

# Statystyki opisowe

Wojciech Kotłowski

Statystyka i analiza danych 2019/2020

17.03.2020

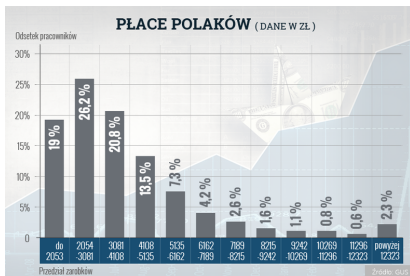
## Miary położenia (tendencji centralnej)

- Dominanta (moda)
- Średnia arytmetyczna
- Mediana
- Kwantyle (m.in. kwartyle)
- Średnia geometryczna
- Średnia harmoniczna
- Średnia ucięta

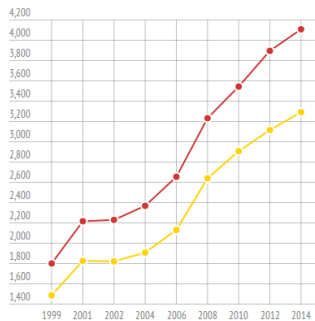
# Średnia a mediana

- Czy średnia i/lub mediana należą do zbioru wartości przyjmowanych przez dane?
- Która z miar jest bardziej odporna na wartości odstające? Jak można temu zaradzić?
- Według Facebooka (2011), średnia liczba znajomych użytkowników portalu wynosi 190, a połowa użytkowników ma więcej niż 100 znajomych. Dlaczego?
- Ile wynosi suma odchyleń od średniej  $\sum_{i=1}^n (x_i - \bar{x})$ ?
- Domy w pewnej okolicy mają średnią 1mln PLN, ale mediana wynosi tylko 600 tys. PLN. Skąd wynika ta asymetria?

# Średnia a mediana: wynagrodzenia (2014)



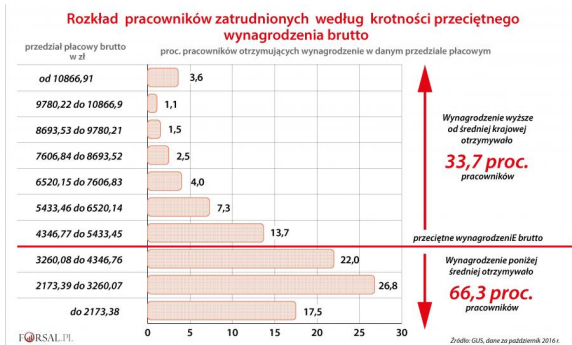
źródło: onet.pl



● Mediana ● Średnia

Źródło: GUS

# Wynagrodzenia: grudzień 2017 (średnie i duże firmy)



źródło: forsal.pl

Średnie wynagrodzenie: 4973,73zł

# Miary rozproszenia

- Rozstęp  $x_{\max} - x_{\min}$
- Wariancja  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  i odchylenie  $s = \sqrt{s^2}$ .
- Rozstęp międzykwartylowy (IRQ)  $Q_3 - Q_1$

# Wariancja i odchylenie standardowe

- Jak zmienia się wariancja gdy do wszystkich danych dodamy stałą?
- Skrócony wzór na wariancję:

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

- Po co nam odchylenie standardowe?

## Pozostałe

- Współczynnik asymetrii  $\frac{M_3}{s^3}$  ( $M_3$  – moment centr. 3. rzędu)
- Współczynnik skośności  $\frac{\mu-d}{s}$  lub  $3\frac{\mu-m}{s}$  ( $d$  – dominanta,  $m$  – mediana)
- Kurtoza  $\frac{M_4}{s^4}$  ( $M_4$  – moment centr. 4. rzędu)