

Zadatak 7: Diskretna slučajna promenljiva

Posmatra se diskretna slučajna promenljiva X koja predstavlja broj dana sa padavinama većim od 1.0 mm u pojedinim mesecima u nekom mestu A:

godina	x	godina	x	godina	x
1961	4	1971	6	1981	7
1962	9	1972	10	1982	7
1963	8	1973	14	1983	3
1964	10	1974	9	1984	7
1965	10	1975	7	1985	6
1966	10	1976	7	1986	8
1967	10	1977	6	1987	9
1968	8	1978	7	1988	9
1969	3	1979	6	1989	11
1970	9	1980	12	1990	7

1. Odrediti elemente empirijske raspodele diskretne promenljive X :

- 1) Izraditi tabele apsolutnih, relativnih i kumulativnih frekvencija, kao i odgovarajuće grafikone.
- 2) Izračunati statistike: srednju vrednost, standardnu devijaciju, indeks disperzije, koeficijente varijacije i asimetrije.

2. Empirijsku raspodelu promenljive X prilagoditi binomnom raspodelom.

3. Izraditi graficki prikaz empirijskih i teorijskih frekvencija.

4. Testirati saglasnost empirijske i teorijske raspodele pomou χ^2 -testa.

5. Odgovoriti na sledeca pitanja:

- 1) Koja je vrednost moda izracunate promenljive (broj kisnih dana)?
- 2) Kolika je verovatnoca da ce se javiti 7 ili vise kisnih dana u zadatom mesecu?
- 3) Kolika je verovatnoca da ce se javiti 5 ili manje kisnih dana u zadatom mesecu?
- 4) Kolika je verovatnoca da nece biti kise ni jednog dana?
- 5) Da li teorijska distribucija ima pozitivnu ili negativnu asimetriju? Na osnovu vrednosti kog parametra se moze na ovo pitanje odmah odgovoriti?

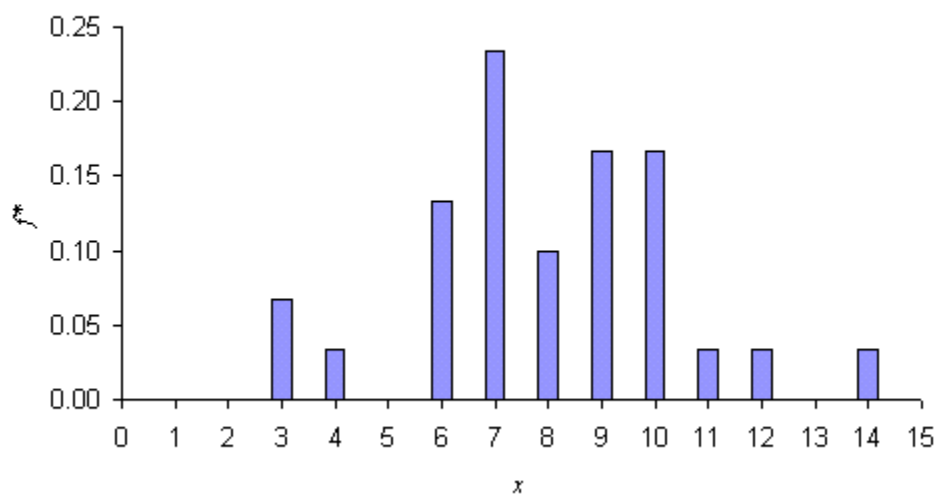
Objasnjenje

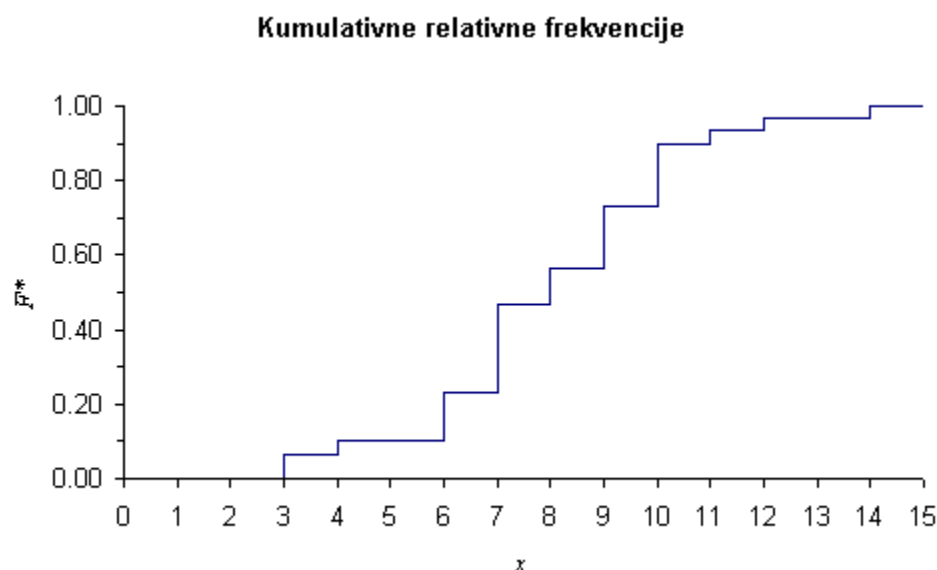
1.1 Proracun frekvencija:

	apsolutna	relativna	kumulativna relativna
--	-----------	-----------	--------------------------

x	frekvencija f	frekvencija f^*	frekvencija F^*
3	2	0.067	0.067
4	1	0.033	0.100
5	0	0.000	0.100
6	4	0.133	0.233
7	7	0.233	0.467
8	3	0.100	0.567
9	5	0.167	0.733
10	5	0.167	0.900
11	1	0.033	0.933
12	1	0.033	0.967
13	0	0.000	0.967
14	1	0.033	1.000
suma	30	1.000	

Relativne frekvencije





1.2 Proracun statistika:

Srednja vrednost:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{30} (3 \cdot 2 + 4 \cdot 1 + 6 \cdot 4 + 7 \cdot 7 + 8 \cdot 3 + 9 \cdot 5 + 10 \cdot 5 + 11 \cdot 1 + 12 \cdot 1 + 14 \cdot 1) = 7.97$$

Standardna devijacija:

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

$$S^2 = \frac{1}{30-1} [(3-7.97)^2 \cdot 2 + (4-7.97)^2 \cdot 1 + (6-7.97)^2 \cdot 4 + (7-7.97)^2 \cdot 7 +$$

$$+ (8-7.97)^2 \cdot 3 + (9-7.97)^2 \cdot 5 + (10-7.97)^2 \cdot 5 + (11-7.97)^2 \cdot 1 + (12-7.97)^2 \cdot 1 +$$

$$+ (14-7.97)^2 \cdot 1] = 6.033$$

$$S = \sqrt{6.033} = 2.46$$

Indeks disperzije:

$$I = \frac{S^2}{\bar{x}} = \frac{6.033}{7.97} = 0.757$$

Koeficijent varijacije:

$$Cv = \frac{S}{\bar{x}} = \frac{2.46}{7.97} = 0.308$$

Koeficijent asimetrije:

$$C_s = \frac{N}{(N-1)(N-2)S^3} \sum_{i=1}^N (x_i - \bar{x})^3$$

$$\sum_{i=1}^N (x_i - \bar{x})^3 = (3-7.97)^3 \cdot 2 + (4-7.97)^3 \cdot 1 + (6-7.97)^3 \cdot 4 + (7-7.97)^3 \cdot 7 +$$

$$+ (8-7.97)^3 \cdot 3 + (9-7.97)^3 \cdot 5 + (10-7.97)^3 \cdot 5 + (11-7.97)^3 \cdot 1 + (12-7.97)^3 \cdot 1 +$$

$$+ (14-7.97)^3 \cdot 1 = 16.50$$

$$C_s = \frac{30}{29 \cdot 28 \cdot 2.46^3} \cdot 16.50 = 0.041$$

2. Binomna raspodela

Prema binomnoj raspodeli, verovatnoca da slucajna promenljiva X uzme vrednost x jednaka je:

$$P\{X = x\} = \binom{n}{x} p^x (1-p)^{n-x}$$

gde je $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ binomni koeficijent.

Parametri binomne raspodele su sledeci:

- ▮ n - broj opita (ovde: broj dana u mesecu); $n = 30$
- ▮ p - verovatnoca uspeha (ovde: verovatnoca da u jednom danu padne kisa veca od 1 mm);
odredjuje se iz prosecnog broja dana sa kisom vecom od 1 mm:

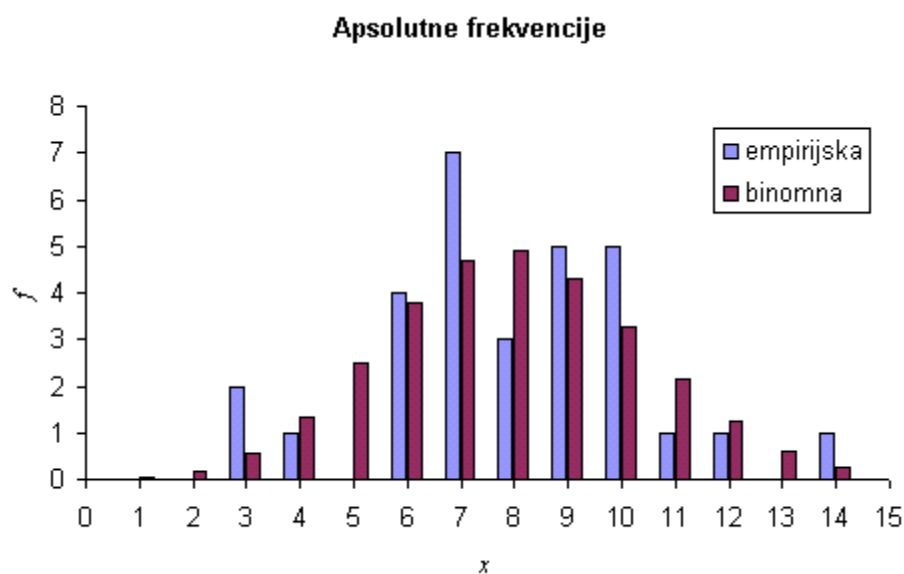
$$p = \frac{\bar{x}}{n} = \frac{7.97}{30} = 0.266$$

Proracun verovatnoca po binomnoj raspodeli i odgovarajucih teorijskih frekvencija dat je u donjoj tabeli. Uporedno su date i empirijske frekvencije sracunate u tacki 1.1.

x	$p_B = P\{X = x\}$	$f_B = p_B \cdot N$	f_e
0	0.0001	0.0	0
1	0.001	0.0	0
2	0.005	0.2	0
3	0.018	0.5	2
4	0.045	1.3	1

5	0.084	2.5	0
6	0.126	3.8	4
7	0.157	4.7	7
8	0.163	4.9	3
9	0.144	4.3	5
10	0.109	3.3	5
11	0.072	2.2	1
12	0.041	1.2	1
13	0.021	0.6	0
14	0.009	0.3	1
15	0.003	0.1	0

3. Graficki prikaz teorijskih i empirijskih frekvencija



4. Testiranje saglasnosti empirijske i teorijske raspodele

Za testiranje se koristi χ^2 (hi-kvadrat) test. U ovom testu porede se teorijske i empirijske frekvencije kroz statistiku:

$$\chi^2 = \sum_x \frac{[f_t(x) - f_e(x)]^2}{f_t(x)}$$

Pri ovom proračunu mora se voditi računa da u jednoj klasi vrednosti slučajne promenljive bude najmanje 5 elemenata. Zbog toga se takve klase spajaju i u njima sabiraju frekvencije.

x	f_e		f_B		c^2
3	2	7	0.5	8.2	0.174
4	1		1.3		
5	0		2.5		
6	4		3.8		
7	7	7	4.7	4.7	1.127
8	3	8	4.9	9.2	0.157
9	5		4.3		
10	5	8	3.3	7.6	0.026
11	1		2.2		
12	1		1.2		
13	0		0.6		
14	1		0.3		
suma:					

Ovako dobijena vrednost za statistiku χ^2 poredi se sa kriticom vrednoscu ove statistike za prag znacajnosti $\alpha = 0.05$. Statistika χ^2 prati istoimenu raspodelu ciji je parametar broj stepeni slobode v . Broj stepeni slobode odredjuje se kao:

$$v = K - p - 1$$

gde je K broj klasa (ovde 4), a p broj parametara u teorijskoj raspodeli (ovde 2). Dakle,

$$v = 4 - 2 - 1 = 1$$

Kriticna vrednost statistike χ^2 za zadati prag znacajnosti i stepen slobode cita se iz statistickih tablica za χ^2 raspodelu, i ona iznosi 3.841. Kako je sracunata vrednost statistike χ^2 manja od kriticke vrednosti:

$$\chi^2 = 1.484 < \chi^2(\alpha, v) = 3.841$$

moze se usvojiti hipoteza da su empirijska i teorijska raspodela saglasne.