

Statystyka i analiza danych - laboratorium 10 - Analiza danych jakościowych - test χ^2

Paweł Misiorek

Instytut Informatyki
Politechnika Poznańska (PP)
Piotrowo 3, 60-965 Poznań, Poland
Email: pawel.misiorek@put.poznan.pl

26/30 maja 2023

- **Definicja i podstawowe cechy rozkładu χ^2**
- Test χ^2 na zgodność rozkładu
- Test χ^2 na niezależność zmiennych
- **Miary siły zależności**
 - ▶ współczynnik Yule'a - ϕ
 - ▶ współczynnik kontyngencji C Pearsona
 - ▶ współczynnik V Crammera

- Definicja i podstawowe cechy rozkładu χ^2 (suma kwadratów zmiennych o rozkładzie normalnym ustandaryzowanym - coś zawsze nieujemnego)
- Zastosowania do analizy danych jakościowych (nienumerycznych)
 - ▶ Test χ^2 na zgodność rozkładu ($k - 1$ stopni swobody, k - liczba możliwych wartości zmiennej skategoryzowanej)
 - ▶ Test χ^2 na niezależność zmiennych ($(k - 1) \cdot (w - 1)$ stopni swobody, k, w - liczba kolumn i wierszy)
 - ▶ wspólny klucz - zdefiniowanie statystyki bazującej na sumie kwadratów odchyłeń
 - ▶ umiejętność obliczania stopni swobody (liczba źródeł błędu - źródeł narastania statystyki)

Miary zależności zmiennych nominalnych

- Współczynnik Yule'a - ϕ :

$$\phi = \sqrt{\frac{\chi^2}{n}},$$

gdzie: χ^2 - wartość statystyki χ^2 dla testu na niezależność zmiennych dla n obserwacji.

- Współczynnik kontyngencji C Pearsona:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}.$$

- Współczynnik V Crammra:

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(k - 1, w - 1)}},$$

gdzie k - liczba kolumn, w - liczba wierszy.

Pytania

Pytania proszę zadawać:

- w czasie konsultacji (zoom),
- na Slacku,
- albo mailowo.

