ANALIZA DWUZMIENNOWA

czyli

ABC KOREALCJI

DZIASIAJ

Pożegnanie ze statystyką:

- Krótko o tym, co to znaczy, że ze sobą korelują
- Jak te korelacje badać
- Kilka ćwiczeń praktycznych

- Skończymy 15 min wcześniej ;)

ANALIZA DWUZMIENNOWA

- Centralne pytanie:

  Czy między badanymi zmiennymi występuje związek?

- WAŻNE:
  - Związek ten nie musi mieć charakteru przyczynowo-skutkowego
  - Jest to po prostu skojarzenie liczbowe pewnego typu

CO TO OZNACZA W PRAKTYCE?

Ćwiczenie:

Proszę na podstawie podanego opisu spróbować przewidzieć (i uzupełnić) brakujące dane w tabeli.

Czas: 3 minuty

CO TO OZNACZA W PRAKTYCE?

Wniosek:

Jeżeli istnieje zależność – jesteśmy w stanie przewidzieć przybliżone rozkłady dla poszczególnych zmiennych;

I odwrotnie:

Na podstawie różnicy między tym, co „oczekiwane" (przy założeniu, że zachodzi zależność), a tym co otrzymane można badać siłę zależności.

Współczynnik ρ - Spearmana

Inna nazwa:

współczynnik korelacji rangowej Spearmana

Zastosowanie:

Zmienne porządkowe i interwałowe.

Logika:

Jeżeli jest zupełna korelacja dodatnia, to kolejność (ustawiana na podstawie wartości danej zmiennej) będzie **taka sama** dla obu zmiennych.

Współczynnik ρ - Spearmana

Zaczynamy od **rangowania** zmiennych

czyli:

Dla każdej z analizowanych zmiennych:

1. Porządkujemy obserwacje wg wartości zmiennej – od najmniejszej do największej
2. Przypisujemy im numer miejsca, na którym się znajdują

Jeżeli kilka obserwacji ma tę samą wartość – przypisujemy im średnią z numerów wszystkich miejsc, które zajmują

Współczynnik ρ - Spearmana

Przykład rangowania

Osoby badane:	*Kolejność:*	*Rangi:*
Wykształcenie zawodowe	Wykształcenie podstawowe	1
Wykształcenie wyższe	Wykształcenie gimnazjalne	2
Wykształcenie średnie	Wykształcenie zawodowe	3,5
Wykształcenie gimnazjalne	Wykształcenie zawodowe	3,5
Wykształcenie podstawowe	Wykształcenie średnie	5
Wykształcenie wyższe	Wykształcenie wyższe	6,5
Wykształcenie zawodowe	Wykształcenie wyższe	6,5

Współczynnik ρ - Spearmana

Obliczenia:

$$\rho = 1 - \frac{6\sum d_i^2}{N(N^2-1)}$$

d – różnica pomiędzy rangą dla zmiennej pierwszej a drugiej

N – liczba obserwacji

Współczynnik ρ - Spearmana

Interpretacja wyniku:

- Kierunek związku:
  - <u>Wartość dodatnia</u> – zależność wprostproporcjonalna
  - <u>Wartość ujemna</u> – zależność odwrotnieproporcjonalna
- Siła związku:
  - $|\rho| < 0{,}3$ – zależność słaba
  - $|\rho| < 0{,}5$ – zależność średnia
  - $|\rho| > 0{,}5$ – zależność silna

Współczynnik V - Cramera

Zastosowanie:

Przynajmniej jedna zmienna nominalna.

Logika:

Jeżeli jest b. silna korelacja mogę poprawnie w przybliżeniu oszacować wartości w poszczególnych polach tablicy krzyżowej.

Współczynnik V - Cramera

Etapy obliczania:

1. Obliczenie wartości oczekiwanych
2. Obliczenie współczynnika pomocniczego – chi kwadrat (χ^2)
3. Obliczenie wartości współczynnika
4. Interpretacja wyniku

Współczynnik V - Cramera

1. Liczebności oczekiwane:

$$E\ (\text{częstość oczekiwana}) = \frac{(\text{suma rzędu})\ (\text{suma kolumny})}{(\text{suma całkowita})}$$

Przykład:

	w1	w2	w3	suma
Z1	21*15/70	24*15/70	25*15/70	15
Z2	21*30/70	24*30/70	25*30/70	30
Z3	21*25/70	24*25/70	25*25/70	25
suma	21	24	25	70

Współczynnik V - Cramera

2. Współczynnik pomocniczy - chi kwadrat (χ^2)

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \sum_{i=1}^{k} \sum_{j=1}^{p} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

E_{ij} – liczebność oczekiwana dla danego pola w tabeli

N_{ij} – liczebność faktyczna dla danego pola w tabeli

Współczynnik V - Cramera

3. Wartość współczynnika:

$$V = \sqrt{\frac{\chi^2}{N \min(k-1, p-1)}}$$

k, *p* – ilość wartości poszczególnych zmiennych

N – ilość badanych jednostek

Współczynnik V - Cramera

4. Interpretacja:

Siła związku:

- V< 0,3 – słaby związek
- V< 0,5 – umiarkowany związek
- V> 0,5 – silny związek

Współczynnik korelacji liniowej Pearsonsa

Zastosowanie:

Dwie zmienne ilorazowe

Wzór:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}},$$

Interpretacja:

Analogiczna

UWAGA! Dotyczy korelacji *liniowej*

ZALICZENIE

- DYŻURY:

  Poniedziałek – CN315; **od 11.30**

  Środa – CN315; **od 13.00**

  Piątek – CN 315; **od 14.00**

- Osoby zapisane – przychodzą koniecznie
  - Jak ktoś nie będzie mógł przyjść z przyczyn losowych – bardzo proszę o informację mailem
- Ok. 10 min na osobę

DZIĘKUJĘ ZA MIŁY SEMESTR ☺

ZAPRASZAM DO WYPEŁNIENIA ANKIET