

Statystyka medyczna

Część 1

Testy diagnostyczne

CZUŁOŚĆ

- zdolność testu do prawidłowego rozpoznania choroby tam, gdzie ona występuje
- liczba wyników prawdziwie dodatnich w badanej populacji w stosunku do liczby przebadanych pacjentów z potwierdzoną chorobą

$$\text{CZUŁOŚĆ} = (\text{osoby z chorobą i dodatnim wynikiem testu} / \text{wszystkie osoby chore}) \times 100\%$$

Testy diagnostyczne

SWOISTOŚĆ

- zdolność testu do wykrywania osobników bez danej cechy (np. rzeczywiście zdrowych)
- liczba wyników prawdziwie ujemnych w badanej populacji w stosunku do liczby przebadanych pacjentów bez choroby

SWOISTOŚĆ = (osoby bez choroby z ujemnym wynikiem testu/wszystkie osoby bez choroby)x100%

Zarówno czułość, jak i swoistość testu są ważnymi wskaźnikami dokładności testu i osobno nie dają pełnego obrazu. Tylko wspólnie dają one pojęcie o stopniu zaufania, jakim można darzyć dany test.

		Stan (np. choroba) Określana jako „złoty” standard	
		Prawdziwy	Fałszywy
Wynik Testu	Dodatni	Prawdziwie dodatni	Fałszywie dodatni
	Ujemny	Fałszywie ujemny	Prawdziwie ujemny
		↓ Czułość	↓ Swoistość

$$\text{czułość} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{prawdziwie pozytywni}}{\text{suma chorych}}$$

$$\text{swoistość} = \frac{\text{TN}}{\text{FP} + \text{TN}} = \frac{\text{prawdziwie negatywni}}{\text{suma zdrowych}}$$

DODATNIA WARTOŚĆ PREDYKCYJNA - oznacza prawdopodobieństwo, że badany miał chorobę mając pozytywny wynik testu.

UJEMNA WARTOŚĆ PREDYKCYJNA - oznacza prawdopodobieństwo, że badany nie ma danej choroby w przypadku negatywnego wyniku testu.

$$\begin{array}{l} \text{wartość} \\ \text{predykcyjna} \\ \text{dodatnia} \end{array} = \frac{PD}{PD + FD} \times 100\%$$

$$\begin{array}{l} \text{wartość} \\ \text{predykcyjna} \\ \text{ujemna} \end{array} = \frac{PU}{PU + FU} \times 100\%$$

Chcemy poznać to:



Populacja



μ , Population Mean

σ , Population Std. Dev.

π , Population Proportion

Parametr

Dobieramy
losowo



Badamy to:



Próbka



$\bar{x}$, Sample Mean

s , Sample Std. Dev.

p , Sample Proportion

Statystyka

Wnioskujemy
(inferujemy)



Miary tendencji centralnych - dane niezgrupowane

1. Średnia arytmetyczna

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

2. Mediana

Mediana jest to środkowy wynik w uporządkowanym zbiorze

$$\text{Me} = \begin{cases} x_{\frac{n+1}{2}} & \text{gdy } n \text{ jest nieparzyste} \\ \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{gdy } n \text{ jest parzyste} \end{cases}$$

3. Moda (dominanta) - najczęściej otrzymywana wartość cechy,

Miary tendencji centralnych - dane niezgrupowane

Q_1 (pierwszy kwartył) – wartość, poniżej której lub na równi mamy 25% obserwacji

Q_2 (drugi kwartył, mediana) – wartość, poniżej której lub na równi mamy 50% obserwacji

Q_3 (trzeci kwartył) – wartość, poniżej której lub na równi mamy 75% obserwacji

Pierwszy kwartył (lub dolny kwartył) Q_1 , drugi kwartył (medianę) $Q_2 = \text{Me}$ i trzeci kwartył (górny kwartył) Q_3 :

$$Q_m = \begin{cases} \frac{1}{2} \left(x_{\frac{mn}{4}} + x_{\frac{mn}{4}+1} \right) & \text{gdy } n|4 \\ \frac{1}{2} \left(x_{\frac{m(n+1)}{4}-0,5} + x_{\frac{m(n+1)}{4}+0,5} \right) & \text{gdy } n+3|4 \\ x_{\frac{mn}{4}-0,5} & \text{gdy } n+2|4 \\ x_{\frac{m(n+1)}{4}} & \text{gdy } n+1|4 \end{cases}$$

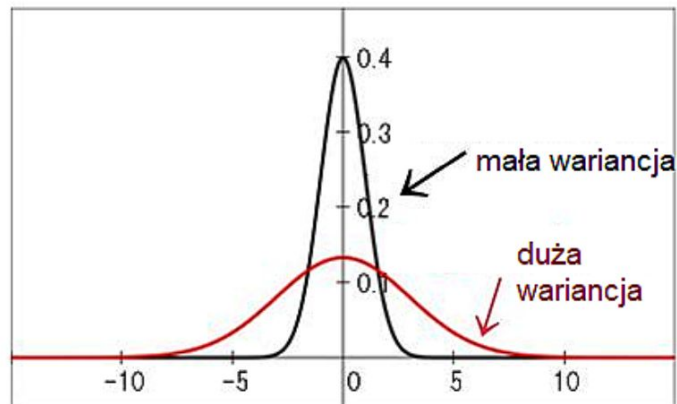
dla $m = 1, 3$.

Miary rozrzutu - dane niezgrupowane

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

Uwaga: wzór ten jest tożsamy z:

$$s^2 = \overline{x^2} - \bar{x}^2, \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$$



Miary rozrzutu - dane niezgrupowane

Odchylenie standardowe to pierwiastek z wariancji

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

Współczynnik zmienności

$$V_s = \frac{s}{\bar{x}}$$

Odchylenie przeciętne

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

Odchylenie ćwiartkowe

$$Q = \frac{Q_3 - Q_1}{2}$$

Średnia w danych zgrupowanych:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_mx_m}{n} = \frac{1}{n} \sum_{i=1}^m f_i x_i$$

(f_i to częstości w klasach, x_i to środki klas, $i=1,2,\dots, m$)

Wariancję dla danych zgrupowanych obliczymy według wzoru:

$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N}$$

gdzie f_i to częstości

x_i to środek (reprezentacja) klasy

$\bar{x}$ to średnia

$N = f_1 + \dots + f_m$ to całkowita liczba danych

Mediana

dana jest wzorem

$$Me = x_{Me} + \frac{\frac{n}{2} - \sum_{i=1}^{k_{Me}-1} n_i}{n_{Me}} i_{Me} ,$$

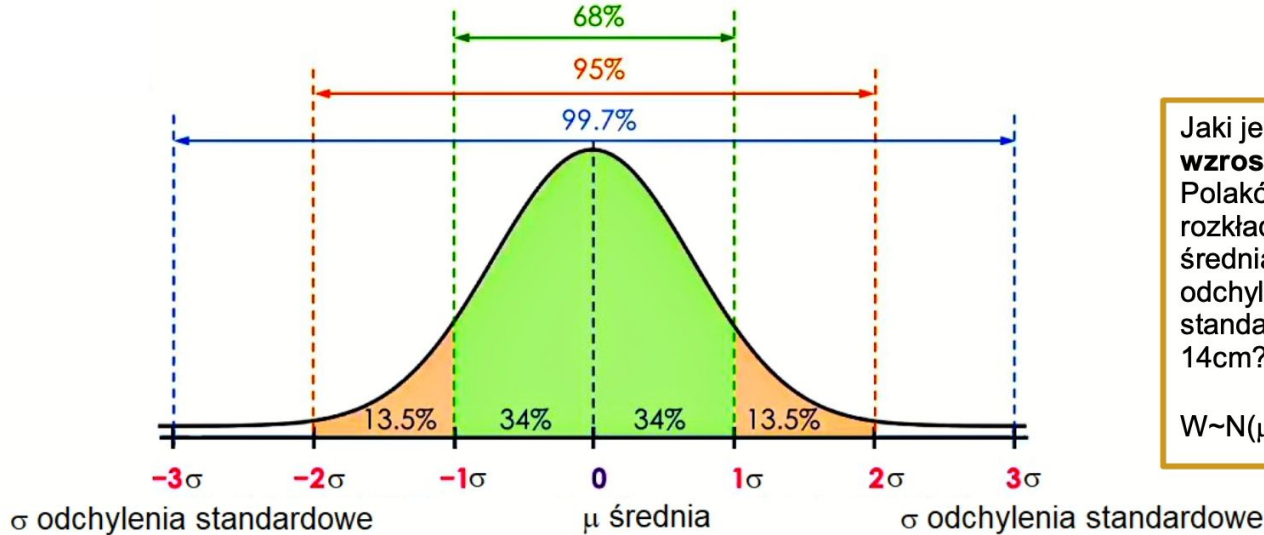
gdzie x_{Me} jest dolnym krańcem przedziału, w którym znajduje się mediana, k_{Me} – numerem przedziału, w którym znajduje się mediana, i_{Me} – szerokością przedziału, w którym znajduje się mediana, n_{Me} – liczba danych w przedziale, w którym jest mediana. Przedział, w którym znajduje się mediana jest to ten przedział, w którym znajduje się obserwacja o numerze równym połowie liczby obserwacji.

Kwartyle obliczane są według wzoru

$$Q_j = x_{Q_j} + \frac{\frac{jn}{4} - \sum_{i=1}^{k_{Q_j}-1} n_i}{n_{Q_j}} i_{Q_j} ,$$

gdzie użyte symbole mają podobne znaczenie.

Rozkład normalny



Jaki jest zakres **wzrostu** dorosłych Polaków, jeśli ma on rozkład normalny ze średnią 170cm i odchyleniem standardowym 14cm?

$$W \sim N(\mu=170, \sigma=14)$$

Zasada 3 sigma: Wszystkie dane mieszczą się w zakresie $\mu \pm 3\sigma$