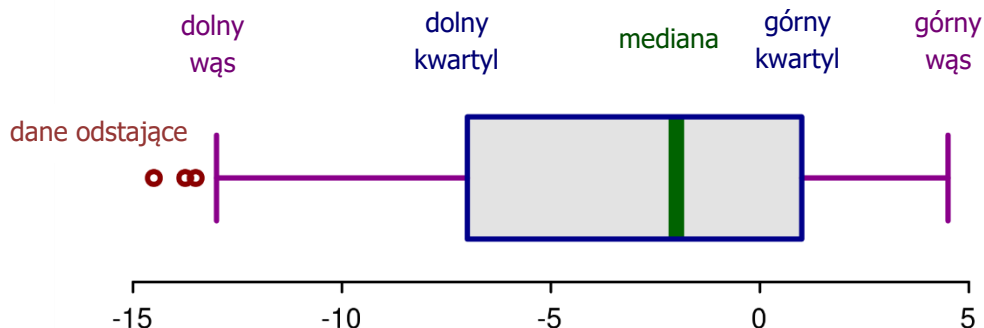


Wykres pudełkowy

- wykres ten służy do prezentacji wyników i zawiera informacje odnośnie położenia i rozproszenia danych
- najczęściej używany do porównywań

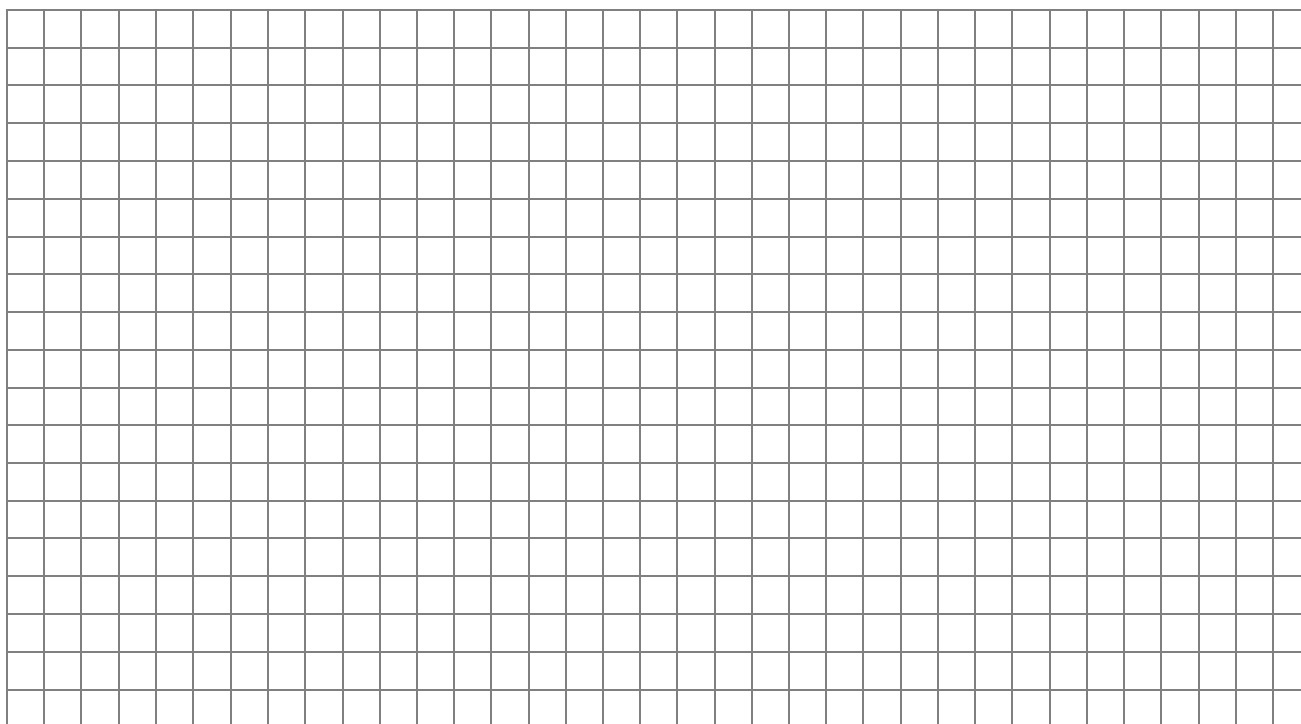


Źródło: https://upload.wikimedia.org/wikipedia/commons/d/d1/Elements_of_a_boxplot_en.svg

- pierwszy kwartyl (Q1) = dolny kwartyl = 25% obserwacji jest położonych poniżej = 25 procent
- drugi kwartyl (Q2) = mediana = dzieli zbiór obserwacji na połowę = 50 procent
- trzeci kwartyl (Q3) = górny kwartyl = dzieli zbiór obserwacji na dwie części odpowiednio po 75% położonych poniżej tego kwartyla i 25% położonych powyżej = 75 procent

Trend liniowy a tendencja

- służy do oszacowania wartości Y gdy zmienną X jest czas
- wówczas mamy do czynienia z danymi, które są uszeregowane względem czasu



***Korelacja = zależność danych**

- określa związek (nie zawsze bezpośrednią zależność) między zmiennymi (dane są ze sobą jakoś powiązane, czyli skorelowane), czyli jak bardzo jedna zmienna wpływa na drugą zmienną (i czy w ogóle istotnie wpływa)
- wartość korelacji (r) jest liczbą z przedziału $[-1, 1]$
- korelacja nie zawsze oznacza bezpośrednią przyczynowość

Rodzaj korelacji:

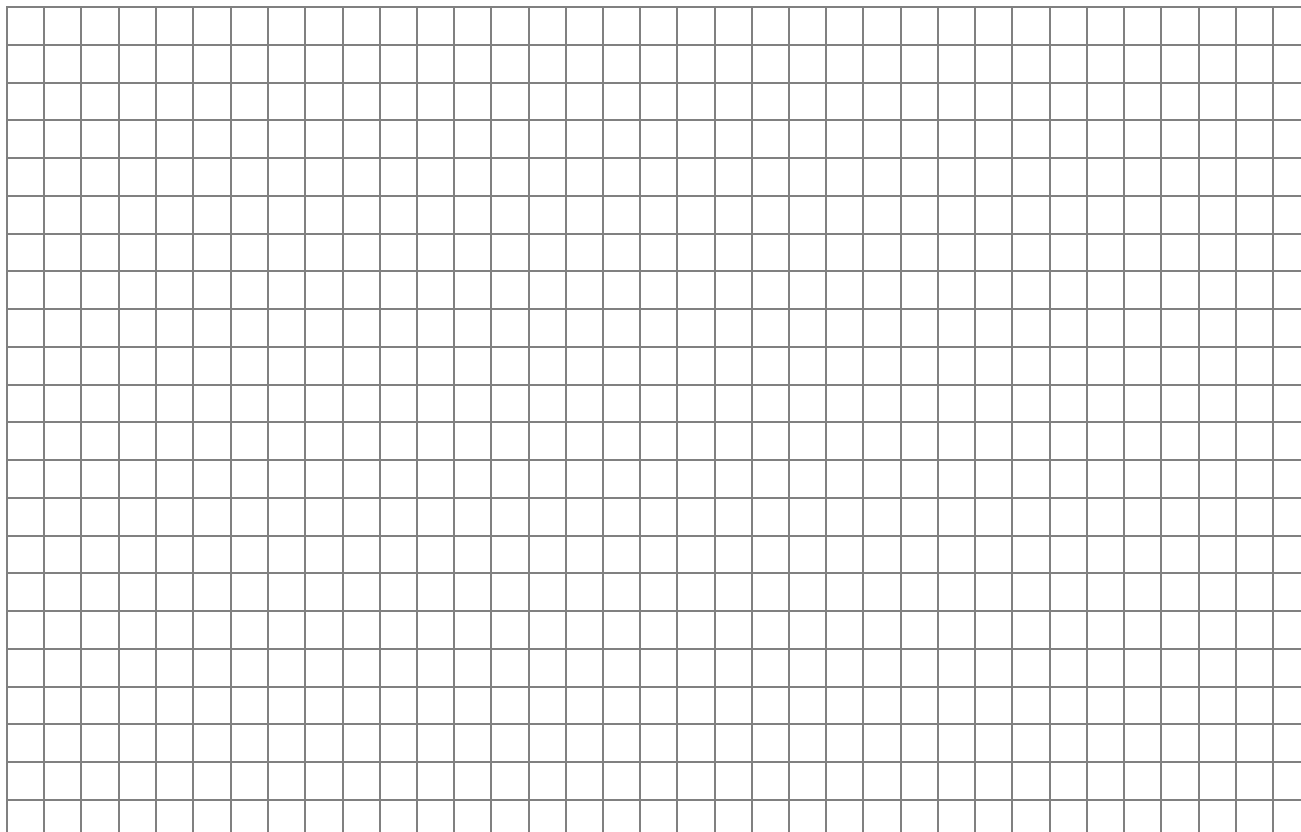
- $r > 0$; dodatnia korelacja, zależność wprost proporcjonalna (rośnie X to rośnie Y albo maleje X to maleje Y)
- $r < 0$; ujemna korelacja, zależność odwrotnie proporcjonalna (rośnie X to maleje Y albo maleje X to rośnie Y)

Siła korelacji:

- $< 0,2$ – brak korelacji
- $0,2 - 0,4$ – słaba korelacja
- $0,4 - 0,7$ – umiarkowana korelacja
- $0,7 - 0,9$ – dość silna korelacja
- $> 0,9$ – bardzo silna korelacja

Zakres i rozstęp danych

W badaniu otrzymano następujące wielkości lęgu kanadyjskiego gatunku gęsi (*Branta canadensis*): 8, 6, 5, 7, 5, 9, 3, 5.



Rozkład normalny i odchylenie standardowe/wariancja

Wariancja

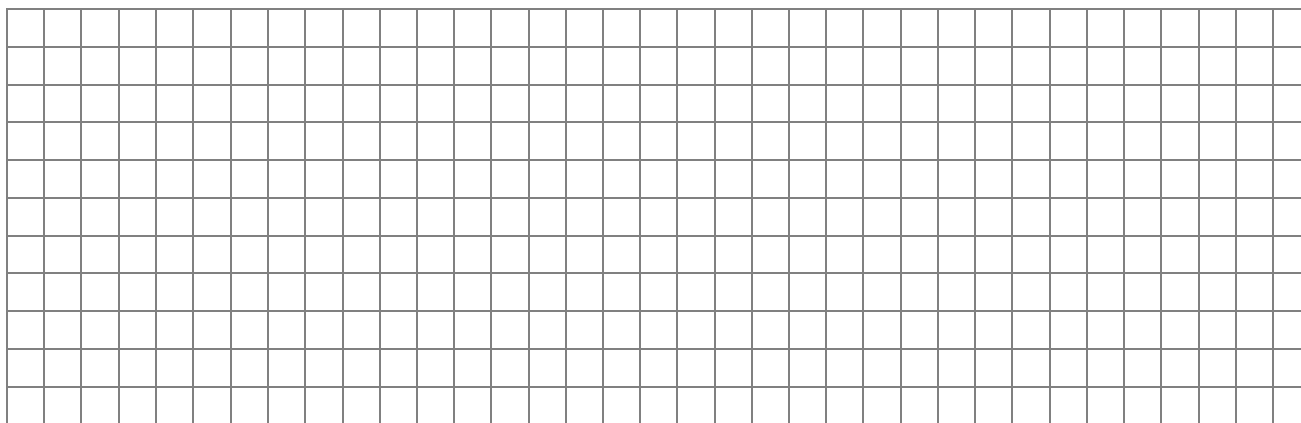
- jest miarą zróżnicowania czyli pozwala nam powiedzieć, czy wyniki leżą blisko średniej, czy są jednak bardzo rozproszone (odległe od średniej)

Odchylenie standardowe

- to pierwiastek z wariancji (ale w ten sam sposób opisuje zróżnicowanie cechy)

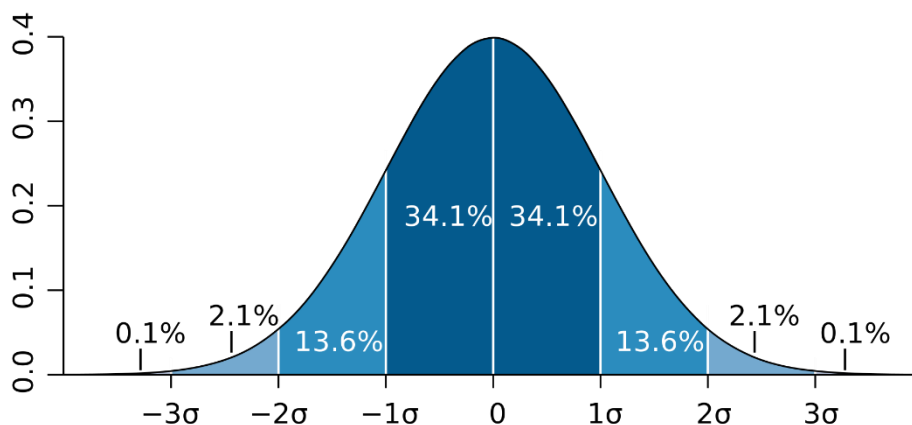
Rozkład normalny

- jest zwany również rozkładem Gaussa lub krzywą Gaussa
- jeden z najważniejszych rozkładów statystycznych – idealny dla biologii
- według tego rozkładu większość obserwacji jest skupiona wokół średniej, obserwacje leżące dalej od średniej występują rzadziej
- jeśli rozkład danych jest rozkładem normalnym to średnia równa się medianie i dominancie



Rozkład normalny ma 2 parametry:

- μ (wartość oczekiwana, czyli w tym przypadku średnia)
- σ (odchylenie standardowe) – im większe odchylenie standardowe tym więcej danych jest bardziej oddalonych od średniej



Źródło: https://pl.wikipedia.org/wiki/Rozk%C5%82ad_normalny#/media/Plik:Standard_deviation_diagram.svg

***Reguła Trzech Sigm**

- jest to reguła wprowadzająca przedziały względem średniej dla danych wykazujących rozkład normalny
 - w przedziale $[\mu - \sigma, \mu + \sigma]$ znajduje się 68,2% wszystkich obserwacji (68%)
 - w przedziale $[\mu - 2\sigma, \mu + 2\sigma]$ znajduje się 95,4% wszystkich obserwacji (95%)
 - w przedziale $[\mu - 3\sigma, \mu + 3\sigma]$ znajduje się 99,7% wszystkich obserwacji
- do czego służy? Po pierwsze można zobaczyć, jak bardzo prawdopodobne jest dane zdarzenie, czyli ile jest procent szansy, że otrzymasz dany wynik. Po drugie można łatwo określić wyniki odstające

Obliczanie i rysowanie odchylenia standardowego (wzór w tablicach CKE)

Chcemy porównać dane z dwóch zbiorów i sprawdzić, czy różnica pomiędzy tymi zbiorami jest istotna statystycznie.

Pomiar	Grupa A	Grupa B
1	32	40
2	37	47
3	41	31
4	25	29
5	28	42
średnia	32,6	37,8
odch.st.	5,82	6,79

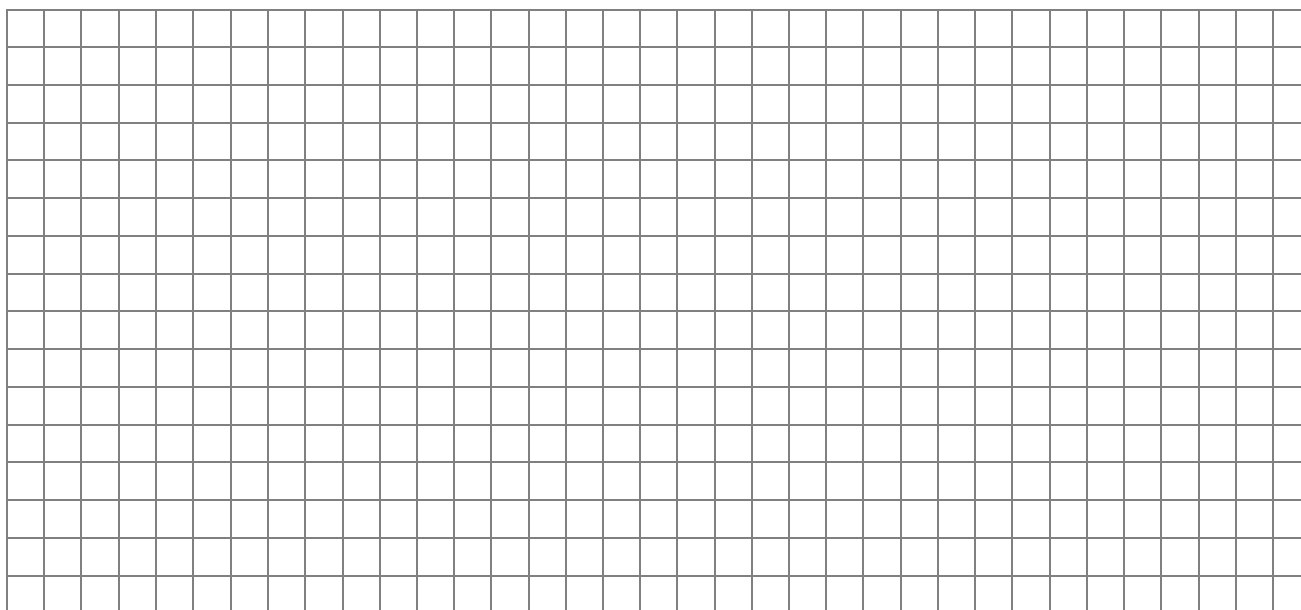
Uwaga:

Wzory matematyczne i przykłady rachunkowe podano w celu lepszego zrozumienia omawianych miar statystycznych. Na egzaminie maturalnym z biologii nie przewiduje się ich obliczania, ale zakłada się wyłącznie interpretację danych przy użyciu podstawowych miar statystycznych.

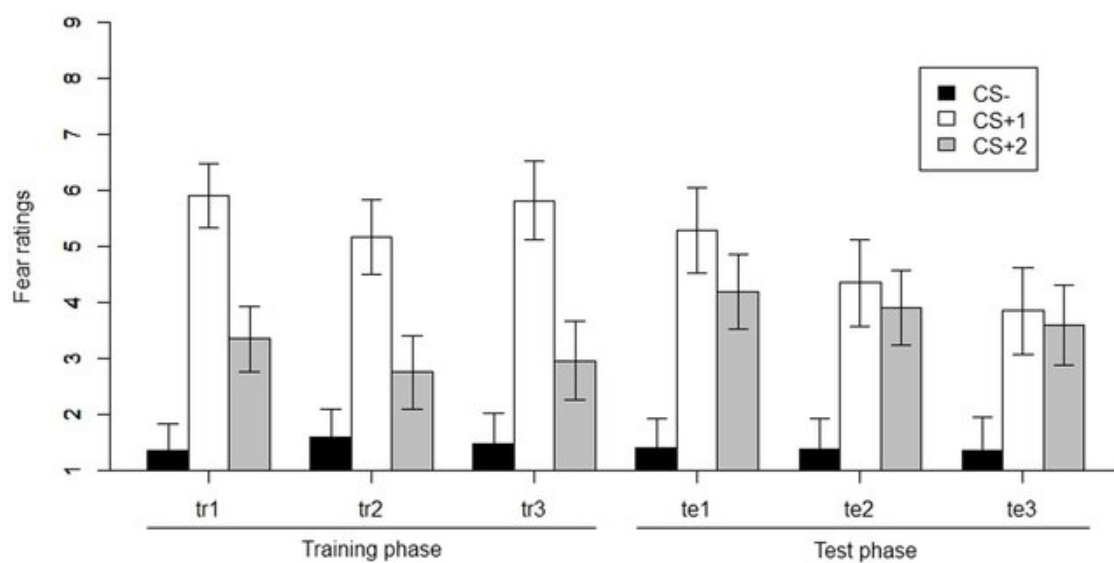
Do obliczenia średniej i odchylenia standardowego służą wzory z tablic:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Narysuj wykres słupkowy z uwzględnieniem odchyłeń standardowych



Interpretacja odchylenia standardowego (rzetelność średniej, istotność różnic)



Źródło: Raes A., De Houwer J., De Schryver M., Kalish R., Do CS-US Pairings Actually Matter? A Within-Subject Comparison of Instructed Fear Conditioning with and without Actual CS-US Pairings, PLoS ONE 9(1):e84888, 2014. Licencja: CC BY 4.0

