

Statystyka opisowa

Rozkłady empiryczne

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

średnia arytmetyczna

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

odchylenie standardowe

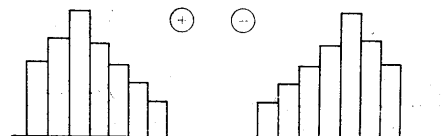
$$a = \frac{1}{(n-1)s^3} \sum_{i=1}^m n_i (x_i - \bar{x})^3$$

współczynnik asymetrii

$a < 0$ – asymetria ujemna, lewostronna,

$a > 0$ – asymetria dodatnia, prawostronna,

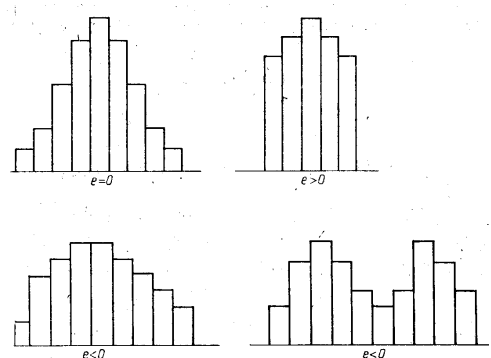
$a = 0$ – szereg symetryczny



$$e = \frac{1}{(n-1)s^4} \sum_{i=1}^m n_i (x_i - \bar{x})^4 - 3$$

eksces

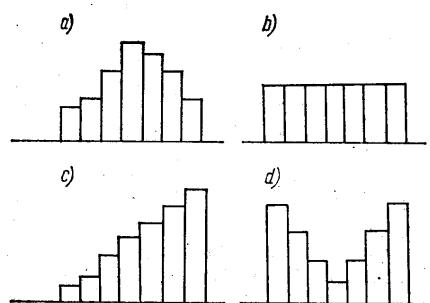
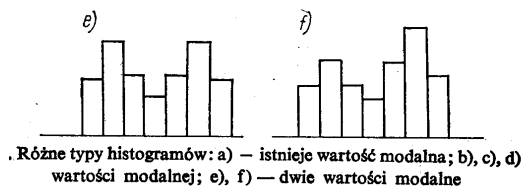
miara skupienia jednostek wokół średniej arytmetycznej. Liczba trzy wprowadzona jest z uwagi na to, że jeżeli badana cecha w populacji generalnej ma rozkład normalny, to wyznaczony analogiczny parametr dla tego rozkładu ma wartość 0. – czyli jest to miara odchylenia rozkładu od rozkładu normalnego



do - moda (dominanta)

wartość cechy występująca najczęściej

UWAGA nie każdy szereg posiada wartość modalną a są szeregi posiadające więcej niż jedna wartość modalna.



me - mediana(50-ty percentyl, drugi kwartyl)

wartość poniżej której znajduje się połowa wyników obserwacji

P-ty percentyl – wartość, poniżej której znajduje się P% wyników.

Miejsce P-tego percentyla określa wzór:

$(n+1)P/100$, gdzie n jest liczbą wyników.

Def. Kwantylem k mędu m , $k=1,2,\dots,m$ zmiennej losowej X nazywa się wartość $x \in \mathbb{R}$ spełniającą równanie:

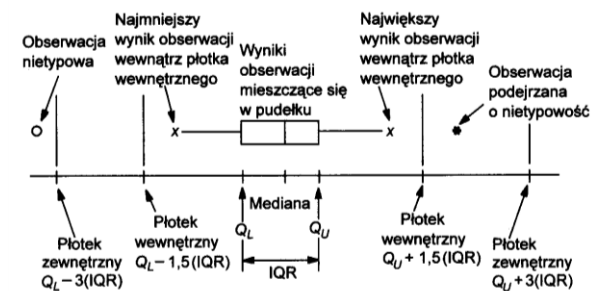
$$P(X \leq x) = \frac{k}{m}; \quad 0 \leq \frac{k}{m} \leq 1$$

$m=4$ $\frac{k}{4}$ $k=1,2,3,4$ - KWARTYLE

$m=10$ $\frac{k}{10}$ $k=1,2,\dots,10$ - DECYLE

$m=100$ $\frac{k}{100}$ $k=1,2,\dots,100$ - CENTYLE
PERCENTYLE

odstęp międzykwartylowy= trzeci kwartyl – pierwszy kwartyl



Szereg rozdzielczy

Aby utworzyć szereg rozdzielczy należy:

- 1) ustalić obszar zmienności badanej cechy, czyli przedział ograniczony najmniejszym i największym elementem próby,
- 2) podzielić obszar zmienności na klasy i ustalić środek przedziału klasowego oraz końce przedziałów, próba o liczności n , liczba przedziałów klasowych m

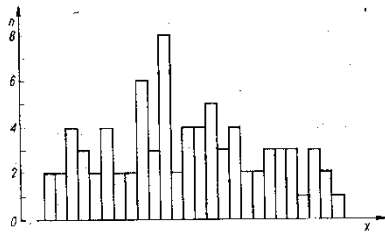
- liczba przedziałów klasowych nie powinna być mniejsza niż 7 i większa niż 15,

- nierówność $0,5\sqrt{n} \leq m \leq \sqrt{n}$

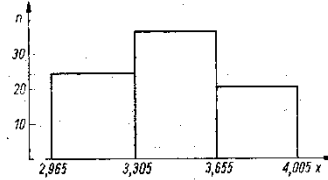
- według Huntsbergera $m = 1 + 3,3 \lg$
- według de Brooksa i Carruthersa $m < 5 \lg n$

oraz licznosc żadnego z przedziałów nie jest mniejsza niż 5

3) policzyć ile elementów próby należy do każdej klasy, czyli określić licznosc każdej z nich.
Poniżej przykłady za dużej i za małej liczby klas:



Rys. 1.8.2. Histogram dla danych liczbowych z przykl. 1.7.3 przy $m = 26$ przedziałach klasowych



Rys. 1.8.3. Histogram dla danych liczbowych z przykl. 1.7.3 przy $m = 3$ przedziałach klasowych

Dla szeregu rozdzielczego:

$$\bar{x} = \frac{\sum_{i=1}^m n_i x_i}{n}$$

średnia arytmetyczna

dla szeregu rozdzielczego,

gdzie x_j oznacza środki przedziałów,

n_j licznosc, m liczbę klas

$$s = \sqrt{\frac{\sum_{i=1}^n n_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n n_i x_i^2 - n(\bar{x})^2}{n-1}}$$

odchylenie standardowe

dla szeregu rozdzielczego

ZADANIA

Zad 1.

Wyniki pomiarów ciśnienia skurczowego w grupie 100 mężczyzn w wieku 20-39 lat przedstawiały się następująco:

100	120	130	155	130	100	95	110	135	135	105	140	145	140	125
135	110	95	160	130	95	140	165	140	120	105	135	115	135	120
140	140	105	110	110	135	165	135	95	115	145	115	140	135	155
165	155	160	135	110	110	145	110	140	130	105	135	155	130	115
140	130	145	125	125	150	150	135	135	120	125	150	130	145	105
160	150	150	150	115	140	135	125	125	125	135	120	130	110	110
140	145	130	115	135	125	115	120	120	115					

- A. Oblicz średnią, odchylenie standardowe dla próby.
- B. Znajdź medianę, odstęp międzykwartylowy oraz percentyle 20-ty, 30-ty, 60-ty i 90-ty
- C. Sporządź szereg rozdzielczy **punktowy** wyników tej próby, a następnie sporządź histogram próby oraz diagram licznosci (częstości), diagram licznosci (częstości) skumulowanej.
- D. Oblicz średnia arytmetyczna, medianę oraz wartość modalną (dominantę) szeregu rozdzielczego punktowego.
- E. Oblicz wariancję empiryczną, odchylenie standardowe dla szeregu rozdzielczego.
- F. Oblicz współczynnik asymetrii oraz eksces.

Zad. 2

Oceniano wymiar lewej komory serca w rozkurczu w grupie 112 pacjentów z dolegliwościami dławicowymi bez przebytego wcześniej zawału serca. Otrzymano następujące wyniki:

57	50	48	48	43	56	65	58	54	66	53	53
48	50	50	51	53	68	52	52	55	55	55	55
49	56	62	56	51	49	62	50	56	61	65	56
56	55	55	59	54	44	51	53	50	49	50	58
50	56	57	51	53	55	58	50	56	60	49	44
55	53	52	53	55	54	51	56	48	54	54	55
53	70	63	62	54	55	56	72	55	61	54	55
59	48	57	53	56	52	52	54	47	50	47	52
50	47	56	51	44	46	51	59	61	55	55	53
66	49	59	67								

- Znajdź medianę, odstęp międzykwartylowy
- Oszacuj liczbę przedziałów, zbuduj szereg **rozdzielczy klasowy**
- Sporządź histogram próby oraz diagram licznosci (częstości), diagram licznosci (częstości) skumulowanej.
- Obliczyć wartości parametrów opisowych dla badanej cechy (średnia, odchylenie standardowe, dominanta dla szeregu rozdzielczego).

Zad.3

W mieście zbadano liczbę zachorowań na anginę dzieci w wieku 6 lat. Uzyskano następujące wyniki:

Liczba zachorowań w ciągu roku	Liczba dzieci
0	60
1	120
2	200
3	80
4	30
5	10

Obliczyć przeciętną liczbę zachorowań na anginę w ciągu roku.