 0,8)$ protiv alternativne $H_1(\sigma^2 > 0,8)$. Kako je obim uzorka mali, test statistika je $\chi_0^2 = \dfrac{(n-1)\widetilde{S}_n^2}{\sigma_0^2}$. Realizovana vrednost test statistike je $\chi_0^2 = \dfrac{24 \cdot 1}{0,8} = 30$. Za kritičnu oblast je $\chi_0^2 \geq \chi^2_{n-1\,;\,1-\alpha} \iff 30 \geq \chi^2_{24;\,0,95} \iff 30 \geq 36,4$ netačno. Nulta hipoteza se prihvata, tj. može se smatrati da mašina radi po standardu.

4.9 PIRSONOV χ^2 TEST

Pirsonov χ^2 test je neparametarski test, tj. njegova test statistika ne zavisi od raspodele posmatranog obeležja.

Kao prvo, on se koristi za ispitivanje saglasnosti podataka sa pretpostavljenom raspodelom.

Kao drugo, koristi se za ispitivanje nezavisnosti dva obeležja, posmatrana na istoj populaciji.

Ispitivanje saglasnosti uzorka sa pretpostavljenom raspodelom

1. Dat je uzorak obima n iz populacije sa obeležjem X. Postavlja se nulta hipoteza H_0 da obeležje ima raspodelu čija je funkcija raspodele F_0, protiv alternativne da nema tu raspodelu. Bez obzira da li je obeležje diskretno ili apsolutno neprekidno, od pomoći može biti neki od grafičkih metoda (poligon, histogram, itd.) da se pretpostavi raspodela.

2. Ako pretpostavljena raspodela ima nepoznate parametre, oni se ocenjuju na osnovu uzorka.

3. Realna prava se deli na K disjunktnih intervala $S_1, S_2, \ldots, S_K$ čija je unija tačno $\mathcal{R}$. Svaki interval mora da sadrži bar 5 elemenata realizovanog uzorka. Ako postoje intervali koji sadrže manje od 5 elemenata, oni se pridružuju susednim intervalima.

4. Pod pretpostavkom da je nulta hipoteza H_0 tačna, izračunavaju se teorijske verovatnoće

$$p_{0i} = P\{X \in S_i\}, \qquad i = 1, 2, \ldots, K.$$

5. U svakom intervalu S_i se izračunavaju teorijske apsolutne frekvencije $\widehat{f}_i$

$$\widehat{f}_i = n \cdot p_{0i}, \qquad i = 1, 2, \ldots, K.$$

Ove frekvencije predstavljaju očekivani broj elemenata u svakom intervalu S_i pod uslovom da je nulta hipoteza tačna.

6. Izračunava se realizovana vrednost test statistike χ_0^2

$$\chi_0^2 = \sum_{i=1}^{K} \frac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i}.$$

Sada je f_i ona frekvencija koja se pridružuju intervalu S_i.

7. Kritična oblast veličine α je

$C: \quad \chi_0^2 \geq \chi^2_{K-1;\,1-\alpha}$ ako nije bilo nepoznatih parametara raspodele, a
$C: \quad \chi_0^2 \geq \chi^2_{K-l-1;\,1-\alpha}$ ako je bilo ocenjeno l nepoznatih parametara raspodele.

8. Donosi se zaključak o prihvatanju ili odbacivanju nulte hipoteze.

Primer 18. Primenom χ^2-testa, sa pragom značajnosti $\alpha = 0,05$ ispitati saglasnost podataka sa normalnom raspodelom. Podaci su dati sledećom tabelom:

[6, 8)	[8, 10)	[10, 12)	[12, 14)	[14, 16)	[16, 18)	[18, 20)	[20, 22)	[22, 24)	[24, 26)
4	5	16	22	21	12	6	3	2	1

Rešenje. Parametri normalne raspodele su nepoznati, zato se prvo oni ocene.

x_i	x_i'	f_i	$x_i' \cdot f_i$	$(x_i')^2 \cdot f_i$
[6, 8)	7	4	28	196
[8, 10)	9	5	45	405
[10, 12)	11	16	176	1936
[12, 14)	13	22	286	3718
[14, 16)	15	21	315	4725
[16, 18)	17	12	204	3468
[18, 20)	19	6	114	2166
[20, 22)	21	3	63	1323
[22, 24)	23	2	46	1058
[24, 26)	25	1	25	625
		$\Sigma = 92$	$\Sigma = 1302$	$\Sigma = 19620$

Ocena za parametar μ je realizovana uzoračka sredina $\overline{x} = \dfrac{1302}{92} = 14,15.$

Ocena za parametar σ^2 je realizovana uzoračka disperzija $\sigma^2 = \dfrac{19620}{92} - (14,15)^2 = 13,04.$

Nulta hipoteza je da obeležje sa datom distribucijom frekvencija ima normalnu raspodelu $N\,(14,15;\ (3,61)^2)$.
Alternativna hipoteza je da ovo obeležje nema normalnu raspodelu.

U sledećem koraku odrede se intervali S_i čija je unija realna prava $\mathcal{R}$. Kako interval $[6,8)$ ima apsolutnu frekvenciju manju od pet pridružićemo ga sledećem pa će biti $S_1 = (-\infty, 10)$. Radi efikasnosti sledeće intervale odmah unosimo u radnu tabelu.

S_i	f_i–za S_i	p_{0i}	$\widehat{f}_i = 92 \cdot p_{0i}$	$\dfrac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i}$
$(-\infty$, 10)	9	0,1251	11,5092	0,5470
[10, 12)	16	0,1491	13,7172	0,3799
[12, 14)	22	0,2098	19,3016	0,3772
[14, 16)	21	0,2110	19,4120	0,1299
[16, 18)	12	0,1627	14,9684	0,5887
[18, 20)	6	0,0897	8,2524	0,6148
[20, $+\infty$)	6	0,0526	4,8392	0,2784
	$\Sigma = 92$			χ_0^2 =2,9159

Pod pretpostavkom da je hipoteza tačna, tj. da je raspodela normalna $N\ :\ (14,15,(3,61)^2)$ računaju se teorijske verovatnoće p_{0i}, i očekivane apsolutne frekvencije $\widehat{f}_i$.

$p_{01} = P\{X \in S_1\} =$
$= P\{-\infty < X < 10\} = P\{-\infty < X^* < -1,15\} = \Phi(-1,15) - \Phi(-\infty)\} = 0,1251$

$p_{02} = P\{X \in S_2\} =$
$P\{10 < X < 12\} = P\{-1,15 < X^* < -0,60\} = \Phi(-0,60) - \Phi(-1,15)\} = 0,1491$

$p_{03} = P\{X \in S_3\} =$
$= P\{12 < X < 14\} = P\{-0,60 < X^* < -0,04\} = \Phi(-0,04) - \Phi(-0,60)\} = 0,2098$

$p_{04} = P\{X \in S_4\} =$
$= P\{14 < X < 16\} = P\{-0,04 < X^* < 0,51\} = \Phi(0,51) - \Phi(-0,04)\} = 0,2110$
$p_{05} = P\{X \in S_5\} = P\{16 < X < 18\} = P\{0,51 < X^* < 1,07\} = \Phi(1,07) - \Phi(0,51)\} = 0,1627$
$p_{06} = P\{X \in S_6\} = P\{18 < X < 20\} = P\{1,07 < X^* < 1,62\} = \Phi(1,62) - \Phi(1,07)\} = 0,0897$
$p_{07} = 1 - p_{01} - p_{02} - p_{03} - p_{04} - p_{05} - p_{06} = 0,0526.$
Realizovana vrednost test statistike je

$$\chi_0^2 = \sum_{i=1}^{7} \frac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i} = 2,9159.$$

Broj intervala S_i je $K = 7$, ocenjena su dva parametra pa je $l = 2$ i kritična oblast C sa pragom značajnosti $\alpha = 0,05$ je

$C: \quad \chi_0^2 \geq \chi^2_{K-l-1;\,1-\alpha}$
$C: \quad \chi_0^2 \geq \chi^2_{7-2-1;\,1-0,05}$
$C: \quad 2,9159 \geq \chi^2_{4;\,0,95}$
$C: \quad 2,9159 \geq 9,49$ Znači kritična oblast veličine $0,05$ je interval $[9,49; +\infty)$.

Pošto je $2,9159 \geq 9,49$ netačno, uzorak ne pripada kritičnoj oblasti C i hipoteza H_0 se prihvata.

Primer 19. Za 1100 slučajno odabranih žena raspodela broja rođene dece je

broj dece	0	1	2	3	4	5	6	7
broj žena	232	313	360	130	52	10	2	1

.

Da li se sa pragom značajnosti $\alpha = 0,05$ može smatrati da raspodela dece ima Puasonovu raspodelu.

Rešenje. Puasonova raspodela ima jedan parametar λ i on je nepoznat. Ocena za parametar λ je realizovana uzoračka sredina.

x_i	f_i	$x_i \cdot f_i$	S_i	f_i–za S_i	p_{0i}	$\widehat{f}_i = 1100 \cdot p_{0i}$	$\frac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i}$
0	232	0	$(-\infty, 1)$	232	0,212	233,2	0,006
1	313	313	$[1, 2)$	313	0,329	361,9	6,607
2	360	720	$[2, 3)$	360	0,255	280,5	22,532
3	130	390	$[3, 4)$	130	0,132	145,2	1,591
4	52	208	$[4, 5)$	52	0,051	56,1	0,300
5	10	50	$[5, +\infty)$	13	0,021	23,1	4,416
6	2	12	-	-	-	-	-
7	1	7	-	-	-	-	-
	$\Sigma = 1100$	$\Sigma = 1700$	-	-	-	-	$\chi_0^2 = 35,452$

Realizovana uzoračka sredina je $\overline{x} = \dfrac{1700}{1100} = 1,55.$

Nulta hipoteza je da raspodela dece ima Puasonovu raspodelu $\mathcal{P}(1,55)$, a alternativna da nema Puasonovu raspodelu.

U sledećem koraku odredićemo intervale S_i, a zatim teorijske verovatnoće p_{0i} pod pretpostavkom da su podaci saglasni sa Puasonovom raspodelom $\mathcal{P}(1,55)$.

$p_{01} = P\{X \in S_1\} = P\{X = 0\} = \dfrac{1,55^0}{0!} \cdot e^{-1,55} = 0,212$

$p_{02} = P\{X \in S_2\} = P\{X = 1\} = \dfrac{1,55^1}{1!} \cdot e^{-1,55} = 0,329$

$p_{03} = P\{X \in S_3\} = P\{X = 2\} = \dfrac{1,55^2}{2!} \cdot e^{-1,55} = 0,255$

$p_{04} = P\{X \in S_4\} = P\{X = 3\} = \dfrac{1,55^3}{3!} \cdot e^{-1,55} = 0,132$

$p_{05} = P\{X \in S_5\} = P\{X = 4\} = \dfrac{1,55^4}{4!} \cdot e^{-1,55} = 0,051$

$p_{06} = 1 - p_{01} - p_{02} - p_{03} - p_{04} - p_{05} = 0,021.$

Realizovana vrednost test statistike je

$$\chi_0^2 = \sum_{i=1}^{7} \frac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i} = 35,452.$$

Broj intervala S_i je $K = 6$, ocenjen je jedan parametar pa je $l = 1$ i kritična oblast C sa pragom značajnosti $\alpha = 0,05$ je

$C: \quad \chi_0^2 \geq \chi^2_{K-l-1;\,1-\alpha}$

$C: \quad \chi_0^2 \geq \chi^2_{6-1-1;\,1-0,05}$

$C: \quad 35,452 \geq \chi^2_{4;\,0,95}$

$C: \quad 35,452 \geq 9,49.$ Znači kritična oblast veličine $0,05$ je interval $[9,49; +\infty)$.

Pošto je $35,452 \geq 9,49$ tačno, uzorak pripada kritičnoj oblasti C i hipoteza H_0 se odbacuje. Raspodela dece iz ovog primera nije saglasna sa Puasonovom raspodelom.

Primer 20. Iz skupa proizvoda na slučajan način se izvlače grupe od po 5 proizvoda. Utvrđuje se broj prvoklasnih nproizvoda u svakoj grupi. Posle izvlačenja 100 takvih grupa dobijeni su sledeći podaci

br. prvoklasnih proizvoda	0	1	2	3	4	5
br. grupa	4	10	18	30	23	15

Koristeći χ^2-test, sa pragom značajnosti $\alpha = 0,01$, proveriti hipotezu da obeležje $X-$broj prvoklasnih proizvoda među 5 izvučenih, ima binomnu raspodelu $B:(5,p)$.

Rešenje. Jedan parametar je poznat $n = 5$. Parametar p je nepoznat. Poznato je da slučajna promenljiva X sa binomnom raspodelom ima matematičko očekivanje $E(X) = n \cdot p$. Odatle je $p = \dfrac{E(X)}{n}$ i ocena za parametar p je $\dfrac{\overline{x}}{n} = \dfrac{\overline{x}}{5}$.

Kako je $\overline{x} = \dfrac{303}{100} = 3,03$ sledi $p = \dfrac{3,03}{5} = 0,606 \approx 0,6.$

Nulta hipoteza je da ova slučajna promenljiva ima binomnu raspodelu $B(5; 0,6)$.

x_i	f_i	$x_i \cdot f_i$	S_i	f_i–za S_i	p_{01}	$\widehat{f}_i = 100 \cdot p_{0i}$	$\dfrac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i}$
0	4	0					
1	10	10	$(-\infty, 2)$	14	0,0870	8,70	3,2287
2	18	36	$[2, 3)$	18	0,2304	23,04	1,1025
3	30	90	$[3, 4)$	30	0,3456	34,56	0,6017
4	23	92	$[4, 5)$	23	0,2592	25,92	0,3290
5	15	75	$[5, +\infty)$	15	0,0778	7,78	6,7003
	$\Sigma = 100$	$\Sigma = 303$		$\Sigma = 100$			$\chi_0^2 = 11,9622$

$p_{01} = P\{X \in S_1\} = P\{X = 0\} + P\{X = 1\} = \binom{5}{0}(0,6)^0(0,4)^5 + \binom{5}{1}(0,6)^1(0,4)^4 =$
$= 0,0102 + 0,0768 = 0,0870$

$p_{02} = P\{X \in S_2\} = P\{X = 2\} = \binom{5}{2}(0,6)^2(0,4)^3 = \dfrac{5 \cdot 4}{2 \cdot 1} \cdot 0,36 \cdot 0,064 = 0,2304$

$p_{03} = P\{X \in S_3\} = P\{X = 3\} = \binom{5}{3}(0,6)^3(0,4)^2 = \dfrac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} \cdot 0,216 \cdot 0,16 = 0,3456$

$p_{04} = P\{X \in S_4\} = P\{X = 4\} = \binom{5}{4}(0,6)^4(0,4)^1 = \binom{5}{1} \cdot 0,1296 \cdot 0,4 = 0,2592$

$p_{05} = 1 - \sum\limits_{i=1} 4 = 0,0778$

Realizovana vrednost test statistike je

$$\chi_0^2 = \sum_{i=1}^{7} \frac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i} = 11,9622.$$

Broj intervala S_i je $K = 5$, ocenjen je jedan parametar pa je $l = 1$ i kritična oblast C sa pragom značajnosti $\alpha = 0,01$ je

$C: \quad \chi_0^2 \geq \chi^2_{K-l-1;\,1-\alpha}$

$C: \quad \chi_0^2 \geq \chi^2_{5-1-1;\,1-0,01}$

$C: \quad 11,9622 \geq \chi^2_{3;\,0,99}$

$C: \quad 11,9622 \geq 11,3$

Znači kritična oblast veličine $0,01$ je interval $[11,3; +\infty)$.
Pošto je $11,9622 \geq 11,3$ tačno, uzorak pripada kritičnoj oblasti C i hipoteza H_0 se odbacuje, tj. odbacuje se hipoteza o binomnoj raspodeli.

Zadatak. Primenom χ^2-testa, sa pragom značajnosti $\alpha = 0,01$ proveriti hipotezu da obeležje X ima normalnu raspodelu $N(0;2^2)$, ako uzorak od 200 elemenata ima vrednosti zadate tabelom

x_i	[-6, -4)	[-4, -2)	[-2, 2)	[2, 3)	[3, 4)
f_i	6	30	140	16	8

.

Rezultat. Parametri raspodele su poznati. Iz $\chi_0^2 \geq \chi^2_{4;\,0,99} \iff 3,2946 \geq 13,3$ sledi da nema razloga da se odbaci nulta hipoteza.

Ispitivanje nezavisnosti dva obeležja

Na populaciji je moguće posmatrati i dva obeležja X i Y, (može i više od dva). Distribucija frekvencija se onda najčešće predstavlja tabelom kontigencije

X/Y	J_1	J_2	$\dots$	J_s	$\sum$
I_1	f_{11}	f_{12}	$\dots$	f_{1s}	$f_{1\bullet}$
I_2	f_{21}	f_{22}	$\dots$	f_{2s}	$f_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I_r	f_{r1}	f_{r2}	$\dots$	f_{rs}	$f_{r\bullet}$
$\sum$	$f_{\bullet 1}$	$f_{\bullet 2}$	$\dots$	$f_{\bullet s}$	n

gde su:

$I_1, I_2, \dots, I_r$ i $J_1, J_2, \dots, J_s$ vrednosti obeležja (konkretni brojevi) ili intervali,
f_{ij} apsolutna frekvencija para (I_i, J_j) u realizovanom uzorku,
$f_{i\bullet} = f_{i1} + f_{i2} + \dots + f_{is}$ i $f_{\bullet j} = f_{1j} + f_{2j} + \dots + f_{rj}$.

Nulta hipoteza je H_0 (X i Y su nezavisna obeležja),
a alternativna je H_1 (X i Y nisu nezavisna obeležja).

Test statistika je

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(f_{ij} - \widehat{f}_{ij}\right)^2}{\widehat{f}_{ij}} = n \cdot \left(\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{f_{ij}^2}{f_{i\bullet} f_{\bullet j}} - 1\right),$$

a kritična oblast $C : \chi_0^2 \geq \chi^2_{(r-1)(s-1);\,1-\alpha}$.

Primer 21. Ispituje se nezavisnost temperature i vlažnosti vazduha. Izdvojeni su podaci za neko mesto, za mesec juli tokom 10 godina:

	visoka temperatura	prosečna temperatura	$\sum$
visoka vlažnost	14 dana	36 dana	$f_{1\bullet} = 50$
prosečna vlažnost	59 dana	201 dan	$f_{2\bullet} = 260$
$\sum$	$f_{\bullet 1} = 73$	$f_{\bullet 2} = 237$	$n = 310$

Kako test statistika sadrži dvostruku sumu najefikasnije je prvo naći sume $\dfrac{f_{ij}^2}{f_{i\bullet} \cdot f_{\bullet j}}$ i to pomoću sledeće tabele

$$\begin{array}{|c|c|}\hline \dfrac{f_{11}^2}{f_{1\bullet} \cdot f_{\bullet 1}} & \dfrac{f_{12}^2}{f_{1\bullet} \cdot f_{\bullet 2}} \\ \hline \dfrac{f_{21}^2}{f_{2\bullet} \cdot f_{\bullet 1}} & \dfrac{f_{22}^2}{f_{2\bullet} \cdot f_{\bullet 2}} \\ \hline \end{array} = \begin{array}{|c|c|}\hline \dfrac{14^2}{50 \cdot 73} & \dfrac{36^2}{50 \cdot 237} \\ \hline \dfrac{59^2}{260 \cdot 73} & \dfrac{201^2}{260 \cdot 237} \\ \hline \end{array} = \begin{array}{|c|c|}\hline 0{,}0537 & 0{,}1094 \\ \hline 0{,}1834 & 0{,}6556 \\ \hline \end{array}$$

Sada je ova dvostruka suma $\sum\limits_{i=1}^{r} \sum\limits_{j=1}^{s} \dfrac{f_{ij}^2}{f_{i\bullet} f_{\bullet j}}$ jednaka zbiru svih brojeva iz prethodne tabele, tj.

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \frac{f_{ij}^2}{f_{i\bullet} f_{\bullet j}} = 0,0537 + 0,1094 + 0,1834 + 0,6556 = 1,0021.$$

Realizovana vrednost test statistike je $\chi_0^2 = 310 \cdot (1,0021 - 1) = 0,651.$

kritična oblast je

$C \ : \ \chi_0^2 \geq \chi_{(2-1)(2-1);\, 1-0,05}$

$C \ : \ \chi_0^2 \geq \chi_{1;0,95}$

$C \ : \ 0,651 \geq 3,84$ Sledi da se nulta hipoteza prihvata, tj. vlažnost i temperatura su nezavisna obeležja.

Zadatak. Na uzorku od 350 slučajno izabranih vlasnika automobila, čije su marke automobila označene sa A, B i C, dobijeni su sledeći podaci:

pol	marka automobila A	B	C
muški	40	80	30
ženski	30	120	50

Da li se sa pragom značajnosti $\alpha = 0,01$ može smatrati da je marka automobila nezavisna od pola vlasnika?

Rezultat. Iz $\chi_0^2 \geq \chi_{2;\,0,99} \iff 22,75 \geq 9,21$ sledi da marka automobila i pol vlasnika nisu nezavisna obeležja.

Primer zadataka za kolokvijum iz predmeta OBRADA I ANALIZA PODATAKA

I GRUPA

1. (9 poena) Pri sistematskom pregledu jednog odeljenja srednje škole dobijeni su sledeći rezultati o telesnoj masi učenika:

x_i−kg	[70, 75)	[75, 80)	[80, 85)	[85, 90)
f_i−br. uč.	4	7	16	5

Za datu distribuciju frekvencija izračunati:
a) mod,
b) interkvartilnu razliku,
c) standardnu devijaciju, tj. standardno odstupanje.

2. (6 poena) Ako slučajna promenljiva ima zakon raspodele:

$$X \ : \ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,05 & 0,2 & 0,4 & 0,25 & 0,1 \end{pmatrix},$$

izračunati:

a) $E(X)$,
b) $D(X)$,
c) $P\{X < 3\}$.

3. (4 poena) Slučajna promenljiva ima binomnu raspodelu $S_{10} : B(10; 2/3)$. Izračunati:
a) $P\{2 \leq S_{10} < 4\}$.
b) Koliki je očekivani broj realizacija, tj. $E(S_{10}) =?$

4. (3 poena) Slučajna promenljiva ima normalnu raspodelu $Z : N(0,1)$. Naći

a) $P\{-1,37 < Z < 2,01\}$,
b) $P\{0,65 < Z < 1,26\}$,
c) $P\{Z > 1,13\}$,

5. (2 poena) Slučajna promenljiva ima normalnu raspodelu $Z : N(0,1)$. Naći t tako da je

$$P\{t < Z < 2\} = 0,3772.$$

6. (6 poena) Slučajna promenljiva ima normalnu raspodelu $X : N(8,16)$. Naći

a) $P\{5 < X < 10\}$,
b) $P\{10 < X < 15\}$,
c) $P\{X < 5\}$.

Primer zadataka za kolokvijum iz predmeta OBRADA I ANALIZA PODATAKA

II GRUPA

1. (9 poena) Pri sistematskom pregledu jednog odeljenja srednje škole dobijeni su sledeći rezultati o telesnoj masi učenika:

x_i–kg	[70, 75)	[75, 80)	[80, 85)	[85, 90)
f_i–br. uč.	4	7	16	5

Za datu distribuciju frekvencija izračunati:
a) medijanu,
b) tridesetpeti percentil P_{35},
c) koeficijent varijacie.

2. (6 poena) Ako slučajna promenljiva ima zakon raspodele:

$$X \ : \ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,05 & 0,2 & 0,4 & p & 0,1 \end{pmatrix},$$

a) izračunati p.
b) Za tako nađeno p izračunati disperziju $D(X)$.
c) Naći funkciju raspodele slučajne promenljive X, tj. $F(x)$.

3. (4 poena) Ako slučajna promenljiva X ima gustinu raspodele:

$$f(x) = \frac{1}{\pi \cdot (1+x^2)} \quad \text{za} \quad x \in \mathcal{R},$$

izračunati verovatnoću $P\{-1 < X < 1\}$.

4. (3 poena) Slučajna promenljiva ima normalnu raspodelu $Z \, : \, N(0,1)$. Naći
a) $P\{-1,73 < Z < 2,25\}$,
b) $P\{-1,79 < Z < -0,54\}$,
c) $P\{Z < 0,91\}$,

5. (2 poena) Slučajna promenljiva ima normalnu raspodelu $Z \, : \, N(0,1)$. Naći t tako da je

$$P\{Z > t\} = 0,1539.$$

6. (6 poena) Slučajna promenljiva ima normalnu raspodelu $X \, : \, N(15,9)$. Naći
a) $P\{12 < X < 17\}$,
b) $P\{21 < X < 30\}$,
c) $P\{X > 16\}$.

Primer zadataka za kolokvijum iz predmeta OBRADA I ANALIZA PODATAKA

III GRUPA

▷**1.** (9 poena) Dati su podaci o broju članova porodice u 80 ispitanih porodica:

broj članova x_i	1	2	3	4	5	6	7	8
broj porodica f_i	5	10	15	20	11	9	6	4

Za ovu distribuciju frekvencija naći:
a) mod,
b) interkvartilnu razliku i
c) standardnu devijaciju.

2. (6 poena) Kolika je verovatnoća da je iz špila (32 karte) izvučena desetka, ako se zna da je izvučena karta u boji trefa.

3. (4 poena) Slučajna promenljiva X je zadata svojom funkcijom raspodele:

$$F(x) = \begin{cases} 0, & x > 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}.$$

Izračunati verovatnoću $P\{X > 2\}$.

4. (3 poena) Slučajna promenljiva ima normalnu raspodelu $Z : N(0,1)$. Naći

a) $P\{-1,25 < Z < 2,73\}$,
b) $P\{0,26 < Z < 2,26\}$,
c) $P\{Z \geq 1,13\}$,

5. (2 poena) Slučajna promenljiva ima normalnu raspodelu $Z : N(0,1)$. Naći t tako da je

$$P\{Z < t\} = 0,6554.$$

6. (6 poena) Slučajna promenljiva ima normalnu raspodelu $X : N(7,9)$. Naći

a) $P\{8 < X < 15\}$,
b) $P\{3 < X < 15\}$,
c) $P\{X < 9\}$.

Primer zadataka za ispit iz predmeta OBRADA I ANALIZA PODATAKA

I GRUPA

1. (7 poena) U slučajnom uzorku od 20 pakovanja jednog proizvoda prosečna težina iznosi $980gr$. Težina pakovanja ima normalan raspored i propisana težina pakovanja je $1000gr$, sa standardnom devijacijom 15. Da li, sa pragom značajnosti $\alpha = 0,05$, možemo konstatovati da težina pakovanja odstupa od standarda?

2. (10 poena) Primenom χ^2-testa proveriti hipotezu da obeležje X ima normalnu raspodelu ako uzorak od 90 elementa ima vrednosti zadate u tabeli:

x_i	[6, 8)	[8, 10)	[10, 12)	[12, 14)	[14, 16)	[16, 18)	[18, 20)	[20, 22)	[22, 24)
f_i	4	5	16	22	21	12	6	3	1

Neka je prag značajnosti $\alpha = 0,01$.

3. (8 poena) Metodom najmanjih kvadrata, krivom $y = a \cdot \sqrt{x} + b$, aproksimirati funkciju datu tabelarno:

x_i	0	1	2	3	4
y_i	1	2	2,41	2,73	3

Pitanja iz usmenog dela ispita

1. (5 poena) Unija, zbir događaja i njihove verovatnoće.

2. (5 poena) Apsolutno neprekidna slučajna promenljiva i njeno matematičko očekivanje.

3. (5 poena) Slučajni uzorak (tablica slučajnih brojeva).

Primer zadataka za ispit iz predmeta OBRADA I ANALIZA PODATAKA

II GRUPA

1.(7 poena) Na placu je izloženo 120 automobila i to 84 iz jedne fabrike, a 36 iz druge. Poznato je da medju proizvodima prve fabrike ima 5% oštećenih, a za drugu fabriku taj procenat je 3. Kolika je verovatnoća da slučajno izabran automobil bude oštećen? Kolika je verovatnoća da bude neoštećen?

2. (7 poena) Obeležje X ima normalnu raspodelu i to $X : N(\mu, 900)$. Posle izvesnog vremena na osnovu uzorka obima $n = 15$ uzoračka standardna devijacija iznosi $\widetilde{S}_{15} = 41$. Da li, sa pragom značajnosti $\alpha = 0,05$, možemo konstatovati da se disperzija populacije promenila?

3. Na osnovu uzorka obima $n = 5$, dobijene vrednosti obeležja X i obeležja Y su:

x_i	1	2	4	5	8
y_i	3	3	6	7	12

a) (5 poena) Izračunati uzorački koeficijent korelacije.

b) (6 poena) Ako postoji bar visoko značajna linearna povezanost izmedju ovih obeležja, tj. ako je $Y = aX + b$, naći koeficijente a i b metodom najmanjih kvadrata.

Pitanja iz usmenog dela ispita

1. (5 poena) Obeležje.

2. (5 poena) Potpun sistem događaja. Bajesova formula.

3. (5 poena) Periodični uzorak.

Primer zadataka za ispit iz predmeta OBRADA I ANALIZA PODATAKA

III GRUPA

1.(5 poena) Ako slučajna promenljiva X ima gustinu raspodele

$$f(x) = \begin{cases} a \cdot x, & 0 < x \leq 1 \\ 0, & \text{inače}, \end{cases}$$

a) naći konstantu a.

b) Izračunati verovatnoću $P\left\{X < \dfrac{1}{2}\right\}$.

2. (10 poena) Iz populacije u kojoj obeležje X ima normalnu raspodelu uzet je sledeći uzorak obima $n = 12$:

x_i	-0,5	-0,4	0	0,2	0,5	0,7
f_i	1	2	1	3	4	1

a) Naći intervalnu ocenu srednje vrednosti sa nivoom poverenja $0,9$.
b) Sa pragom značajnosti $\alpha = 0,05$ testirati hipotezu $H_0(\sigma^2 = 1)$ nasuprot hipotezi $H_1(\sigma^2 < 1)$.

3. (10 poena) Na svaka dva sata kontrolor uzima po pet proizvoda jedne mašine i registruje broj neispravnih. Posle 100 izvršenih kontrola dobio je sledeće podatke

x_i-broj neispravnih	0	1	2	3	4	5
f_i	10	22	27	26	11	4

Sa pragom značajnosti $\alpha = 0,01$ ispitati saglasnost dobijenih podataka sa binomnom raspodelom $B(5; 0,4)$.

Pitanja iz usmenog dela ispita

1. (5 poena) Klasična definicuja verovatnoće.

2. (5 poena) Binomna raspodela.

3. (5 poena) Tačkaste ocene parametara raspodele.

OZNAKE

$\mathcal{N}$ – skup prirodnih brojeva

$\mathcal{R}$ – skup realnih brojeva

N – broj elemenata populacije

n – broj elemenata uzorka

k – broj različitih vrednosti obeležja

μ – aritmetička sredina obeležja populacije

$\overline{X}_n$ – uzoračka sredina

K – broj intervala vrednosti obeležja

x'_i – sredina intervala vrednosti obeležja

f_i – apsolutna frekvencija vrednosti obeležja

f_i^* – relativna frekvencija vrednosti obeležja

$\mathcal{Z}_i$ – zbirna frekvencija vrednosti obeležja

M_o – modus ili mod

L_{mo} – početak modalnog intervala

f_{mo} – frekvencija modalnog intervala

f_{mo-1} – frekvencija intervala pre modalnog

f_{mo+1} – frekvencija intervala posle modalnog

Δ – dužina intervala vrednosti obeležja

M_e – medijana

K_{me} – broj koji označava koji je po redu medijalni interval

L_{me} – početak medijalnog intervala

f_{me} – frekvencija medijalnog intervala

I_v – interval varijacije

I_q – interkvartilna razlika

Q_i – i-ti kvartil

K_{kv} – broj koji označava koji je po redu kvartilni interval

L_{kv} – početak kvartilnog intervala

f_{kv} – frekvencija kvartilnog intervala

D_i – i-ti decil

P_i – i-ti percentil

$AD(\mu)$ – srednje apsolutno odstupanje od aritmetičke sredine

σ^2 – disperzija ili varijansa populacije

$\overline{S}_n^2$ – uzoračka disperzija

$\widetilde{S}_n^2$ – popravljena uzoračka disperzija

σ – standardna devijacija populacije

$\overline{S}_n$ – uzoračka standardna devijacija

$\widetilde{S}_n$ – popravljena uzoračka standardna devijacija

C_v – koeficijent varijacije

Z_i – normalizovano standardno odstupanje

$\rho_{X,Y}$ – koeficijent korelacije slučajnih promenljivih X i Y.

$R_{X,Y}$ – uzorački koeficijent korelacije

$r_{X,Y}$ – realizovana vrednost uzoračkog koeficijenta korelacije

$E(X)$ – matematičko očekivanje slučajne promenljive X

$D(X)$ – disperzija slučajne promenljive X

$B(n, p)$ – binomna (Bernulijeva) raspodela

$P(\lambda)$ – Puasonova raspodela

$N(\mu, \sigma^2)$ – normalna (Gausova) raspodela

$\mathcal{U}[a, b]$ – uniformna (ravnomerna) raspodela

S_n – slučajna promenljiva sa binomnom raspodelom

S_∞ – slučajna promenljiva sa Puasonovom raspodelom

$F(x)$ – funkcija raspodele verovatnoća

$f(x)$ – gustina raspodele verovatnoća

$P\{X < a\},\ a \in \mathcal{R}$ – verovatnoća da slučajna promenljiva X uzme vrednost manju od a

$\Phi(x)$ – Laplasova funkcija (vrednosti ove funkcije su u tablici za normalnu raspodelu)

H_0 – nulta hipoteza

H_1 – alternativna hipoteza

LITERATURA

[1] Popović B., Blagojević B.: **Matematička statistika sa primenama u hidrotehnici**, Univerzitet u Nišu, Niš, 1997.

[2] Popović B., Ristić M.: **Statistika u psihologiji**, Mrlješ, Beograd, 2001.

[3] Ristić M., Popović B., Đorđević M.: **Statistika za studente geografije**, PMF Univerziteta u Nišu, (Niš: "Vuk Karadžić"), 2006.

[4] Ivković Z.,Banjević D.: **Verovatnoća i matematička statistika** za III razred usmerenog obrazovanja matematičko-tehničke struke, Naučna knjiga, Beograd, 1988.

[5] Ivković Z.: **Matematička statistika**, Naučna knjiga, Beograd, 1992.

[6] Stojaković M.: **Uvod u teoriju verovatnoće i matematičke statistike**, Fakultet tehničkih nauka Univerziteta u Novom Sadu, (Novi Sad: STYLOS), 1995.

[7] Mališić J.: **Zbirka zadataka iz teorije verovatnoće sa primenama**, Univerzitet u Beogradu, (Beograd: IP Građevinska knjiga), 1975.

[8] Vukadinović S.: **Elementi teorije verovatnoće i matematičke statistike**, Privredni pregled, Beograd, 1981.

[9] Joksimović D.: **Zbirka zadataka iz poslovne statistike**, Megatrend univerzitet primenjenih nauka, Beograd, 2004.

[10] Mekić S., Žižović M., Šekarić M., Prlinčević G.: **Poslovna statistika kroz primere**, Viša poslovna škola, Blace, 2004.

[11] Stanković J., Ralević N., Ljubanović-Ralević I.: **Statistika sa primenom u poljoprivredi**, MLADOST BIRO, Beograd 2002.

[12] Bulatović J., **Merni sistemi i instrumenti. Statistička obrada rezultata merenja**, (Bor: Bakar), Niš, 1982.