



# Wydział Matematyki i Nauk Informatycznych

POLITECHNIKA WARSZAWSKA

## RPiESM, Omówienie laboratorium 3

Julian Zalewski, Antoni Zasada

21 maja 2025

**Zadanie 1.** Ornitolog, badający określony gatunek, pobrał próbę losową 10 dorosłych ptaków i zmierzył ich wagę, otrzymując następujące wyniki (w kg): 5.21, 5.15, 5.20, 5.48, 5.19, 5.25, 5.09, 5.17, 4.94, 5.11. Zakładamy, że waga ptaków badanego gatunku ma rozkład normalny.

- Utworzyć 95% przedział ufności dla średniej wagi ptaków badanego gatunku.
- Czy na poziomie istotności 0,05 można stwierdzić, że średnia waga ptaków badanego gatunku wynosi 5,15 kg?
- Czy na poziomie istotności 0,05 można stwierdzić, że średnia waga ptaków badanego gatunku jest mniejsza niż 5,20 kg?
- Z jakim prawdopodobieństwem test, przeprowadzony w pkt. c), przyjmie na poziomie istotności 0,05 hipotezę, że średnia waga ptaków badanego gatunku jest mniejsza niż 5,20 kg, w sytuacji, gdy w rzeczywistości ta średnia waga wynosi 5,15 kg?
- Ile by musiała wynosić średnia waga ptaków tego gatunku, by test z pkt. c) z prawdopodobieństwem 0,8, na poziomie istotności 0,05, przyjmował hipotezę, że średnia waga jest mniejsza niż 5,20 kg?
- Założmy, że rzeczywista średnia waga ptaków jest równa 5,15 kg. Wyznaczyć minimalną liczbę próby, która zagwarantuje, że test na poziomie istotności 0,05, z prawdopodobieństwem nie mniejszym niż 0,8, będzie przyjmował hipotezę, że średnia waga jest mniejsza niż 5,20 kg.
- Utworzyć 95% przedział ufności dla wariancji wagi ptaków badanego gatunku.
- Utworzyć 95% przedział ufności dla odchylenia standardowego wagi ptaków badanego gatunku.
- Czy na poziomie istotności 0,05 można stwierdzić, że odchylenie standardowe wagi ptaków badanego gatunku wynosi 0,20 kg?

**Rozwiązanie:** Wektor `birds` tworzymy poprzez skopiowanie wartości wag z treści zadania i połączenie ich za pomocą funkcji `c`.

- Przedział ufności dla średniej wagi tworzymy korzystając z funkcji `t.test` i podobnie jak na poprzednich laboratoriach przedział ufności pobieramy za pomocą `$conf`. Otrzymujemy `[5.08; 5.28]`.
- Ponownie używamy funkcji `t.test`. Hipotezą zerową jest  $\mu = \mu_0 = 5.15$ , a alternatywną  $\mu \neq \mu_0$ . Jako argumenty przekazujemy wektor wartości, wartość średniej oraz hipotezę alternatywną - w tym przypadku `two.sided`, ponieważ w naszym przypadku hipoteza alternatywna nie rozróżnia, czy waga jest mniejsza, czy większa niż  $\mu_0$ .

Sytuacja w tym podpunkcie wpisuje się w model II (zmienna losowa dana rozkładem normalnym, nieznane  $\mu, \sigma$  i hipoteza zerowa  $\mu = \mu_0$ ). Za pomocą funkcji `t.test` wyznaczamy p-value, które wynosi około 0.5188 - więcej niż przyjęty poziom istotności, czyli nie ma podstaw do odrzucenia hipotezy zerowej.

- Nadal korzystamy z modelu II, tym razem hipotezą alternatywną jest  $\mu \geq 5.2$ . Zmieniamy zatem parametr funkcji `t.test` na `less` i ponownie sprawdzamy p-value. Otrzymujemy p-value równe około 0.7, czyli zdecydowanie więcej niż  $\alpha$  - więc nie ma podstaw do odrzucenia hipotezy zerowej.
- W tym podpunkcie chcemy obliczyć wartość

$$P(\text{odrzucaamy } H_0 \mid \mu = 5.15).$$

Używamy do tego `power.t.test` (obliczamy moc testu), jako argumenty podajemy kolejno: liczbę obserwacji, różnicę średnich ( $\delta = 5.2 - 5.15$ ), typ testu, hipotezę alternatywną taką samą jak w poprzednim podpunkcie oraz poziom istotności równy 0.05. Uzyskane prawdopodobieństwo wynosi około 0.28.

- e) Korzystamy jeszcze raz z funkcji `power.t.test`, z taką różnicą, że tym razem dana jest moc testu równa 0.8, natomiast nieznana jest średnia waga ptaków. Jako parametry funkcji wstawiamy zatem znaną wartość `power`, a w jednym z pól wyniku otrzymujemy  $\delta \approx 0.12$ , czyli rzeczywista średnia waga musiałaby wynosić  $\mu_0 - \delta \approx 5.08$  kg.
- f) Podpunkt ten rozwiązujemy analogicznie do poprzedniego, tym razem dana jest moc testu i  $\delta = 5.2 - 5.15 = 0.05$ , a szukamy liczności próby  $n$ . Obliczenia zwracają  $n \approx 47.51533$ , zatem minimalna potrzebna liczba obserwacji wynosi 48.
- g) Do wyznaczenia przedziału ufności dla wariancji danej populacji korzystamy z funkcji `sigma.test` z biblioteki `TeachingDemos`. Argumenty są podobne jak w przypadku użytej wcześniej funkcji `t.test`. Wyznaczony przedział to  $[0.00882574; 0.06217251]$ .
- h) Aby obliczyć 95% przedział ufności odchylenia standardowego wagi ptaków należy wziąć pierwiastki z obu krańców wyznaczonego w poprzednim punkcie przedziału ufności dla wariancji - wynik to  $[0.09394541; 0.24934416]$ .
- i) Hipoteza zerowa:  $\sigma = 0.2$ , alternatywna:  $\sigma \neq 0.2$ . Wykonujemy `sigma.test` z parametrem  $\sigma = 0.2$ . Uzyskane p-value wynosi 0.2041 - więcej niż  $\alpha = 0.05$ , więc nie mamy podstaw do odrzucenia hipotezy zerowej.

Pełny kod rozwiązania:

```
1 rm(list = ls())
2
3 alpha = 0.05
4
5 birds = c(5.21, 5.15, 5.20, 5.48, 5.19, 5.25, 5.09, 5.17, 4.94, 5.11)
6
7 range = t.test(birds, conf.level=1-alpha)$conf
8
9 t = t.test(birds, mu=5.15, alt="two.sided")
10 print(t$p.value)
11
12
13 t = t.test(birds, mu=5.2, alt="less")
14 print(t$p.value)
15
16 power.t.test(n=length(birds), delta=5e-2, sd=sd(birds),
17             type="one.sample", alternative="one.sided", sig.level=alpha)$power
18
19
20 power.t.test(n=length(birds), power=8e-1, sd=sd(birds), type="one.sample", alternative="
21             one.sided", sig.level=alpha)
22
23 power.t.test(delta=5e-2, power=8e-1, sd=sd(birds), type="one.sample", alternative="one.
24             sided", sig.level=alpha)
25
26 library(TeachingDemos)
27 sigma.test(birds, conf.level=1-alpha)$conf
28 sqrt(sigma.test(birds, conf.level=1-alpha)$conf)
29 sigma.test(birds, sigma=2e-1)
```

**Zadanie 2.** W kolumnie `WeightInitial` w pliku `goats.txt` zapisano wagę (w kg) losowo wybranych młodych kóz hodowanych w Australii. Wiadomo, że rozkład badanej cechy jest normalny.

- a) Utworzyć 95% przedział ufności dla wartości oczekiwanej wagi młodych kóz hodowanych w Australii.
- b) Na poziomie istotności 0,05 przetestować hipotezę, że średnia waga młodych kóz hodowanych w Australii przekracza 23 kg.
- c) Zakładając, że rzeczywista średnia waga młodych kóz hodowanych w Australii wynosi 24 kg, wyznaczyć prawdopodobieństwo, że przeprowadzając test na poziomie istotności 0,05 i na podstawie 40 obserwacji, błędnie uznamy, że średnia waga takich kóz nie przekracza 23 kg.
- d) Załóżmy, że rzeczywista średnia waga młodych kóz hodowanych w Australii wynosi 24 kg. Ile trzeba by zbierać pomiarów wag takich kóz, by test (przeprowadzony na poziomie istotności 0,05) wykrywał, z prawdopodobieństwem nie mniejszym niż 0,8, że średnia waga takich kóz przekracza 23 kg?
- e) Utworzyć 90% przedział ufności dla wariancji wagi młodych kóz hodowanych w Australii.
- f) Czy można przyjąć, że wariancja wagi młodych kóz hodowanych w Australii wynosi 20 kg<sup>2</sup>? Zweryfikować odpowiednią hipotezę na poziomie istotności 0,1.
- g) Na poziomie istotności 0,1 zweryfikować hipotezę, że odchylenie standardowe wagi młodych kóz hodowanych w Australii przekracza 3 kg.

**Zadanie 3.** Pełnomocnik rządu Alfalandii ds. równego statusu kobiet i mężczyzn podejrzewa, że udział mężczyzn wśród pracowników przedszkoli jest niższy niż minimum przewidziane w ustawie wynoszące 35%.

- a) Czy na poziomie istotności 0,05 można uznać to stwierdzenie za uzasadnione, jeśli wśród losowo zbadanych 400 pracowników przedszkoli było 128 mężczyzn?
- b) Utworzyć 95% przedział ufności dla odsetka mężczyzn wśród pracowników przedszkoli wykorzystując wyniki badania z punktu a).
- c) Czy odpowiedź uzyskana w pkt. a) zmieniłaby się, gdyby pełnomocnik pobrał reprezentatywną próbkę 10 pracowników przedszkoli i 3 z nich okazałoby się mężczyznami?
- d) Utworzyć 95% przedział ufności dla odsetka mężczyzn wśród pracowników przedszkoli dla danych z punktu c).

**Zadanie 4.** Badano staż pracy osób zatrudnionych w pewnej dużej sieci handlowej. Na 150 losowo wybranych pracowników, 118 pracowało w tej sieci mniej niż 5 lat. Czy na tej podstawie można twierdzić, że 80% pracowników tej sieci legitymuje się stażem pracy mniejszym niż 5 lat? Zweryfikować odpowiednią hipotezę przyjmując poziom istotności 0,05.

**Zadanie 5.**

- a) Napisać funkcję, która dla dużej próby losowej ( $n \geq 100$ ) z dowolnego rozkładu, zwraca przedział ufności dla średniej na zadanym poziomie ufności. Zadbać, by funkcja zwracała błąd w przypadku jej użycia do próby o liczności mniejszej niż 100.
- b) W pakiecie MASS znajduje się zbiór danych `geyser` zawierający kolumnę `duration` z czasami trwania (w minutach) wybuchów gejzeru Old Faithful w Parku Narodowym Yellowstone w USA. Na poziomie ufności 0,95 wyznaczyć przedział ufności dla średniego czasu trwania wybuchu tego gejzera.

## Uwagi

- Różne funkcje mogą różnie reagować na wartość NA. `t.test` domyślnie je ignoruje, `sigma.test` zwróci NA jako wynik, podobnie jak `min`, `max`, `mean`. Niektóre funkcje posiadają parametry do usuwania wartości NA - `min`, `max`, `mean` posiadają parametr `na.rm`, który może przyjąć wartości TRUE albo FALSE. Warto też wspomnieć funkcje:
  - `na.omit` - usuwa wartości NA
  - `na.fail` zwróci dany obiekt o ile ten nie posiada żadnych brakujących wartości