

CZY STATYSTYKA POTRZEBUJE NARRACJI?

Mirostław Szreder

Statystyka w służbie polityki społeczno-gospodarczej: nowe badania, metody i narzędzia
23-24 marca 2023, Warszawa



„Liczby potrzebują narracji”

(s. 277)

„Dane w zasadzie nigdy nie funkcjonują bez opowieści. Żeby nadać sens temu, co widzimy, automatycznie wpisujemy liczby w proste schematy narracyjne”

(s. 276)

Przykład:

oceny ucznia A: 6, 6, 5, 4, 3, 2, 1, 1;

oceny ucznia B: 1,1, 2, 3, 4, 5, 6, 6.

Wszystkie charakterystyki liczbowe (miary statystyczne) tego zbioru liczb są identyczne.

Uważam, że można wyróżnić 3 zasadnicze podejścia do kwestii natury i roli danych statystycznych w poznawaniu rzeczywistości:

- 1) przekonanie o tym, że same dane statystyczne są najlepszym, niezmaconym interpretacjami i kontekstem źródłem wiedzy
→ *Let data speak for themselves;*
- 2) traktowanie całego procesu gromadzenia danych, krytycznej analizy i ścisłej interpretacji zgodnej z teorią statystyki, jako nierozdzielnej i spójnej całości;
- 3) uznanie interpretacji danych za konieczny, lecz niewystarczający etap w opisie oraz rozumieniu rzeczywistości, a w efekcie postulowanie wzbogacenia interpretacji danych liczbowych narracją
nie tylko *fact-checking*, lecz także *fact-telling*

Ad. 1) *Let data speak for themselves*

Zwolennicy tego podejścia:

- nieakceptujący podejścia Bayesowskiego w statystyce, w szczególności wiedzy a priori oraz subiektywnych rozkładów prawdopodobieństwa a priori,
- część badaczy, entuzjastów *big data*, którzy nieufnie odnoszą się do tradycyjnej metodyki opisu i wnioskowania statystycznego.



Tytuł jednego z podrozdziałów: **Niech przemówią dane**

„Nie zawsze musimy znać przyczyny jakiegoś zjawiska, możemy po prostu pozwolić danym mówić za siebie”

(s. 30).

Potrzeba rozumienia? Pozorne współzależności i korelacje (*spurious correlations*).

To nie dane mówią za siebie, lecz my mówimy za dane: dobieramy odpowiednie dane, wypełniamy je treścią, nadajemy im interpretację.

Ad. 2) Traktowanie całego procesu gromadzenia danych, krytycznej analizy i ścisłej interpretacji zgodnej z teorią statystyki jako nierozdzielnej, spójnej całości

Z jednej strony poprawna interpretacja, a z drugiej dbałość o właściwe zrozumienie przez odbiorców (pobudzenie u odbiorcy pragnienia dobrego zrozumienia danych i wyników analiz statystycznych).

Przykładami są miary przeciętne w charakterystyce wynagrodzeń w gospodarce.

Mediana *versus* średnia arytmetyczna

Dominanta i jej wartość poznawcza w rozkładzie ciągłym. Wątpliwy sens szacowania punktowej jej wartości.

Przykład innej cechy o rozkładzie ciągłym (emerytury):

Z raportu

„Struktura wysokości świadczeń wypłacanych przez ZUS po waloryzacji w marcu 2022 roku”

wynika, że największy (dominujący) odsetek emerytur – **16,6%** – mieścił się w przedziale od 2 000,01 zł do 2 400,00 zł.

Oprócz interpretacji warto zadbać o możliwie pełne zrozumienie przez odbiorców zalet i wad danej miary statystycznej.



Rozdz. XI

**Odmiany biedy,
Czyli Gini nie mówi wszystkiego**

DLACZEGO

JEDNI SĄ BOGACI
A INNI NIE



N2
MIŁOCZYWIŚCIE

1.3 Dochód narodowy nie jest miarą dochodów osobistych

„... warto wspomnieć o wskaźniku, który odnosi się do dochodów, ale w rzeczywistości posługiwanie się nim nie jest ani poprawne, ani wskazane i jest swoistym uproszczeniem. Chodzi mi o wskaźnik dochodu narodowego, a ściślej produktu krajowego brutto przypadającego na jednego mieszkańca”.

(s. 35)

Fakty liczbowe a percepcja

„Mind The Gap Between Perception and Reality”

Rzeczywista a postrzegana wysokość inflacji.


Wykres 1. Percepcja inflacji w Niemczech oraz krajach strefy euro po wprowadzeniu wspólnej waluty



Ad. 3) uznanie interpretacji danych za konieczny, lecz niewystarczający etap w opisie oraz rozumieniu rzeczywistości, a w efekcie postulowanie wzbogacenia interpretacji danych liczbowych narracją.

The Median Isn't the Message


by Stephen Jay Gould



AMA Journal of Ethics[®]
Illuminating the Art of Medicine

Home Issues Articles Cases Art Multimedia CME

Search



PERSONAL NARRATIVE
JAN 2013

The Median Isn't the Message

Stephen Jay Gould, PhD

[Citation](#) | [PDF](#) | [Altmetric](#)

My life has recently intersected, in a most personal way, two of Mark Twain's famous quips. One I shall defer to the end of this essay. The other (sometimes attributed to Disraeli) identifies three species of mendacity, each worse

Call for Papers

Ad. 3) uznanie interpretacji danych za konieczny, lecz niewystarczający etap w opisie oraz rozumieniu rzeczywistości, a w efekcie postulowanie wzbogacenia interpretacji danych liczbowych narracją.

My life has recently intersected, in a most personal way, two of Mark Twain's famous quips.

(....)

This is a personal story of statistics, properly interpreted, as profoundly nurturant and life-giving.

(...)



Stephen Jay Gould (ur. 10 września 1941 w Nowym Jorku, zm. 20 maja 2002) – amerykański paleontolog, popularyzator ewolucjonizmu.

W eseju tym Autor pisze o medianie w rozkładzie ciągłym – statystycznym rozkładzie czasu życia, jaki pozostał mu po wykryciu w jego organizmie złośliwego nowotworu w 1982 roku.

Wikipedia:

In July 1982 Gould was diagnosed with peritoneal mesothelioma, a deadly form of cancer

Z eseju:

„The literature couldn't have been more brutally clear: mesothelioma is incurable, with a median mortality of only eight months after discovery”.

„When I learned about the eight-month median, my first intellectual reaction was: fine, half the people will live longer; now what are my chances of being in that half”.

Cechą mediany jest to, że nie informuje w ogóle o zróżnicowaniu wartości (długości życia) tych jednostek, które znalazły się po jej prawej stronie, czyli chorych, którym dane było przeżyć więcej niż 8 miesięcy.

*But all evolutionary biologists know that **variation itself is nature's only irreducible essence**. Variation is the hard reality, not a set of imperfect measures for a central tendency. Means and medians are the abstractions.*

S.J. Gould

Gdyby było odwrotnie:

If the median is the reality and variation around the median just a device for its calculation, the "I will probably be dead in eight months" may pass as a reasonable interpretation.

Autor przeżył jeszcze prawie 20 lat, a dokładnie 230 miesięcy ponad wielkość mediany wynoszącą 8 miesięcy.

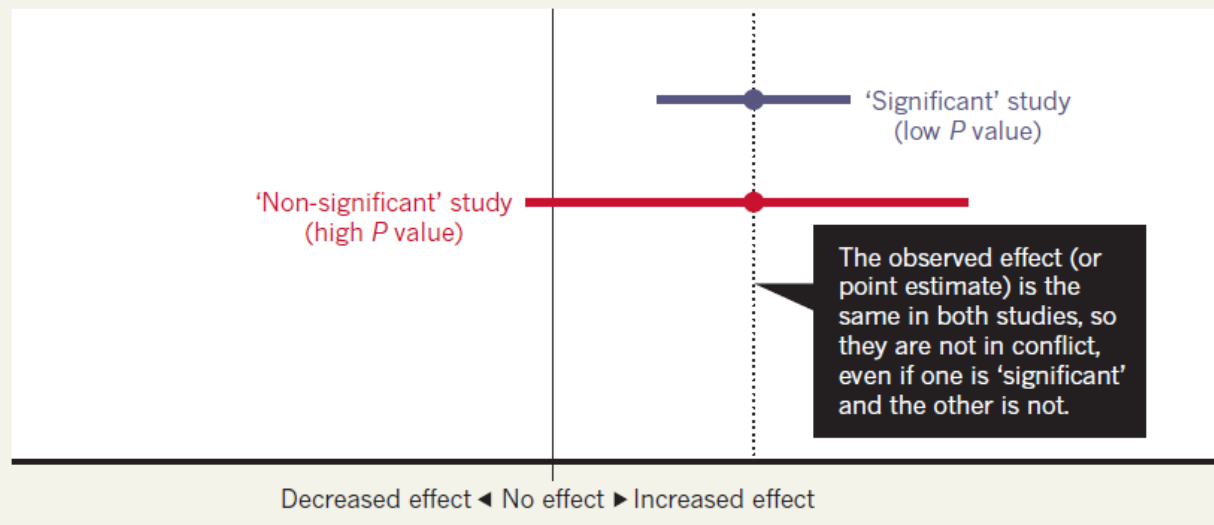
Variation itself is nature's only irreducible essence

Statystyczna istotność i wskaźnik p -value – współczesna krytyka

Wynik statystycznie istotny nie zawsze oznacza ważny
significant \neq important”, “influential”

BEWARE FALSE CONCLUSIONS

Studies currently dubbed ‘statistically significant’ and ‘statistically non-significant’ need not be contradictory, and such designations might cause genuine effects to be dismissed.

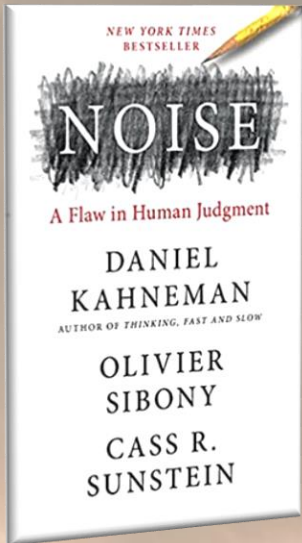


Wskaźnik *p-value* powinien być traktowany jako **jedna z przesłanek** dowodzenia nieprawdziwości hipotezy zerowej.

Tym bardziej, że *p-value* odnosi się bezpośrednio nie tylko do hipotezy zerowej, jak przyjęto się uważać.

Odnosi się także do całego modelu wnioskowania i jego założeń, a więc także do wszystkich tych okoliczności i zakłóceń (błędów nielosowych), które miały wpływ na przyjęcie takiej, a nie innej wartości statystyki testowej.

Test statystyczny nie jest narzędziem, które potrafi przekształcić niepewność towarzyszącą wnioskowaniu w pewność decyzji dotyczącej prawdziwości lub nieprawdziwości hipotezy.

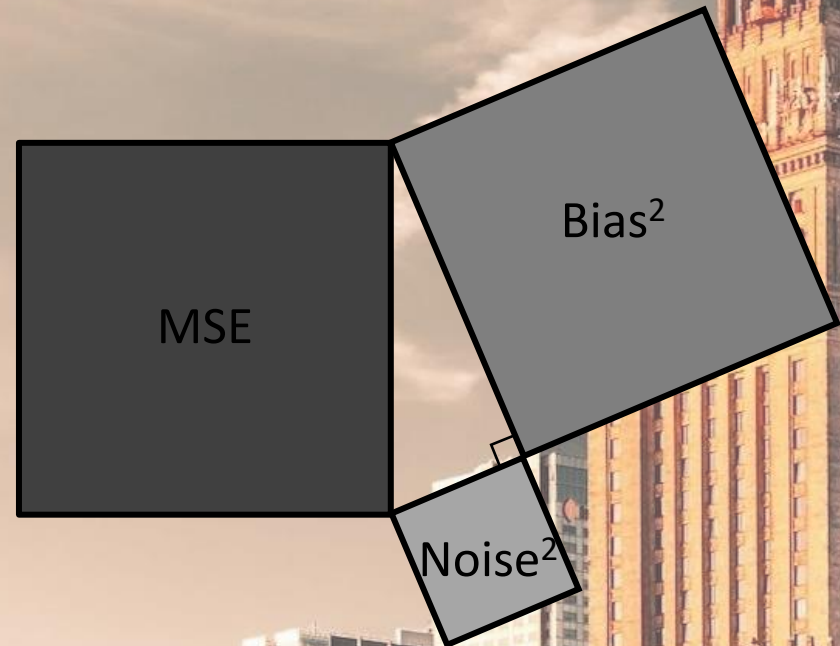
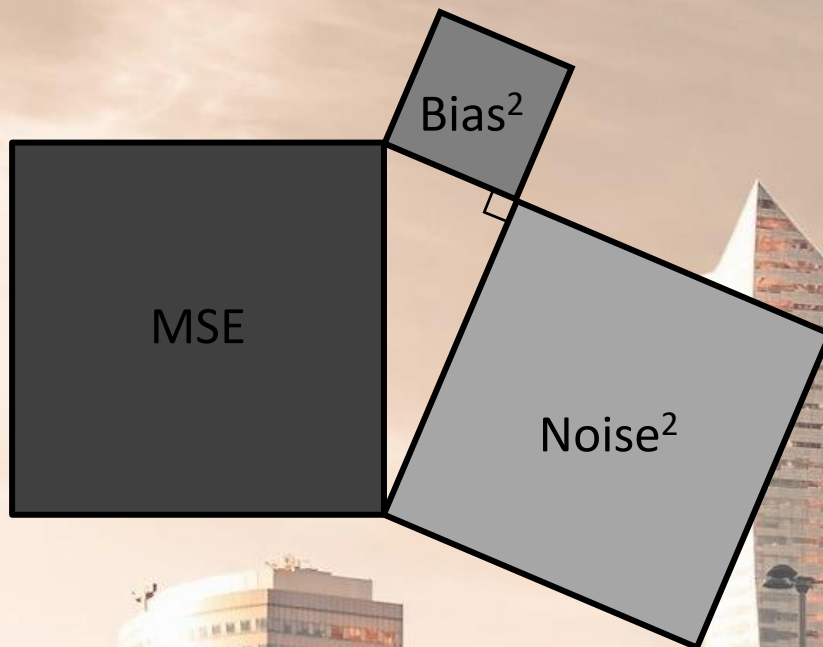


NOISE



„... tendencyjność i szum odgrywają identyczną rolę w równaniu błędu. Są od siebie niezależne i mają jednakową wagę w określeniu ogólnego błędu”

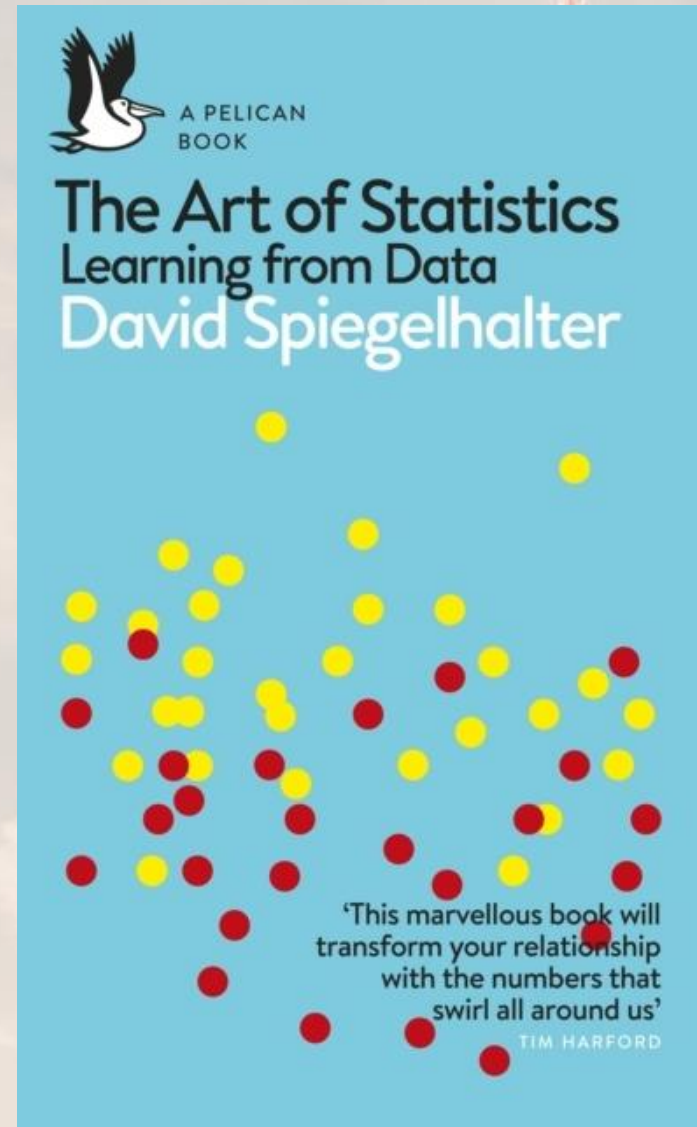
s. 79



Podsumowanie

Właściwe rozumienie mierników statystycznych oraz wyników wnioskowania (testowania hipotez) wymaga nie tylko ścisłej, poprawnej interpretacji, ale także scharakteryzowania pełniejszego kontekstu rozważanego zagadnienia.

„We have to acknowledge we are telling a story, and it is inevitable that people will make comparisons and judgements, no matter how much we only want to inform and not to persuade” (s. 69).



DZIĘKUJĘ ZA UWAGĘ

Statystyka w służbie polityki społeczno-gospodarczej: nowe badania, metody i narzędzia
23-24 marca 2023, Warszawa