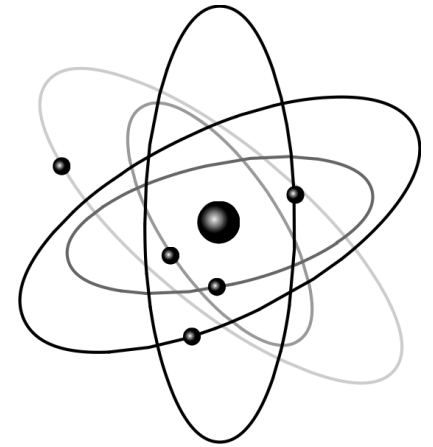




**Seminarium Wydziałowe**  
**10 czerwca 2022**



# **Wybór modelu dopasowania - testy statystyczne i kryterium informacyjne Akaike'go**

***Andrzej Zięba***

*Wydział Fizyki i Informatyki Stosowanej, AGH Kraków*

Referat przygotowany w związku  
z przygotowaniem  
manuskryptu podręcznika

A. Zięba

**ANALYSIS OF EXPERIMENTAL DATA**  
**in science and technology**

Cambridge Scholar Publishing, Newcastle

# Potrzeba wyboru modelu dopasowania

## ♣ Dawniej

realizacja algorytmu dopasowanie funkcji wymagało wielkiego wysiłku. Metoda najmniejszych kwadratów stosowana była nieczęsto, przy wykorzystaniu **jednej**, dobrze uzasadnionej **funkcji** (zwykle liniowej).

## ◆ Współcześnie

skomputeryzowane algorytmy metody najmniejszych kwadratów stosuje się powszechnie do **każdych danych** doświadczalnych, do których próbuje się dopasować **różne modele** teoretyczne. Umiejętność **wyboru modelu** staje się ważniejsza, niż kiedykolwiek przedtem.

# Plan

- Weryfikacja hipotez statystycznych
- Kryterium informacyjne Akaike'go
- Przykład 1: cechowanie miernika CH<sub>4</sub>
- Przykład 2: dobór stopnia wielomianu
- Przykład 3: analiza danych TOPR
- Wnioski

# Weryfikacja hipotez statystycznych

Przykład: rzucamy kostką  $n = 240$  razy. Szóstkę wyszła nam  $m = 62$  razy, gdy średnio powinno być:  $1/6 \times 240 = 40$ .  
Podejrzewamy, że kostka jest fałszywa.

- **Hipoteza zerowa:**  $H_0: P(x=6) = 1/6$
- **Hipoteza alternatywna\*:**  $H_1: P(x=6) > 1/6$
- **Zmienna testowa:** zależy wyłącznie od  $H_0$  wzór asymptotyczny 
$$t = \frac{m - \mu}{\sqrt{\mu}}$$
- Binarny wynik testu zależy od **poziomu istotności**  $\alpha$
- **Błąd I rodzaju:** test zaprzecza prawdziwej  $H_0$   
**Błąd II rodzaju:** test potwierdza fałszywą  $H_0$
- **Moc testu:** prawdopodobieństwo „wykrycia” fałszywej  $H_0$

\* - kolor - pojęcia wprowadzone w teorii N-P

# TESTY - RYS HISTORYCZNY

- od końca XVIII w - intuicyjne stosowanie testów, w oparciu o właściwości rozkładu normalnego. Test Chauvenet'a.
- testy dla małej liczebności próby:  
Karl Pearson - test chi-kwadrat  
William Gosset - test t Studenta
- Raymond Fisher - pierwsza ogólna teoria testów
- Egon S. Pearson i Jerzy Neyman - kompletna logicznie (?) teoria weryfikacji hipotez statystycznych

Jerzy Neyman and Egon S. Pearson. **1928**. "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part I" *Biometrika* **20A**, 175-240. Part II – *ibid.* 263-94.

Jerzy Neyman and Egon S. Pearson. **1936**. "Contributions to the Theory of Testings Statistical Hypotheses. *Statistical Research Memorandum* **1**, 1-37. *Ibid.*, 113-37: "Sufficient Statistics and Uniformly Most Powerful Tests of Statistical Hypotheses."

# Jerzy Neyman (1894 - 1981)



Courtesy of Berkeley Statistics Department Archives

*J. Neyman*





- Sejm 1875: nadanie szlachectwa Mateuszowi Neymanowi z Brodów, z przydaniem herbu Splawa. 21 potomków w Polskim Słowniku Biograficznym

- Jerzy Neyman – syn Czesława i Kazimiery z Lutosławskich. Ur. 1984 w mieście

**Bendery** (obecnie postsowiecka „**Naddniewstrzańska Republika Mołdawska**”).

- studia w **Charkowie**. Tamże początek pracy naukowej.

- 1921 - przybycie do Polski w ramach „wymiany jeńców” po pokoju ryskim

- w Polsce międzywojennej:

- doktorat na Uniwersytecie Warszawskim 1924

- praca jako statystyka w różnych instytucjach

- 40 publikacji, w tym 3 książki

- współpraca z Egonem Pearsonem

- od 1938 - Uniwersytet Kalifornijski w Berkeley

- kontynuacja pracy naukowej (np. statystyka gromad galaktyk)

- stworzenie Wydziału Statystycznego

- liczni doktoranci, najwybitniejszy – E. L. Lehmann

- konferencje: *Berkeley Symposium on Mathematical Statistics and Probability*

- związki z nauką w Polsce

# Wybór modelu: test F Snedecora-Fishera:

porównanie dwóch modeli 1 i 2, o liczbie parametrów  $m_1$  i  $m_2$

i wartościach funkcji kryterialnej  $M_1$  i  $M_2$

Przypadek uproszczony:  $m_1 = m$ ,  $m_2 = m + 1$

zmienna testowa: 
$$F = \frac{M(m+1) - M(m)}{M(m+1)} (n - m - 1)$$

Dzięki wykorzystaniu *stosunków* f. kryterialnych  $M$  nie jest potrzebna znajomość wartości bezwzględnych  $u(y_i)$

**Resume:** dane wejściowe: liczba parametrów funkcji  $m_1, m_2$   
wartości f. kryterialnej  $M_1, M_2$   
liczba punktów danych  $n$

# Kryterium informacyjne Akaike'go

## Informacja a teoria największej wiarygodności

- Boltzmann 1877. Entropia:  $S = k_B \ln \Omega$
- Lata 20-te: pierwsze wyniki t. informacji. Nyquist.
- Shannon, *Mathematical theory of communication* (1948): wykorzystanie entropii do sformułowania ilości informacji.
- Kullback and Leibler (1951). Odległość K-L – informacja  $I(g, f)$  stracona gdy funkcję prawdziwą  $f$  aproksymujemy przy pomocy funkcji  $g$ .
- Akaike (1974): związek między  $I(g, f)$  i funkcją największej wiarygodności  $L$

$$\text{AIC} = -2 \ln L + 2K$$



# Zastosowanie do dopasowania metodą najmniejszych kwadratów

- Model statystyczny:  $y = y(x_i, a_1, \dots, a_m) + e_i$

błąd pomiaru  $e_i$  ma rozkład Gaussa, liczba parametrów modelu:  
zwykle  $K = m + 1$

- Funkcja największej wiarygodności

$$L(a_1, \dots, a_m) = \prod_{i=1}^n g_i[y_i - y(x_i; a_1, \dots, a_m)].$$

gdzie

$$g_i(y_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ - \frac{[y_i - y(x_i)]^2}{2\sigma_i^2} \right\}$$

# Zastosowanie do dopasowania metodą najmniejszych kwadratów

- Podstawienie  $L$  do wzoru ogólnego  $AIC = -2 \ln L + 2K$  daje **wzór roboczy metody AIC**:

$$AIC = -n \ln M + 2K$$

gdzie  $M$  jest zwykłą funkcją kryterialną metody najmniejszych kwadratów

$$M(a_1, \dots, a_m) = \sum_{i=1}^n [y_i - y(x_i)]^2$$

- **Poprawka na małą liczebność próby**

$$\text{AIC}_c = \text{AIC} + \frac{2K(K+1)}{n-K-1}$$

- **Wagi Akaike'go** (prawdopodobieństwo, że model jest dobry w zbiorze testowanych modeli),

$$w_j = \frac{\exp[-\Delta_j / 2]}{\sum_{j=1}^V \exp[-\Delta_j / 2]}$$

gdzie różnice AIC:  $\Delta_j = \text{AIC} - \text{AIC}^{(min)}$ .



# Hirotoogu Akaike 1927 - 2009



# Hirotsugu Akaike i jego teoria

- Ur. w Fujinomiya, **syn hodowcy jedwabników**.
- Studia, doktorat, praca w Japonii & staże zagr.
- Teoria AIC: najpierw na konferencji, główny artykuł **1974, w *IEEE Transactions of Automatic Control***.
- Przez ćwierćwiecze rozwijana głównie w Japonii.
- **Monografia Burhama i Andersona (1998, 2002)** początkiem akceptacji AIC w „reszcie świata”.
- Obecnie, **artykuł z 1974 jest w pierwszej setce najczęściej cytowanych prac nauk. wszechczasów**.
- Podobne podejścia:
  - Bayesian Information Criterium (BIC)
  - Kryterium informacyjne Hanna-Quinna .... Etc.

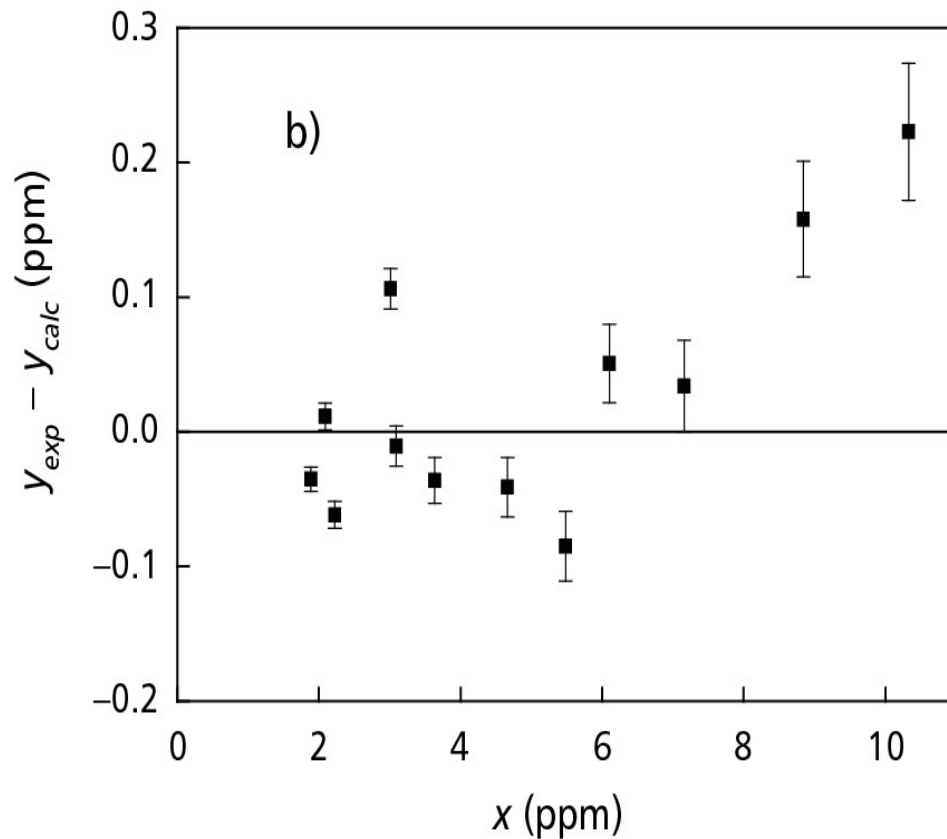
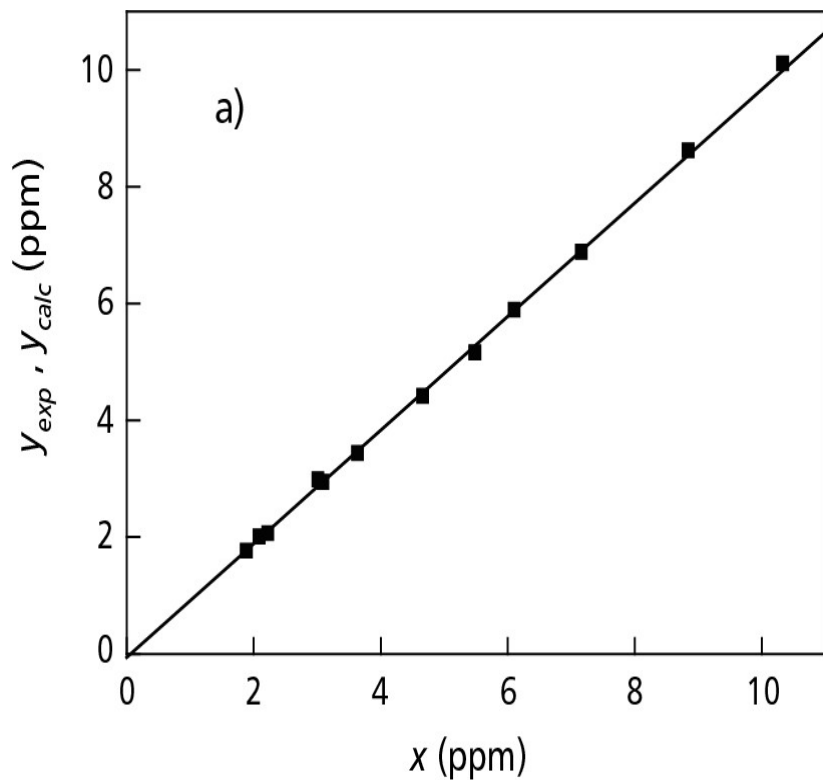
Przykłady

# Przykład 1: cechowanie miernika CH<sub>4</sub>

**Table 12.2.** Input data. All values in ppm of methane. The data from Nęcki (2019).

$x$	$y$	$u(y)$	$x$	$y$	$u(y)$
1.884	1.768	0.009	4.662	4.421	0.022
2.087	2.009	0.010	5.485	5.165	0.026
2.223	2.066	0.010	6.105	5.894	0.029
3.013	2.990	0.015	7.16	6,887	0.034
3.09	2.947	0.015	8.841	8.62	0.043
3.635	3.443	0.017	10.33	10.11	0.051

# Czy prosta kalibracji ma przechodzić przez początek układu?



Trzy dopasowania;  $n = 12$

<b>Model</b>	<b>M</b>	<b>AIC</b>	<b>AIC<sub>c</sub></b>	<b><math>w_i</math></b>
<b><math>y = ax</math></b>	<b>160,6</b>	<b>64,95</b>	<b>66,28</b>	<b>0,59</b>
<b><math>y = ax + b</math></b>	<b>128,4</b>	<b>64,26</b>	<b>67,26</b>	<b>0,36</b>
<b><math>y = ax^2 + bx + c</math></b>	<b>120,8</b>	<b>65,53</b>	<b>71,24</b>	<b>0,05</b>

Model finalny:  $y = a x$

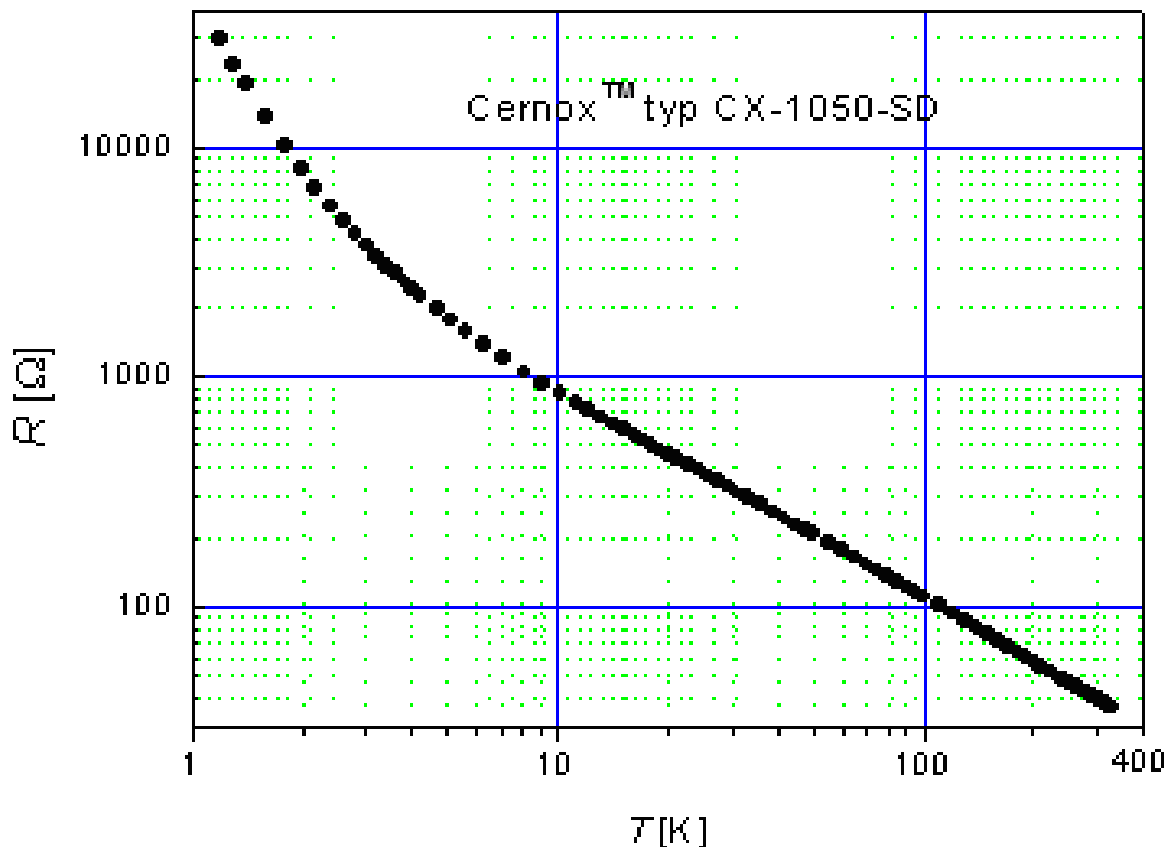
$$a = 0,9571$$

$$u(a) = 0,0053$$

Ponieważ  $1 - a = 0,0429 \gg u(a)$ , poprawka jest istotna, tj. surowy wynik pomiaru należy mnożyć przez czynnik poprawkowy  $a = 0,9571$

# Przykład 2: funkcja kalibracji sensora temperatury Cernox™

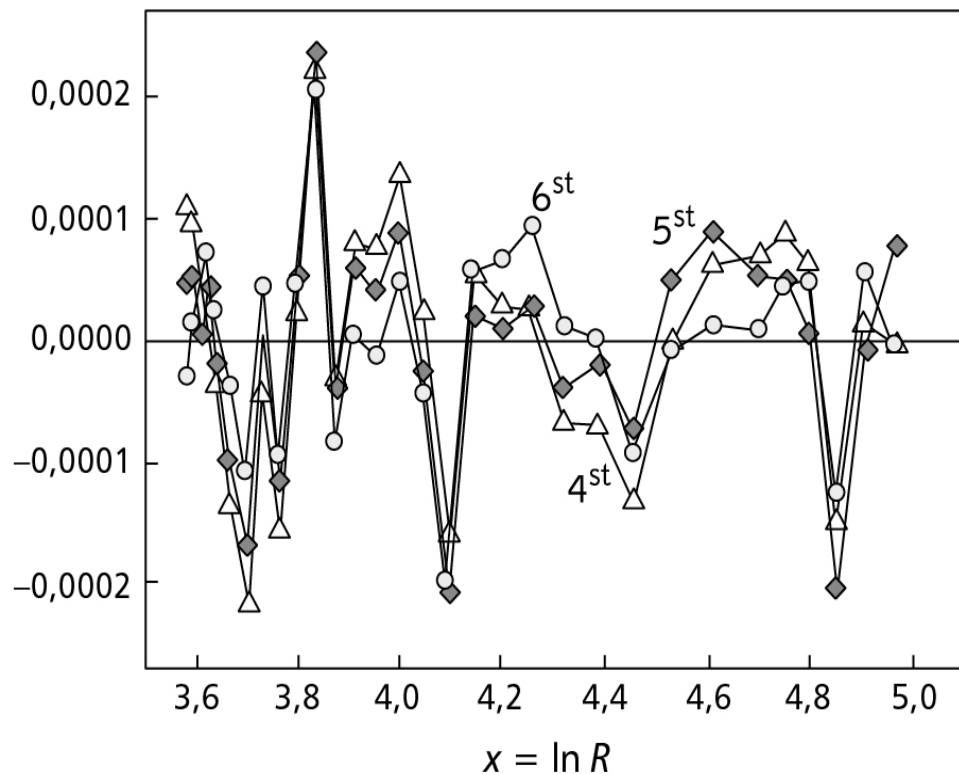
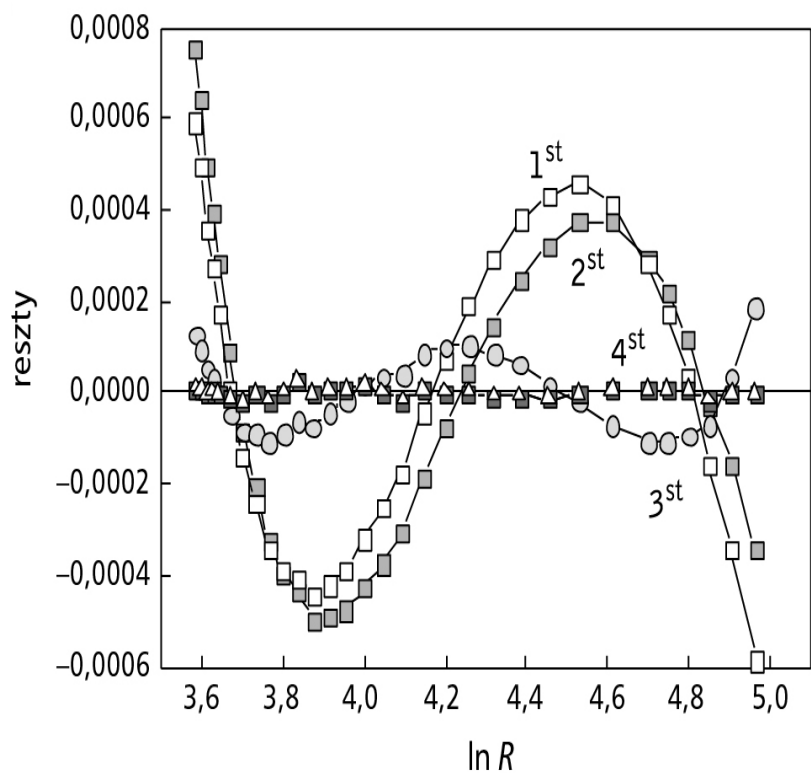
Lake Shore  
Cryogenics,  
używany w  
IFJ Kraków



dopasowanie wielomianami  
różnych stopni do zależności:  $\ln T = f(\ln R)$ ;  
w zakresie temp. 74 - 330 K



# Wykresy reszt dopasowania wielomianami różnego stopnia, liczba punktów $n = 31$ .



## Tabela: porównanie testu $F$ i metody AIC

$k$	$M$	$F$	$p$ -value	AIC	AIC <sub>c</sub>	$w_i$
4	301	5,12	0,033	188,9	192,4	0,008
5	249,8	9,84	0,004	185,1	190,0	0,025
6	177,2	0,01	<b>0,920</b>	<b>176,5</b>	<b>183,0</b>	<b>0,831</b>
7	177,1	0,96	0,338	178,5	187,1	0,111
8	169,7			179,1	190,2	0,024

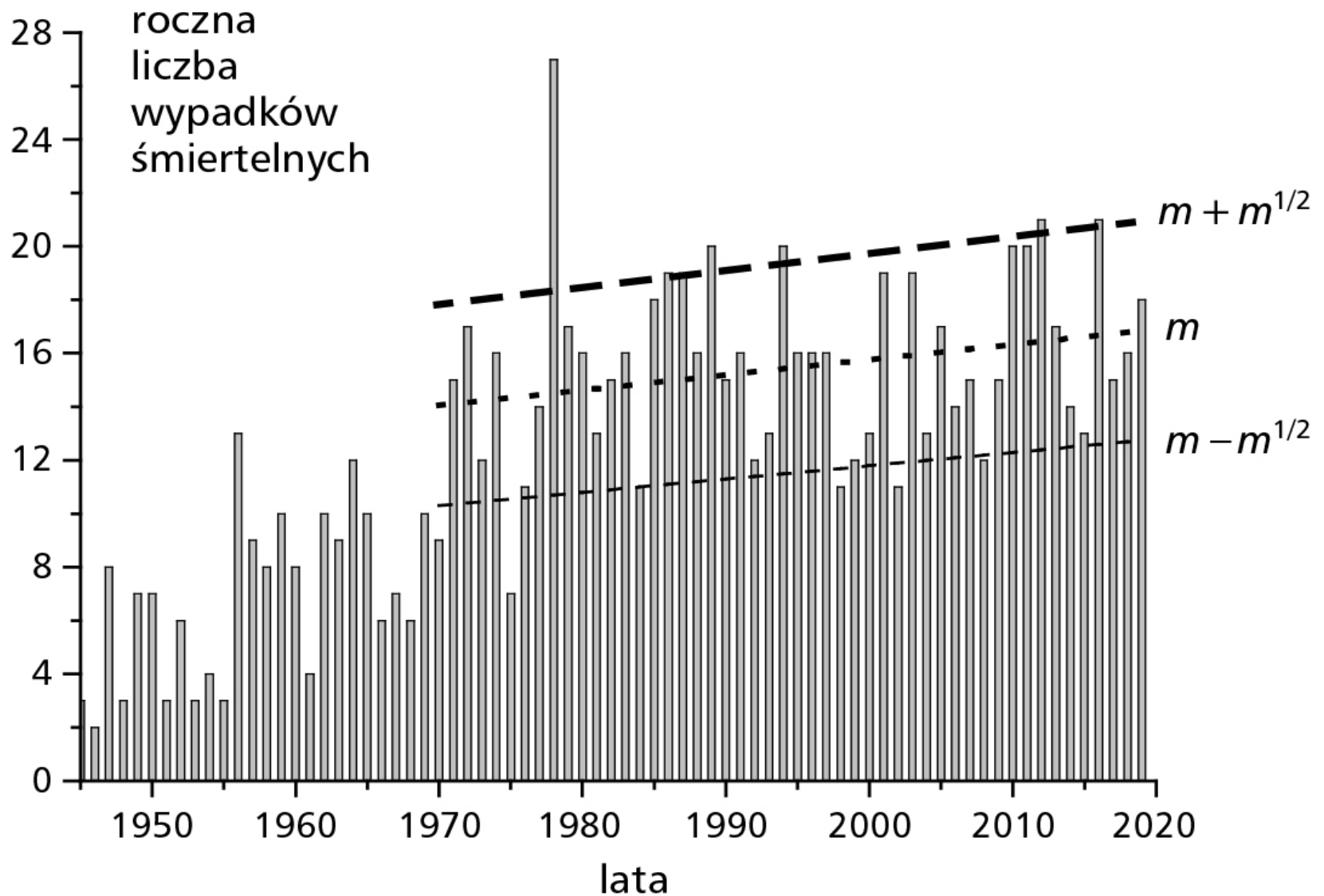
# Przykład 3: statystyka wypadków śmiertelnych na terenie TPN

Dane:

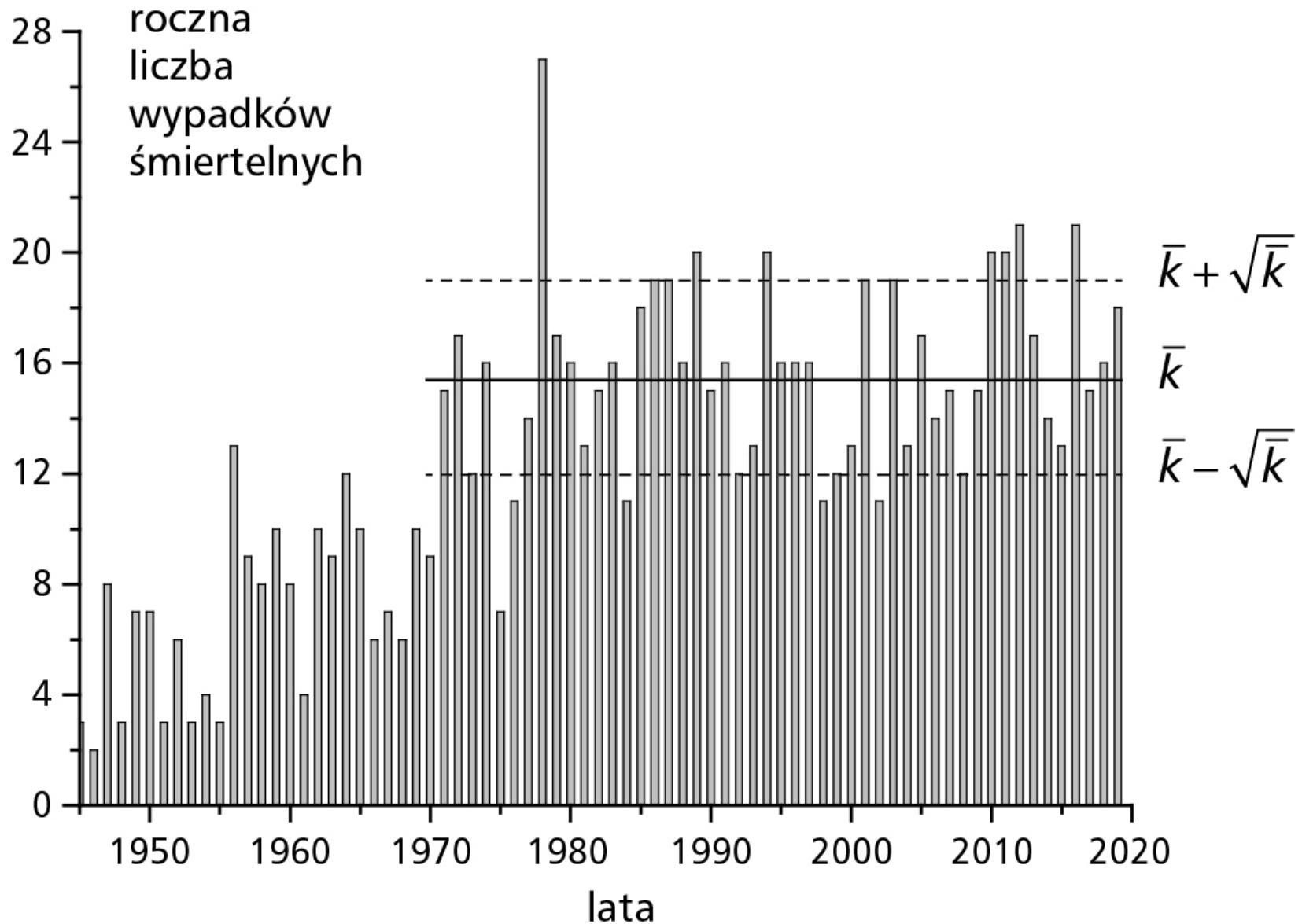
- Marasek A. & Marasek A. (2011). Żywym ku przestrodze. Wypadki śmiertelne na terenie działania TOPR w latach 1909–2009. *Tatry TPN*, nr. 1/2011, 32–39.
- Info Adama Maraska nt. lat 2010 – 2019.

Pytanie: czy liczba wypadku ma w ostatnim półwieczu tendencję wzrostową?

# Model 1: $y = ax + b$



# Model 2: $y = \text{const.}$



## Tabela - metoda AIC

<b>Model</b>	<b><math>K</math></b>	<b><math>M</math></b>	<b>AIC</b>	<b><math>AIC_c</math></b>	<b><math>w_i</math></b>
$y = const$	2	630,3	<b>326,31</b>	<b>326,57</b>	<b>0,47</b>
$y = ax + b$	3	606,4	326,38	326,90	0,40
$y = ax^2 + bx + c$	4	605,2	328,28	329,17	0,13

## Wnioski końcowe:

- Te same dane wejściowe, co w przypadku testu F:  
**M** (minimum sumy kwadratów), **m** (liczba parametrów),  
oraz **n** (liczba punktów eksp).
- Test F porównuje 2 modele, model definiujący  $H_0$  jest wyróżniony. W metodzie Akaike'go modele są równoważne, ich liczba dowolna.
- Tylko dwa pojęcia: **AIC** oraz **wagi Akaike'go**
- Wzory dla tych wielkości elementarne. Nie trzeba tabeli (czy „kalkulatora”) zmiennej  $F$ , ani wyboru poziomu istotności.
- Metoda AIC nie może zastąpić tych testów stat., gdzie badamy pojedynczy model (np. jakości dopasowania).

**Dziękuję za uwagę**

**[andrzej.zieba@fis.agh.edu.pl](mailto:andrzej.zieba@fis.agh.edu.pl)**