

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 wariancji
14. Analiza kowariancji
15. Podsumowanie materiału, wspólna analiza przykładów, dyskusja

1. Testowanie hipotez

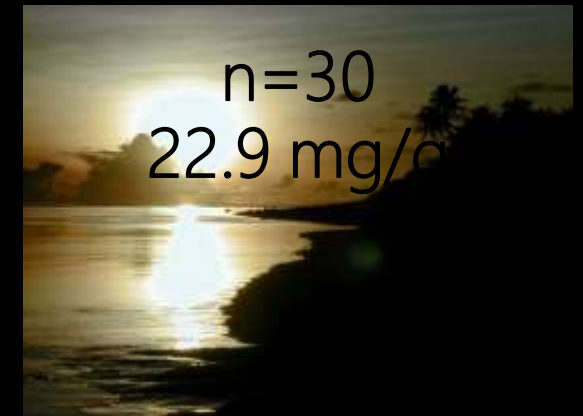
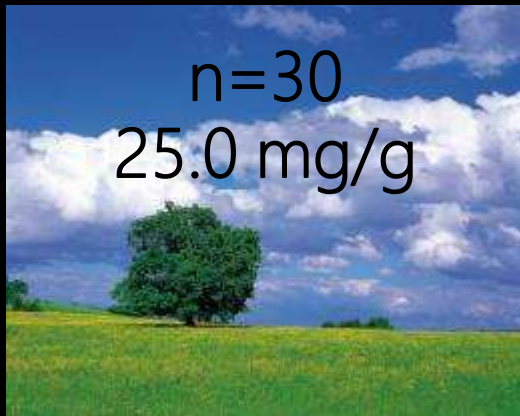
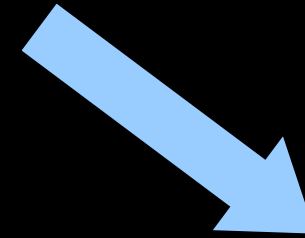
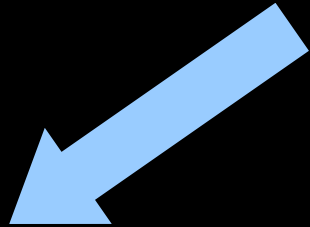
- Błędy związane z testowaniem hipotez
- Etapy testowania hipotez
- Testowanie wielokrotne

2. Estymacja parametrów

- Błąd standardowy
- Przedział ufności

Wnioskowanie statystyczne

Omulek słodkowodny
Hyridella menziesi



1. „Na oko” – różnica
2. UWAGA !!!
3. Błąd próbkowania
4. Estymatory różne
5. Parametry różne/równe
6. ???
7. Wnioskowanie statystyczne

Wnioskowanie statystyczne



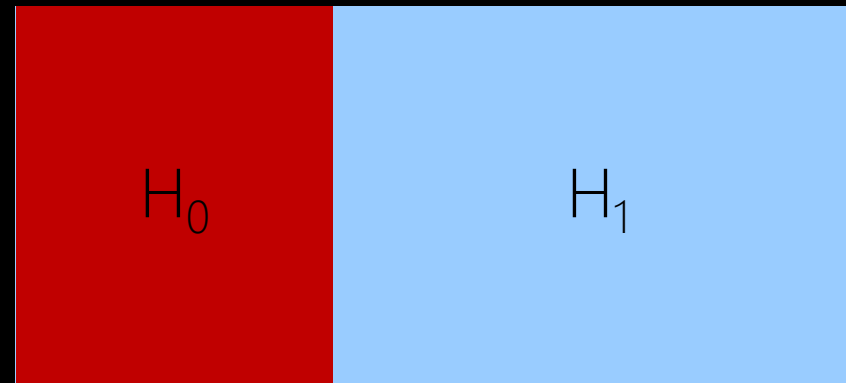
testowanie hipotez

estymacja parametrów

Testowanie hipotez

H_0 - hipoteza zerowa

H_1 - hipoteza alternatywna



- H_1 jest odwrotnością H_0
- H_0 zakłada „brak zmiany”, „brak efektu”
- Testowanie hipotez dotyczy przyjęcia lub odrzucenia H_0
- Sekwencyjne testowanie hipotez: $H_0 \rightarrow H_{1a} \rightarrow H_{1b}$

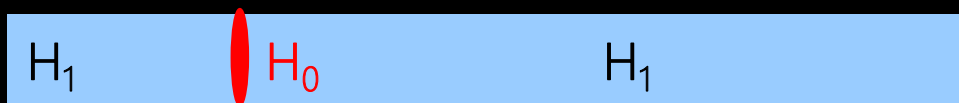
Testowanie hipotez

np.

1 PARAMETR

H_0 : koncentracja lipidów w grupie doświadczalnej wynosi 25.0 mg/g

H_1 : koncentracja lipidów w grupie doświadczalnej jest inna



$$H_0: k = 25.0$$

$$H_1: k \neq 25.0$$

H_0 : koncentracja lipidów w grupie doświadczalnej przekracza 25.0 mg/g

H_1 : koncentracja lipidów w grupie doświadczalnej jest mniejsza lub równa 25.0 mg/g



$$H_0: k > 25.0$$

$$H_1: k \leq 25.0$$

Testowanie hipotez

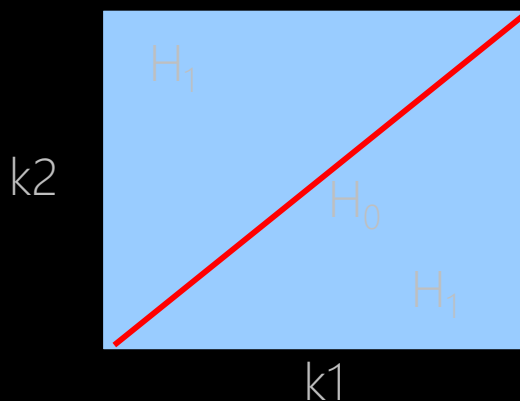
np.

2 PARAMETRY

H_0 : koncentracja lipidów w grupie doświadczalnej jest równa koncentracji w grupie kontrolnej

H_1 : koncentracja lipidów w grupach doświadczalnej i kontrolnej są różne

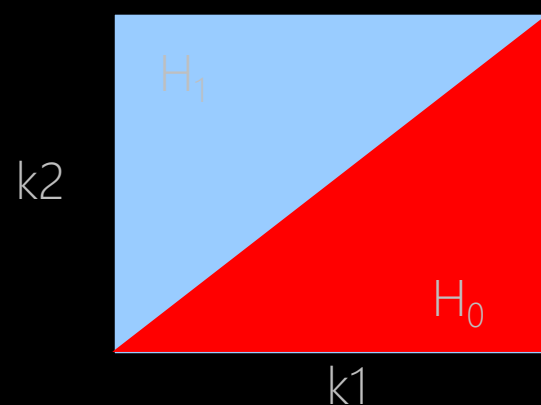
$H_0: k1 = k2$ $H_1: k1 \neq k2$



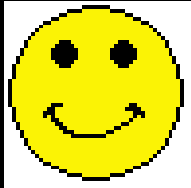


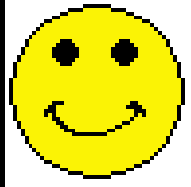
H_0 : koncentracja lipidów w grupie doświadczalnej jest wyższa niż w grupie kontrolnej

H_1 : koncentracja lipidów w grupie doświadczalnej jest niższa lub równa grupie kontrolnej

$H_0: k1 > k2$ $H_1: k1 \leq k2$



Błędy związane z testowaniem hipotez


błędy		prawdziwa hipoteza	
		H_0	H_1
przyjęta hipoteza	H_0		β 
	H_1	α 	

α - błąd I-go rodzaju (*type I error*)

β - błąd II-go rodzaju (*type II error*)

Błędy związane z testowaniem hipotez


błąd I-go rodzaju

błędy	prawdziwa hipoteza	
	H_0	H_1
przyjęta hipoteza	H_0	
	H_1	

- prawdopodobieństwo błędnego odrzucenia prawdziwej H_0
- prawdopodobieństwo zaobserwowania próby danych | H_0 jest prawdziwa
- poziom istotności testu (*significance level*)
- P wartość (*P value*)
- np. jeżeli $\alpha=0.05$ to na 100 testów w 5 niepotrzebnie odrzucono H_0
- kontrolujemy w czasie testowania

Błędy związane z testowaniem hipotez


błąd I-go rodzaju

błędy	prawdziwa hipoteza	
	H_0	H_1
przyjęta hipoteza	H_0	
	H_1	

```
for(i in 1:10) {  
  x=c(rnorm(500, mean=0, sd=1))  
  r=t.test(x, mu=0)$p.value  
  print(r)  
  i=i+1  
}
```

Błędy związane z testowaniem hipotez


błąd II-go rodzaju

błędy		prawdziwa hipoteza	
		H_0	H_1
przyjęta hipoteza	H_0		
	H_1		

- prawdopodobieństwo odrzucenie prawdziwej H_1
- $1-\beta$ – moc testu (power)
- określenie mocy: symulacje, numerycznie
- zależy od informatywności próby danych: liczebność, struktura, różnica pomiędzy H_0 a H_1

Błędy związane z testowaniem hipotez

błąd II-go rodzaju

błędy	prawdziwa hipoteza	
	H_0	H_1
przyjęta hipoteza	H_0	
	H_1	

```
for(i in 1:10) {  
  x=c(rnorm(500, mean=10, sd=1))  
  r=t.test(x, mu=0)$p.value  
  print(r)  
  i=i+1  
}
```

Etapy testowana hipotez

1. Określenie hipotez H_0 i H_1

$$H_0: k_1 \leq k_2$$

$$H_1: k_1 > k_2$$

2. Ustalenie poziomu istotności

- teoretycznego maksymalnego błędu α pozwalającego na odrzucenie H_0
- określenie maksymalnego poziomu nieściłości wnioskowania statystycznego

$$\alpha_{MAX}=0.05$$

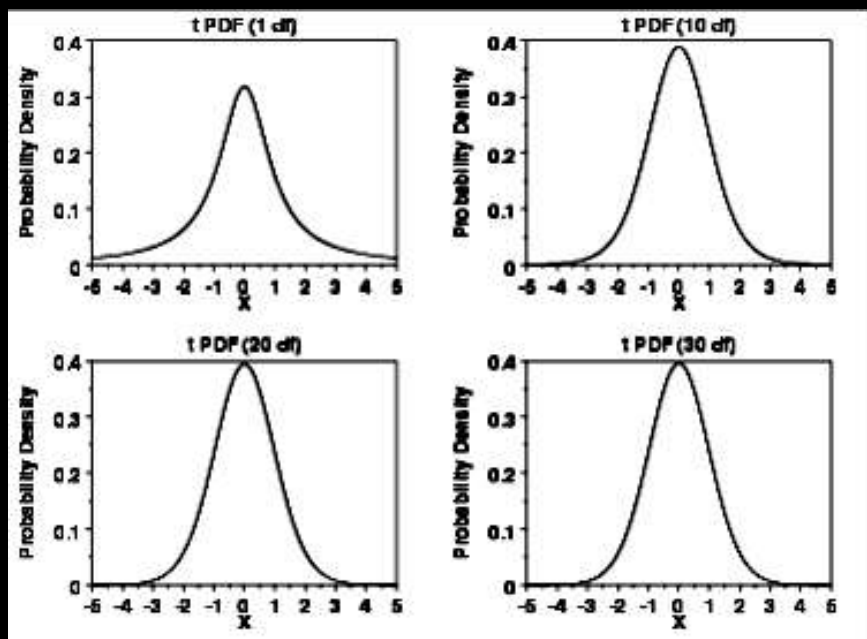
3. Wybór i obliczenie wartości testu statystycznego

$$t = \frac{\hat{k}_1 - \hat{k}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_2^2}{N_2}}}$$

Etapy testowana hipotez

4. Określenie rozkładu testu statystycznego

- asymptotyczny (\sim przybliżony) rozkład opisujący prawdopodobieństwo uzyskania danej wartości testu przy założeniu, że H_0 jest prawdziwa
- potrzebne do obliczenia błędu I-go rodzaju α
- kształt rozkładu zależy od: rodzaju testu, liczby stopni swobody



$$t = \frac{\hat{k}_1 - \hat{k}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_2^2}{N_2}}} \sim t_{N_1 + N_2 - 2}$$

Etapy testowana hipotez

5. Na podstawie 3. i 4. obliczenie poziomu błęd α dla danego eksperymentu (wartość P , P value) \rightarrow określa prawdopodobieństwo otrzymania wartości testu równej lub wyższej niż wartość uzyskanej w danym eksperymencie \rightarrow przy założeniu, że H_0 jest prawdziwa

$$t \rightarrow \alpha_T = P = 0.02$$

6. **Decyzja dotycząca przyjęcia lub odrzucenia H_0** \rightarrow im niższa wartość P tym większe przesłanki do odrzucenia H_0

np.	$\alpha_{MAX} = 0.05$	$t = 2.8$	$\rightarrow \alpha_T = P = 0.02$	H_0	H_1	??
np.	$\alpha_{MAX} = 0.05$	$t = 1.3$	$\rightarrow \alpha_T = P = 0.21$	H_0	H_1	??

Testowanie wielokrotne




$H_0: k_1 \leq k_2 / H_1: k_1 > k_2 \rightarrow \alpha_{MAX}=0.05 \rightarrow t \rightarrow \alpha_T \rightarrow H_0/H_1 \rightarrow 5%$ 



$H_0: k_1 \leq k_2 / H_1: k_1 > k_2 \rightarrow \alpha_{MAX}=0.05 \rightarrow t \rightarrow \alpha_T \rightarrow H_0/H_1 \rightarrow 5%$ 



$H_0: k_1 \leq k_2 / H_1: k_1 > k_2 \rightarrow \alpha_{MAX}=0.05 \rightarrow t \rightarrow \alpha_T \rightarrow H_0/H_1 \rightarrow 5%$ 

...



$H_0: k_1 \leq k_2 / H_1: k_1 > k_2 \rightarrow \alpha_{MAX}=0.05 \rightarrow t \rightarrow \alpha_T \rightarrow H_0/H_1 \rightarrow 5%$ 

całkowity błąd I-go rodzaju $\alpha_{max} \approx 0.05 * 10 = 50%$



Testowanie wielokrotne

Jak temu zaradzić ① ?

Korekta Bonferroniego

→

testy niezależne od siebie



$$\alpha_{MAX^*} = \alpha_{MAX} / N$$

$$\rightarrow \alpha_{MAX^*} = 0.05 / 10$$

$$\rightarrow \alpha_{MAX^*} = 0.005$$



$$\alpha_{MAX^*} = \alpha_{MAX} / N$$

$$\rightarrow \alpha_{MAX^*} = 0.05 / 10$$

$$\rightarrow \alpha_{MAX^*} = 0.005$$

...



$$\alpha_{MAX^*} = \alpha_{MAX} / N$$

$$\rightarrow \alpha_{MAX^*} = 0.05 / 10$$


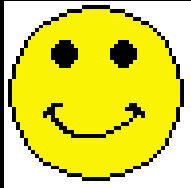
$$\rightarrow \alpha_{MAX^*} = 0.005$$

całkowity błąd I-go rodzaju $\alpha_{max} \approx 0.005 * 10 = 5\%$

Testowanie wielokrotne

Jak temu zaradzić ② ?

Korekta False Discovery Rate (FDR) → testy zależne od siebie

błędy	prawdziwa hipoteza	
	H_0	H_1
przyjęta hipoteza	H_0	
	H_1	 

$$FDR = E \left[\frac{\text{sad face emoji}}{\text{sad face emoji} + \text{happy face emoji}} \right]$$

Testowanie wielokrotne

Jak temu zaradzić ② ?

Korekta False Discovery Rate (FDR) → testy zależne od siebie

test	alfa(t)	FDR=(ranking/N)*alfa(max)	alfa(t)<FDR	alfa(max)=
1	0.001	0.005	H1	0.05
2	0.009	0.010	H1	
3	0.165	0.015	H0	
4	0.205	0.020	H0	
5	0.396	0.025	H0	
6	0.450	0.030	H0	
7	0.641	0.035	H0	
8	0.781	0.040	H0	
9	0.900	0.045	H0	
10	0.993	0.050	H0	

Journal List > Cell J > v.20(4); Winter 2019 > PMC6099145



Cell J. 2019 Winter; 20(4): 604–607.
Published online 2018 Aug 7. doi: [10.22034/201810001](https://doi.org/10.22034/201810001)

Why, When and How to

[Mohieddin Jafari, Ph.D¹](#) and [Naser Ansari](#)

▶ [Author information](#) ▶ [Article notes](#) ▶

 **BMC** Part of Springer Nature


Search

Genome Biology

[Home](#) [About](#) [Articles](#) [Submission Guidelines](#)

Method | [Open Access](#) | [Published: 11 October 2021](#)

Clipper: p -value-free FDR control on high-throughput data from two conditions

[Xinzhou Ge](#), [Yiling Elaine Chen](#), [Dongyuan Song](#), [MeiLu McDermott](#), [Kyla Woysner](#), [Antigoni Manousopoulou](#), [Ning Wang](#), [Wei Li](#), [Leo D. Wang](#) & [Jingyi Jessica Li](#) 

Genome Biology **22**, Article number: 288 (2021) | [Cite this article](#)

1470 Accesses | **36** Altmetric | [Metrics](#)

Wnioskowanie statystyczne

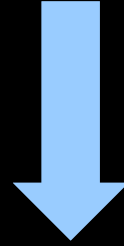


testowanie hipotez

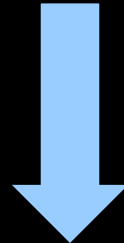
estymacja parametrów

Błąd standardowy

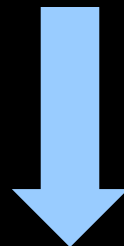
Jak dokładny jest dany estymator ???



błąd standardowy

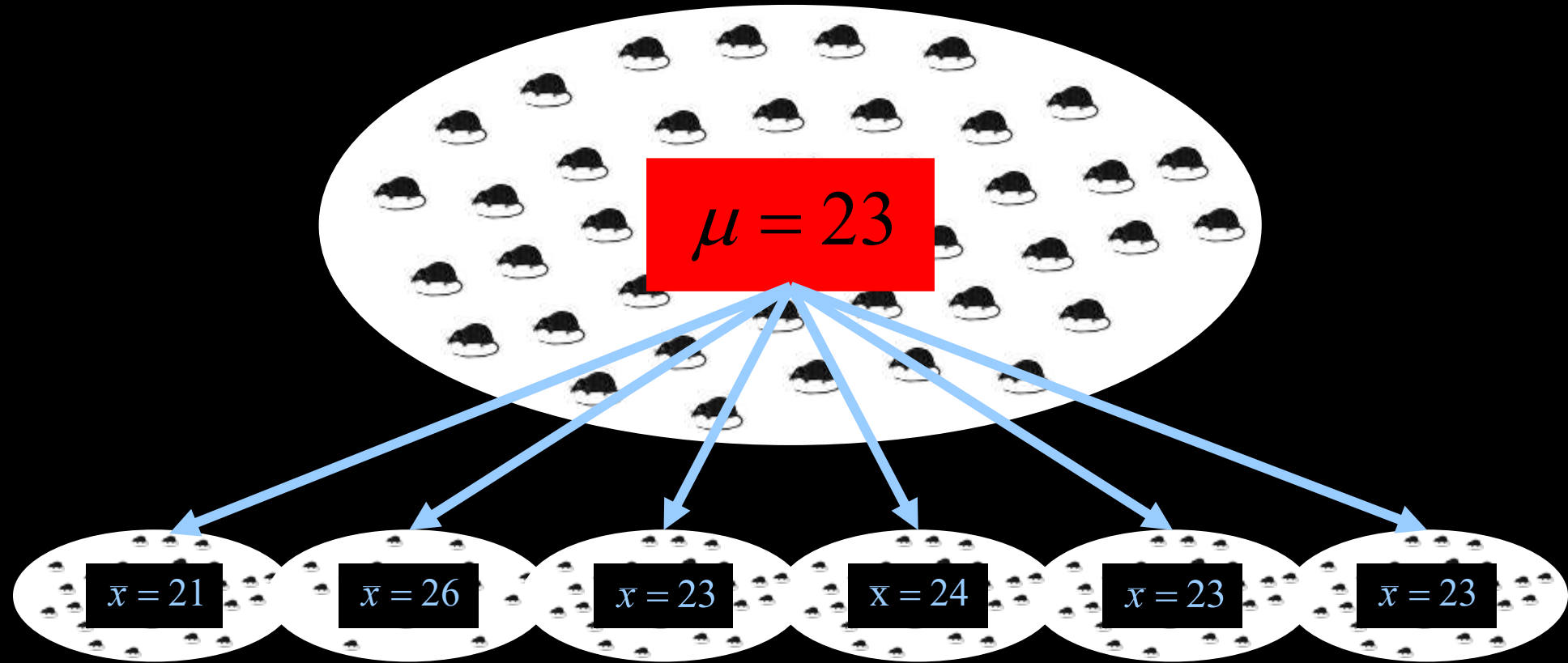


przedział ufności



\bar{X}

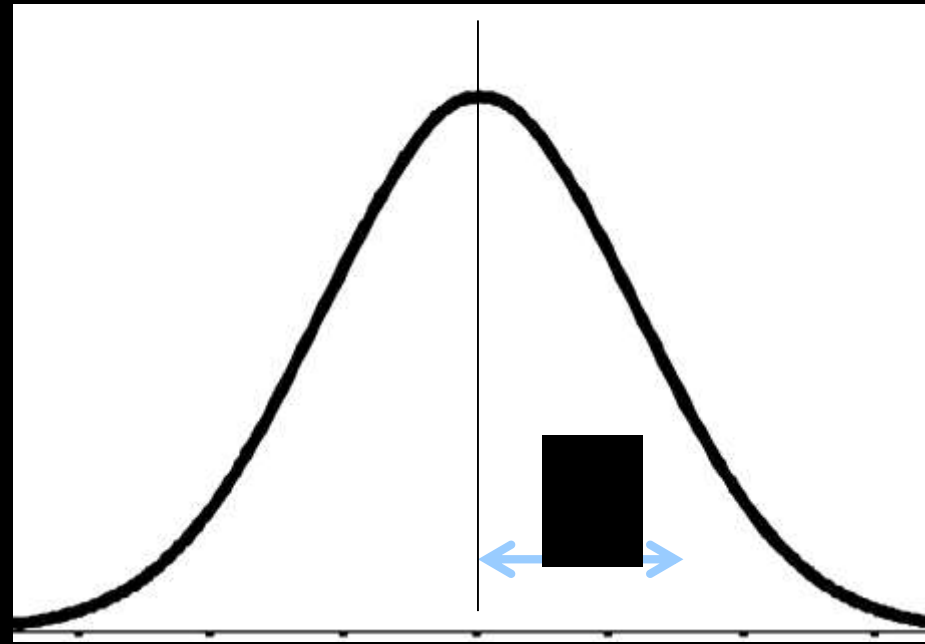
Błąd standardowy



Błąd standardowy estymatora:

odchylenie standardowe rozkładu estymatora

Błąd standardowy estymatora średniej arytmetycznej



\bar{x}

Jaki jest rozkład \bar{x} ?

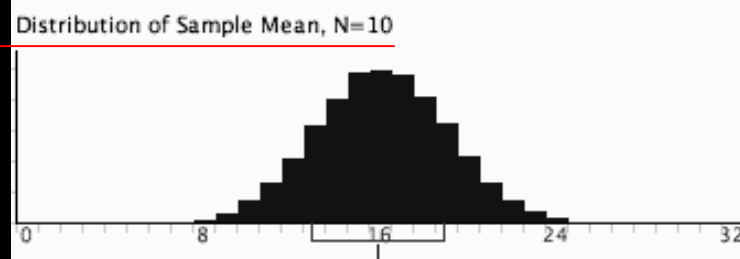
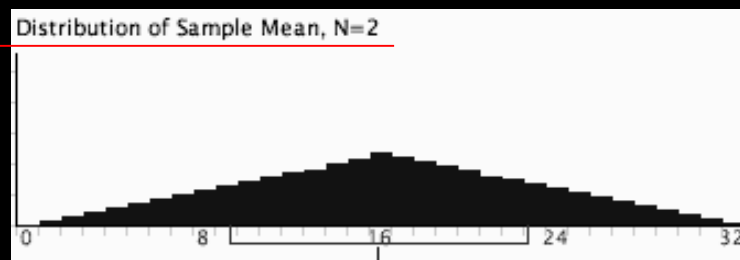
Jak obliczyć $s_{\bar{x}}$?

Błąd standardowy estymatora średniej arytmetycznej

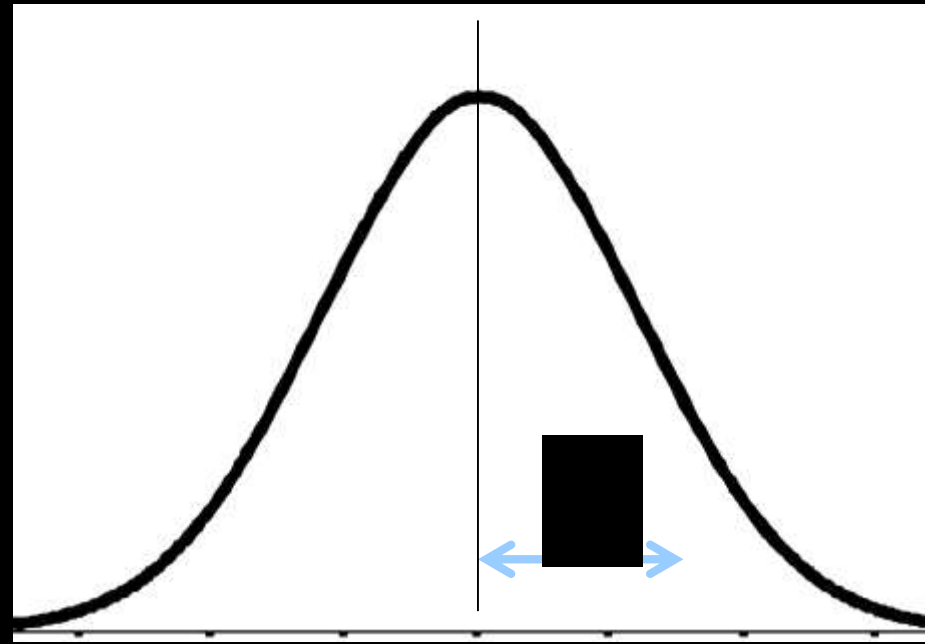
Jaki rozkład ma estymator średniej?

Dla dużych prób danych (N):

- rozkład estymatora średniej zbliża się do rozkładu Normalnego
- estymator średniej zbliża się do prawdziwej wartości parametru próby
- niezależnie od rozkładu obserwacji w próbie danych 😊



Błąd standardowy estymatora średniej arytmetycznej



\bar{x}

Jaki jest rozkład \bar{x} ?

Jak obliczyć $s_{\bar{x}}$?

Błąd standardowy estymatora średniej arytmetycznej

Jak obliczyć odchylenie standardowe rozkładu estymatora średniej (bez konieczności pobierania wielu prób danych) ?

$$S_{\bar{x}} = \frac{S_x}{\sqrt{N}}$$

Odchylenie standardowe w próbie danych:

$$S_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$$

Liczebność próby danych

błąd standardowy estymatora średniej

Błąd standardowy innych estymatorów

Błąd standardowy estymatora prawdopodobieństwa

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{N}}$$

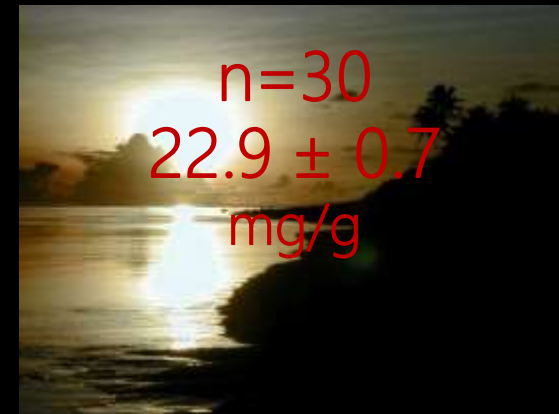
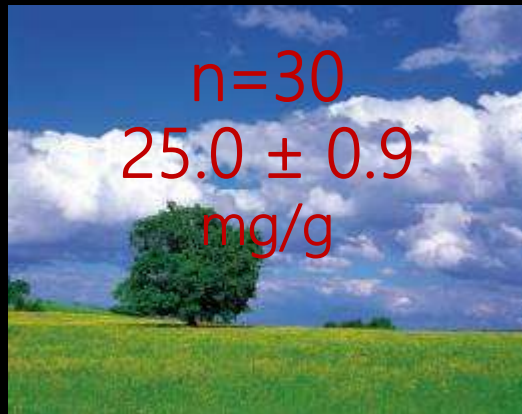
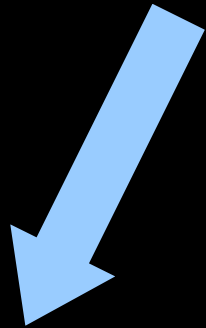
Błąd standardowy współczynnika regresji

$$S_{b_1} = \frac{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N - 2}}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Co wpływa na wielkość błędu standardowego?

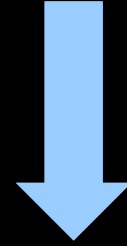
Wnioskowanie statystyczne

Omulek słodkowodny
Hyridella menziesi

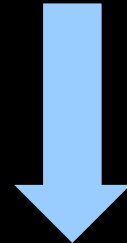


Przedział ufności estymatora średniej arytmetycznej

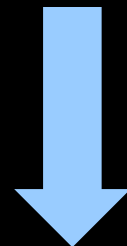
Jak dokładny jest dany estymator ???



błąd standardowy



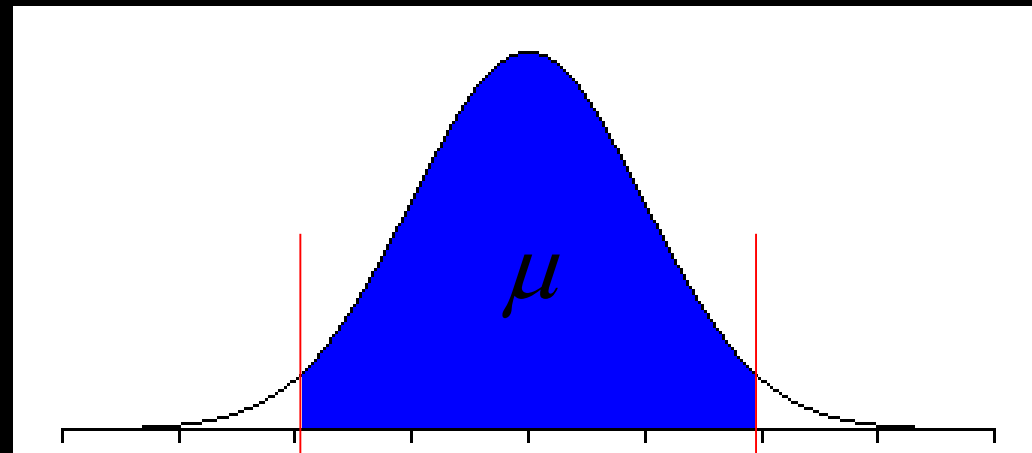
przedział ufności



\bar{X}

Przedział ufności

Przedział ufności estymatora: przedział w jakim z określonym prawdopodobieństwem znajduje się prawdziwa wartość parametru



\bar{X}_{\min}

\bar{X}_{\max}

↓ ↓
granice przedziału ufności

Przedział ufności estymatora średniej arytmetycznej

Jak obliczyć granice przedziału ufności ?

1. Wariancja próby znana lub próba bardzo liczna

$$\bar{X}_{\min} = \bar{X} - \underline{z_{\alpha}} S_{\bar{X}} \qquad \bar{X}_{\max} = \bar{X} + \underline{z_{\alpha}} S_{\bar{X}}$$

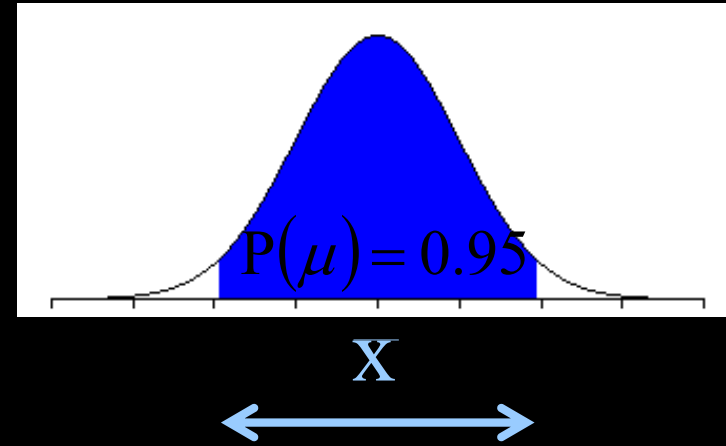
2. Wariancja próby nieznaną = obliczana na podstawie obserwacji w próbie

$$\bar{X}_{\min} = \bar{X} - \underline{t_{\alpha, N-1}} S_{\bar{X}} \qquad \bar{X}_{\max} = \bar{X} + \underline{t_{\alpha, N-1}} S_{\bar{X}}$$

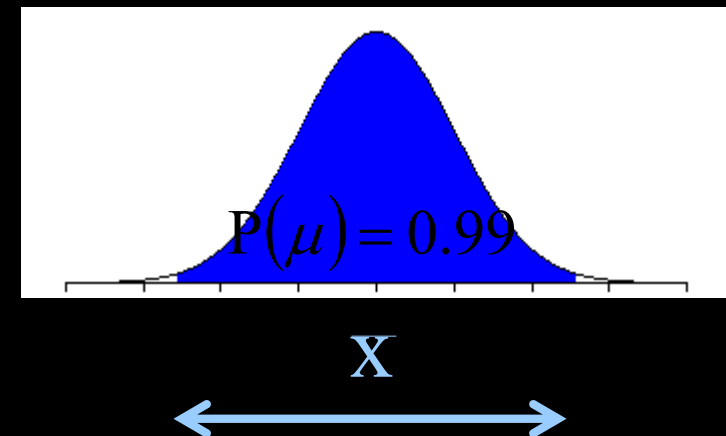
Przedział ufności estymatora średniej arytmetycznej

Prawdopodobieństwo wystąpienia prawdziwej średniej w przedziale ufności, a długość przedziału

1. Przedział ufności 95%

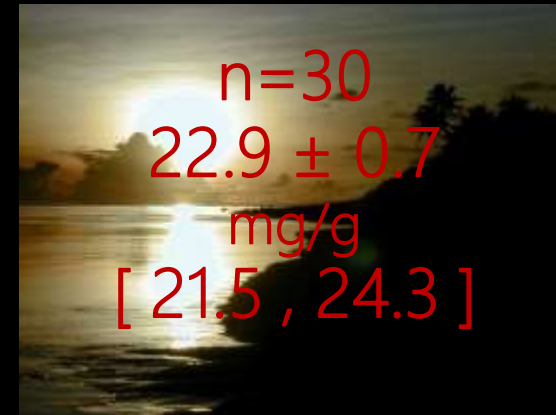
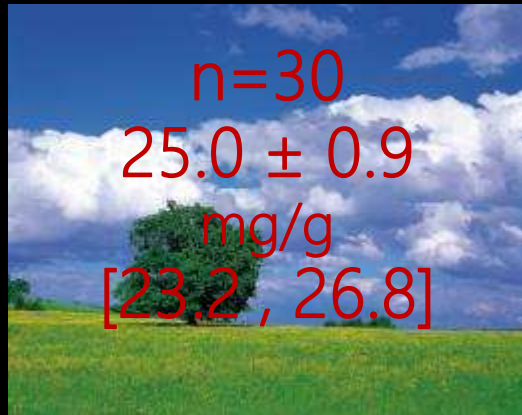
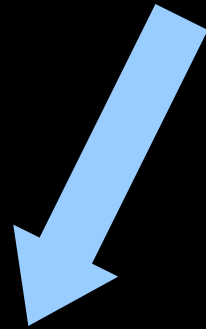


2. Przedział ufności 99%



Wnioskowanie statystyczne

Omulek słodkowodny
Hyridella menziesi





BMC Part of Springer Nature Search

Genome Biology

[Home](#) [About](#) [Articles](#) [Submission Guidelines](#)

Method | [Open Access](#) | [Published: 11 October 2021](#)

Clipper: p -value-free FDR control on high-throughput data from two conditions

[Xinzhou Ge](#), [Yiling Elaine Chen](#), [Dongyuan Song](#), [MeiLu McDermott](#), [Kyla Woysner](#), [Antigoni Manousopoulou](#), [Ning Wang](#), [Wei Li](#), [Leo D. Wang](#) & [Jingyi Jessica Li](#) 

Genome Biology **22**, Article number: 288 (2021) | [Cite this article](#)

1470 Accesses | **36** Altmetric | [Metrics](#)

— Bonferroni

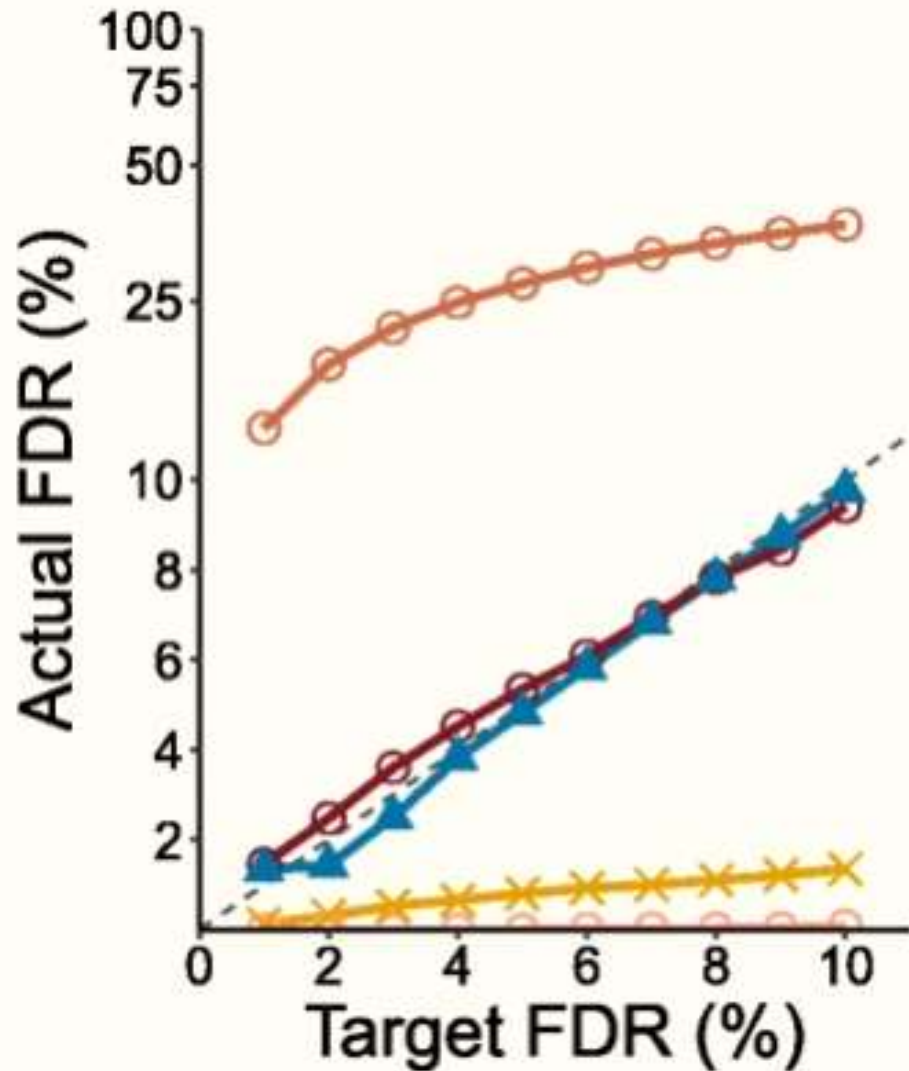
— **FDR** \rightarrow **False Discovery Rate** $\rightarrow = \frac{N(\text{False Positives})}{N(\text{Positives})} \rightarrow = \frac{N(\text{False } H_1)}{N(H_1)}$

— Bayesian FDR

— local FDR

— local false sign rate

Analiza danych symulowanych



- Clipper
- BH-pool
- BH-pair-correct
- BH-pair-2as1
- BH-pair-mis
- locfdr-emp
- locfdr-swap

1. Testowanie hipotez

- Błędy związane z testowaniem hipotez
- Etapy testowania hipotez
- Testowanie wielokrotne

2. Estymacja parametrów

- Błąd standardowy
- Przedział ufności