

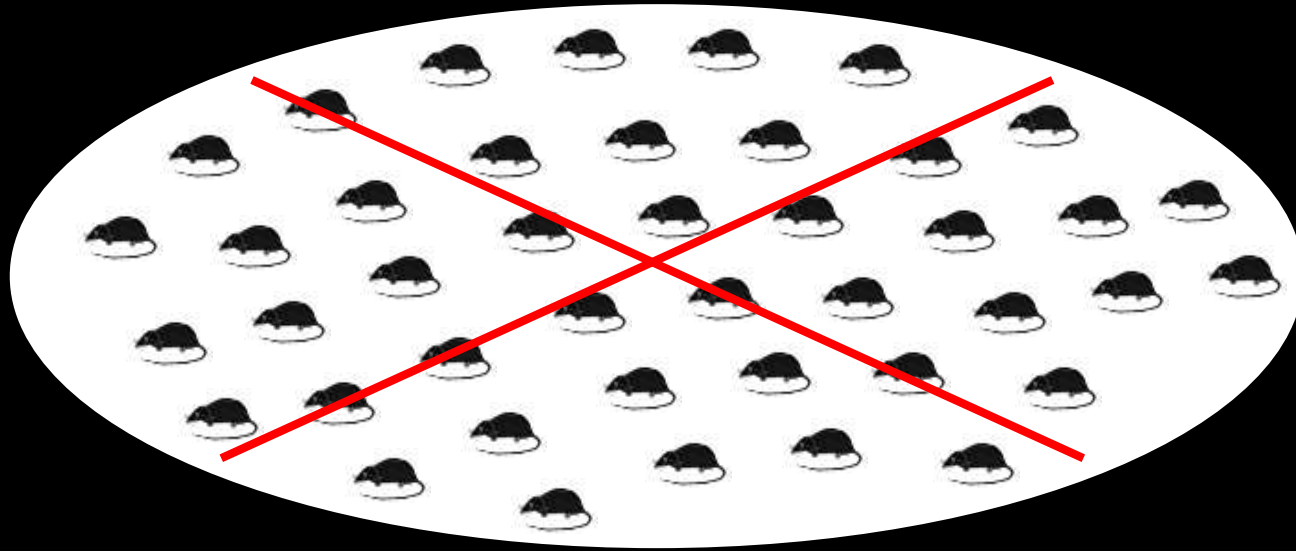
Metody Statystyczne w Biologii

1. Wykład wstępny
2. **Populacje i próby danych**
3. Testowanie hipotez i estymacja parametrów
4. Planowanie eksperymentów biologicznych
5. Najczęściej wykorzystywane testy statystyczne I
6. Najczęściej wykorzystywane testy statystyczne II
7. Regresja liniowa
8. Regresja nieliniowa
9. Określenie jakości dopasowania równania regresji liniowej i nieliniowej
10. Korelacja
11. Elementy statystycznego modelowania danych
12. Porównywanie modeli
13. Analiza wariacji
14. Analiza kowariancji
15. Podsumowanie materiału, wspólna analiza przykładów, dyskusja

Podstawowe pojęcia statystyczne

1. Populacja i próba danych
2. Parametr populacji i estymator
3. Ważne aspekty związane z tworzeniem prób danych i estymacją parametrów
4. Błąd próbkowania

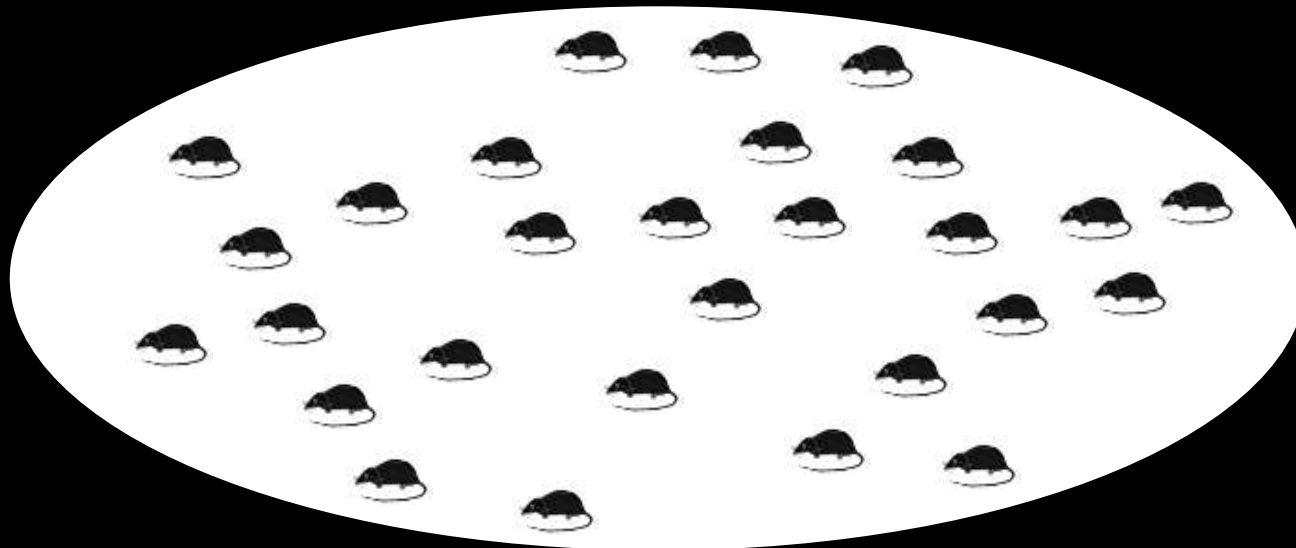
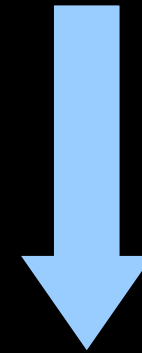
Populacja i próba danych



POPULACJA

w rozumieniu statystycznym

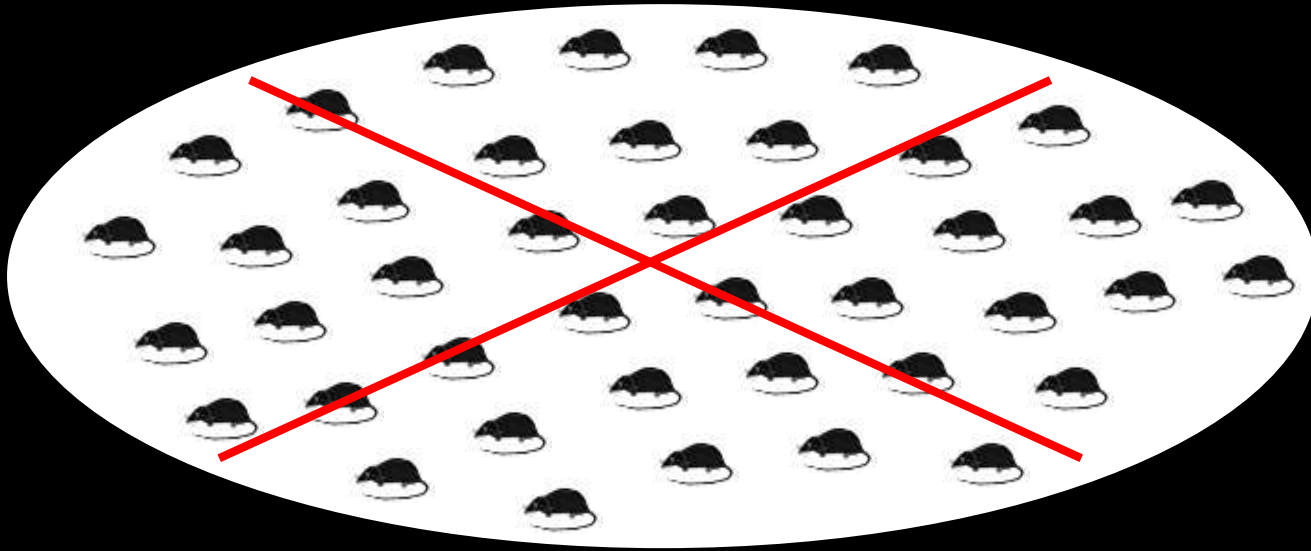
Obserwacje dla wszystkich osobników danego gatunku / rasy



PRÓBA DANYCH

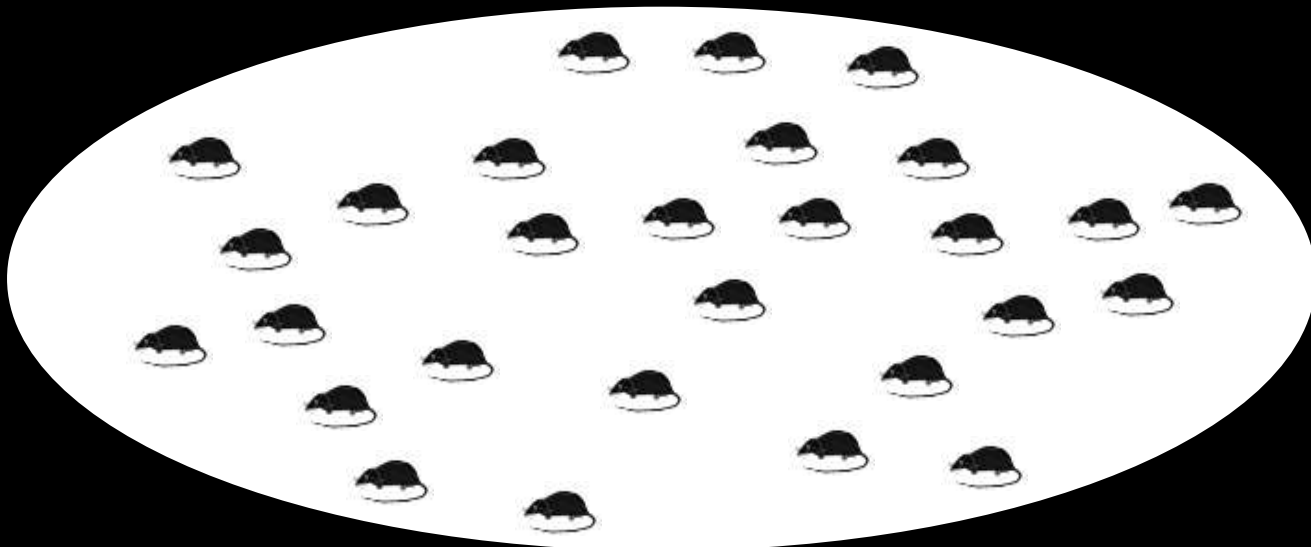
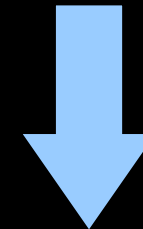
Obserwacje dotyczące wycinka populacji

Populacja i próba danych



populacja

Należy ją zdefiniować przed rozpoczęciem eksperymentu ...



próba danych

... ponieważ wielkości obliczone na podstawie uzyskanej próbie danych odnoszą się do populacji, z której została pozyskana



[Ind Psychiatry J](#). 2010 Jan-Jun; 19(1): 60–65.

doi: [10.4103/0972-6748.77642](https://doi.org/10.4103/0972-6748.77642)

PMCID: PMC3105563

PMID: [21694795](https://pubmed.ncbi.nlm.nih.gov/21694795/)

Statistics without tears: Populations and samples

[Amitav Banerjee](#) and [Suprakash Chaudhury](#)¹

▶ [Author information](#) ▶ [Copyright and License information](#) ▶ [Disclaimer](#)

This article has been [cited by](#) other articles in PMC.

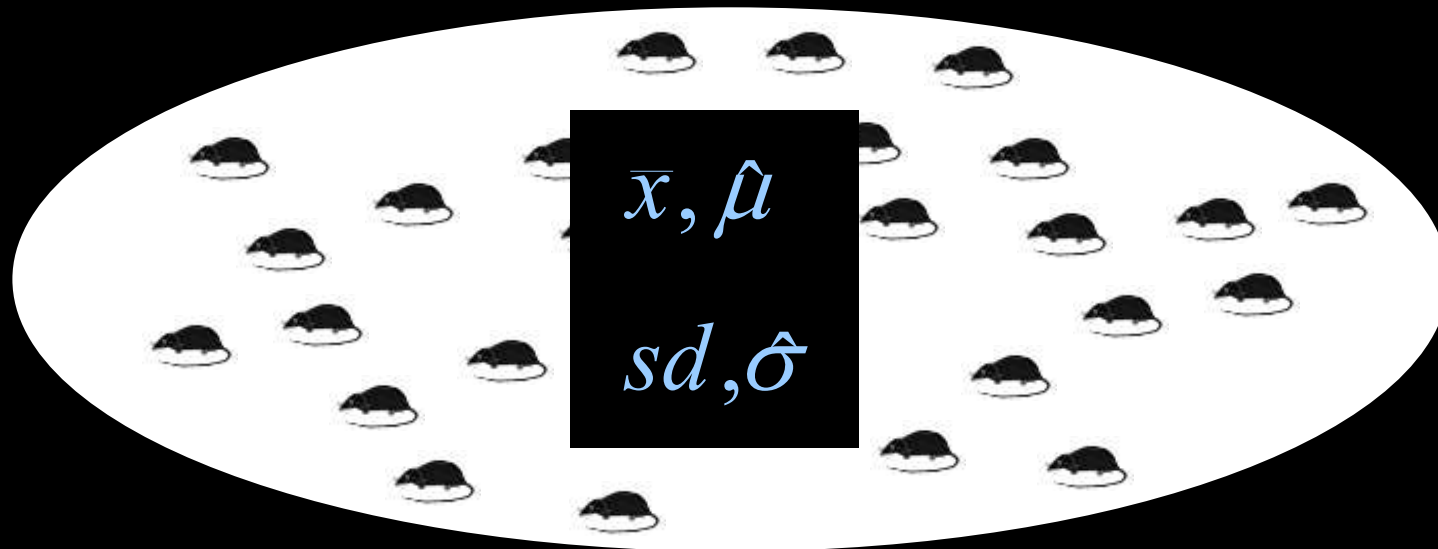
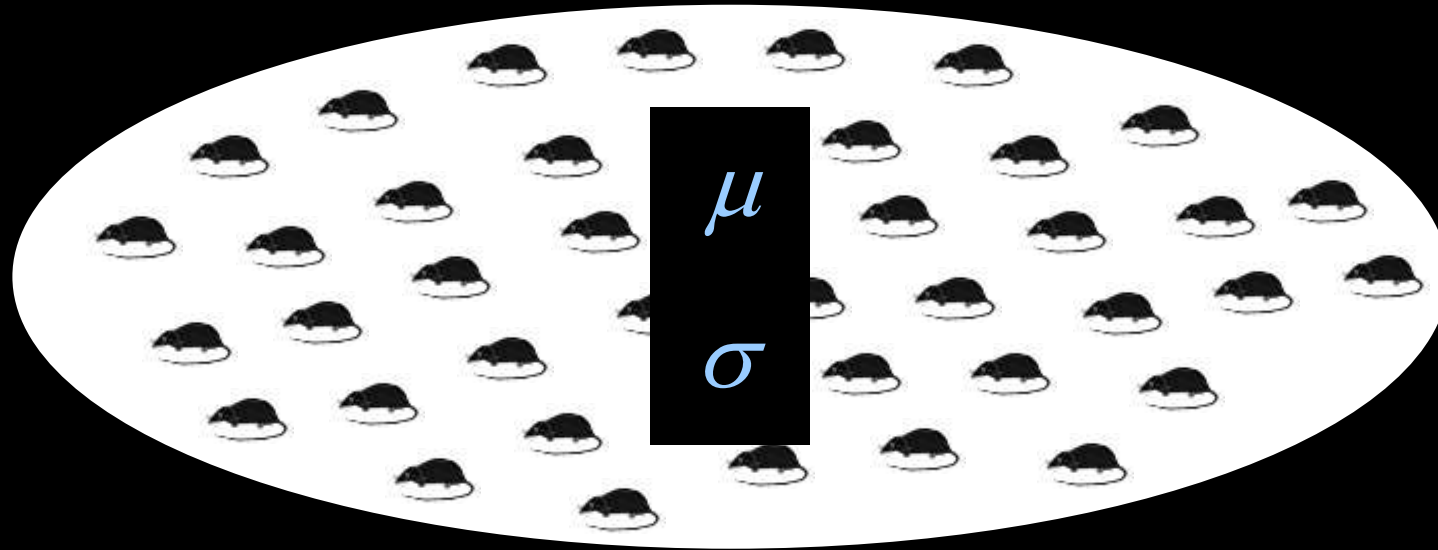
Abstract

Go to:

Research studies are usually carried out on sample of subjects rather than whole populations. The most challenging aspect of fieldwork is drawing a random sample from the target population to which the results of the study would be generalized. In actual practice, the task is so difficult that some sampling bias occurs in almost all studies to a lesser or greater degree. In order to assess the degree of this bias, the informed reader of medical literature should have some understanding of the population from which the sample was drawn. The ultimate decision on whether the results of a particular study can be generalized to a larger population depends on this understanding. The subsequent deliberations dwell on sampling strategies for different types of research and also a brief description of different sampling methods.

Keywords: Methods, population, sample

Parametr populacji i estymator



Populacja → Parametr

- nieznan
- dotyczy całej populacji

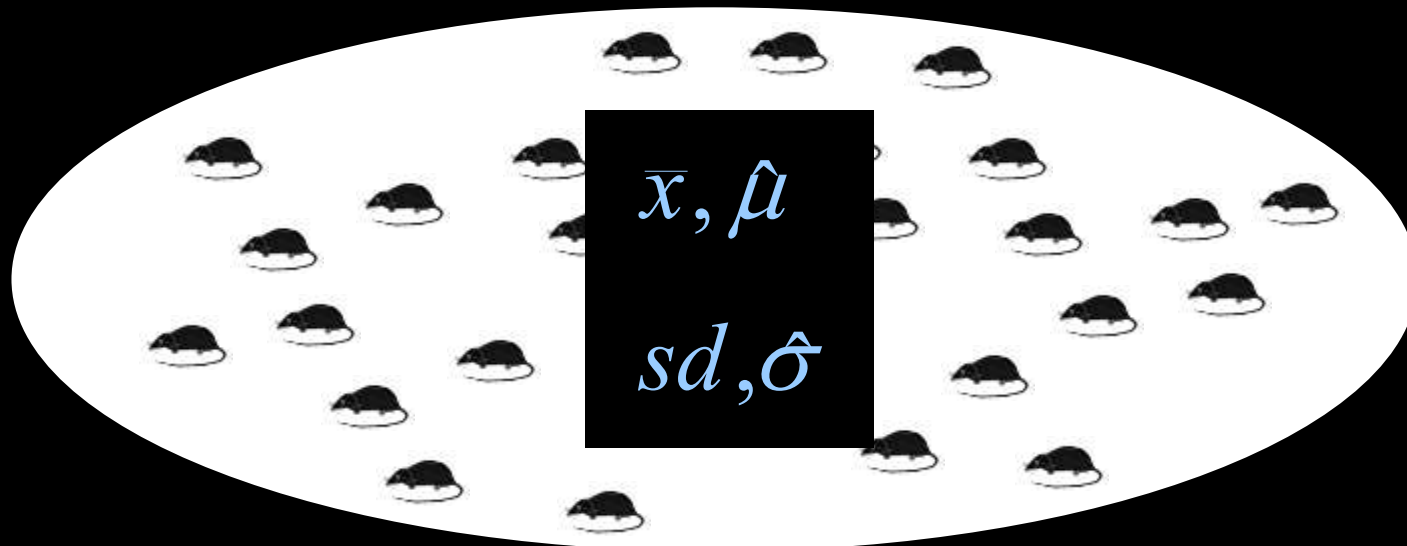
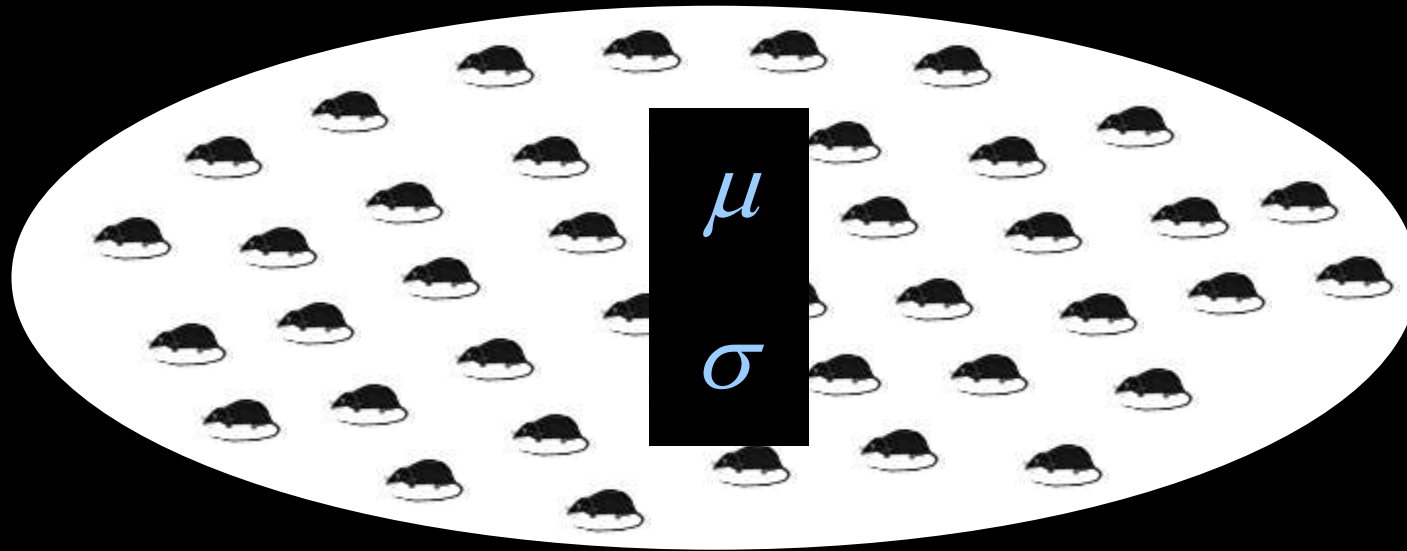
np. **rzeczywista** średnia masa ciała i jej odchylenie standardowe w **populacji** ryjówek

próba danych → Estymator

- znany
- wartość obliczona dla danej próby

np. **obliczona** średnia masa ciała i jej odchylenie standardowe w **próbie danych**

Parametr populacji i estymator



estymator

... powinien być


- nieobciążony
 $E(\text{estymator}) = \text{parametr}$

- zgodny
ze wzrostem N estymator \rightarrow parametr




- efektywny
najmniejsza wariancja


najlepszy estymator

Click the lever to pick a random ball. Then click Reset balls.
In 12 'picks', how many times is each colour picked?



The machine is green with a purple lever on the left. The lever has a red ball on it. A sign on the machine says "Reset balls". The machine's body is open, revealing a transparent globe containing several balls: two red, one yellow, and two blue. A blue ball is currently being dispensed into a purple tray at the bottom of the machine.

			Total
0	1	2	3

Reset 

wybór próby danych z populacji

1. Liczebność próby danych

2. Replikacje

3. Wybór osobników z populacji do próby danych

Ważne aspekty związane z tworzeniem prób danych i estymacją parametrów

liczebność próby danych

populacja	próba 1	próba 2	próba 3	próba 4		średni wzrost	różnica
156						w. prawdziwa 173.04	
178						próba 1	#DZIEL/0!
179						próba 2	#DZIEL/0!
156						próba 3	#DZIEL/0!
155						próba 4	#DZIEL/0!
189							
178							
167							
154							
178							
179							
190							
178							
176							
189							
180							
184							
160							
170							
175							
176							

liczebność próby danych

1. Liczba obserwacji w próbie
2. Im większa próba danych tym dokładniej można oszacować wartości parametrów populacji
3. Wielkość próby danych zależy od: łatwości zdobywania obserwacji, kosztów, dostępności obserwacji
4. Czynniki wpływające na wymaganą wielkość próby danych – zmienność obserwacji = odchylenie standardowe, dokładność pomiarów, hipoteza badawcza

liczebność próby danych

Analiza mocy - technika statystyczna pozwalająca na określenie minimalnej wielkości próby danych dla danego eksperymentu

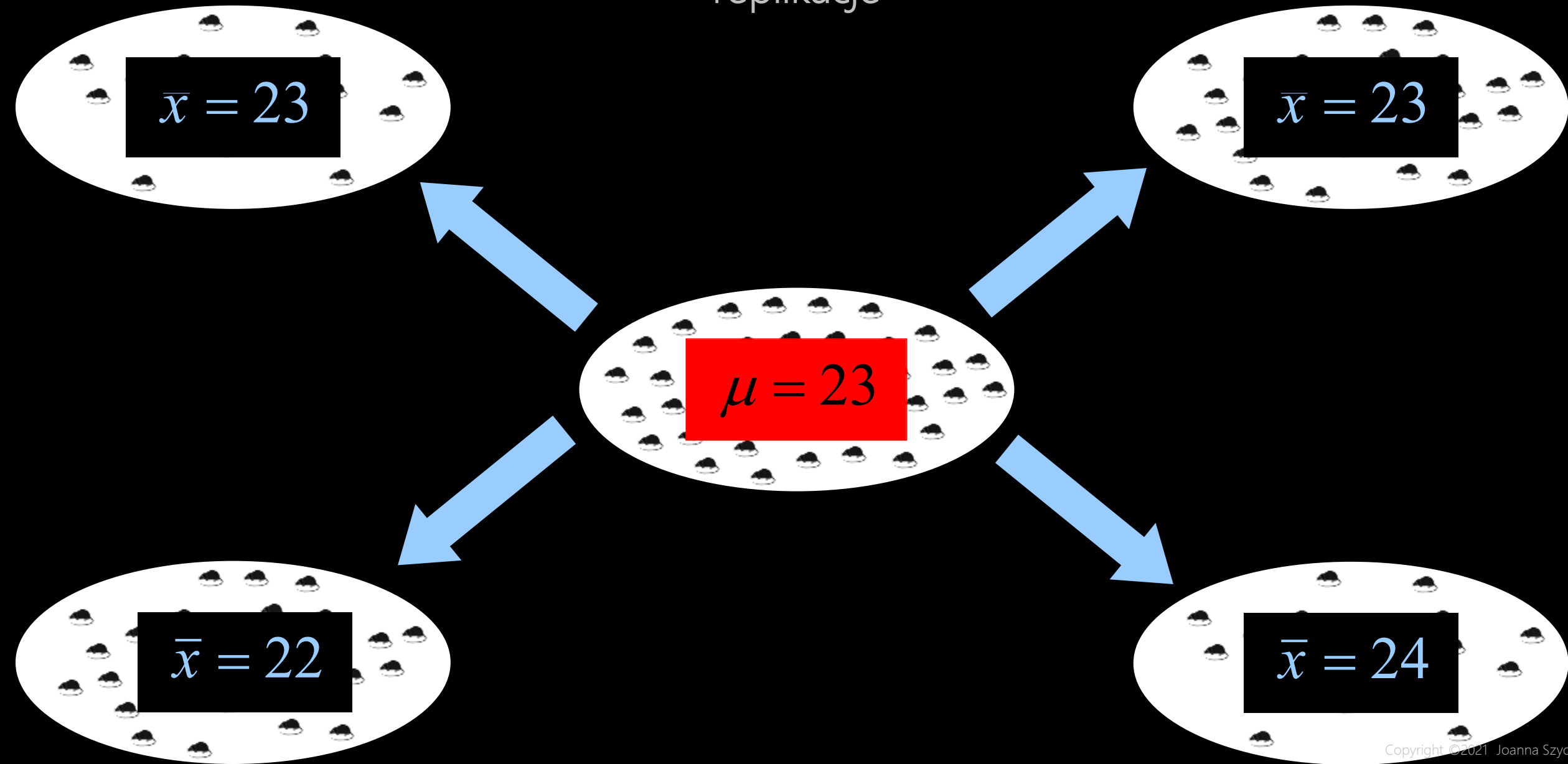
- np. moc 80% prawdopodobieństwo przyjęcia prawdziwej hipotezy alternatywnej np. wykrycie istniejącej różnicy = 0.8
- metody empiryczne: symulacje Monte Carlo
- metody numeryczne

wybór próby danych z populacji

1. Liczebność próby danych
 2. Replikacje
-
3. Wybór osobników z populacji do próby danych

Ważne aspekty związane z tworzeniem prób danych i estymacją parametrów

replikacje



replikacje

1. Ten sam rodzaj pomiaru / eksperymentu wykonany na różnych próbach danych
2. Stosowanie replikacji znacznie zwiększa wiarygodność wyciągniętych wniosków
3. Wybór osobników do każdej replikacji powinien być losowy
4. Idealnie → każda próba danych powinna zawierać różne osobniki
5. Jeżeli technicznie trudno pobrać osobniki do replikacji i niektóre powtarzają się w poszczególnych próbach danych → pseudoreplikacje

wybór próby danych z populacji

1. Liczebność próby danych
 2. Replikacje
 3. Wybór osobników z populacji do próby danych
-

Ważne aspekty związane z tworzeniem prób danych i estymacją parametrów

tworzenie próby danych

populacja	próba 1	próba 2		średnia		różnica		
150				w. prawdziwa	173.04			
150				próba 1	#DZIEL/0!	#DZIEL/0!		
154				próba 2	#DZIEL/0!	#DZIEL/0!		
154								
155								
155								
156								
156								
156								
156								
156								
159								
159								
159								
159								
160								
160								
160								
160								
160								
160								
160								

tworzenie próby danych

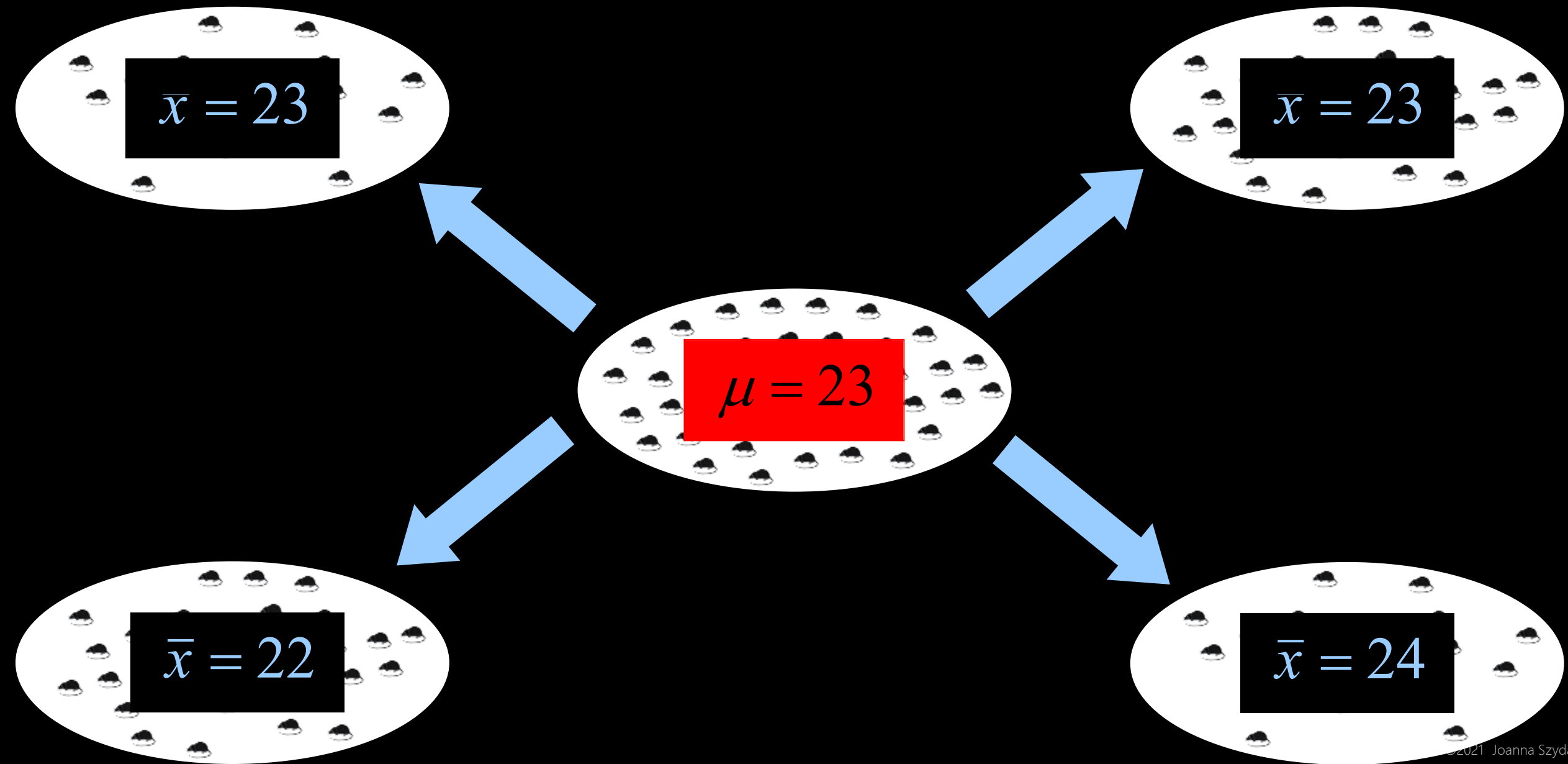


losowy

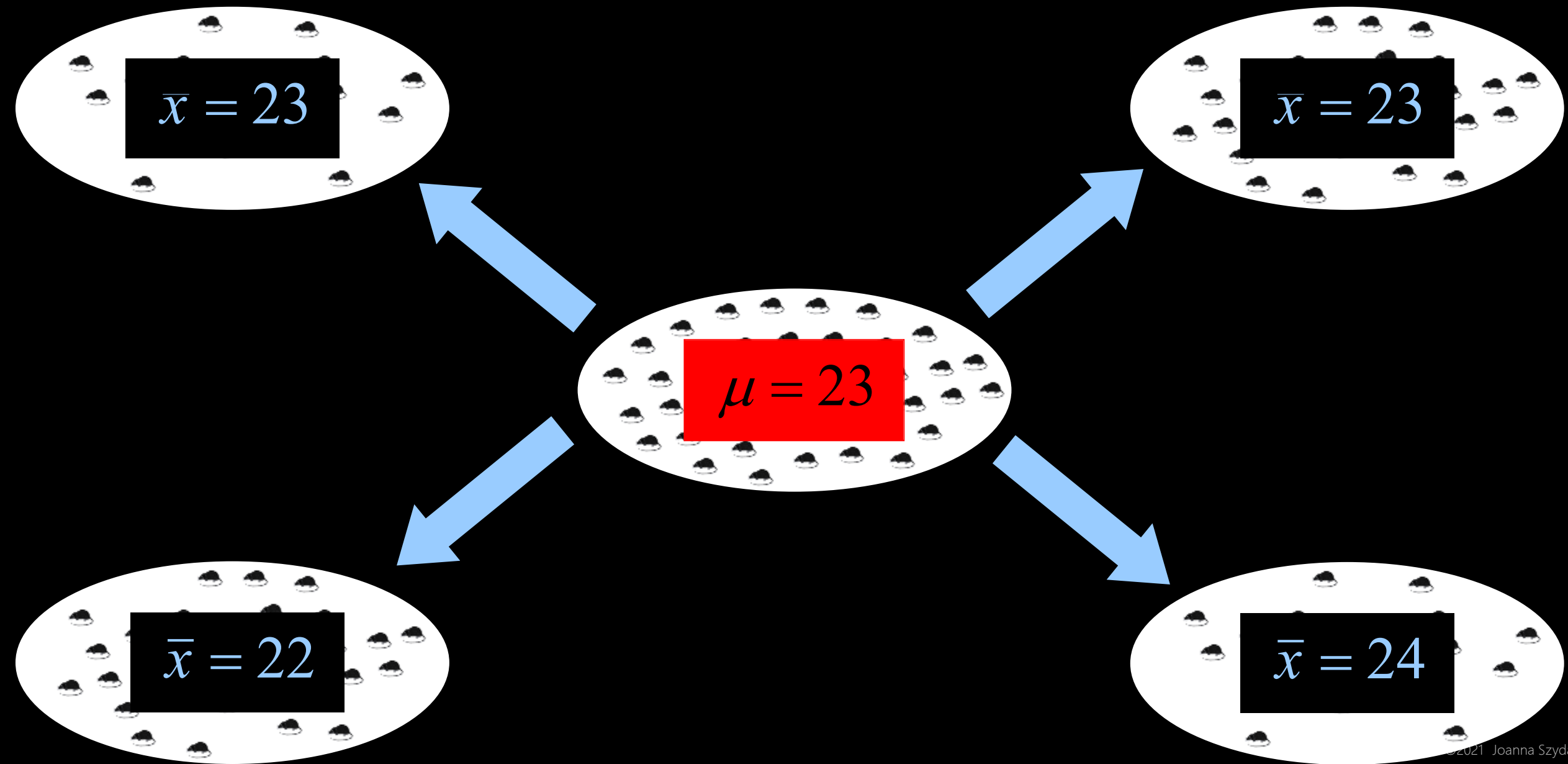
- Osobniki wybierane są bez stosowania selekcji
- Osobniki muszą reprezentować wszystkie warianty efektów działających na populację np. lokalizacja, wiek, płeć, ekspozycja na warunki środowiskowe, spokrewnienie

niełosowy

- Wybór wg spokrewnienia
- Wybór wg wartości cechy
- Eksperyment



1. Błąd próbkowania / błąd estymacji wynika z faktu, że obserwujemy próbę danych zamiast populacji
2. Nie możemy precyzyjnie określić wielkości błędu próbkowania
3. Możemy w przybliżeniu określić dokładność estymatora
 - Błąd standardowy estymatora
 - Przedziały ufności estymatora



Observational Study > [Surg Endosc. 2022 Jul;36\(7\):4893-4902. doi: 10.1007/s00464-021-08841-z.](#)

Epub 2021 Nov 1.

[REDACTED] in the diagnosis of colorectal cancer is associated with delay to surgery: a retrospective cohort study

Garrett G R J Johnson ^{1 2}, Olivia Hershorn ¹, Harminder Singl

Affiliations + expand

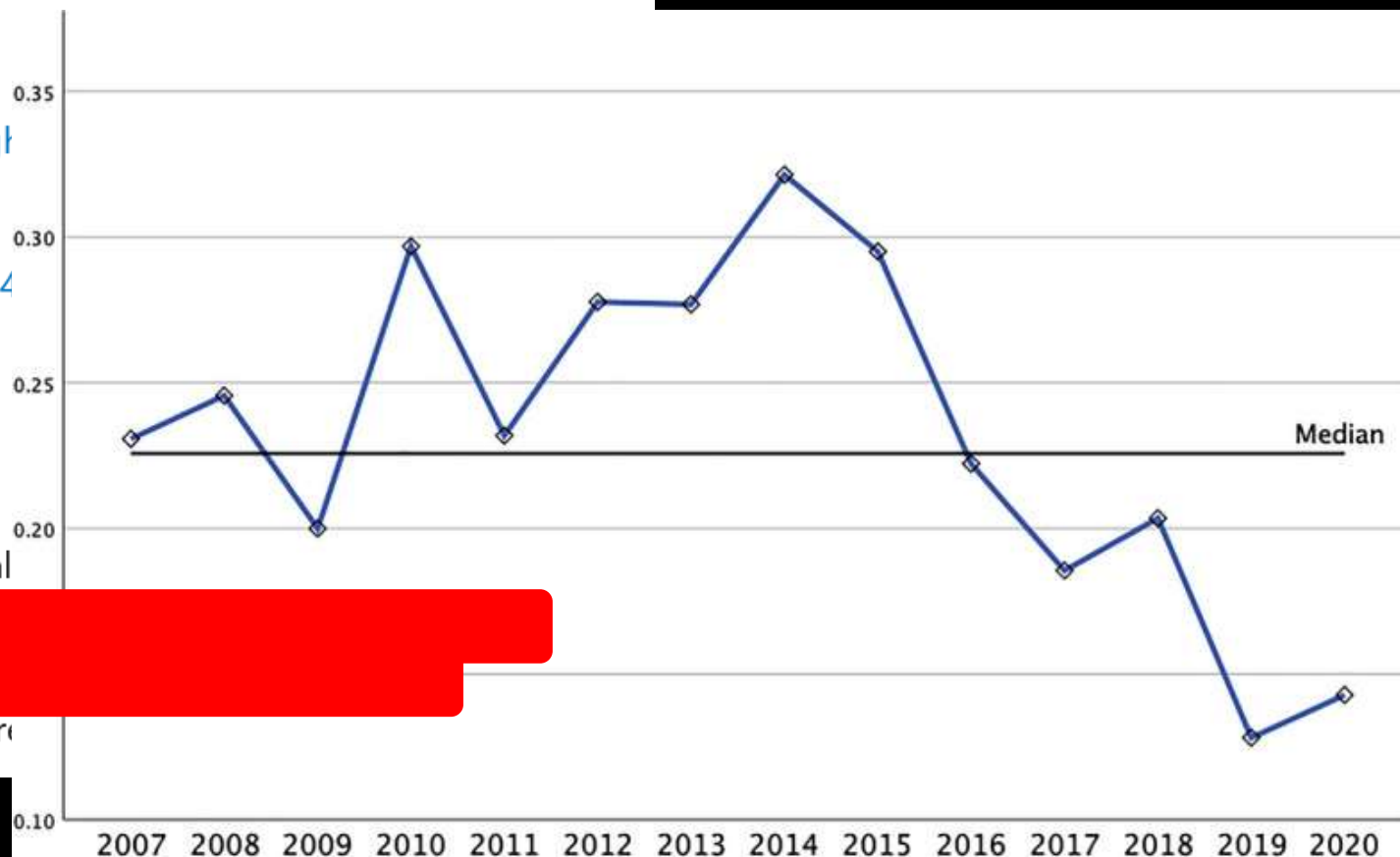
PMID: 34724583 PMCID: [PMC8559691](#) DOI: [10.1007/s00464](#)

[Free PMC article](#)

Abstract



Background: Accurate histopathologic diagnosis of colorectal decision-making and timely care. [REDACTED]

[REDACTED] determine whether these events lead to a delay in surgical care



Review

10 Years of GWAS Discovery: Biology, Functional Genomics, and Translational Research

Peter M. Visscher^{1, 2}  , Naomi R. Wray^{1, 2}, Qian Zhang¹, Pamela Sklar³, Brown⁷, Jian Yang^{1, 2}

 [Show more](#)

<https://doi.org/10.1016/j.ajhg.2017.06.005>

Under an Elsevier user license

Application of the experimental design of genome-wide association studies (GWAS) 10 years old (young), and here we review the remarkable range of genetic architecture facilitated in population and complex-trait genetics, the biology of complex traits, and the path toward new therapeutics. We predict the likely discoveries in the next 10 years of GWASs will be based on millions of samples with array data imputed to a large reference panel and on hundreds of thousands of samples with whole-genome sequencing data.

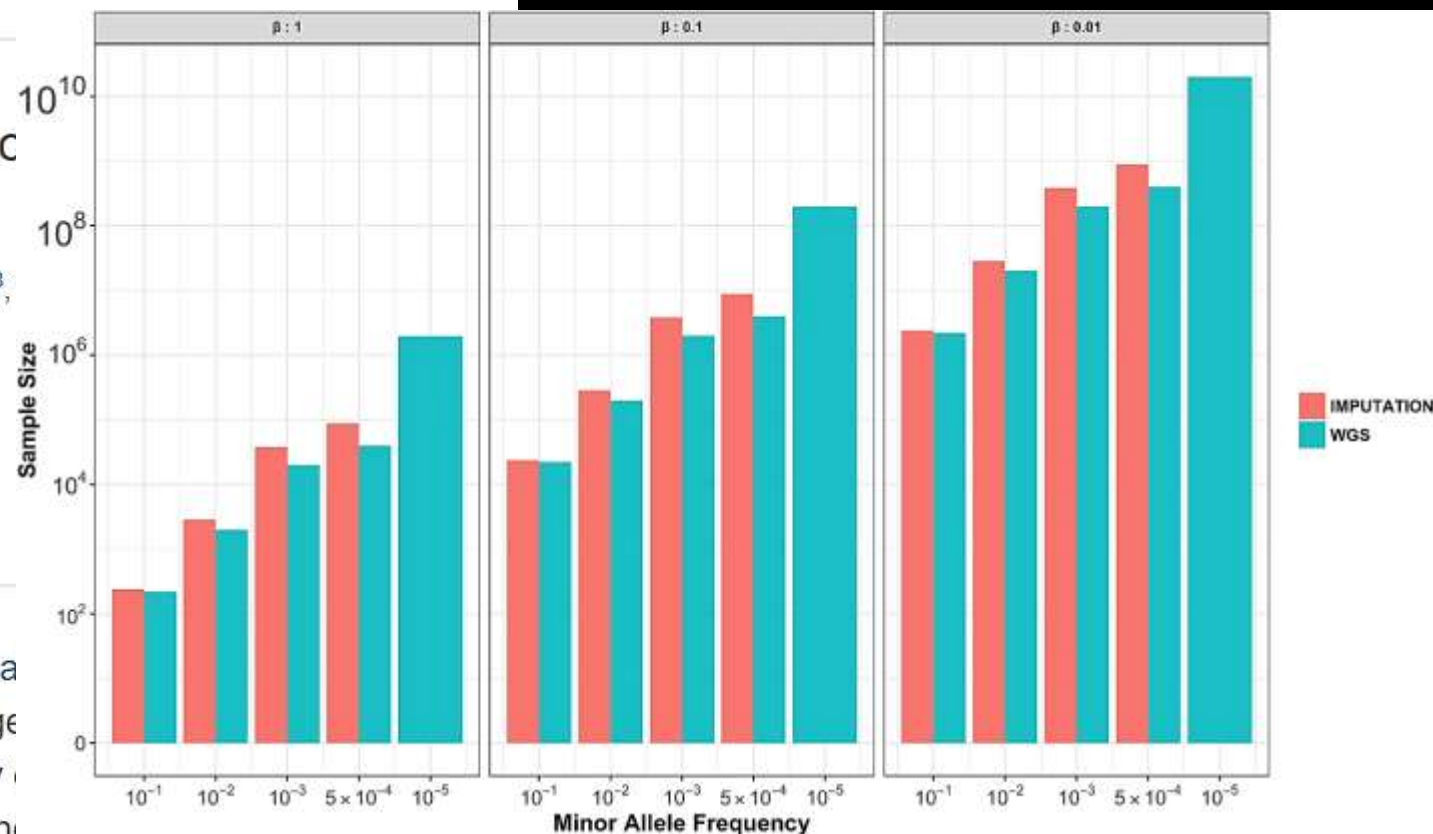


Figure 1. Minimum Sample Sizes for Detecting Trait-SNP Associations from Imputed and WGS Data

1. Populacja i próba danych
2. Parametr populacji i estymator
3. Ważne aspekty związane z tworzeniem prób danych i estymacją parametrów
4. Błąd próbkowania