Seminar 10 - Teoria Probabilităţilor şi Statistică Matematică

Legi ale numerelor mari:

http://math.ubbcluj.ro/~tradu/geologi/lnmtlc_article.pdf

Selecţie şi statistici:

La o staţie meteorologica temperaturile (în grade Celsius) înregistrate în ultimii 8 ani, la ora 12:00 din data de 1 august, au fost: 30,24,35,36,32,23,31,37.

1. Să se determine: repartiţia empirică.
2. Să se calculeze amplitudinea, media şi dispersia de selecţie, precum şi funcţia empirică de repartiţie a selecţiei.
3. Determinaţi dispersia de selecţie corectată.

SELECTIE SI STATISTICI

Exerciţiul 1.10 *Să presupunem că un aparat de măsurare este utilizat pentru a citi o distanţă de 12 de ori. Se obţin valorile:*

$$0.20, 0.10, 0.35, 0.25, 0.13, 0.20, 0.10, 0.20, 0.25, 0.20, 0.30, 0.35.$$

Datele sunt colectate in tabelul de mai jos:

$$\begin{pmatrix} 0.10 & 0.13 & 0.20 & 0.25 & 0.30 & 0.35 \\ 2 & 1 & 4 & 2 & 1 & 2 \end{pmatrix}$$

Obţinem deci

$$X^* : \begin{pmatrix} 0.10 & 0.13 & 0.20 & 0.25 & 0.30 & 0.35 \\ 2/12 & 1/12 & 4/12 & 2/12 & 1/12 & 2/12 \end{pmatrix}$$

Amplitudinea este $0.35 - 0.10 = 0.25$.

Mediana este o valoare situată între a şasea şi a şaptea, adică media aritmetică $\frac{0.20+0.25}{2} = 0.225$.

Moda (valoarea modală) este 0.20.

Media de selecţie (sau media aritmetică(este dată de

$$\bar{x} = \sum_{i=1}^{6} f_i x_i = 0.22$$

sau echivalent

$$\bar{x} = \frac{\sum_{i=1}^{6} n_i x_i}{20}.$$

Dispersia (sau varianţa) empirică este dată de formula

$$s^2 \stackrel{def}{=} \frac{1}{12}\sum_{i=1}^{6} n_i\,(x_i-\bar{x})^2=\sum_{i=1}^{6} f_i\,(x_i-\bar{x})^2=0.00643,$$

deci abaterea empirică este $s=\sqrt{s^2}\simeq 0.0802$.

Pe de altă parte dispersia empirică modificată este numărul

$$(s^*)^2=\frac{12}{11}\,s^2=\frac{12}{11}\,0.00643=0.0070244\,.$$

În plus abaterea empirică modificată este

$$s^* \stackrel{def}{=} \sqrt{(s^*)^2}\simeq 0.083811.$$

Mai trebuie făcut graficul poligonului frecvenţelor relative.

Se poate scrie şi funcţia empirică de repartiţie $F\,(x)$ *(care este o funcţie în scară).*

Exerciţiul 1.11 *Să presupunem că un aparat de măsurare este utilizat pentru a citi o distanţă de 20 de ori. Datele sunt colectate in tabelul de mai jos:*

(1.9)

22.7	25.4	22.0	20.5	22.5
22.3	24.2	24.7	23.5	23.1
25.5	24.7	23.1	22.0	23.8
23.8	24.4	23.7	23.8	22.6

Aceste citiri reprezintă mulţimea de date. O primă analiză a lor din punct de vedere numeric poate fi făcută calculând amplitudinea. Vedem din tabel ca amplitudinea este $25.5-20.5=5.0$.

Să considerăm în continuare datele de mai sus puse în ordine crescătoare.

(1.10)

20.5	22.0	22.0	22.3	22.5
22.6	22.7	23.1	23.1	23.5
23.7	23.8	23.8	23.8	24.2
24.4	24.7	24.7	25.4	25.5

Putem determina imediat mediana. În cazul nostru mediana este dată de o valoare situată între a zecea şi a unsprezecea valoare, adică media aritmetică $\frac{23.5+23.7}{2}$ *(se poate considera drept mediană şi una dintre cele două valori).*

Moda este valoarea 23.8 *(valoarea cu frecvenţa cea mai mare).*

Variabila empirică (de selecţie) X^* va avea tabloul

$$X^* : \begin{pmatrix} 20.5 & 22.0 & 22.3 & 22.5 & 22.6 & 22.7 & 23.1 & 23.5 & 23.7 \\ 0.05 & 0.1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.1 & 0.05 & 0.05 \\ & & & 23.8 & 24.2 & 24.4 & 24.7 & 25.4 & 25.5 \\ & & & 0.15 & 0.05 & 0.05 & 0.1 & 0.05 & 0.05 \end{pmatrix} \tag{1.11}$$

Momentul empiric de ordin 1 (sau media empirică sau media aritmetică) este dată de

$$\begin{aligned} \bar{x} &= \sum_{i=1}^{15} f_i x_i \\ &= 0.05 \cdot 20.5 + 0.1 \cdot 22.0 + 0.05 \cdot 22.3 + 0.05 \cdot 22.5 + 0.05 \cdot 22.6 + 0.05 \cdot 22.7 \\ &\quad +0.1 \cdot 23.1 + 0.05 \cdot 23.5 + 0.05 \cdot 23.7 + 0.15 \cdot 23.8 + 0.05 \cdot 24.2 + 0.05 \cdot 24.4 \\ &\quad +0.1 \cdot 24.7 + 0.05 \cdot 25.4 + 0.05 \cdot 25.5 \\ &= 23.415 \end{aligned}$$

sau echivalent

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^{15} n_i x_i}{20} \\ &= \frac{20.5 + 2 \cdot 22.0 + 22.3 + 22.5 + 22.6 + 22.7 + 2 \cdot 23.1 + 23.5 + 23.7 + 3 \cdot 23.8}{20} \\ &\quad + \frac{24.2 + 24.4 + 2 \cdot 24.7 + 25.4 + 25.5}{20} = \frac{468.3}{20} = 23.415 \end{aligned}$$

unde x_i sunt valorile citite din tabloul (1.11).

Echivalent $\bar{x}$ este media aritmetică a tuturor valorilor citite (valori ce se pot repeta),

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{20},$$

unde x_i sunt valorile din tabelul (1.9).

Dispersia (sau varianţa) empirică este dată de formula

$$s^2 \stackrel{def}{=} \frac{1}{20} \sum_{i=1}^{15} n_i (x_i - \bar{x})^2 = \sum_{i=1}^{15} f_i (x_i - \bar{x})^2 .$$

Este util să scriem mai întâi un tabel cu diferenţele[1] $x_i - \bar{x}$ *şi* $(x_i - \bar{x})^2$:

x_i	*Frecvenţa abs.* n_i	*Frecvenţa rel.* f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
20.5	1	$0.05 = 1/20$	-2.915	8.4972
22.0	2	$0.1 = 2/20$	-1.415	2.0022
22.3	1	$0.05 = 1/20$	-1.115	1.2432
22.5	1	$0.05 = 1/20$	-0.915	0.8372
22.6	1	$0.05 = 1/20$	-0.815	0.6642
22.7	1	$0.05 = 1/20$	-0.715	0.5112
23.1	2	$0.1 = 2/20$	-0.315	0.0992
23.5	1	$0.05 = 1/20$	0.085	0.0072
23.7	1	$0.05 = 1/20$	0.285	0.0812
23.8	3	$0.15 = 3/20$	0.385	0.1482
24.2	1	$0.05 = 1/20$	0.785	0.6162
24.4	1	$0.05 = 1/20$	0.985	1.97
24.7	2	$0.1 = 2/20$	1.285	1.6512
25.4	1	$0.05 = 1/20$	1.985	3.9402
25.5	1	$0.05 = 1/20$	2.085	4.3472
	20	$1 = 20/20$		

Deci, calculând obţinem valoarea dispersiei empirice

$$s^2 = 1.4832$$

iar abaterea medie pătratică empirică este

$$s = \sqrt{s^2} = \sqrt{1.4832} = 1,2178\,.$$

Pe de altă parte dispersia empirică modificată este numărul

$$(s^*)^2 = \frac{n}{n-1}\, s^2 = \frac{20}{19} 1.4832 = 1.5612\,.$$

În plus abaterea empirică modificată este

$$s^* \stackrel{def}{=} \sqrt{(s^*)^2} = 1.2494\,.$$

Remarca 1.12 *În toate tabele şi formulele de mai sus putem lăsa toate valorile x_i chiar dacă se repetă (deci $n = 20$ în acest caz). Atunci frecvenţa relativă a fiecărei valori va fi aceeaşi $f_i = 1/20 = 0.05$ şi frecvenţa absolută a fiecărei valori va fi aceeaşi $n_i = 1$. Formula pentru s^2 devine*

$$s^2 \stackrel{def}{=} \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x})^2$$

iar

$$\begin{aligned}(s^*)^2 &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}x_i^2 - 2\sum_{i=1}^{n}x_i\bar{x} + \sum_{i=1}^{n}\bar{x}^2}{n-1}\\ &= \frac{\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n\bar{x}^2}{n-1} = \frac{\sum_{i=1}^{n}x_i^2 - 2n\bar{x}\cdot\bar{x} + n\bar{x}^2}{n-1}\\ &= \frac{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}{n-1} = \frac{\sum_{i=1}^{20}x_i^2 - 20\bar{x}^2}{n-1}\,,\end{aligned}$$

adică obţinem următoarea ***formulă de calcul a dispersiei empirice modificate*** *(vezi şi formula (1.8)):*

$$(s^*)^2 = \frac{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}{n-1}\,. \tag{1.12}$$

TEMA:

Legi ale numerelor mari:

`http://math.ubbcluj.ro/~tradu/geologi/lnmtlc_article.pdf`

Selecţie şi statistici:

La o staţie meteorologica temperaturile (în grade Celsius) înregistrate în ultimii 8 ani, la ora 12:00 din data de 1 august, au fost: 30,24,35,36,32,23,31,37.

1. Să se determine: repartiţia empirică.
2. Să se calculeze amplitudinea, media şi dispersia de selecţie, precum şi funcţia empirică de repartiţie a selecţiei.
3. Determinaţi dispersia de selecţie corectată.