W04. Psychometria obliczeniowa

Psychoinformatyka | Wykład 04

2026-11-01

Cele wykładu

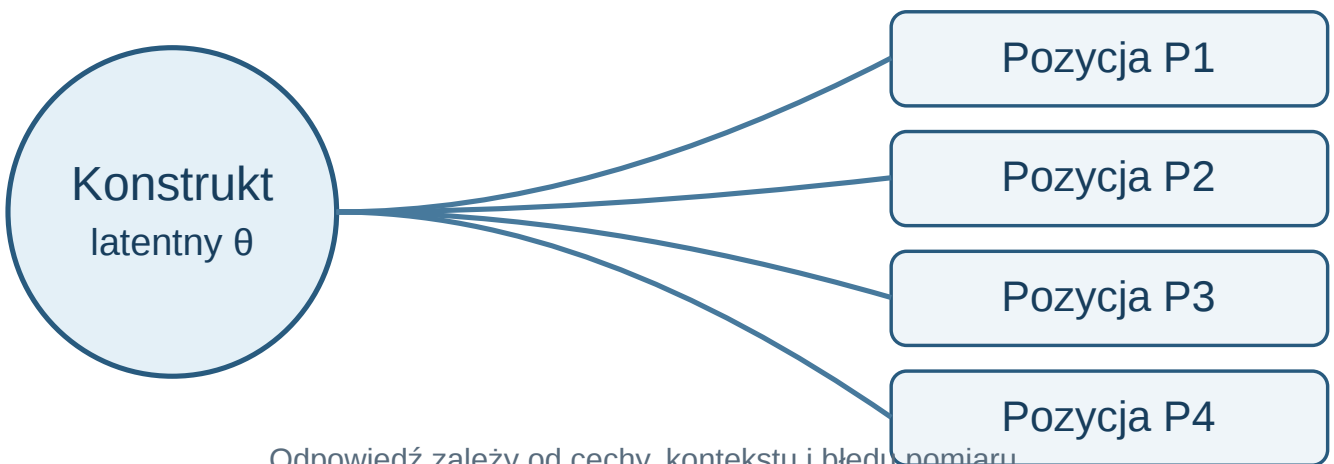
- Zrozumieć model pomiarowy: konstrukt → pozycje → wynik.
- Odróżnić spójność wewnętrzną, rzetelność i trafność.
- Rozpoznać zastosowania CTT, IRT i testowania adaptacyjnego.

Czym zajmuje się psychometria?

- Projektuje i ocenia narzędzia pomiaru cech latentnych.
- Bada własności wyników w określonej populacji i zastosowaniu.
- Wykorzystuje statystykę, modele probabilistyczne i analizę danych.

Cecha latentna a wskaźnik

- Stres jest konstruktem; pojedyncza odpowiedź to wskaźnik pośredni.
- Na odpowiedź wpływa cecha, kontekst, zrozumienie pytania i błąd.
- Model zakłada określony związek między zmienną ukrytą a obserwacjami.



Rysunek 1: Model pomiarowy: konstrukt latentny i pozycje kwestionariusza

Od teorii do skali

- Zdefiniuj konstrukt i docelową populację.
- Stwórz bank pozycji, przeprowadź ocenę treści i pilotaż.
- Zbierz dowody rzetelności, trafności i sprawiedliwości pomiaru.

Jak pisać pozycje?

- Jedna treść poznawcza lub emocjonalna na jedną pozycję.
- Unikaj podwójnych pytań, niejasnego okresu odniesienia i sugestii.
- Pozycje odwrócone mogą generować efekt metody, nie tylko redukować przytakiwanie.

Skala Likerta i dane porządkowe

- Odpowiedzi od 1 do 5 kodują uporządkowane kategorie.
- Odstępy między kategoriami nie muszą być psychologicznie równe.
- Modelowanie danych porządkowych wymaga jawnych założeń.

Odwracanie punktacji

- Dla zakresu od 1 do 5 stosujemy x_odw = 6 − x.
- Rekodowanie wykonujemy przed sumowaniem i analizą spójności.
- Braku danych nie wolno zamienić na zero podczas rekodowania.

$$x_{odw} = x_{min} + x_{max} - x = 6 - x$$

Klasyczna teoria testów (CTT)

- W modelu CTT obserwowany wynik X = T + E.
- T to oczekiwana wartość pomiaru przy powtarzaniu w określonym modelu.
- Nie obserwujemy bezpośrednio T; rozkład błędu zależy od założeń.

Wynik surowy i normy

- Suma pozycji jest wynikiem surowym zależnym od kodowania.
- Wynik standaryzowany wymaga grupy odniesienia i procedury normowania.
- Nie wolno określać wyniku jako „wysokiego klinicznie„ bez danych normatywnych.

Efekty podłogi i sufitu

- Skupienie odpowiedzi na minimum lub maksimum ogranicza rozdzielczość pomiaru.
- Rozkład wyników zależy od populacji i trudności pozycji.
- Średnia może maskować asymetrię i skupienia przy granicach.

Skorygowana korelacja pozycja–wynik

- Koreluj pozycję z sumą pozostałych pozycji.
- Słaba korelacja sygnalizuje możliwy problem, ale nie wyznacza automatycznie decyzji o usunięciu.
- Rozważ też treść pozycji i strukturę czynnikową.

Rzetelność i standardowy błąd pomiaru

- Rzetelność opisuje udział wariancji wyniku prawdziwego w wariancji obserwowanej pod założeniami CTT.
- SEM można oszacować jako $SD(X)\cdot\sqrt{(1-r)}$ w określonym modelu.
- Niepewność pojedynczego wyniku musi uwzględniać adekwatność oszacowania r.

Alfa Cronbacha

- Alfa wyliczamy z liczby pozycji oraz wariancji pozycji i ich sumy.
- Alfa zależy m.in. od liczby pozycji oraz współzmienności odpowiedzi.
- Wysoka alfa nie dowodzi jednowymiarowości ani trafności.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_X^2}\right)$$

Omega i modele kongeneryczne

- Omega wykorzystuje oszacowane ładunki czynnikowe i wariancje błędów.
- Może być odpowiedniejsza niż alfa przy nierównych ładunkach, jeśli model pomiarowy jest adekwatny.
- Nieprawidłowy model czynnikowy daje mylące omega.

Trafność interpretacji wyniku

- Dowody obejmują treść, strukturę wewnętrzną, relacje z innymi zmiennymi i konsekwencje stosowania.
- Trafność dotyczy konkretnej interpretacji i użycia wyniku, nie abstrakcyjnej etykiety testu.
- Spójność wewnętrzna stanowi tylko jeden element dowodów.

5 minut przerwy

Pytania do samodzielnego przemyślenia

1. Czym różni się rzetelność od trafności?
2. Dlaczego odwraca się punktację wybranych pozycji?
3. Co można, a czego nie można wnioskować z alfa = 0,85?
4. Czy wysoka alfa dowodzi jednowymiarowości skali?

EFA: odkrywanie struktury

- Eksploracyjna analiza czynnikowa bada wspólną strukturę współzmienności pozycji.
- Dobór liczby czynników wymaga uzasadnienia, np. analizy równoległej.
- Dla danych porządkowych rozważ odpowiednie korelacje i estymatory.

CFA: sprawdzanie modelu

- Konfirmacyjna analiza czynnikowa testuje z góry określony układ ładunków.
- Analiza obejmuje dopasowanie, reszty i interpretowalność parametrów.
- Dobre dopasowanie samo w sobie nie gwarantuje trafności.

Jednowymiarowość i zależność lokalna

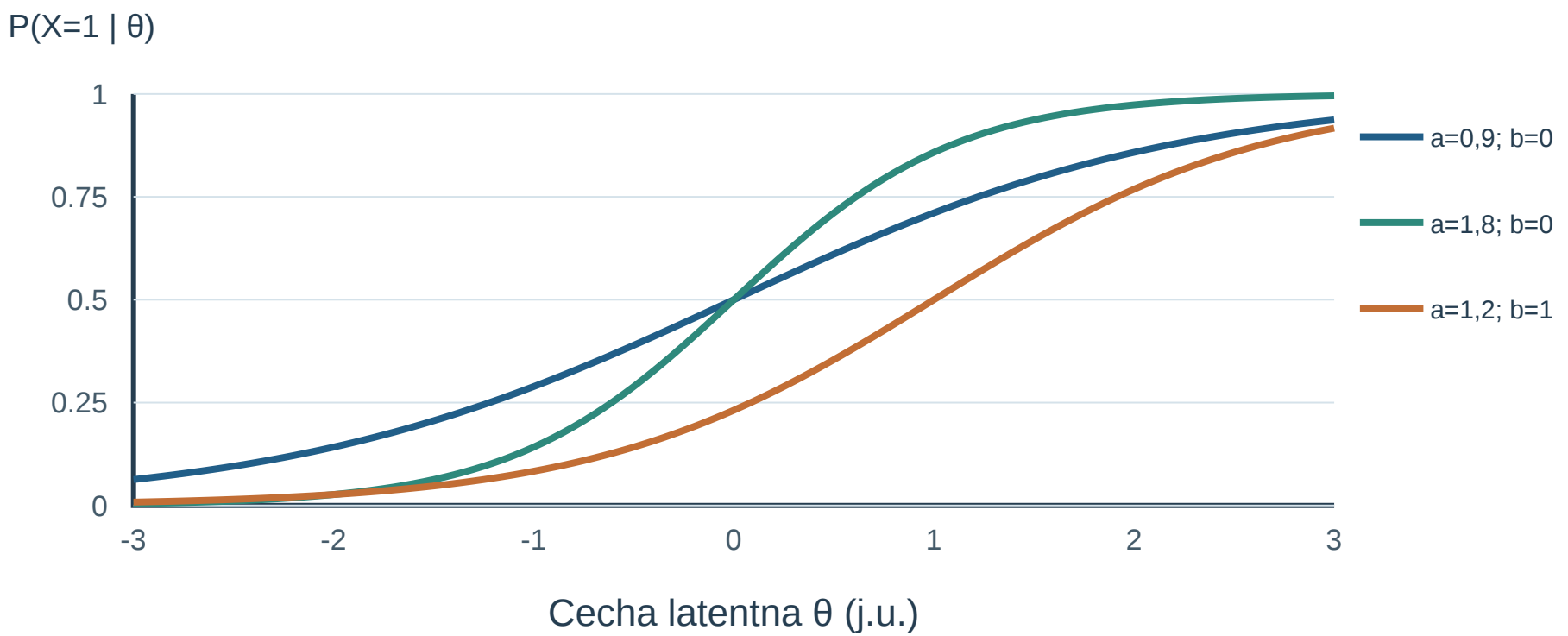
- Jednowymiarowość dotyczy liczby dominujących zmiennych latentnych.
- Lokalna niezależność oznacza brak zależności odpowiedzi po uwzględnieniu cechy w modelu.
- Korelowane błędy lub efekty metody mogą łamać założenia.

Teoria odpowiedzi na pozycję (IRT)

- IRT opisuje prawdopodobieństwo odpowiedzi jako funkcję cechy latentnej.
- Parametry pozycji i osoby są szacowane w wybranym modelu.
- Modele dla pozycji dychotomicznych i porządkowych różnią się.

Model logistyczny 2PL

- Dla odpowiedzi 0/1: $P(X=1 \mid \theta) = 1/(1+\exp[-a(\theta-b)])$.
- a to parametr dyskryminacji, b to lokalizacja/trudność na skali θ.
- Parametry zależą od identyfikacji i kalibracji modelu.



Rysunek 2: Krzywe 2PL dla pozycji binarnych (symulacja)

Komputerowe testowanie adaptacyjne (CAT)

- CAT dobiera kolejną pozycję na podstawie dotychczasowych odpowiedzi i niepewności θ.
- Bank pozycji wymaga kalibracji, kontroli ekspozycji i reguły zatrzymania.
- Porównywalność wyników wymaga sprawdzenia modelu i sprawiedliwości pomiaru.

Python: syntetyczne odpowiedzi

- Notebook W04 generuje odpowiedzi P1–P6 w skali 1–5 dla 240 fikcyjnych osób.
- P5 sformułowano odwrotnie, P6 celowo słabo wiąże się z konstruktem.
- Losowe braki pozwalają omówić regułę przypadków kompletnych.

```
rng = np.random.default_rng(42)
# Kompletny kod generowania: pi_w04.ipynb
```

Python: rekodowanie i suma

- P5_po = 6 − P5_surowa.
- Użyj kompletnego przypadku do porównania wariancji pozycji.
- Suma poprawnie zakodowanych sześciu odpowiedzi mieści się w [6,30].

```
kompletne = dane.dropna(subset=pozycje).copy()
kompletne['P5'] = 6 - kompletne['P5']
kompletne['wynik'] = kompletne[pozycje].sum(axis=1)
```

Python: alfa i analiza pozycji

- Oblicz alfa przy ddof=1 dla wariancji pozycji i sumy.
- Korelacja item–rest wyklucza badaną pozycję z sumy.
- Porównaj alfa z P6 i bez P6 bez automatycznej selekcji skali.

```
k = len(pozycje)
suma = kompletne[pozycje].sum(axis=1)
alfa = k/(k-1)*(1-kompletne[pozycje].var(ddof=1).sum()/suma.var(ddof=1))
```

Python: rozkłady i korelacje

- Histogram pokazuje rozkład sum skali na osi 6–30 pkt.
- Wykres item–rest przedstawia skorygowane korelacje dla P1–P6.
- Wszystkie osie i wartości odnoszą się do syntetycznych obserwacji.

```
# Skorygowana korelacja item–rest
p = 'P6'
r = kompletne[p].corr(kompletne[pozycje].drop(columns=p).sum(axis=1))
```

Python: charakterystyki 2PL

- Wygeneruj θ w zakresie od −3 do +3.
- Porównaj trzy krzywe P(X=1|θ) przy różnych a i b.
- Krzywe opisują przykładowe pozycje binarne, nie wyniki P1–P6.

```
# Prawdopodobieństwo odpowiedzi 1 w binarnym modelu 2PL
p = 1 / (1 + np.exp(-a * (theta - b)))
```

Raport psychometryczny

- Raportuj liczebność, braki, kodowanie, model oraz przedziały niepewności.
- Podaj rozkład wyników, wskaźniki spójności i dowody dotyczące struktury.
- Wyraźnie zaznacz ograniczenia interpretacji i populację odniesienia.

Stronniczość i niezmienniczość pomiaru

- Niezmienniczość bada porównywalność modelu między grupami lub czasem.
- DIF wskazuje różną charakterystykę pozycji po kontroli poziomu cechy.
- Prywatność i sprawiedliwość pomiaru są częścią projektu cyfrowego narzędzia.

Literatura i narzędzia

- AERA, APA, NCME (2014). Standards for Educational and Psychological Testing.
- Cronbach (1951). Coefficient alpha and the internal structure of tests.
- Embretson i Reise (2000). Item Response Theory for Psychologists; dokumentacja pandas/NumPy.

Wnioski i samodzielna praca

- Trafność interpretacji i rzetelność są odrębnymi własnościami narzędzia.
- Odwracaj kodowanie wskazanych pozycji przed obliczeniem wyniku skali.
- Wysoka alfa nie dowodzi jednowymiarowości; porównania grup wymagają kontroli porównywalności pomiaru.
- Obliczenia wykonaj w `pi_w04.ipynb` i odnieś do laboratorium L03.