

Ilu poddanych będzie miał książę #Wilmon (jak już zostanie królem)

Marek Galewski



Politechnika Łódzka

18 kwietnia, 2023

**Cel badawczy wykładu:** Estymowanie wielkości populacji Szwecji w perspektywie krótkookresowej (20-30lat).

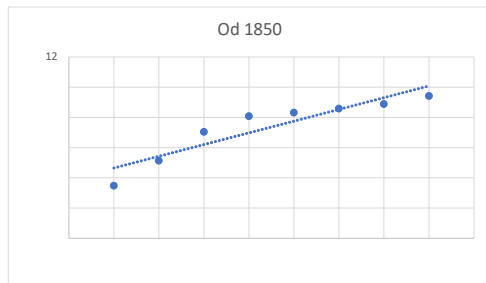
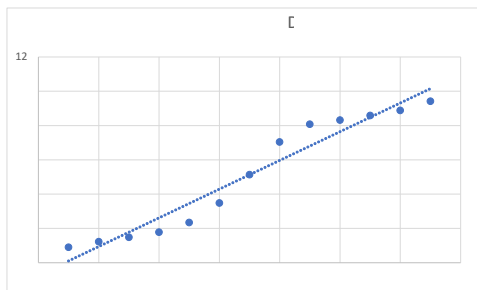
**Uzasadnienie:** Po prawdzie książkę Wilhelm (serial Netflix „Young Royals”) to postać fikcyjna, ale jest następcą tronu jak najbardziej rzeczywistego kraju -Szwecji. Kiedyś zostanie królem (o ile nie zostanie nim August) i będzie miał swoich poddanych. Ciekawe zatem ilu? I w jaki sposób można dostępne dane demograficzne opracować tak, by otrzymać w miarę prosty i w miarę sensownie prognozujący model matematyczny.



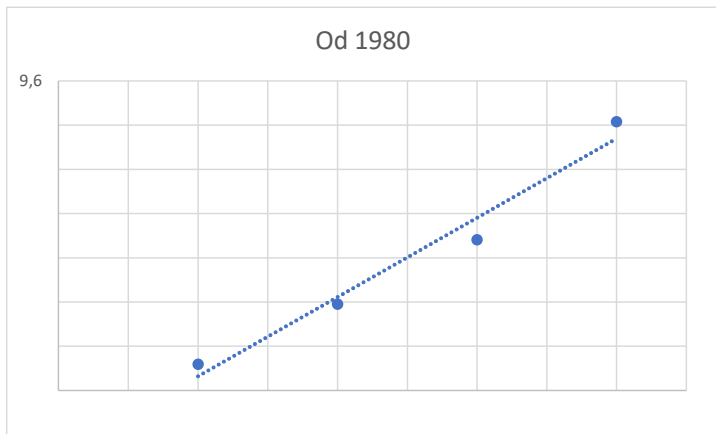
# Dane demograficzne

Rok	Pop.	Rok	Pop.	Rok	Pop.
1570	900,000	1850	3,482,500	1980	8,318,000
1650	1,225,000	1900	5,136,400	1990	8,590,600
1700	1,485,000	1950	7,041,900	2000	8,882,800
1750	1,780,700	1970	8,081,300	2010	9,415,600
1800	2,347,300			2020	10,099,295

# Kilka wykresów



# Model graficzny i jak go wyznaczyć analitycznie



Funkcja liniowa (linia trendu) "prawie przybliża" badane przez nas dane populacyjne. Jak tę funkcję wyznaczyć (z najmniejszym możliwym błędem, czyli najlepszą możliwą dokładnością)? Jak zweryfikować wynik? Czy jest sens znajdować wielomian, którego wykres przechodziłby przez uzyskane

## Trochę o danych i zarysie metody

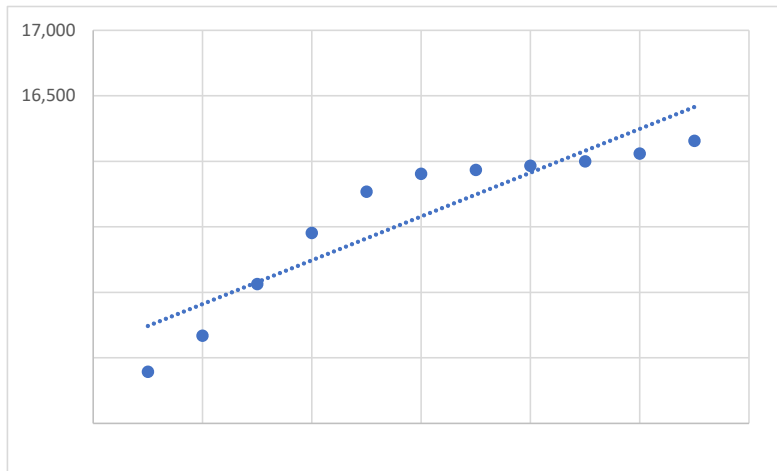
Metoda ta służy do dopasowywania układów danych najczęściej przedstawionych graficznie i będących wynikami pomiarów do postaci krzywej o której sądzi się, że dane najlepiej do niej pasują. Najprościej więc "dopasować dane" do linii prostej (co zresztą odpowiada naszej sytuacji).

Dane są to pary  $(t_1, s_1), (t_2, s_2), \dots, (t_n, s_n)$  – czas i wynik pomiaru. Jeżeli dane "nie pasują" wizualnie do wykresu linii prostej to liczymy ich logartymany naturalne

$$(t_1, \ln s_1), (t_2, \ln s_2), \dots, (t_n, \ln s_n)$$

i do dla nich szukamy wzoru interesującej nas funkcji. Pojedyncze (nieliczne) "odstające od schematu" dane bądź uśredniamy, bądź pomijamy.

# Wykres kontr-atakuje, czyli zlogarytmowane dane

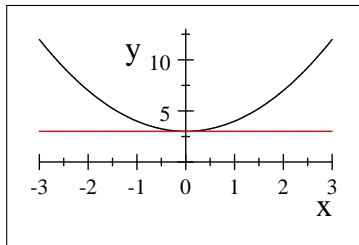


Niestety w naszym przypadku logarytmowanie nie pomaga - raczej w doborze zakresu danych pomoże nam historia.

# Jak znaleźć tę "linię prostą"-czyli o optymalizacji

Funkcja  $f : \mathbb{R} \rightarrow \mathbb{R}$  posiada w punkcie  $x_0$  argument minimum:

$$f(x) \geq f(x_0) \text{ dla wszystkich } x \in \mathbb{R}$$



Wykres funkcji  $f(x) = x^2 + 3$  i jej stycznej w  $x_0 = 0$

Argmin  $x_0 = 0$ ,  $f'(x_0) = 0$ , styczna  $y = 3$ . **Pochodna**

$f'(x) = (x^2)' = 2x$ . Ogólniej:  $(x^n)' = nx^{n-1}$  dla  $n = 1, 2, 3, \dots$ ,

$(const)' = 0$ . **Styczna do wykresu  $f$  w  $(x_0, f(x_0))$ :**

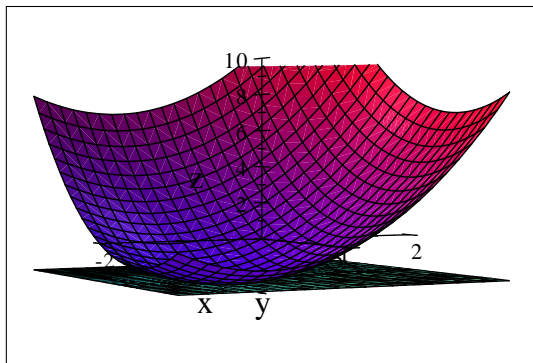
$$y = f'(x_0)(x - x_0) + f(x_0)$$



## Więcej zmiennych- co wtedy?

Funkcja  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  posiada w punkcie  $(x_0, y_0)$  argument minimum:

$$f(x, y) \geq f(x_0, y_0) \text{ dla wszystkich } (x, y) \in \mathbb{R} \times \mathbb{R}$$



Wykres  $(x - 1)^2 + (y + 1)^2 - 2$  oraz płaszczyzny  $z = -2$

Argmin  $(x_0, y_0) = (1, -1)$ ,  $f'_x(x_0, y_0) = f'_y(x_0, y_0) = 0$ , styczna  $z = -2$   
równoległa do płaszczyzny  $OXY$ .

# Metoda najmniejszych kwadratów

Dla zbioru danych  $(t_1, s_1), (t_2, s_2), \dots, (t_n, s_n)$  poszukujemy funkcji

$$r(t) = b + at$$

takiej, że suma

$$(r(t_1) - s_1)^2 + (r(t_2) - s_2)^2 + \dots + (r(t_n) - s_n)^2$$

jest możliwie najmniejsza. Kilka pytań:

- czemu brakuje pierwiastka?
- czemu kwadraty nie moduły?
- czym się ta funkcja różni od poznanych funkcji kwadratowych?
- skąd wiemy, że taka suma będzie możliwie najmniejsza?
- ile jest zmiennych?

# Kilka odpowiedzi:

- a) poszukajmy argumentu minimum funkcji  $\sqrt{(x-3)^2 + 5}$  oraz  $(x-3)^2 + 5$  i porównajmy;
- b) to samo zadanie dla funkcji  $|x-3| + 5$  oraz  $(x-3)^2 + 5$ ;
- c)-d) dowiedcie się tego na studiach na kierunkach *Matematyka Stosowana*, *Aktuariat i Analiza Finansowa* lub **Mathematical Methods in Data Analysis** na najlepszym wydziale Politechniki Łódzkiej -**FTIMS**ie:-)



# To jaki w końcu jest ten wzór (wariacie)?

Podamy teraz wzory na współczynniki  $a$ ,  $b$  we wzorze aproksymacji średniokwadratowej

$$r(t) = b + at.$$

Oznaczmy

$$\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i, \quad \bar{s} = \frac{1}{n} \sum_{i=1}^n s_i$$

Wtedy

$$a = \frac{\frac{1}{n} \sum_{i=1}^n t_i s_i - \bar{t} \bar{s}}{\frac{1}{n} \sum_{i=1}^n t_i^2 - (\bar{t})^2}$$
$$b = \bar{s} - a \bar{t}$$

# Ilu poddanych będzie mieć Wilhelm?

**Obserwacja:** żeby na to pytanie dobrze odpowiedzieć musimy stwierdzić (przy pomocy naszego modelu) ilu poddanych miała w 2022 królowa Kristina. Model musimy wytestować na dostępnych danych. Wybierzmy dane:

1980=1	8,318,000	.....	2010=4	9,415,600	<b>2020=5</b>	10,099,238
--------	-----------	-------	--------	-----------	---------------	------------

$$\bar{t} = \frac{5}{2}, \quad \bar{s} = \frac{1}{4} (8318 + 8590.6 + 8882.8 + 9415.6) = 8801.8$$

$$a = \frac{\frac{1}{4} (1 \cdot 8318 + 2 \cdot 8590.6 + 3 \cdot 8882.8 + 4 \cdot 9415.6) - \frac{5}{2} \cdot 8801.8}{\frac{1}{4} (1^2 + 2^2 + 3^2 + 4^2) - \left(\frac{5}{2}\right)^2} = 358.4$$

$$b = 8801.8 - \frac{5}{2} \cdot 358.4 = 7905.8$$

$$r(t) = 7905.8 + 358.4 \cdot t \text{ oraz } r(5) = 7905.8 + 358.4 \cdot 5 = 9697.8$$

Jaki jest błąd?

$$\frac{10099 - 9697.8}{10099} \text{ czyli } 3.97\%$$

# Jak poprawić otrzymany wynik?

- 1 Brać "gęstsze wyniki" - spis roczny; zwiększyć dokładność obliczeń?
- 2 Oddzielnie estymować prowincje oraz grupy narodowościowe.
- 3 Uwzględnić migrację.
- 4 Zastanowić się, czy uzyskany tak prostymi metodami model jest już optymalny.

# Zastosujmy rozwiązanie nr 1

Dane:

<https://www.worldometers.info/world-population/sweden-population/>

<b>2020=12</b>	10,099,295
----------------	------------

$$r(t) = 9.2852 \times 10^6 + 68209 \cdot t$$

$$r(12) = 9.2852 \times 10^6 + 68209 \cdot 12 = 1.0104 \times 10^7$$

Jaki jest błąd?

$$\frac{|10099295 - 1.0104 \times 10^7|}{10099295} \text{ czyli } 4.7\%$$

Czyli trzeba szukać lepszego doboru danych, albo pozostać przy otrzymanym wcześniej oszacowaniu.

# Ilu poddanych będzie miał #Wilmon (wg modelu)

Przypomnijmy wzór

$$r(t) = 7905.8 + 358.4 \cdot t,$$

gdzie  $t = 1$  oznacza 1980,  $t = 2$  oznacza 1990. Założyliśmy, że interesuje nas perspektywa 20 lat, a zatem

$$r(7) = 7905.8 + 358.4 \cdot 7 = 10\ 415$$

Założywszy, że Wilhelm wstąpi na tron w 2060 (królowe żyją dość długo), to

$$r(9) = 7905.8 + 358.4 \cdot 9 = 11\ 131.$$



