BIOSTATYSTYKA

Wykład 3: Testowanie hipotez i estymacja

dr inż. Bartosz Czech

Katedra Genetyki · Uniwersytet Przyrodniczy we Wrocławiu

bartosz.czech@upwr.edu.pl

Plan wykładu

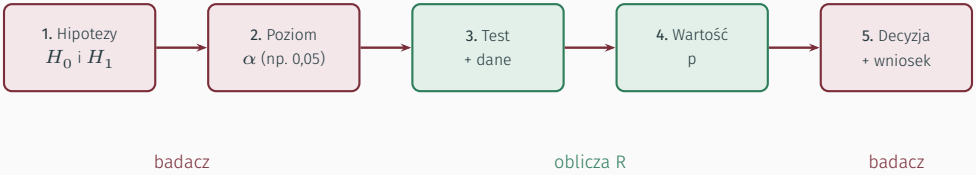
1. Schemat testu hipotez i statystyka testowa
2. Wartość p — co mówi, a czego nie mówi
3. Błędy I i II rodzaju, poziom istotności i moc testu
4. Estymacja parametrów i przedział ufności

Cel wykładu

Zrozumieć, jak statystyka podejmuje decyzję o hipotezie (wartość p, błędy, moc) i jak szacuje wielkość efektu (estymacja, przedział ufności).

1. Schemat testu hipotez

Pięć kroków testu



Ten sam schemat obowiązuje w teście t, chi-kwadrat i ANOVA — zmienia się tylko **krok 3** (wybór testu).

> Program obliczy p zawsze. Za to, czy pytanie i test mają sens, odpowiada **badacz**.

Hipotezy statystyczne

Hipoteza statystyczna (*statistical hypothesis*)

Twierdzenie o parametrze lub rozkładzie populacji, które można zweryfikować na podstawie próby.

Hipoteza zerowa H_0 (*null hypothesis*)

Brak efektu lub różnicy, zapisywany jako równość: $H_0 : \mu_1 = \mu_2$.

Hipoteza alternatywna H_1 (*alternative hypothesis*)

Przeczy H_0: $H_1 : \mu_1 \neq \mu_2$; zwykle wyraża to, co badacz chce wykazać.

- Para H_0, H_1 — przypomnienie z W1 (nokaut genu a masa ciała myszy), teraz w zapisie symbolicznym: μ_1, μ_2 to średnie masy w populacjach myszy z nokautem i typu dzikiego
- Hipotezy dotyczą **parametrów populacji** (μ), a nie statystyk z próby ($\bar{x}$): różnica $\bar{x}_1 - \bar{x}_2$ jest prawie zawsze niezerowa, pytamy o $\mu_1 - \mu_2$
- H_0 i H_1 wykluczają się wzajemnie i razem obejmują wszystkie możliwości

Statystyka testowa — sygnał do szumu

Statystyka testowa (*test statistic*)

Liczba obliczana z danych, która mierzy, jak bardzo wynik odbiega od H_0 w stosunku do spodziewanego szumu. Jej rozkład przy założeniu prawdziwości H_0 (**rozkład zerowy**, *null distribution*) jest znany, dlatego można obliczyć p.

- Typowa postać w testach jednego parametru (średnia, różnica średnich, nachylenie): $\frac{\text{oszacowanie} - \text{wartość z } H_0}{\text{błąd standardowy oszacowania}}$, czyli **sygnał** podzielony przez **szum**
- Dla dwóch średnich: $t = (\bar{x}_1 - \bar{x}_2)/\mathrm{SE}$, gdzie SE jest błędem standardowym tej różnicy
- Statystyki chi-kwadrat (W5) i F (W8, W10) mierzą odstępstwo od H_0 inaczej, ale także względem szumu
- Duża co do wartości bezwzględnej statystyka oznacza wynik mało prawdopodobny przy H_0, a więc małe p

2. Wartość p

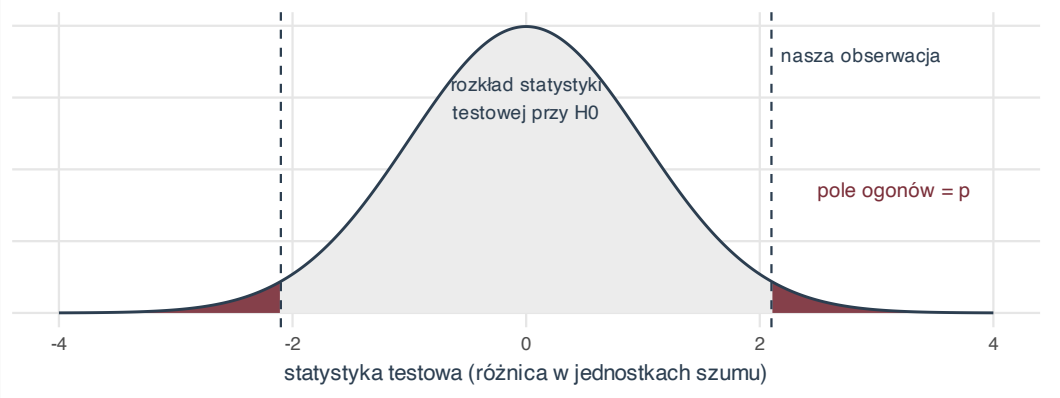
Czym jest wartość p

Wartość p (*p-value*)

Prawdopodobieństwo uzyskania wartości statystyki testowej **co najmniej tak skrajnej** jak zaobserwowana, przy założeniu, że H_0 **jest prawdziwa** i spełnione są założenia testu.

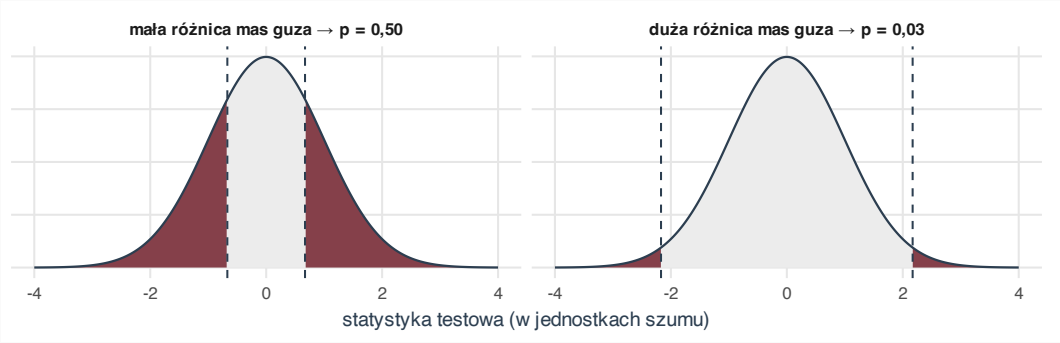
- Zapis formalny: $p = P(\text{statystyka co najmniej tak skrajna jak obserwowana} \mid H_0)$
- Małe p → dane byłyby **zaskakujące**, gdyby efektu nie było → argument przeciw H_0
- Duże p → taki wynik łatwo powstaje przy prawdziwej H_0 → brak podstaw do jej odrzucenia
- p jest **warunkowe**: całe rozumowanie zakłada prawdziwość H_0 i założeń testu

Jak test „widzi” wynik



p to **pole ogonów** rozkładu zerowego poza naszą obserwacją — im dalej w ogon, tym mniejsze p.

Przykład: p = 0,03 kontra p = 0,50



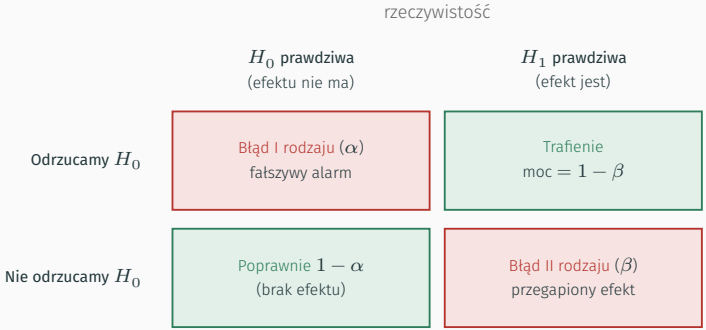
Ten sam rozkład zerowy, dwie obserwacje. Im dalej obserwacja od zera, tym mniejsze pole ogonów i mniejsze p.

Mity o wartości p

Mit	Jak jest naprawdę
„p to prawdopodobieństwo, że H_0 jest prawdziwa”	p *zakłada*, że H_0 jest prawdziwa; nie mówi o jej prawdopodobieństwie
„p to prawdopodobieństwo, że wynik jest przypadkowy”	p oblicza się przy założeniu, że przypadek (H_0) jest jedynym wyjaśnieniem; nie jest prawdopodobieństwem tego wyjaśnienia
„$p < 0{,}05$ znaczy, że efekt na pewno istnieje”	to tylko mała szansa takich danych przy H_0, nie pewność
„małe p = duży efekt”	przy dużym n nawet znikomy efekt daje małe p
„$1 - p$ to szansa, że wynik się powtórzy”	nie; powtarzalność zależy od mocy badania
„$p > 0{,}05$ znaczy, że efektu nie ma”	brak dowodu efektu $\neq$ dowód braku efektu

3. Błędy, poziom istotności i moc

Dwa rodzaje pomyłek



- α — prawdopodobieństwo odrzucenia H_0, gdy jest prawdziwa; β — prawdopodobieństwo nieodrzucenia H_0, gdy prawdziwa jest H_1
- Przy stałym n zmniejszenie α zwiększa β

Poziom istotności i reguła decyzyjna

Poziom istotności α (*significance level*)

Dopuszczalne prawdopodobieństwo błędu I rodzaju, czyli odrzucenia H_0, gdy jest ona prawdziwa; ustalane przed analizą.

Wynik	Decyzja	Poprawne sformułowanie
$p \leq \alpha$	odrzucamy H_0	różnica jest istotna statystycznie
$p > \alpha$	brak podstaw do odrzucenia H_0	nie wykazano różnicy

Niepoprawne: „przyjmujemy H_0" lub „H_0 jest prawdziwa". Brak podstaw do odrzucenia H_0 nie jest dowodem braku efektu.

Poziom istotności a wartość p

	Poziom istotności α	Wartość p
Kto ustala	badacz	dane
Kiedy	przed analizą	po analizie
Rola	próg decyzji; kontroluje błąd I rodzaju	miara zgodności danych z H_0
Zmienia się z próbą	nie	tak

Decyzja powstaje z porównania obu liczb: $p \leq \alpha$. Mylenie ich prowadzi do zdań w rodzaju „α wyniosło 0,03”.

Test dwustronny i jednostronny

	Dwustronny (*two-sided*)	**Jednostronny** (*one-sided*)
Hipoteza H_1	$\mu_1 \neq \mu_2$	$\mu_1 > \mu_2$ lub $\mu_1 < \mu_2$
Wartość p	pole obu ogonów	pole jednego ogona
Kiedy	domyślnie	tylko gdy kierunek jest uzasadniony przed analizą

Przykład: nokaut genu *Lep* (leptyny) prowadzi u myszy do otyłości, więc kierunek zmiany masy jest znany z literatury, a mimo to stosuje się dwustronną H_1: spadek masy też byłby ważnym wynikiem.

Kierunku nie wybiera się po zobaczeniu danych

Wybór ogona zgodnego z wynikiem podwaja rzeczywisty poziom błędu I rodzaju: nominalne 0,05 w teście jednostronnym odpowiada wtedy 0,10.

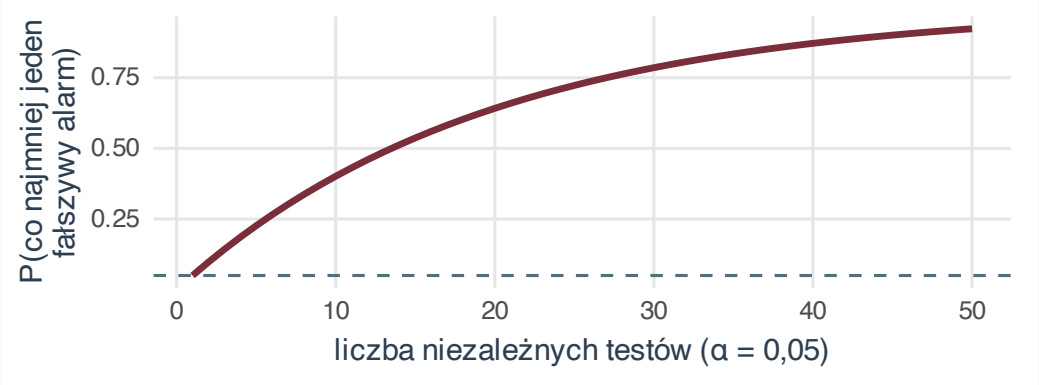
Próg istotności jako konwencja

- Próg $\alpha = 0{,}05$ zaproponował **Ronald Fisher** — to **umowa**, nie prawo natury
- Nic szczególnego nie dzieje się między $p = 0{,}049$ a $p = 0{,}051$ — to praktycznie ten sam wynik
- Próg ustala się **przed** analizą; w niektórych dziedzinach bywa $0{,}01$ lub $0{,}005$

Pułapka: kult progu 0,05

Traktowanie $0{,}05$ jak ostrej granicy „jest / nie ma efektu” to nadużycie. Wartość p to liczba ciągła — należy ją raportować, a nie tylko etykietę „istotne / nieistotne”.

Wielokrotne testowanie



- Przy 20 niezależnych testach prawdziwych H_0 prawdopodobieństwo co najmniej jednego błędu I rodzaju wynosi $1 - 0{,}95^{20} \approx 64\%$

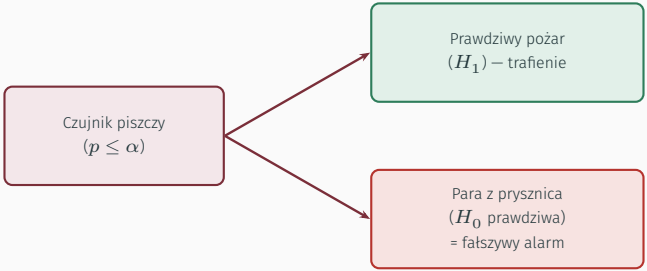
p-hacking — definicja i środki zaradcze

p-hacking

Wielokrotne analizowanie danych (podgrup, zmiennych, testów, usuwanie obserwacji) aż do uzyskania $p \leq \alpha$, bez ujawnienia liczby prób.

- Nie musi to być świadome oszustwo: często jest to skutek swobody w prowadzeniu analizy
- Środki zaradcze: analiza zaplanowana z góry, korekta na wielokrotne testowanie (np. Bonferroniego: próg α/k dla k testów), raportowanie wszystkich wykonanych testów

Analogia: p jako sygnał alarmu



Małe p to **sygnał**, nie dowód. Alarm bez pożaru = błąd I rodzaju; pożar bez alarmu = błąd II rodzaju.

Wielkość efektu

Wielkość efektu (*effect size*)

Miara siły zjawiska, niezależna od liczebności próby. Dla dwóch średnich: d Cohena — różnica średnich w jednostkach wspólnego odchylenia standardowego, $d = (\bar{x}_1 - \bar{x}_2)/s_p$, gdzie s_p to wspólne (łączone) odchylenie standardowe obu grup.

- Progi umowne (Cohen, 1988): $0{,}2$ mały, $0{,}5$ średni, $0{,}8$ duży; ich sens zależy od dziedziny
- Przykład: różnica mas 3 g przy wspólnym SD równym 6 g daje $d = 0{,}5$
- Inne miary: różnica bezwzględna w jednostkach pomiaru, krotność zmiany, iloraz szans

Wartość p zależy od n, wielkość efektu nie. Dlatego sama istotność nie mówi, czy efekt jest duży.

Moc testu

Moc testu (*statistical power*)

Prawdopodobieństwo odrzucenia H_0, gdy prawdziwy jest efekt o określonej wielkości; moc $= 1 - \beta$. Zależy od wielkości efektu, n, zmienności i α.

- Przyjęty cel w planowaniu badań: moc $\geq 80\%$
- Moc 50% = rzut monetą, czy w ogóle zobaczymy realny efekt
- Niska moc to częsty powód, dla którego prawdziwe efekty „nie wychodzą istotne”

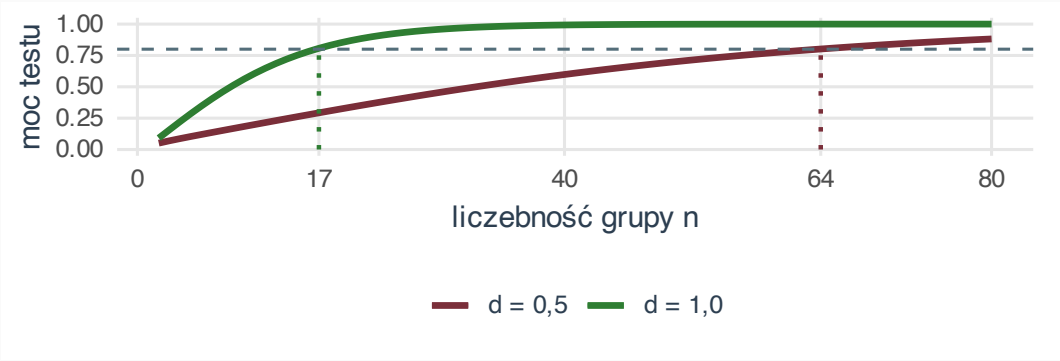
> Wynik nieistotny przy małej próbie może oznaczać zbyt słaby test, a nie brak efektu (por. mity o wartości p).

Czynniki wpływające na moc

- **Liczebność** n — więcej niezależnych jednostek (zwierząt, roślin) w każdej grupie
- **Wielkość efektu** — silniejszą różnicę łatwiej wykryć; zwykle narzuca ją biologia
- **Zmienność (szum)** — kontrola warunków, dokładniejszy pomiar, jednorodny materiał
- **Poziom** α — wyższe α daje większą moc, ale więcej fałszywych alarmów
- **Plan doświadczenia i rodzaj testu** — pomiary sparowane i bloki usuwają część zmienności; test jednostronny, jeśli kierunek uzasadniono z góry

> Moc rośnie, gdy **sygnał** rośnie względem **szumu**. Badacz kontroluje bezpośrednio n, α i plan doświadczenia; wielkości efektu zwykle nie wybiera.

Krzywa mocy



- Moc rośnie z liczebnością i z wielkością efektu
- Moc 80% (linia przerywana) osiąga się przy $d = 1{,}0$ dla ok. 17 osobników na grupę, a przy $d = 0{,}5$ dla ok. 64 (linie kropkowane)

Analiza mocy przed eksperymentem

Przed zebraniem danych: **ile obserwacji potrzeba, żeby mieć realną szansę wykryć spodziewany efekt?**

- Wejście: spodziewana wielkość efektu, zmienność (SD), żądana moc (np. 80%), α
- Wyjście: minimalne n na grupę

Wymiar etyczny (badania na zwierzętach)

Za **mała** próba = zwierzęta wykorzystane bez szans na wniosek. Za **duża** = niepotrzebne cierpienie i koszt. Komisje etyczne wymagają uzasadnienia liczby zwierząt (zasada *3R*: *reduction*).

4. Estymacja parametrów

Testowanie a estymacja

Test hipotez

Odpowiada **tak / nie**: czy są podstawy, by odrzucić H_0? „Czy nawóz zmienia plon?”

Estymacja (*estimation*)

Szacowanie parametru populacji na podstawie próby: **ile** wynosi efekt i z jaką niepewnością? „O ile ton na hektar nawóz zwiększa plon?”

> Samo „istotne” rzadko wystarcza. Recenzent i praktyk chcą wiedzieć, **jak duży** jest efekt — to daje estymacja.

Estymator i oszacowanie

Estymator (*estimator*)

Reguła (funkcja próby) służąca do szacowania parametru populacji, np. średnia z próby $\bar{x}$ dla μ.

Oszacowanie (*estimate*)

Wartość estymatora obliczona z konkretnej próby, np. $\bar{x} = 24$ g.

- W kolejnych próbach estymator daje różne oszacowania — ma więc swój **rozkład z próbkowania** (wykład 2)
- **Wartość oczekiwana estymatora** (*expected value*) — średnia jego rozkładu z próbkowania, czyli średnia oszacowań z bardzo wielu prób
- **Wariancja estymatora** — wariancja tego rozkładu, czyli kwadrat błędu standardowego (SE^2)

> Własności estymatora (następne slajdy) opisują jego rozkład z próbkowania, a nie jedno oszacowanie.

Obciążenie estymatora — ilustracja



nieobciążony

oszacowania rozłożone wokół parametru



obciążony

systematyczne przesunięcie

Obciążenie (*bias*)

Różnica między wartością oczekiwaną estymatora a prawdziwą wartością parametru.

> Rysunek ilustruje jedną z sześciu własności estymatora; pełny zestaw — następny slajd.

Własności estymatora

Własność	Znaczenie
Nieobciążoność (*unbiasedness*)	wartość oczekiwana estymatora jest równa parametrowi przy każdej liczebności próby
Asymptotyczna nieobciążoność (*asymptotic unbiasedness*)	obciążenie maleje do zera, gdy liczebność próby rośnie
Zgodność (*consistency*)	gdy próba rośnie, prawdopodobieństwo, że oszacowanie odbiega od parametru o więcej niż przyjęta tolerancja, spada do zera — dla każdej tolerancji, nawet bardzo małej
Efektywność (*efficiency*)	spośród estymatorów nieobciążonych ma najmniejszą wariancję
Asymptotyczna efektywność (*asymptotic efficiency*)	wraz ze wzrostem n stosunek jego wariancji do najmniejszej osiągalnej dąży do 1
Dostateczność (*sufficiency*)	wykorzystuje całą informację o parametrze zawartą w próbie

Zgodność i własności asymptotyczne opisują zachowanie przy rosnącej próbie; przy małej niczego nie gwarantują. Estymator nieobciążony, którego wariancja maleje do zera, jest zgodny.

Własności estymatora — przykłady

- **Średnia z próby** — nieobciążona i zgodna; dla rozkładu normalnego także efektywna i dostateczna dla μ
- **Wariancja z dzielnikiem** n — obciążona (zaniża σ^2), lecz asymptotycznie nieobciążona; dzielnik $n - 1$ usuwa obciążenie
- **Mediana** przy rozkładzie normalnym — zgodna, lecz mniej efektywna od średniej, także asymptotycznie: przy dużej próbie jej wariancja jest ok. **1,57 raza** większa
- **Estymatory największej wiarygodności** — wartości parametru, przy których zaobserwowane dane są najbardziej prawdopodobne; przy typowych założeniach zgodne i asymptotycznie efektywne
- **Liczba wykiełkowanych nasion** spośród n wysianych — statystyka dostateczna dla prawdopodobieństwa kiełkowania

Estymator obciążony bywa użyteczny, jeśli ma wyraźnie mniejszą wariancję: liczy się łączny błąd, czyli **błąd średniokwadratowy** = kwadrat obciążenia + wariancja.

Estymator punktowy a przedziałowy

Estymator punktowy (*point estimator*)

Daje jedną liczbę jako oszacowanie parametru.

„Kiełkuje 82% nasion."

Estymator przedziałowy (*interval estimator*)

Daje zakres wartości z określonym poziomem ufności. „Kiełkuje 82% nasion (95% przedział ufności, *confidence interval*, CI: 76–88%; definicja za chwilę)."

Reguła raportowania

Liczba bez zakresu jest jak pomiar bez błędu. **Zawsze** podawać oszacowanie razem z jego niepewnością.

Przedział ufności — definicja i budowa

Przedział ufności (*confidence interval*, CI)

Przedział wyznaczony z próby metodą, która w długiej serii powtórzeń obejmuje prawdziwą wartość parametru w $(1-\alpha)\cdot 100\%$ przypadków. Liczbę $1-\alpha$ (zwykle 95%) nazywamy **poziomem ufności**.

- Dla dużych prób: $\bar{x} \pm 1{,}96 \cdot \mathrm{SE}$ — w rozkładzie normalnym 95% pola leży w odległości 1,96 SD od środka (reguła 68–95–99,7), a SD rozkładu średniej to SE
- Dla małych prób mnożnik t^* jest **większy od 1,96**, bo uwzględnia dodatkową niepewność oszacowania SD z próby; pochodzi z rozkładu t (wykład 4)
- Przykład: dane 20, 22, 23, 25, 30 — $\bar{x} = 24$, $\mathrm{SE} = 3{,}81/\sqrt{5} \approx 1{,}70$, dla $n = 5$ mnożnik $t^* = 2{,}78$

Wynik

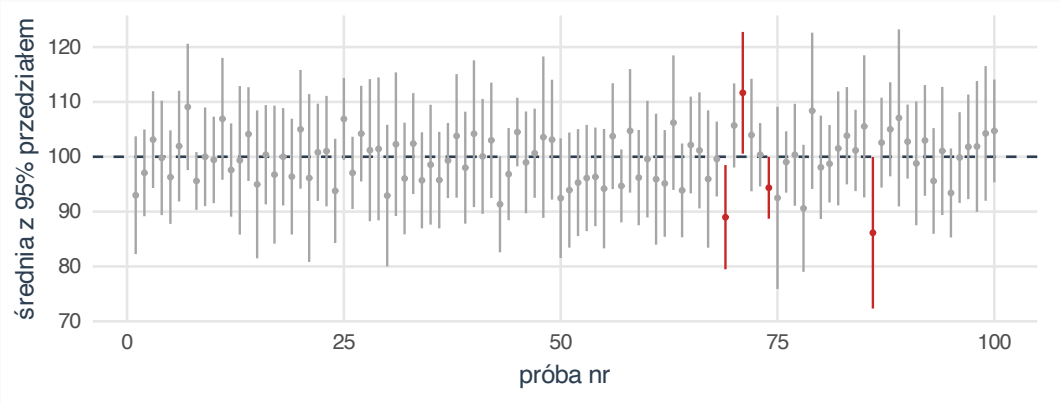
95% przedział ufności dla średniej: $24 \pm 4{,}73$, czyli od $19{,}3$ do $28{,}7$.

Interpretacja przedziału ufności

Poprawnie	Błędnie
Metoda daje przedziały, z których 95% obejmuje prawdziwy parametr	Z prawdopodobieństwem 95% parametr leży w tym konkretnym przedziale
Przedział wskazuje wartości parametru zgodne z danymi	W przedziale mieści się 95% obserwacji
Średnia z nowej próby trafia w ten przedział rzadziej (przy dużych próbach ok. 83%)	95% średnich z kolejnych prób wpadnie w ten przedział

Parametr jest stały, losowy jest przedział. Po wyznaczeniu konkretnego przedziału parametr albo w nim jest, albo nie.

Pokrycie przedziałów ufności — symulacja



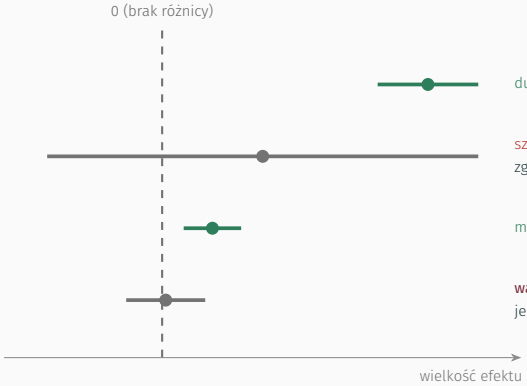
Przedziały, które nie obejmują prawdziwej średniej (kolor czerwony; pozostałe szare): 4 ze 100. W długiej serii nie obejmuje jej ok. 5% przedziałów.

Szerokość przedziału ufności

Czynnik	Zmiana	Szerokość przedziału
liczebność próby n	rośnie	maleje
rozrzut danych (SD)	rośnie	rośnie
poziom ufności	90% $\longrightarrow$ 95% $\longrightarrow$ 99%	rośnie

- Mnożniki dla dużych prób: $1{,}645$ (90%), $1{,}96$ (95%), $2{,}576$ (99%)
- Czterokrotne zwiększenie n zawęża przedział mniej więcej dwukrotnie
- Wyższa pewność, że metoda obejmie parametr, kosztuje szerszy przedział

Przedział ufności — odczytywanie



duży, wąski CI — istotny i precyzyjny

szeroki CI obejmujący 0 — wynik niepewny: zgodny i z brakiem efektu, i z dużym efektem

mały, wąski — istotny, lecz niewielki

wąski CI wokół 0 — efekt, jeśli istnieje, jest mały (nie: „brak efektu”)

CI pokazuje trzy rzeczy naraz: **ile**, **w którą stronę** i **jak precyzyjnie**.

Przedział ufności a istotność

Reguła spójna z wartością p

Jeżeli dwustronny 95% przedział ufności dla **różnicy** nie zawiera 0, test tej samej metody da $p < 0{,}05$.
Dla ilorazów (krotność zmiany, iloraz szans) wartością braku efektu jest **1**.

- Różnica mas: $+3{,}1$ g (95% CI: $1{,}2$ do $5{,}0$) → nie obejmuje 0 → istotna
- Różnica mas: $+1{,}0$ g (95% CI: $-0{,}6$ do $2{,}6$) → obejmuje 0 → nieistotna
- CI niesie **więcej** informacji niż samo p: pokazuje też precyzję i kierunek

Istotność statystyczna a biologiczna

To nie to samo

Przy $n = 20\,000$ na grupę nawet różnica ekspresji o $0{,}1\%$ *może* okazać się istotna ($p < 0{,}05$), jeśli zmienność jest mała, choć biologicznie jest bez znaczenia. Przy $n = 20$ realny, duży efekt może nie osiągnąć istotności.

> Dlatego samo p nie wystarcza: zapis wyniku musi pokazywać i istotność, i wielkość efektu (przykład w części 5, na wydruku testu z R).

Ćwiczenie: interpretacja wyników

Dla każdego wyniku sformułować wniosek:

- $p = 0{,}03$ przy $n = 8$ na grupę, $d = 1{,}2$.
- $p = 0{,}09$ przy $n = 8$ na grupę; 95% CI różnicy: od $-0{,}5$ do $6{,}5$ g.
- $p = 0{,}003$ przy $n = 20\,000$ na grupę, $d = 0{,}03$.
- Z 20 analiz podgrup jedna dała $p = 0{,}04$.

Ćwiczenie: interpretacja wyników

Dla każdego wyniku sformułować wniosek:

- $p = 0{,}03$ przy $n = 8$ na grupę, $d = 1{,}2$. $\rightarrow$ **istotne, efekt duży**, ale oszacowanie niepewne przy małej próbie
- $p = 0{,}09$ przy $n = 8$ na grupę; 95% CI różnicy: od $-0{,}5$ do $6{,}5$ g. $\rightarrow$ **brak podstaw do odrzucenia** H_0; szeroki przedział dopuszcza i brak efektu, i duży efekt
- $p = 0{,}003$ przy $n = 20\,000$ na grupę, $d = 0{,}03$. $\rightarrow$ **istotne, lecz efekt znikomy**
- Z 20 analiz podgrup jedna dała $p = 0{,}04$. $\rightarrow$ **wynik zgodny z oczekiwaną liczbą fałszywych alarmów** ($20 \cdot 0{,}05 = 1$); bez korekty nie stanowi dowodu

5. W praktyce: R

Test t: wartość p i przedział ufności w wydruku

Wzrost siewek kontrolnych i nawożonych (cm):

```r
kontrola <- c(12.1, 11.8, 13.0, 12.4, 11.5, 12.9)
nawozone <- c(14.2, 15.1, 13.8, 14.9, 15.4, 14.0)
t.test(nawozone, kontrola)
```

```
    Welch Two Sample t-test

data:  nawozone and kontrola
t = 6.3153, df = 9.9232, p-value = 9.036e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.476892 3.089775
sample estimates:
mean of x mean of y
 14.56667  12.28333
```

Jak to czytać

- `t`, `df`, `p-value`: statystyka testowa, stopnie swobody, wartość p (`e-05` oznacza $\cdot 10^{-5}$, czyli $p < 0{,}001$) → odrzucamy H_0
- `alternative`: test dwustronny (H_1: różnica $\neq 0$)
- Przedział ufności [1,48; 3,09] cm nie zawiera 0
- Różnica średnich: ok. 2,3 cm

Planowanie liczebności przed eksperymentem

```r
power.t.test(delta = 2, sd = 1.5,
             power = 0.80, sig.level = 0.05)
```

```
     Two-sample t test power calculation

              n = 9.889068
          delta = 2
             sd = 1.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Jak to czytać

Aby z prawdopodobieństwem 80% wykryć różnicę 2 cm przy odchyleniu 1,5 cm (test dwustronny, `alternative = two.sided`), potrzeba ok. 10 osobników na grupę: `NOTE` informuje, że `n` dotyczy każdej grupy. Liczy się to przed zbieraniem danych.

Kompletny zapis wyniku

Komplet, który należy podać

Różnica średnich 2,28 cm, 95% CI [1,48; 3,09], test t (wariant Welcha, W4) $t(9{,}9) = 6{,}32$, $p < 0{,}001$, $n = 6$ na grupę.

- **Wielkość efektu**: różnica średnich w jednostkach pomiaru (lub d)
- **Niepewność**: 95% przedział ufności dla tej różnicy
- **Test**: nazwa testu, statystyka i stopnie swobody oraz dokładna wartość p, do 2–3 miejsc po przecinku (poniżej 0,001: $p < 0{,}001$), a nie tylko „$p < 0{,}05$”
- **Liczebność**: n w każdej grupie

Czego unikać

Samego „istotne” lub „$p < 0{,}05$” bez wielkości efektu i przedziału ufności. Taki zapis nie pozwala ocenić ani wielkości, ani precyzji wyniku.

Podsumowanie

Do zapamiętania

- Schemat testu jest **zawsze ten sam** — zmienia się tylko wybór testu (krok 3)
- Wartość p mówi o **danych przy założeniu** H_0, nie o prawdziwości hipotezy
- **Błąd I** = fałszywy alarm (α); **błąd II** = przegapiony efekt (β); moc $= 1 - \beta$ dla określonego efektu
- α ustala badacz przed analizą: $p \leq \alpha$ oznacza odrzucenie H_0, a $p > \alpha$ brak podstaw do jej odrzucenia
- Wiele testów zwiększa ryzyko fałszywego alarmu; sama istotność nie mówi o wielkości efektu
- Analizę mocy wykonuje się **przed** eksperymentem — także z powodów etycznych
- Estymator to reguła, oszacowanie to liczba; własności estymatora opisują jego rozkład z próbkowania
- 95% przedział ufności jest własnością metody; CI różnicy bez 0 odpowiada $p < 0{,}05$
- Istotność statystyczna $\neq$ biologiczna — raportować efekt + CI + p + n

Na następny wykład

Wykład 4: Testy statystyczne I (testy t)

- Od schematu do **konkretnego testu**
- Rozkład t, stopnie swobody i wartość krytyczna
- Test jednej próby i dwóch prób niezależnych
- Czytanie wyniku funkcji `t.test()` w R

Dziękuję za uwagę

dr inż. Bartosz Czech · bartosz.czech@upwr.edu.pl