

Metody badań społecznych

dr Michał Bujalski

ISNS UW

IPiN

Badania ilościowe i jakościowe

Badania ilościowe

- Nacisk na **wyjaśnienie, obiektywność**, standaryzację, pomiar.
- Duże próby badawcze, **losowy** dobór próby.
- Analiza statystyczna, wnioskowanie na podstawie danych ilościowych - **liczb**.
- **Testowanie** teorii, hipotez badawczych.
- Wyniki mogą być **uogólniane** na populację.
- **Oddzielenie** badacza od procesu badawczego.

Badania jakościowe

- Nacisk na **zrozumienie, interpretację**, pogłębioną analizę.
- Małe, **celowo** dobierane próby.
- Analiza tekstów, obserwacji, wywiadów indywidualnych lub grupowych, wnioskowanie na podstawie **znaczeń**.
- Budowanie teorii, **eksploracja** nowych zagadnień.
- Wyniki trudne do uogólnienia, zazwyczaj ograniczone do badanego **kontekstu**.
- Badacz znajduje się w środku procesu badawczego.



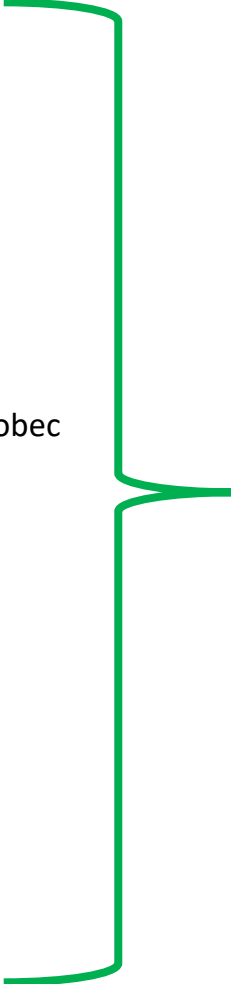
Typy badań ilościowych

- **Badania przekrojowe**

- Jednorazowy pomiar
- „Obraz” sytuacji w danym czasie
- Zastosowanie:
 - Określenie częstości występowania zjawiska
 - Badanie zależności między zmiennymi w danym momencie
 - Porównywanie różnych grup w tym samym czasie, np. postawy wobec prawa

- **Badania podłużne**

- Wielokrotne pomiary tej samej grupy
- Obserwacja zmian w czasie
- Zastosowanie:
 - Śledzenie trendów i zmian
 - Badanie procesów rozwojowych
 - Analiza związków przyczynowo-skutkowych

- 
- Badania panelowe (CAWI)
 - Badania telefoniczne (CATI)
 - Badania F2F (CAPI, P&P)

- Ilościowe badania dyskursu
- Ilościowe badania źródeł

Proces badawczy

1. Formułujemy **cel** badania

- np. ustalenie, czy i w jakim stopniu posiadanie profesjonalnego pełnomocnika wpływa na wynik postępowania w sprawach o eksmisję w sądach rejonowych w latach 2022-2023

2. Formułujemy **pytanie badawcze** na podstawie naszych obserwacji

- Czy istnieje związek między posiadaniem profesjonalnego pełnomocnika a korzystnym dla pozwanego rozstrzygnięciem w sprawach o eksmisję?

3. Formułujemy **hipotezy**, która przewiduje, jakiego układu wyników oczekujemy.

- Pozwani reprezentowani przez profesjonalnego pełnomocnika mają statystycznie wyższą szansę na korzystne rozstrzygnięcie (np. dłuższy okres na opuszczenie lokalu, przyznanie lokalu socjalnego, etc.).

4. Ustalamy taki **schemat badania**, który będzie najlepszy, by odpowiedzieć na zadane pytanie.

- Analiza ilościowa akt sądowych spraw o eksmisję
 - Dobór losowy warstwowy z uwzględnieniem wielkości miasta (małe/średnie/duże)
 - Okres badania: sprawy zakończone w latach 2022-2023
 - Wielkość próby: $n=300$ spraw z 3 różnych okręgów sądowych

5. Wybieramy **narzędzia** pomiaru

- np. arkusz kodowy zawierający informacje nt. reprezentacji prawnej (tak/nie, typ pełnomocnika), wyniku sprawy, przyznania lokalu socjalnego, czasu na opuszczenie lokalu, okoliczności, podstawowe dane demograficzne

6. Dokonujemy **pomiaru** - analiza akt sądowych, wprowadzenie danych do arkusza kalkulacyjnego

- Analizujemy dane, np. test Chi-kwadrat dla zbadania zależności między reprezentacją prawną a przyznaniem lokalu socjalnego lub/oraz rodzajem rozstrzygnięcia

7. Opisujemy **wyniki** ...

Zmienne

- Zmienna - każda właściwość, cecha, która może przybierać różne (przynajmniej dwie) wartości.
 - Mogą być obserwowalne i nieobserwowalne
 - Mogą być wyjaśniane i wyjaśniające
 - Mogą być zależne i niezależne

Wskaźniki i operacjonalizacja

- Wskaźnik
 - Obserwowalny i mierzalny przejaw zjawiska, na podstawie którego wnioskujemy o danej zmiennej, np. liczba spraw z art. XYZ.
- Operacjonalizacja
 - Projektowanie sposobu pomiaru zmiennej teoretycznej - ustalamy, co będzie wskaźnikiem zmiennej, np. liczba spraw zakończona wyrokiem w danym roku.

Jakościowe skale pomiaru zmiennych

- **Nominalne** - klasyfikacja
 - np. płeć, zawód
- **Porządkowe** (rangowe) - uporządkowanie od najmniejszych do największych
 - np. wykształcenie, wielkość msc
 - skala Likerta
 - Zdecydowanie s. zgadzam
 - Raczej się zgadzam
 - Ani się zgadzam ani się nie zgadza
 - Raczej się nie zgadzam
 - Zdecydowanie się nie zgadzam
- Skala nominalna
 - ✓ Można zobaczyć, czy X jest **różny** od Y
- Skala porządkowa
 - ✓ Można zobaczyć, czy X jest **wiekszy** lub **mniejszy** od Y

Ilościowe skale pomiaru zmiennych

- **Przedziałowe** - równe jednostki
 - np. temperatura lub współczynnik osób zatrudnionych w sądownictwie
- **Ilorazowe** - zero bezwzględne
 - np. dochód, liczba spraw w danym roku
- Skala przedziałowa
 - Można zobaczyć, **o ile** X jest większy lub mniejszy od Y.
- Skala ilorazowa
 - Można zobaczyć, **ile razy** X jest większy od Y

Rodzaje pomiaru/próby

- **Dobór losowy** - dobieramy osoby w sposób losowy odpowiednio do cech badanej populacji
 - **Prosty dobór losowy** - Każdy element populacji ma równe szanse wyboru
 - **Dobór warstwowy** - uwzględniamy warstwy losowania, np. województwo, płeć, grupy wieku, etc.
 - **Dobór systematyczny** - dobieramy n-tą osobę z operatu losowania.
- **Dobór nielosowy**
 - **Próba kwotowa** - szukamy osób spełniających określone warunki/założenia w taki sposób, aby był proporcjonalny do rzeczywistego rozkładu cech w populacji.
 - **Dobór celowy** - arbitralna decyzja badacza.
 - Próba **celowa** - szukamy osób spełniających określone warunki/założenia
 - Próba **uznaniowa** - **okolicznościowa** (*opportunity sample*) - dobór ze wzgl. na okoliczność, która może wiązać się przedmiotem badania, np. przystanek autobusowy, jeśli badamy opinie pasażerów WTP.
 - Próba **uznaniowa** - **dogodna** (*convenience sample*) - dobór z wzgl. na możliwość rekrutacji osób (np. ze wzgl. na cechę X), których doświadczenia są zróżnicowane.
 - **Kula śnieżna** - zwiększamy liczebność próby, rekrutując respondentów wskazanych przez os. badane.

Błąd próby

- **Przedział ufności (CI)** - odpowiada poziomowi istotności: $p < 0,05$ (lub $p < 0,01$; $p < 0,001$)
 - Przy $p < 0,05$ zostawiamy sobie 5% szans na to, że jest otrzymany wynik badanej zależności jest dziełem przypadku.
- **Błąd statystyczny / margines błędu (MoE)** - odpowiedzi w naszej próbie różnią się od populacji.
 - Im większa próba, tym lepiej, choć... błąd statystyczny nie maleje proporcjonalnie do wielkości próby (...tylko do pierwiastka liczebności).
 - Dla populacji Polski ($n = 37,64$ mln) - populacja nieskończona
 - Dla sądów rejonowych w Polsce ($n = 318$) - populacja skończona
 - Granica: 20 tys. / 50 tys. jednostek

Jaka wielkość próby?

	Sprawy w sądach w 2022 r. (n=14,6 mln)	Wnioski o tymczasowe aresztowanie w 2022 r. (n=2226)	Sądy rejonowe w 2022 r. (n=318)
95% CI; 3% MoE	1068	722	246
95% CI; 1% MoE	9598	1808	308
99 %CI; 3% MoE	1849	1011	272
99 %CI; 1% MoE	16623	1964	313

Miary tendencji centralnej

Modalna (M_o) - wartość, która **występuje najczęściej** w zbiorze danych. Jest to miara wykorzystywana dla danych jakościowych (nominalnych) lub danych liczbowych z wyraźnymi wartościami odstającymi, np. w zbiorze {2, 3, 5, 6, 6, 6, 8, 9} moda wynosi 6, ponieważ ta wartość pojawia się najwięcej razy.

Mediana (M_e) - to wartość środkowa, która **dzieli zbiór danych na dwie równe części**. Połowa obserwacji ma wartości mniejsze od mediany, a druga połowa większe. Jest to miara odporna na wartości odstające i często używana dla danych poziomu przedziałowego/ilorazowego, np. w zbiorze {5, 7, 9, 11, 16} mediana wynosi 9, ponieważ jest to środkowa wartość w posortowanym ciągu.

Średnia (M) - to suma wszystkich wartości podzielona przez ich liczbę. Jest to najbardziej popularna miara centralnej tendencji, ale jest wrażliwa na wartości odstające. Wykorzystywana dla danych poziomu ilorazowego, np. w zbiorze {2, 4, 5, 7, 12} średnia wynosi $(2+4+5+7+12)/5 = 6$.

Miary rozproszenia

Rozstęp - różnica między największą a najmniejszą wartością w zbiorze danych. Pokazuje maksymalny zakres zmienności, np. dla zbioru {3, 5, 7, 8, 12}, rozstęp wynosi $12 - 3 = 9$.

Wariancja - średni kwadrat odchyłeń poszczególnych wartości od średniej arytmetycznej całego zbioru. Im większa wariancja, tym bardziej dane są rozproszone wokół średniej, np. dla zbioru {5, 6, 8, 9, 12} ze średnią 8, wariancja wynosi

$$\Sigma(x - \text{średnia})^2 / n = (5-8)^2 + (6-8)^2 + \dots + (12-8)^2 / 5 = 6.$$

Odchylenie standardowe (SD) - pierwiastek kwadratowy z wariancji. Daje przybliżony obraz rozproszenia wartości wokół średniej, np. dla zbioru z powyższego przykładu $SD = \sqrt{6} \approx 2,45$.

Pytania i testy

- **Pytania o związek** - badana cecha wykazuje **podobieństwa** w populacji A i B
 - Analiza korelacji - określa siłę i kierunek liniowego związku między dwiema cechami ilościowymi w obrębie populacji/grupy.
- **Pytania o różnice** - badana cecha **różni się** w populacji A i B
 - Test Chi-kwadrat - używany do oceny zależności między dwiema cechami jakościowymi w różnych grupach.
 - Test t-Studenta dla prób niezależnych - porównuje średnie wartości cechy w dwóch niezależnych grupach.
 - Test U Manna-Whitneya - nieparametryczny odpowiednik testu t, służący do analizy różnic w rozkładach cechy między dwiema grupami (gdy dane nie spełniają założeń o normalności rozkładu i homogeniczności wariancji).
 - Jednoczynnikowa analiza wariancji (ANOVA) - umożliwia porównanie średnich więcej niż dwóch grup i sprawdzenie, czy różnią się one istotnie statystycznie
 - Test Kruskala-Wallisa - nieparametryczny odpowiednik analizy wariancji (gdy dane nie spełniają założeń o normalności rozkładu i homogeniczności wariancji).

DZIEKUJĘ ZA UWAGĘ!