

ORGANIZACJA BADANIA STATYSTYCZNEGO

Określenie:

- a) populacji
- b) jednostki populacji
- c) cechy populacji

Metody badania statystycznego

- 1) Badanie pełne (badanie obejmuje całą populację)
- 2) Badanie częściowe (badanie odbywa się na pewnych losowo wyodrębnionych elementach populacji, czyli **próbie losowej**)
 - a) metoda reprezentacyjna
 - b) metoda monograficzna
 - c) metoda ankietowa

OPRACOWANIE MATERIAŁU STATYSTYCZNEGO

Charakterystyki położenia

- Średnia arytmetyczna:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- inne średnie:
 średnia harmoniczna
 średnia geometryczna

- Mediana

$$Me = \begin{cases} \frac{x_{n+1}}{2} & \text{dla } n \text{ nieparzystych} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{dla } n \text{ parzystych} \end{cases}$$

Charakterystyki rozproszenia

- Odchylenie przeciętne

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

- Wariancja

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

- Odchylenie standardowe

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

wartości typowe: $(\bar{x} - s, \bar{x} + s)$

- **współczynnik zmienności**

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

- kwartyle, decyle, centyle

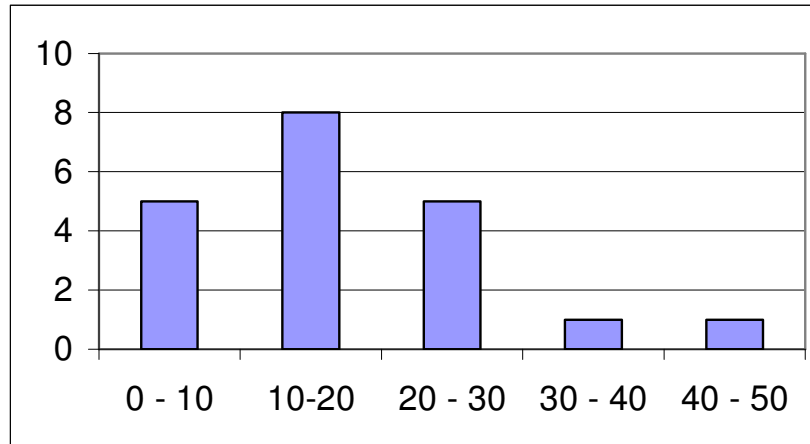
Grupowanie danych

- proste (jedna cecha np. wg wieku)

- złożone (wiele cech np. wg wieku i płci)

Wartości cechy (np. wiek)	Liczebność	Częstość
0 - 10	5	0.25
10 - 20	8	0.40
20 - 30	5	0.25
30 - 40	1	0.05
40 - 50	1	0.05

Przedstawianie graficzne za pomocą **histogramu**



Estymacja - to dział wnioskowania statystycznego będący zbiorem metod pozwalających na uogólnianie wyników badania próby losowej na nieznaną postać i parametry rozkładu zmiennej losowej całej populacji oraz szacowanie błędów wynikających z tego uogólnienia.

Metody estymacji parametrycznej można w zależności od sposobu szacowania szukanego parametru podzielić na dwie grupy:

- **estymacja punktowa** (szacowanie wartości)
- **estymacja przedziałowa** (szacowanie przedziałów ufności)

Estymatory punktowe

estymator wariancji

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

suma kwadratów odchyłeń od średniej

$$\text{var } X = \sum_{i=1}^n (x_i - \bar{x})^2$$

estymator odchylenia standardowego

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

Estymatory przedziałowe

Przedział ufności dla średniej

$$\left(\bar{X} - t_{(\alpha; n-1)} \frac{s}{\sqrt{n}}, \bar{X} + t_{(\alpha; n-1)} \frac{s}{\sqrt{n}} \right)$$

$t_{(\alpha; n-1)}$: wartość krytyczna rozkładu t - Studenta

$n-1$ lub ν - stopnie swobody

α - poziom istotności (zazwyczaj przyjmujemy $\alpha=0,05$)

Poziom ufności: $1 - \alpha$ ustalone z góry
prawdopodobieństwo z jakim ten przedział pokrywa
nieznaną wartość parametru np. średniej

Przedział ufności dla wariancji

$$\left(\frac{\text{var } X}{\chi^2_{\left(\frac{\alpha}{2}; n-1\right)}}, \frac{\text{var } X}{\chi^2_{\left(1-\frac{\alpha}{2}; n-1\right)}} \right)$$

$\chi^2(\alpha; n - 1)$ jest wartością krytyczną rozkładu chi-kwadrat z ν stopniami swobody.

Przedział ufności dla odchylenia standardowego

$$\left(\sqrt{\frac{\text{var } X}{\chi^2_{\left(\frac{\alpha}{2}; n-1\right)}}}, \sqrt{\frac{\text{var } X}{\chi^2_{\left(1-\frac{\alpha}{2}; n-1\right)}}} \right)$$