

Metody probabilistyczne i statystyka - wykład ósmy¹

dr Jarosław Kotowicz

Instytut Informatyki Uniwersytet w Białymstoku

10 kwietnia 2020

¹©J.Kotowicz, 2020

Spis treści

1 Centralne twierdzenie graniczne.

- Zbieżność według rozkładu
- Centralne twierdzenie graniczne

2 Wstęp do statystyki

3 Statystyka opisowa

- Rozkład cechy i jego prezentacja (statystyki graficzne)

4 Charakterystyki liczbowe struktury zbiorowości

- Miary położenia

Słaba zbieżność dystrybuant i zbieżność według rozkładu. I

Niech (Ω, Σ, P) będzie ustaloną przestrzenią probabilistyczną, $(X_n)_{n \in \mathbb{N}}$ ciągiem jednowymiarowych zmiennych losowych określonych na niej, $(F_n)_{n \in \mathbb{N}}$ ciągiem ich dystrybuant.

Uwaga 1

Możemy też rozważać ciąg rozkładów prawdopodobieństwa na \mathbb{R} i odpowiadający mu ciąg dystrybuant tych rozkładów.

Definicja 1

Mówimy, że ciąg dystrybuant $(F_n)_{n \in \mathbb{N}}$ jest słabo zbieżny do dystrybuanty F wtedy i tylko wtedy, gdy w każdym punkcie x ciągłości dystrybuanty F mamy $\lim_{n \rightarrow \infty} F_n(x) = F(x)$. Oznaczamy $F_n \xrightarrow{\omega} F$.

Słaba zbieżność dystrybuant i zbieżność według rozkładu.

II

Definicja 2

Mówimy, że ciąg zmiennych losowych $(X_n)_{n \in \mathbb{N}}$ (odpowiednio ciąg rozkładów prawdopodobieństw na \mathbb{R}) jest zbieżny według rozkładu do zmiennej losowej X (odpowiednio rozkładu prawdopodobieństwa na \mathbb{R}) o dystrybuancie F wtedy i tylko wtedy, gdy ciąg dystrybuant tego ciągu zmiennych losowych (dystrybuant rozkładów) jest słabo zbieżny do dystrybuanty F . Oznaczamy wtedy $X_n \xrightarrow{D} X$.

Wstęp

Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną, a $(X_n)_{n \in \mathbb{N}}$ ciągiem niezależnych zmiennych losowych określonych na niej.

Twierdzenie 1

Jeżeli zmienne losowe w ciągu $(X_n)_{n \in \mathbb{N}}$ posiadają ten sam rozkład oraz $\mathbb{E}(X_1) = 0$ i $\mathbb{D}^2(X_1) = 1$, to

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, 1). \quad (1)$$

Centralne twierdzenie graniczne. I

Będziemy rozważać ciąg podwójny – tablicę zmiennych losowych $\{X_{n,k} : n \in \mathbb{N} \wedge k \in \overline{1, n}\} \equiv (X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$.

Definicja 3

Mówimy, że ciąg $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ jest schematem serii wtedy i tylko wtedy, gdy dla dowolnego n ciąg $(X_{n,k})_{k \in \overline{1, n}}$ jest ciągiem zmiennych niezależnych.

Centralne twierdzenie graniczne. II

Definicja 4

Niech $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ będzie schematem serii. Załóżmy, że dla dowolnych $n \in \mathbb{N}$ i $k \in \overline{1, n}$ zmienne losowe $X_{n,k}$ są całkownalne z kwadratem.

Mówimy, że schemat serii $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ jest znormalizowanym schematem serii wtedy i tylko wtedy, gdy spełnione są warunki

- Ⓐ dla dowolnego n i k zachodzi $\mathbb{E}(X_{n,k}) = 0$,
- Ⓑ dla dowolnego n zachodzi $\sum_{k=1}^n \mathbb{D}^2(X_{n,k}) = 1$.

Uwaga 2

Będziemy oznaczać $s_n^2 \stackrel{\text{ozn}}{=} \sum_{k=1}^n \mathbb{D}^2(X_{n,k})$.

Centralne twierdzenie graniczne. III

Uwaga 3

Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie całkowalnych z kwadratem. Wówczas określając dla dowolnego $k \in \overline{1, n}$ zmienne losowe $X_{n,k} := \frac{X_k}{\sigma\sqrt{n}}$, gdzie $\sigma^2 = \mathbb{D}^2(X_1)$ otrzymujemy schemat serii. Jeżeli dodatkowo dla dowolnej liczby naturalnej n spełniony jest $\mathbb{E}(X_n) = 0$, to otrzymujemy znormalizowany schemat serii.

Definicja 5

Mówimy, że schemat serii $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ spełnia warunek Lindeberga wtedy i tylko wtedy, gdy

$$\forall n \in \mathbb{N} \forall k \in \overline{1, n} \mathbb{E}(X_{n,k}^2) < +\infty \quad (2)$$

$$\forall r > 0 \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}((X_{n,k} - \mathbb{E}(X_{n,k}))^2 \mathbb{I}_{\{\omega: |X_{n,k}(\omega) - \mathbb{E}(X_{n,k})| > r s_n\}}) = 0 \quad (3)$$

Centralne twierdzenie graniczne. IV

Uwaga 4

W przypadku znormalizowanego schematu serii warunek Lindeberga (3) ma postać

$$\forall r > 0 \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}(X_{n,k}^2 \mathbb{I}_{\{\omega: |X_{n,k}(\omega)| > r\}}) = 0. \quad (4)$$

Twierdzenie 2

Jeżeli znormalizowany schemat serii spełnia warunek Lindeberga, to

- 1 $\lim_{n \rightarrow \infty} \max_{k \in \overline{1, n}} \mathbb{D}^2(X_{n,k}) = 0,$
- 2 $\forall r > 0 \lim_{n \rightarrow \infty} \max_{k \in \overline{1, n}} P(\{\omega : |X_{n,k}(\omega)| > r\}) = 0.$

Centralne twierdzenie graniczne. V

Twierdzenie 3

Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie posiadających zerową wartość oczekiwaną i jednostkową wariancję. Niech $X_{n,k} := \frac{X_k}{\sqrt{n}}$ dla $k \in \overline{1, n}$. Wówczas schemat serii $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ spełnia warunek Lindeberga.

Centralne twierdzenie graniczne. VI

Twierdzenie 4

Niech $(X_n)_{n \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych jednostajnie ograniczonych^a całkowalnych z kwadratem oraz

$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{D}^2(X_k) = +\infty$. Niech $X_{n,k} := X_k$ dla $k \in \overline{1, n}$. Wtedy schemat serii $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ spełnia warunek Lindeberga.

^aMówimy, że ciąg zmiennych $(X_n)_{n \in \mathbb{N}}$ jest jednostajnie ograniczony wtedy i tylko wtedy, gdy

$$\exists M > 0 \forall n \in \mathbb{N} \forall \omega \in \Omega |X_n(\omega)| \leq M. \quad (5)$$

Centralne twierdzenie graniczne. VII

Twierdzenie 5 (Lindeberga - Lévy'ego)

Niech znormalizowany schemat serii $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ spełnia warunek

Lindeberga wtedy. Wtedy $\sum_{k=1}^n X_{n,k} \xrightarrow{D} \mathcal{N}(0, 1)$

Centralne twierdzenie graniczne. VIII

Uwaga 5

W dowodzie postępujemy się następującymi faktami

$$\forall x \in [-\frac{1}{2}, \frac{1}{2}] \quad |e^x - 1 - x| \leq x^2 \quad (6)$$

$$\forall x \in \mathbb{R} \quad |e^{ix} - 1 - ix| \leq \frac{x^2}{2} \quad (7)$$

$$\forall x \in \mathbb{R} \quad \left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \leq \frac{|x|^3}{6} \quad (8)$$

$$\forall n \in \mathbb{N} \forall 1 \leq k \leq n \forall \{a_1, \dots, a_n, b_1, \dots, b_n\} \subset \mathbb{C} \quad |a_k| \leq 1 \wedge |b_k| \leq 1 \Rightarrow$$

$$|a_1 \cdot \dots \cdot a_n - b_1 \cdot \dots \cdot b_n| \leq \sum_{k=1}^n |a_k - b_k| \quad (9)$$

Warunek Lapunowa i jego związek z warunkiem Lindeberga.

Definicja 6

Mówimy, że schemat serii $(X_{n,k})_{n \in \mathbb{N}, k \in \overline{1, n}}$ spełnia warunek Lapunowa wtedy i tylko wtedy, gdy

$$\exists \delta > 0 \forall n \in \mathbb{N} \forall k \in \overline{1, n} \mathbb{E}(|X_{n,k}|^{2+\delta}) < +\infty \quad (10)$$

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}(|X_{n,k} - \mathbb{E}(X_{n,k})|^{2+\delta}) = 0 \quad (11)$$

Twierdzenie 6

Jeżeli ciąg serii spełnia warunek Lapunowa, to spełnia warunek Lindeberga.

Spis treści

- 1 Centralne twierdzenie graniczne.
 - Zbieżność według rozkładu
 - Centralne twierdzenie graniczne
- 2 **Wstęp do statystyki**
- 3 Statystyka opisowa
 - Rozkład cechy i jego prezentacja (statystyki graficzne)
- 4 Charakterystyki liczbowe struktury zbiorowości
 - Miary położenia

Podstawowe pojęcia statystyczne. I

- Zbiorowość generalna (populacja generalna) – kompletny zbiór elementów lub możliwych wyników procesu generującego dane.
- Element (jednostka) zbiorowości – indywidualna składowa zbiorowości mogąca być jednostką fizyczną lub wynikiem procesu.
- Cecha statystyczna – właściwość elementów zbiorowości.
- Rozkład cechy – zestawienie różnych wariantów lub wartości, które pojawiają się w zbiorowości i podanie częstości ich występowania.
- Parametry populacji – charakterystyki liczbowe populacji.
- Obserwacja statystyczna – polega na dokonywaniu pomiaru, zliczeniu faktów, ankietowaniu bezpośrednim lub pośrednim albo na samorejestracji; rozumiana też jest jako indywidualny pomiar lub zarejestrowany wariant cechy.
- Badanie statystyczne – proces pozyskiwania danych do analizy poprzez dokonywanie obserwacji statystycznych.

Podstawowe pojęcia statystyczne. II

- Próba – podzbiór elementów populacji podlegający badaniu częściowemu.
- Próba losowa – próba dobrana poprzez losowanie elementów.
- Statystyka z próby – charakterystyki rozkładu cechy obliczone na podstawie próby.
- Losowe błędy badania – losowe odchylenia statystyk z próby od parametrów populacji.
- Badanie reprezentacyjne – badanie statystyczne oparte na próbie losowej.

Podział cech statystycznych. I

- 1 niemierzalna (jakościowe) – cechy, których warianty są kategoriami jakościowymi
 - 1 geograficzne,
 - 2 inne.
- 2 mierzalna (ilościowe) – cechy, które określa się liczbowo,
 - 1 typu skokowego (dyskretnego) – cechy, które mogą przyjmować co najwyżej przeliczana ilość wartości,
 - 2 typu ciągłego – cechy, które mogą przyjmować każdą wartość liczbową z pewnego przedziału warianty są kategoriami jakościowymi.

Uwaga 6

Do cech mierzalnych zalicza się też cechy quasi-ilościowe zwane porządkowymi. Cechy te kwalifikują natężenie badanej właściwości przedstawionej w sposób opisowy

Klasyfikacja Stanleya Stevensa zmiennych w analizie danych:

Podział cech statystycznych. II

- 1 zmienne jakościowe (inaczej wyliczeniowe, czynnikowe lub kategoryczne) to zmienne przyjmujące określoną liczbę wartości (najczęściej nieliczbowych). Zmienne te można podzielić na:
 - binarne (dwumianowe, dychotomiczne), np. płeć (poziomy: kobieta/mężczyzna),
 - nominalne (jakościowe nieuporządkowane), np. marka samochodu, pomiędzy wartościami tych zmiennych nie ma żadnego porządku,
 - porządkowe (jakościowe uporządkowane), np. wykształcenie (poziomy: podstawowe/średnie/wyższe).
- 2 zmienne ilościowe, które można podzielić na:
 - licznikowe (liczba wystąpień pewnego zjawiska, opisywana liczbą naturalną), np. liczba lat nauki,
 - przedziałowe (interwałowe), mierzone w skali, w której odejmowanie wartości ma sens, ale iloraz już nie, np. temperatura w stopniach Celsjusza lub rok naszej ery,

Podział cech statystycznych. III

- ilorazowe, czyli zmienne przedziałowe, mierzone w skali, w której dodatkowo zachowane są proporcje, wartości można dzielić, skala ma zero absolutne, np. temperatura w Kelvina, wzrost w centymetrach itp.

Cechy mierzalne – stymulanta, destymulanta, nominanta

W badaniach społeczno-ekonomicznych cechy mierzalne dzielimy na

- 1 stymulanty – cechy, których wyższe wartości pozwalają zakwalifikować daną jednostkę statystyczną jako lepszą z punktu widzenia realizowanego badania,
- 2 destymulanty – cechy, których wyższe wartości świadczą o niskiej pozycji jednostki w zbiorze,
- 3 nominata – pożądane poziomy zmiennych (normatywne).

Podział zbiorowości generalnej ze względu na ilość badanych cech statystycznych

- 1 jednowymiarowe (jednocechowe),
- 2 wielowymiarowe (wielocechowe).

Skale pomiaru. I

- **nominalna** – odpowiada jej relacja równe lub różne; pozwala jedynie określić przynależność badanych elementów do wyróżnionych dla danej cechy wyróżnionych kategorii jakościowych; polega na zastosowaniu liczb jako nazw; szczególnym przypadkiem tej skali jest skala dychotomiczna – dwupunktowa,
- **porządkowa** – odpowiada jej relacja większe lub mniejsze; pozwala na porządkowanie jednostek według stopnia natężenia danej cechy; każdemu ze stanów można przypisać liczbę, a ten proces nazywamy rangowaniem,

Skale pomiaru. II

- **interwałowa (przedziałowa)** – odpowiada jej relacja większe o tyle; skala liczbowa z ustalonymi jednostkami pomiaru, co pozwala na określenie porządku jednostek oraz różnic (dystansu) między nimi; w skali interwałowej, w przeciwieństwie do skali ilorazowej, nie ma naturalnego początku, który uprawniałby obliczanie ilorazów wartości cechy; jest początek umowny,
- **ilorazowa (stosunkowa)** – odpowiada jej relacja większe tyle razy; skala liczbowa z ustalonymi jednostkami pomiaru, co pozwala na określenie porządku jednostek oraz różnic między nimi, w skali ilorazowej jest naturalny początek, który pozwala na obliczanie ilorazów wartości cechy.

Podział badań statystycznych

- 1 pełne – badanie wszystkich jednostki danej zbiorowości (spis statystyczne, rejestr statystyczny itp.),
- 2 częściowe – badanie części jednostek populacji (ankietowe, monograficzne, reprezentacyjne),
- 3 szacunki
 - 1 interpolacyjne (szacowanie nieznanymi wartości cechy na podstawie znanych wartości sąsiednich),
 - 2 ekstrapolacyjne (szacowanie wartości wykraczających poza przedział wartości znanych).

Spis treści

- 1 Centralne twierdzenie graniczne.
 - Zbieżność według rozkładu
 - Centralne twierdzenie graniczne
- 2 Wstęp do statystyki
- 3 **Statystyka opisowa**
 - Rozkład cechy i jego prezentacja (statystyki graficzne)
- 4 Charakterystyki liczbowe struktury zbiorowości
 - Miary położenia

Motywacja

Pierwotny zbiór danych uzyskanych z badania ma na ogół postać danych indywidualnych. Tworzą go wszystkie wartości (warianty) cech zaobserwowanych u zbadanych jednostek.

Materiał liczbowy otrzymany w wyniku obserwacji statystycznej należy odpowiednio usystematyzować i pogrupować w postaci szeregów statystycznych.

Typy grupowania

- 1 grupowanie typologiczne – celem wyodrębnienie grup różnych jakościowo,
- 2 wariacyjne – celem uporządkowanie badanej zbiorowości i poznanie jej struktury polegające na łączeniu w klasy jednostek statystycznych o odpowiednich wartościach cechy statystycznej.

Szeregi statystyczne. I

Szereg statystyczny – ciąg wielkości statystycznych uporządkowanych według określonego kryterium.

Wyróżniamy następujące główne typy szeregów statystycznych

- ❶ szereg szczegółowy (uporządkowany ciąg wartości badanej cechy statystycznej),
- ❷ szereg rozdzielczy - zbiorowość statystyczna podzielona na klasy według określonej cechy jakościowej lub ilościowej z podaniem liczebności lub częstotliwości każdej z wyodrębnionych klas
 - ❶ z cechą mierzalną
 - punktowe,
 - przedziałowe,
 - ❷ z cechą niemierzalną
 - geograficzne,
 - inne.

Szeregi statystyczne. II

- 3 szereg czasowy (dynamiczny, chronologiczny) – grupowanie typologiczne i wariacyjne, gdy podstawą grupowania jest zmiana badanego zjawiska w czasie.

Szeregi statystyczne. III

Uwaga 7

- *Szeregi punktowe i przedziałowe można jeszcze podzielić na szeregi proste i skumulowane.*
- *Szeregi przedziałowe buduje się dla cech ciągłych i cech dyskretnych w dużą ilość obserwacji.*
- *Szeregi punktowe buduje się dla cech dyskretnych z niewielką ilością obserwacji.*
- *Szeregi czasowe dzielimy na*
 - *szeregi okresów – szereg zawiera informacje o rozmiarach zjawiska w krótszych lub dłuższych okresach,*
 - *szeregi momentów – szereg ujmuje wielkość zjawiska w danych momentach, najczęściej na początku lub na koniec okresu.*

Rozkład empiryczny cechy – zestawienie wyników w postaci szeregu rozdzielczego z cechą mierzalną.

Szeregi statystyczne. IV

Wskaźnik struktury szeregu punktowego to częstotliwość występowania danego wariantu cechy.

Przykład

Przykład 1 ([1])

Liczba pasażerów: 2, 0, 2, 1, 2, 1, 0, 3, 1, 3, 3, 1, 0, 3, 4, 2, 2, 1, 3, 1, 1, 2, 1, 4, 2.

Prędkość (km/h): 64, 77, 51, 70, 69, 50, 72, 47, 93, 52, 60, 56, 63, 59, 58, 82, 60, 63, 65, 67, 61, 71, 66, 62, 68.

Płeć kierowcy: K, M, M, M, K, M, M, K, M, K, M, K, M, K, M, M, K, K, M, M, K, M, K, K, M.

Materiały dodatkowe z wykorzystaniem R

Spis treści

- 1 Centralne twierdzenie graniczne.
 - Zbieżność według rozkładu
 - Centralne twierdzenie graniczne
- 2 Wstęp do statystyki
- 3 Statystyka opisowa
 - Rozkład cechy i jego prezentacja (statystyki graficzne)
- 4 **Charakterystyki liczbowe struktury zbiorowości**
 - Miary położenia

Wszechstronna analiza struktury

Na wszechstronną analizę struktury składają się cztery zagadnienia

- 1 analiza tendencji centralnej (średniego poziomu cechy),
- 2 analiza zróżnicowania,
- 3 analiza asymetrii,
- 4 analiza koncentracji.

Parametry statystyczne i ich podział. I

Analiza danych statystycznych prowadzi do przedstawienia wyników badań za pomocą charakterystyk liczbowych zwanych parametrami statystycznymi.

Podział parametrów statystycznych

- 1 miary położenia,
- 2 miary zmienności (rozproszenia, dyspersji, zróżnicowania),
- 3 miary asymetrii (skośności),
- 4 miary koncentracji (spłaszczenia).

Parametry statystyczne i ich podział. II

Uwaga 8

- *Miary położenia wyznaczają przeciętną wartość cechy statystycznej.*
- *Miary zmienności wyznaczają siłę zróżnicowania wartości cechy statystycznej. Pozwalają określić jakie jest zróżnicowanie wartości cechy statystycznej w zbiorze obserwacji (jak mocno „rozproszone” są poszczególne obserwacje).*
- *Miary asymetrii wyznaczają siłę skupienia wartości cechy statystycznej bliżej dolnej lub górnej granicy zbioru wartości.*
- *Miary koncentracji wyznaczają siłę skupienia wartości cechy statystycznej wokół wartości przeciętnej.*

Miary położenia i przeciętne oraz ich podział. I

Podział miar położenia

- 1 miary przeciętne,
- 2 kwantyle.

Podział miar przeciętnych

- 1 klasyczne,
 - 1 średnia arytmetyczna,
 - 2 średnia harmoniczna,
 - 3 średnia geometryczna,
- 2 pozycyjne,
 - 1 moda (dominanta, modalna),
 - 2 kwantyle
 - kwartyle,
 - mediana (kwartył drugi),
 - decyle itp,

Miary położenia i przeciętne oraz ich podział. II

- 3 wartości skrajne.

Uwaga 9

Miary tendencji centralnej to szczególne miary położenia, a mianowicie: średnia arytmetyczna, mediana, modalna.

Średnia arytmetyczna. I

Dla szeregów szczegółowych – średnia arytmetyczna (nieważona, prosta)

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i.$$

Dla szeregów rozdzielczych punktowych i przedziałowych – średnia arytmetyczna ważona

$$\bar{x} := \frac{1}{n} \sum_{i=1}^k x_i n_i \quad (\text{dla szeregu punktowego})$$

$$\bar{x} := \frac{1}{n} \sum_{i=1}^k \dot{x}_i n_i \quad (\text{dla szeregu przedziałowego}),$$

gdzie k jest liczbą klas, a \dot{x}_i środkiem i -tego przedziału klasowego.

Średnia arytmetyczna. II

Uwaga 10

- 1 Jeżeli znamy średnie w każdej z klas tzn. \bar{x}_i , gdzie $i \in \overline{1, k}$, to dla szeregu rozdzielczego przedziałowego można liczyć średnią ważoną zgodnie z następującym wzorem

$$\bar{x} := \frac{1}{n} \sum_{i=1}^k \bar{x}_i n_i.$$

- 2 Od tej chwili nie będziemy rozróżniać szeregów rozdzielczych punktowych i przedziałowych i \bar{x}_i będzie oznaczać dla nich albo środek przedziału, albo wartość punktu.

Własności średniej

- 1 średnia jest z przedziału $]x_{\min}, x_{\max}[$,
- 2 suma odchyłeń poszczególnych wartości cechy od średniej równa się zero,
- 3 suma kwadratów odchyłeń poszczególnych wartości cech od średniej jest minimalna,
- 4 średnia arytmetyczną oblicza się w zasadzie dla szeregów o zamkniętych przedziałach klasowych,
- 5 średniej arytmetycznej nie oblicza się w szeregach, w których udział liczebności w przedziałach klasowych otwartych jest duży,
- 6 średnia arytmetyczna z próby reprezentatywnej jest dobrym przybliżeniem wartości przeciętne w populacji generalnej,
- 7 średnia arytmetyczna jest wrażliwa na skrajne wartości cechy (tzw. wartości przypadkowe).

Średnia harmoniczna

$$\bar{x}_H := \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (\text{dla szeregu szczegółowego}),$$

$$\bar{x}_H := \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k \frac{n_i}{x_i}} \quad (\text{dla szeregu rozdzielczego}).$$

Uwaga 11

Średnia harmoniczną stosujemy wtedy, gdy wartości cechy są podane w przeliczeniu na stałą jednostkę, czyli w postaci wskaźników natężenia, wagi natomiast w jednostkach liczników tych cech.

Średnia geometryczna

$$\bar{x}_G := \sqrt[n]{\prod_{i=1}^n x_i} \quad (\text{dla szeregu szczegółowego}),$$

$$\bar{x}_G := \sqrt[n]{\prod_{i=1}^k \dot{x}_i^{n_i}} \quad (\text{dla szeregu rozdzielczego}),$$

gdzie k jest ilością klas, a $n = \sum_{i=1}^k n_i$.

Uwaga 12

Średnia harmoniczną stosujemy przy badaniu średniego tempa zmian zjawiska.

Moda (modalna, dominanta). I

Definicja 7

Moda jest to wartość cechy statystycznej występującej w rozkładzie empirycznym najczęściej (M_o).

Uwaga 13

Moda musi być pojedynczą wartością. Jeśli w zbiorze obserwacji nie istnieje pojedyncza wartość cechy statystycznej występująca najczęściej, to moda nie istnieje w tej zbiorowości lub też mówimy, że cecha ma rozkład wielomodowy.

Modę można zdefiniować też zmiennej ciągłej.

Moda (modalna, dominanta). II

Definicja 8

Modą dla zmiennej losowej o rozkładzie ciągłym jest to wartość, dla której funkcja gęstości prawdopodobieństwa ma wartość największą.

Dla szeregu rozdzielczego modalną wyznaczamy według wzoru

$$Mo := x_{0m} + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} h_m, \quad (12)$$

gdzie

n_i liczebność przedziału i - tego ($i = m$ zawierającego modalną, $i = m - 1$ ($i = m + 1$) poprzedzającej przedział (następującym po przedziale) z modalną,

x_{0m} – granica dolna przedziału, w którym znajduje się modalna,

h_m – rozpiętość przedziału, w którym znajduje się modalna.

Mediana. I

Definicja 9

Mediana jest to wartość cechy statystycznej, dzieląca zbiór obserwacji na dwie liczebnie równe części (zbiór obserwacji o wartościach mniejszych lub równych oraz zbiór obserwacji o wartościach większych lub równych od wartości mediany).

$$\text{Me} := \begin{cases} x_{\frac{n+1}{2}} & \text{dla } n \notin \mathbb{N} \setminus 2\mathbb{N} \\ \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{dla } n \in 2\mathbb{N} \end{cases} \quad (\text{szereg szczegółowy}),$$

$$\text{Me} := x_{0\text{Me}} + \frac{N_{\text{Me}} - \sum_{i=1}^{m-1} n_i}{n_m} h_m \quad (\text{szereg rozdzielczy}),$$

Mediana. II

gdzie

m – numer klasy w której znajduje się mediana,

x_{0Me} – granica dolna przedziału w którym znajduje się mediana,

n_m – liczebność przedziału mediany,

$\sum_{i=1}^{m-1} n_i$ – liczebność skumulowana,

h_m – rozpiętość przedziału mediany,

N_{Me} – pozycja mediany (przyjmuje się, że $N_{Me} = \frac{n}{2}$).

Bibliografia



J. Podgórski. *Statystyka dla studiów licencjackich*. Warszawa: Polskie Wydawnictwo Ekonomiczne, 2010.