W07. Statystyczna analiza danych psychologicznych

Psychoinformatyka | Wykład 07

2026-11-03

Cele i plan wnioskowania

- Dopasować metodę do pytania badawczego, jednostki analizy i schematu próby.
- Rozróżnić opis danych od wnioskowania o populacji.
- Interpretować łącznie oszacowanie, niepewność, wielkość efektu i wartość p.

Przykład: różnica czasu reakcji między warunkami zgodnym i niezgodnym w zadaniu Stroopa.

Trzy rodzaje pytań badawczych

- Różnica: czy warunek niezgodny wydłuża średni czas reakcji?
- Związek: czy większe obciążenie koreluje z wynikiem zadania?
- Predykcja: na ile dokładnie przewidujemy wynik nowej osoby?

Każde pytanie określa inny cel analizy i wymaga odpowiedniej walidacji.

Populacja, próba i jednostka analizy

- Populacja docelowa: zbiór osób, na które chcemy uogólniać.
- Próba: osoby faktycznie biorące udział w badaniu.
- Poziom obserwacji: próba zadaniowa, sesja, uczestnik lub grupa.

30 osób po 40 prób = 1200 obserwacji, ale nadal 30 uczestników.

EDA: kontrola przed testowaniem

Exploratory Data Analysis obejmuje:

1. strukturę zbioru i typy zmiennych;
2. braki, duplikaty i wartości niemożliwe;
3. rozkłady, wartości odstające i relacje między zmiennymi;
4. zgodność liczby uczestników i prób z protokołem.

Wyniki EDA nie zastępują uprzednio zdefiniowanej analizy potwierdzającej.

Miary opisowe: centrum i rozproszenie

- Średnia: wrażliwa na obserwacje skrajne.
- Mediana: odporna na silne ogony rozkładu.
- Odchylenie standardowe (SD): opisuje zmienność obserwacji.
- Rozstęp międzykwartylowy (IQR): szerokość centralnej połowy rozkładu.

Raportuj jednostki: np. średni czas reakcji [ms].

Asymetria czasu reakcji

- Czasy reakcji mają naturalną dolną granicę i często długi prawy ogon.
- W pojedynczej próbie możliwe są przedwczesne odpowiedzi i długie opóźnienia.
- Nie usuwaj punktów tylko dlatego, że zwiększają wariancję.

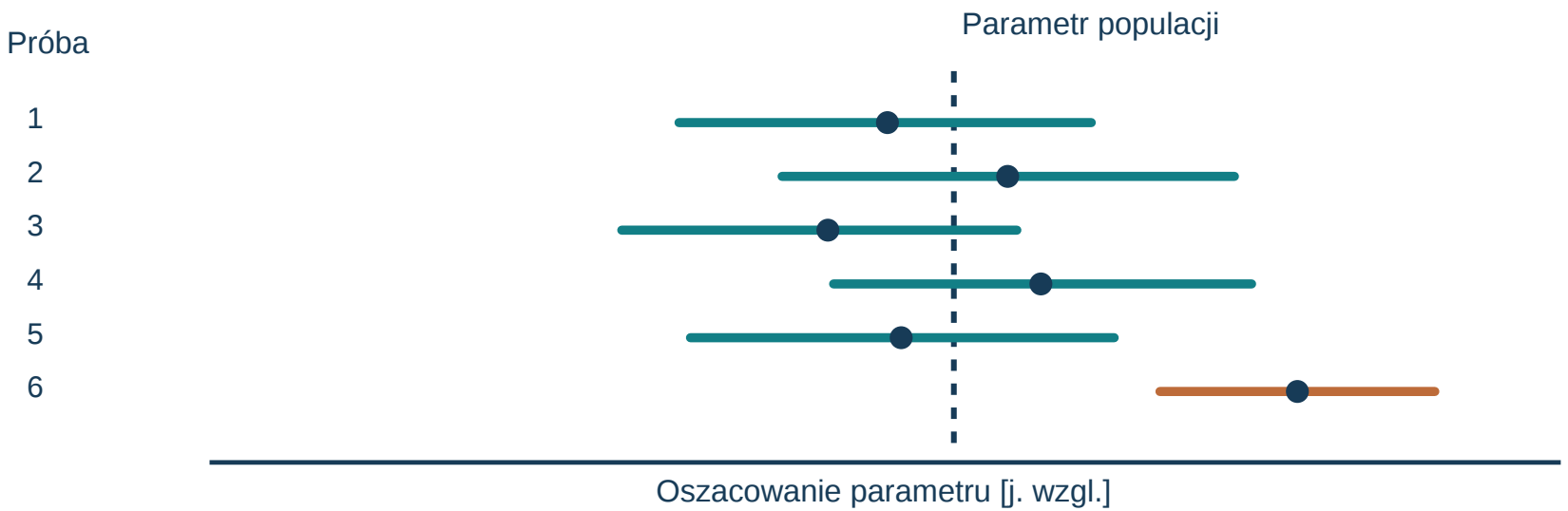
Kontroluj rozkład różnic wewnątrzosobniczych, gdy testujesz różnicę sparowaną.

Wizualizacja danych sparowanych

- Histogram: kształt rozkładu i ogony.
- Boxplot: mediana, kwartyle i potencjalne obserwacje odstające.
- Wykres sparowany: linia łączy dwie obserwacje tej samej osoby.

Dla testu Stroopa kluczowy jest rozkład różnic, a nie samo porównanie słupków średnich.

SD, SE i niepewność próbkowania



Rysunek 1: Wielokrotne przedziały z prób przy stałym parametrze

- SD: zmienność wyników w próbie.
- SE: zmienność estymatora między hipotetycznymi powtórzeniami próbkowania.
- Przy niezależnych obserwacjach: (SE_{x}=s/).

Przedział ufności jako estymacja

Dla średniej sparowanych różnic (d_i) i założeń testu t:

[CI_{95%}=dt_{0.975,,n-1}]

- Pokazuje precyzję oszacowania w jednostkach badanej zmiennej.
- Procedura o nominalnym poziomie 95% ma 95% pokrycia w powtarzanych próbach przy spełnieniu założeń.

Hipoteza zerowa i alternatywna

- H_0: średnia populacyjna różnica czasów reakcji jest równa zero.
- H_1: średnia populacyjna różnica jest różna od zera.
- Test dwustronny: decyzję o kierunku interpretacji podejmujemy po obejrzeniu oszacowania.

Zdefiniuj kontrast: (d_i=RT_{,i}-RT_{,i}).

Wartość p: poprawna interpretacja

Wartość p to prawdopodobieństwo wyniku co najmniej tak ekstremalnego jak zaobserwowany, zakładając H_0 i model testu.

Wartość p nie jest:

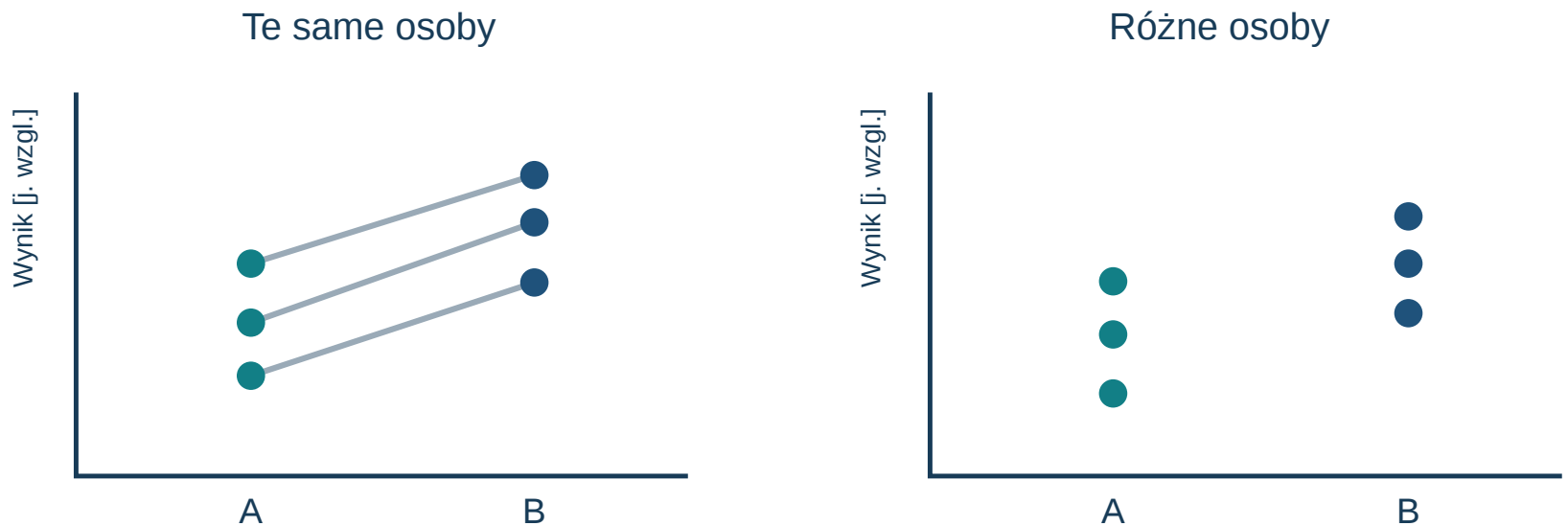
- prawdopodobieństwem prawdziwości H_0;
- wielkością efektu;
- miarą prawdopodobieństwa replikacji wyniku.

Wielkość efektu: średnia różnica i d_z

- Efekt w jednostkach: (d) w milisekundach.
- Standaryzacja dla par: (d_z=d/s_d).
- Wartość (d_z) zależy od zmienności różnic wewnątrzosobniczych.

Przedział ufności średniej różnicy odpowiada bezpośrednio na pytanie o precyzję efektu w ms.

Próby sparowane i niezależne



Rysunek 2: Porównanie schematów sparowanego i niezależnego

- Sparowane: ten sam uczestnik dostarcza oba pomiary.
- Niezależne: pomiary w grupach pochodzą od różnych uczestników.

Schemat badania decyduje o właściwej strukturze błędu standardowego.

Dobór metody do struktury danych

Cel i schemat	Przykładowa metoda
Dwa warunki, te same osoby	test t sparowany / Wilcoxon
Dwie niezależne grupy	test t Welcha
Więcej warunków powtarzanych	ANOVA powtarzanych pomiarów
Zależności i powtarzane próby	model mieszany

Dobór poprzedza kontrola założeń i estymandu.

Badanie Stroopa: plan analizy

Projekt: 30 osób wykonuje próby zgodne i niezgodne.

1. Wykluczenia według uprzednio zdefiniowanych kryteriów jakości.
2. Średni RT na osobę × warunek.
3. Różnica (d_i) dla każdej osoby.
4. Oszacowanie (d), przedziału ufności i testu sparowanego.

Dane demonstracyjne w dalszej części będą syntetyczne.

5 minut przerwy

Pytania do samodzielnego sprawdzenia

1. Czym różni się odchylenie standardowe od błędu standardowego?
2. Jak interpretować 95-procentowy przedział ufności?
3. Dlaczego w zadaniu Stroopa wybieramy test sparowany?
4. Czy p = 0,001 oznacza, że H_0 ma 0,1% szansy być prawdziwa?

Agregacja prób i różnice wewnątrz osób

- Najpierw oblicz (RT_{i,A}) i (RT_{i,B}) dla każdej osoby.
- Następnie (d_i=RT_{i,B}-RT_{i,A}).
- Analizuj rozkład 30 różnic, nie wszystkie pojedyncze próby jak niezależne osoby.

Kryteria agregacji (średnia, mediana, model prób) wybierz zgodnie z pytaniem badawczym.

Test t-Studenta dla par

[t=,df=n-1]

- Licznik: oszacowana średnia różnica względem H_0.
- Mianownik: jej błąd standardowy.
- Dodatnie t przy konwencji (B-A) oznacza wyższy RT w warunku B.

Przy analizie par nie stosuj testu t dla niezależnych prób.

Założenia i analiza wrażliwości

- Niezależność uczestników i właściwe parowanie.
- Rozkład różnic adekwatny do użycia testu t, szczególnie przy małej próbie.
- Brak dominacji kilku skrajnych różnic.
- Stabilność wyniku po uzasadnionych wariantach kontroli jakości.

Analiza wrażliwości: porównaj oszacowanie ze skrajną obserwacją i bez niej; usunięcie musi mieć uzasadnienie niezależne od wartości p.

Wykres QQ i wykres różnic pomagają diagnozować rozkład, ale nie zastępują rozumowania o projekcie.

Test Wilcoxona dla różnic sparowanych

- Wykorzystuje rangi niezerowych różnic sparowanych.
- Testuje symetrię rozkładu różnic wokół zera w typowej interpretacji.
- Nie jest automatycznie testem średniej różnicy.

W przykładzie zestawiamy wyniki testu t i Wilcoxona, zachowując różnice w ich hipotezach i założeniach.

Dobór zależy od pytania: średnia różnica i test t nie zawsze mają ten sam cel co test rang.

Więcej niż dwa warunki: ANOVA

- ANOVA powtarzanych pomiarów: te same osoby badane w kilku warunkach.
- Ocenia, czy średnie warunków różnią się globalnie.
- Następnie analizuje się zaplanowane kontrasty i ich niepewność.

Kontroluj założenia, w tym sferyczność dla odpowiednich efektów wielopoziomowych, lub stosuj uzasadnione korekty.

Korelacja Pearsona i Spearmana

- Pearson r: siła i kierunek zależności liniowej.
- Spearman ρ: monotoniczne uporządkowanie rang.
- Sprawdź wykres rozrzutu, obserwacje wpływowe i zakres zmienności.

Korelacja między uczestnikami nie musi opisywać zależności wewnątrz tej samej osoby.

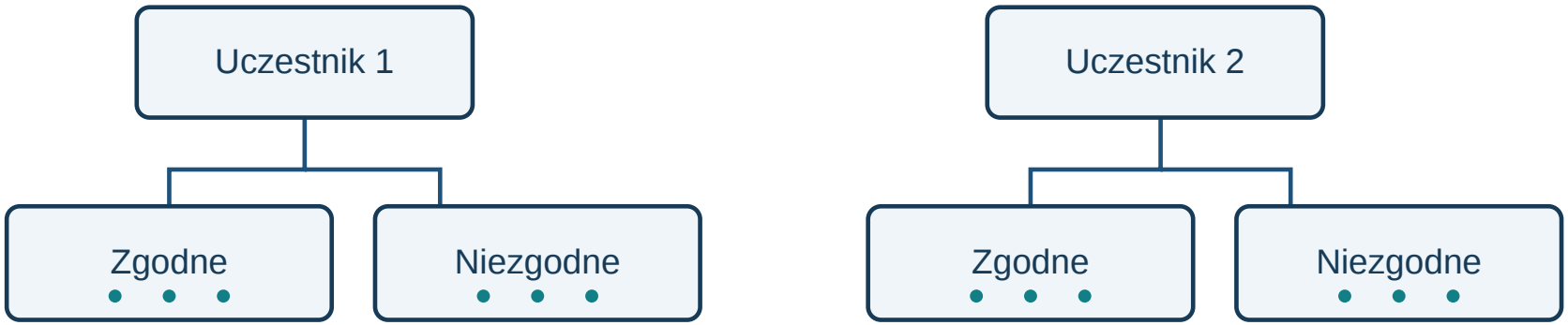
Regresja: oszacowanie zależności

[Y_i=_0+_1X_i+_i]

- Współczynnik (_1): oczekiwana zmiana Y dla jednostkowej zmiany X w modelu.
- Analizuj reszty, współliniowość i obserwacje wpływowe.
- Uwzględniaj zmienne kontekstowe zgodnie z teorią i projektem.

Regresja obserwacyjna nie ustanawia przyczynowości.

Modele mieszane i dane zagnieżdżone



Rysunek 3: Zagnieżdżenie prób w osobach

- Efekt stały: średni kontrast warunków w populacji.
- Efekt losowy: różnice między uczestnikami w poziomie bazowym lub efekcie warunku.
- Dobór modelu wynika z projektu i liczby dostępnych niezależnych jednostek.

Python: syntetyczne dane Stroopa

```
import numpy as np
from scipy import stats

rng = np.random.default_rng(42)
n = 30
zgodne = rng.normal(520, 60, n)
roznice = rng.normal(65, 40, n)
niezgodne = zgodne + roznice
```

Jednostką tablic jest uczestnik, nie pojedyncza próba zadania. Wartości w ms zostały wygenerowane syntetycznie.

Python: rozkład sparowanych różnic

```
import matplotlib.pyplot as plt

d = niezgodne - zgodne
fig, ax = plt.subplots(figsize=(6, 4))
ax.hist(d, bins=8, edgecolor="black")
ax.axvline(d.mean(), linestyle="--", label="średnia")
ax.set(xlabel="Niezgodne − zgodne [ms]", ylabel="Liczba osób",
       title="Różnice czasu reakcji: dane syntetyczne")
ax.legend()
plt.show()
```

Rozkład dotyczy 30 różnic wewnątrzosobniczych.

Python: test sparowany, CI i d_z

```
wynik = stats.ttest_rel(niezgodne, zgodne)
sd = d.std(ddof=1)
se = sd / np.sqrt(n)
ci = stats.t.interval(0.95, df=n-1, loc=d.mean(), scale=se)
dz = d.mean() / sd

print(f"Różnica: {d.mean():.1f} ms; 95% CI {ci}")
print(f"t({n-1}) = {wynik.statistic:.2f}; p = {wynik.pvalue:.3g}")
print(f"d_z = {dz:.2f}")

wilcoxon = stats.wilcoxon(d)
d_skrajne = d.copy()
d_skrajne[0] += 400  # celowo wprowadzona obserwacja skrajna
print(f"Wilcoxon p = {wilcoxon.pvalue:.3g}")
print(f"Średnia przed/po skrajnej: {d.mean():.1f} / {d_skrajne.mean():.1f} ms")
```

Raportuj oszacowanie i 95% CI nawet wtedy, gdy test nie przekracza progu istotności.

Raportowanie wyniku i jego ograniczeń

W raporcie należy podać:

- liczbę uczestników i liczebność użytych prób na warunek;
- definicję różnicy oraz jej oszacowanie [ms];
- 95% CI, statystykę t, stopnie swobody, p i miarę efektu;
- kryteria wykluczeń, założenia i zakres uogólnienia.

Dla danych syntetycznych nie formułuj wniosku o rzeczywistych procesach poznawczych.

Wielokrotne testy i rzetelność analizy

- P-hacking: wybieranie analiz pod pożądany wynik po obejrzeniu danych.
- Wielokrotne testowanie: kontrola rodziny hipotez i jawne raportowanie kontrastów.
- Analiza prerejestrowana: odrębna od analiz eksploracyjnych.
- Reprodukowalność: kod, wersje bibliotek, dane i reguły wykluczeń.

Transparentność pozwala ocenić wiarygodność wniosków.

Literatura naukowa i narzędzia

- Wasserstein, R. L., Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician, 70(2), 129–133. DOI.
- Cumming, G. (2012). Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. Routledge.
- Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. Behavior Research Methods, 44, 158–175. DOI.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18(6), 643–662. DOI.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4, 863. DOI.

Dokumentacja narzędzi: SciPy stats, NumPy, Matplotlib.

Wnioski i zadanie po wykładzie

- Najpierw projekt i estymand, następnie wykresy, założenia i właściwy test.
- Nie utożsamiaj p ze znaczeniem praktycznym ani SD z SE.
- W projekcie powtarzanych pomiarów zachowuj sparowanie i strukturę osób.
- Zmień ziarno generatora, porównaj CI i zastanów się, które wnioski są stabilne.