

Statystyka

Defenicja – nauka, której przedmiotem zainteresowania są metody pozyskiwania i prezentacji, a przede wszystkim analizy danych opisujących zjawiska, w tym masowe.

Funkcje statystyki:

- Informacyjna
- Analityczna
- Prognostyczna

Zbiorowość statystyczna – zbiór dowolnych elementów podobnych, ale nie identycznych/ Może być **generalną** (wszystkie elementy) i **probną** (wybrane elementy). Elementem zbiorowości statystycznej jest **jednostka statystyczna**.

Budowa szeregów przedziałowych

Etap I

Porządkowanie – ustawianie wartości wg kolejności rosnącej.

Ustalenie liczby przedziałów k , na które podzielony zostanie pierwotny szereg wyliczający, na podstawie wzorów:

$$\begin{aligned}k &\sim \sqrt{n} \\k &\sim 1 + 3.322 \log n \\k &\sim 5 \log n\end{aligned}$$

Etap II

Określenie rozpiętości R wyjściowego szeregu wyliczającego:

$$R = X_{\max} - X_{\min}$$

Gdzie

- X_{\max} – Wartość maksymalna szeregu wyliczającego
- X_{\min} – Wartość minimalna szeregu wyliczającego

Etap III

Ustalenie rozpiętości h pojedynczej klasy szeregu przedziałowego

$$h \geq R / k$$

Etap IV

Ustalenie granic przedziałów szeregu przedziałowego i określenie liczby jednostek statystycznych, których wartości mieszczą się w poszczególnych przedziałach.

Dolna granica pierwszego przedziału:

$$X_{ld} = X_{\min} - 0.5 \cdot (k \cdot h - R)$$

Górna granica pierwszego przedziału jest o h większa od dolnej granicy i jednocześnie stanowi dolną granicę kolejnego przedziału.

$$X_{lg} = X_{ld} + h$$

Charakterystyki opisowe – wartości liczbowe, jakie pozwalają syntetyczny opisać rozkłady zmiennych ilościowych. Charakterystyki opisowe dla populacji próbnych nazywają się statystykami, a dla populacji generalnych parametrami.

Średnia arytmetyczna:

$$\frac{a_1 + a_2 + \dots + a_n}{n}$$

Zalety:

- Prostota: łatwe do obliczenia i zrozumienia.
- Retywności: Często odzwierciedla typową wartość w zbiorze danych.
- Zastosowanie: Nadaje się do różnych typów danych liczbowych.
- Właściwości matematyczne: Przydatne do analiz statystycznych.

Wady:

- Wrażliwy dla osób odstających: zniekształcone przez ekstremalne wartości, wpływając na dokładność.
- Dystrybucje nienarodzone: mniej dokładne dla zniekształconych lub niesymetrycznych danych.
- Nie dla danych kategorycznych: Nie ma zastosowania do danych nienumerycznych lub zwyczajowych.
- Niewybuchowy: może być niewiarygodny z małymi rozmiarami próbek.
- Ograniczony wskaźnik insight: Nie dostarcza informacji o rozprzestrzenianiu danych lub zmienności.

Powinieneś rozważyć użycie średniej arytmetycznej (średnia), gdy:

- Typ danych: Masz do czynienia z danymi numerycznymi (skala międzystawiania lub współczynnika).
- Prostota jest kluczem: potrzebujesz prostej, łatwej do zrozumienia miary tendencji centralnej.
- Dystrybucja danych: Dane są w przybliżeniu normalnie rozmieszczone lub symetryczne, a odstające nie są znacząco przeszkławane dane.
- Wartości równoważenia: Wartości ekstremalne są zrównoważone innymi wartościami w zbiorze danych, minimalizując ich wpływ na średnią.
- Analiza porównawcza: Chcesz porównać centralną tendencję wielu zestawów danych lub grup.
- Obliczenia matematyczne: Środek jest potrzebny do dalszej analizy matematycznej, takiej jak obliczenia wariancji lub modele regresji.

Pozycyjne parametry analizy struktury zbiorowości

Decyle wo wartości dzielące zbiorowość na 10 części.

Wszystkich decyli jest więc 9, przy czym warto pamiętać że decyl 5 jest medianie.

Percentyle (centyle)

jest ich 99, wartości te dzielą zbiorowość statystyczną na 100 części.
Percentyl 50 jest równy medianie.

Dominanta D=Mo

- **szereg rozdzielczy przedziałowy**
 - **wskazanie przedziału**

$$D = x_d + \frac{n_d - n_{d-1}}{(n_d - n_{d-1}) - (n_d - n_{d+1})} * h_d$$

Grupowanie danych wpływa na wartość średnią przedziałową

Pozycyjne miary położenia

Poza medianą kwartylami są także kwartył pierwszy, nazywany także dolnym (Q_1), mediana nazywa się kwartyłem drugim I kwartył trzeci, nazywamy także górnym (Q_3):

- kwartył pierwszy jest mieduaną pierwszej połowy szeregu, zatem dzieli zbiorowość w ten sposóbm że $\frac{1}{4}$ (25%) obserwacji ma wa...
- ...

Wzor na medianę w szeregu przedziałowym:

$$N/2 - \sum n_i$$

$$M = x_{Mk} + \frac{N/2 - \sum n_i}{\dots}$$

(13/11/2023)

X	N	XN	Nskum	$(x_i - \bar{x})^2 \cdot n$
1	8	8	0+8=8	$(1-2.52)^2 \cdot 8 = 18.48$
2	6	12	8+6=14	$(1-2.52)^2 \cdot 6 =$
3	4	12	14+4=18	
4	4	16	18+4=22	
5	3	15	22+3=25	
	25	63		

$$\bar{x} = 2.52$$

$$N/2 = 12.5 \quad Me = 2$$

$$n/4 = 6.25 \quad Q1 = 1$$

$$3/4n = 18.7 \quad Q3 = 4$$

$$S^2x = 1.93$$

$$R = 4 \quad Sx = 1.39$$

Analiza struktury - badanie asymetrii I koncentracji.

(20/11/2023)

Miary asymetrii zdarza się, że badanie średniego poziomu cechy i rozproszenia jej wartości nie wskazuje na istnienie różnic między porównywanymi zbiorowościami. Okazuje się bowiem, że istotne są przeciętny poziom i wewnętrzne zróżnicowanie cechy, ale także to czy przeważająca liczba badanych jednostek ma wartości cechy powyżej czy też poniżej przeciętnego poziomu cechy.

Miary asymetrii (skośności) określają kierunek rozkładu cech zmiennych w zbiorowości (rozkład może być symetryczny lub asymetryczny lewostronnie lub prawostronnie) oraz **stopień odchylenia rozkładu cechy zmiennej** od rozkładu symetrycznego. **Asymetria oznacza deformację rozkładu cechy zmiennej** w związku z wydłużeniem ramienia krzywej liczebności w prawo/lewo w stosunku do dominanty. Im asymetria rozkładu jest większa, tym mniejsza jest wartość poznawcza średniej arytmetycznej oraz pozostałych miar klasycznych i odwrotnie.