Spis treści

1 Wstęp

1.1 Wstęp

STATYSTYKA
nauka poświęcona metodom badania (analizowania) zjawisk masowych; polega na systematyzowaniu obserwowanych cech ilościowych i jakościowych oraz przedstawianiu wyników w postaci zestawień tabelarycznych, wykresów, itp.; posługuje się rachunkiem prawdopodobieństwa.

STATYSTYKA MATEMATYCZNA
dział matematyki stosowanej oparty na rachunku prawdopodobieństwa; zajmuje się badaniem zbiorów na podstawie znajomości własności ich części.

1.2 Literatura

Literatura

1. **Feller W.,** *Wstęp do rachunku prawdopodobieństwa,* T. 1. PWN, Warszawa 1966

2. **Feller W.,** *Wstęp do rachunku prawdopodobieństwa,* T. 2. PWN, Warszawa 1969

3. **Fisz M.,** *Rachunek prawdopodobieństwa i statystyka matematyczna,* PWN, Warszawa 1958

4. **Jasiulewicz H., Kordecki W.,** *Rachunek prawdopodobieństwa i statystyka matematyczna, przykłady i zadania,* Oficyna Wydawnicza GiS, Wrocław 2002

5. **Jakubowski J., Sztencel R.,** *Wstęp do teorii prawdopodobieństwa,* SCRIPT, Warszawa 2001

Literatura

6. **Jakubowski J., Sztencel R.,** *Rachunek prawdopodobieństwa dla (prawie) każdego,* SCRIPT, Warszawa 2002

7. **Krysicki W. (i inni),** *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, część I Rachunek prawdopodobieństwa,* PWN 1995

8. **Krzyśko M.,** *Wykłady z teorii prawdopodobieństwa,* UAM, Poznań 1997

9. **Niemiro W.,** *Rachunek prawdopodobieństwa i statystyka matematyczna,* Szkoła Nauk Ścisłych, Warszawa 1999

10. **Plucińska A., Pluciński E.,** *Probabilistyka,* WNT, Warszawa 2000

2 Elementy rachunku prawdopodobieństwa

2.1 Podstawy

Przestrzeń zdarzeń elementarnych

Pojęciem pierwotnym w rachunku prawdopodobieństwa jest **przestrzeń zdarzeń elementarnych** Ω.

Zdarzenie losowe
Podzbiory zbioru Ω nazywamy zdarzeniami losowymi

Prawdopodobieństwo

Definicja
Dowolną funkcję określoną na zdarzeniach losowych taką, że

- Dla każdego zdarzenia A zachodzi $P(A) \geq 0$
- $P(\Omega) = 1$
- Jeśli A i B są takimi zdarzeniami, że $A \cap B = \emptyset$ dla $i \neq j$, to

$$P\left(A \cup B\right) = P(A) + P(B)$$

2.2 Prawdopodobieństwo warunkowe

Prawdopodobieństwo warunkowe

Definicja
Prawdopodobieństwem warunkowym zajścia zdarzenia A pod warunkiem zajścia zdarzenia B, gdzie $P(B) > 0$, nazywamy liczbę

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2.3 Prawdopodobieństwo całkowite

Prawdopodobieństwo całkowite

Prawdopodobieństwo całkowite

Jeżeli $B_1, \ldots, B_n$ są takimi zdarzeniami losowymi, że

- $P(B_i) > 0$ dla $i = 1, \ldots, n$,
- $B_1 \cup \cdots \cup B_n = \Omega$,
- $B_i \cap B_j$ dla $i \neq j$,

to dla dowolnego zdarzenia A zachodzi

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_n)P(B_n)$$

Prawdopodobieństwo całkowite - przykład

Zadanie

Przedsiębiorstwo zawarło umowy z zakładami Z_1, Z_2 oraz Z_3 na dostawę podzespołów. Zakład Z_1 dostarcza 50%, zakład Z_2 dostarcza 35% natomiast zakład Z_3 dostarcza 15% potrzebnych podzespołów. Wiadomo, że 95% dostaw zakładu Z_1, 80% dostaw zakładu Z_2 oraz 85% dostaw zakładu Z_3 odpowiada wymaganiom technicznym. Jakie jest prawdopodobieństwo, że jeden wylosowany podzespół odpowiada wymaganiom technicznym?

Prawdopodobieństwo całkowite - przykład

Rozwiązanie

B_1 - zdarzenie polegające na wylosowaniu podzespołu wyprodukowanego w zakładzie Z_1.

B_2 - zdarzenie polegające na wylosowaniu podzespołu wyprodukowanego w zakładzie Z_2.

B_3 - zdarzenie polegające na wylosowaniu podzespołu wyprodukowanego w zakładzie Z_3.

A - zdarzenie polegające na wylosowaniu podzespołu odpowiadającemu wymogom technicznym.

Prawdopodobieństwo całkowite - przykład

Rozwiązanie

$$P(A|B_1) = 0.95 \quad P(B_1) = 0.50$$

$$P(A|B_2) = 0.80 \quad P(B_2) = 0.35$$

$$P(A|B_3) = 0.85 \quad P(B_3) = 0.15$$

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\ &= 0.8825 \end{aligned}$$

Prawdopodobieństwo całkowite

Wzór Bayesa
Jeżeli $B_1, \ldots, B_n$ są takimi zdarzeniami losowymi, że

- $P(B_i) > 0$ dla $i = 1, \ldots, n$,
- $B_1 \cup \cdots \cup B_n = \Omega$,
- $B_i \cap B_j$ dla $i \neq j$,

oraz niech $P(A) > 0$. Wówczas dla dowolnego $j \in \{1, \ldots, n\}$ mamy

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)}$$

Prawdopodobieństwo całkowite - przykład cd

Zadanie
Do punktu serwisowego zgłasza się klient z urządzeniem, w którym uszkodzony jest podzespół. Jakie jest prawdopodobieństwo, że producentem zepsutego podzespołu był zakład Z_1?

Prawdopodobieństwo całkowite - przykład cd

Rozwiązanie

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = 0.5382$$

2.4 Niezależność zdarzeń

Niezależność zdarzeń

Definicja
Zdarzenia A oraz B nazywamy niezależnymi, gdy

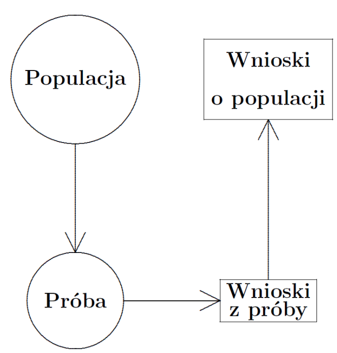
$$P(B|A) = P(B), \quad P(A) > 0.$$

Zdarzenia A oraz B są niezależne, gdy

$$P(A \cap B) = P(A)P(B)$$

3 Wnioskowanie statystyczne

Idea wnioskowania statystycznego



Pojęcia

Populacja
Zbiór obiektów z wyróżnioną cechą (cechami)

Próba
Wybrana część populacji podlegająca badaniu

Cecha
Wielkość losowa charakteryzująca obiekty danej populacji

Pojęcia

Cecha jakościowa
Cecha przyjmująca wartości nie będące liczbami (np. *kolor*, *płeć*, *smakowitość*)

Cecha skokowa (dyskretna)
Cecha przyjmująca pewne wartości liczbowe i nie przyjmująca wartości pośrednich (np. *liczba bakterii*, *liczba pracowników*, *liczba pasażerów*).

Cecha ciągła
Cecha przyjmująca wartości z pewnego przedziału liczbowego (np. *wzrost*, *waga*, *plon*)

4 Podstawowe modele statystyczne

4.1 Model dwumianowy

Doświadczenie Bernoulliego

Określenie
Wykonujemy dwuwynikowe doświadczenie. Wyniki nazywane są umownie *sukces* oraz *porażka*. Prawdopodobieństwo sukcesu wynosi p (porażki: $1 - p$).

Zmienna losowa
Zmienną losową X jest uzyskanie sukcesu.

Schemat Bernoulliego

Określenie
Doświadczenie Bernoulliego wykonujemy n krotnie w sposób niezależny. Zmienną losową X jest liczba sukcesów.

Rozkład dwumianowy
Zmienna losowa X ma rozkład $B(n, p)$:

$$P_{n,p}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

Schemat Bernoulliego

Przykład
Wadliwość procesu produkcyjnego wynosi 10%. Obliczyć prawdopodobieństwo, że na osiem wylosowanych produktów będą co najwyżej dwa złe.

Schemat Bernoulliego

Rozwiązanie
Doświadczenie Bernoulliego: wylosowanie jednego elementu.

Sukces: element wadliwy
Porażka: element dobry

Prawdopodobieństwo sukcesu: 0.1
Prawdopodobieństwo porażki: 0.9

Schemat Bernoulliego

Rozwiązanie
Schemat Bernoulliego: wylosowanie dziesięciu elementów.

Prawdopodobieństwo wylosowania co najwyżej dwóch złych=
prawdopodobieństwo wylosowania żadnego złego+
prawdopodobieństwo wylosowania jednego złego+
prawdopodobieństwo wylosowania dwóch złych.

$$\binom{10}{0}0.1^0 0.9^{10} + \binom{10}{1}0.1^1 0.9^9 + \binom{10}{2}0.1^2 0.9^8$$

4.2 Model gaussowski

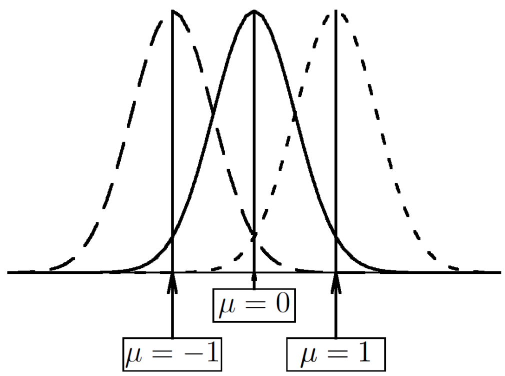
Rozkład normalny

Określenie
Zmienna losowa X ma rozkład normalny $N(\mu, \sigma^2)$:

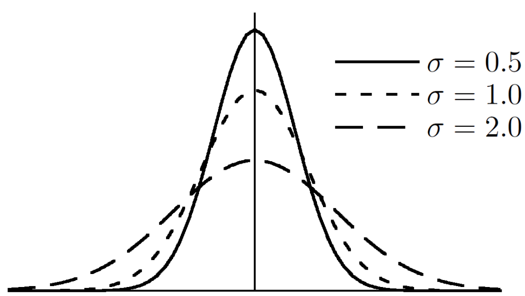
$$f_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \ -\infty < x < \infty.$$

$$EX = \mu \qquad D^2X = \sigma^2.$$

Rozkład normalny



Rozkład normalny



Rozkład normalny

UWAGA
Oznaczenie rozkładu normalnego

$$N(\text{wartość średnia, wariancja})$$

lub

$$N(\text{wartość średnia, odchylenie standardowe})$$

Rozkład normalny

UWAGA
Nasze oznaczenie rozkładu normalnego

$$N(\text{wartość średnia, wariancja})$$

Rozkład normalny

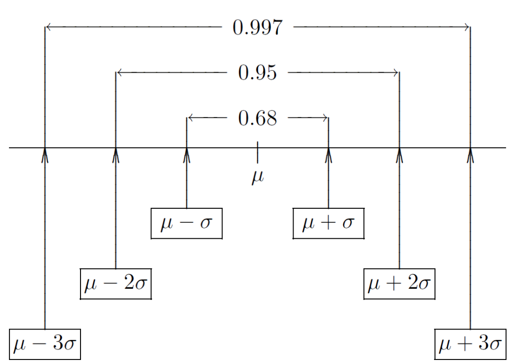
Prawo trzech sigm

$$P\{|X-\mu|<\sigma\} = 0.68268 \approx 0.68$$

$$P\{|X-\mu|<2\sigma\} = 0.95450 \approx 0.95$$

$$P\{|X-\mu|<3\sigma\} = 0.99730 \approx 0.997$$

Rozkład normalny



Rozkład normalny

Przykład
Przyjmując, że waga (w kilogramach) noworodka jest zmienną losową o rozkładzie $N(3, 0.25)$ określić procent noworodków o wadze:

- z przedziału $(3, 3.25)$,
- z przedziału $(2.5, 3.5)$.

Rozkład normalny

Rozwiązanie
Niech X oznacza wagę (w kilogramach) noworodka. Jest to zmienna losowa o wartości średniej 3 i o odchyleniu standardowym 0.5.

Mamy obliczyć:

- $P\{X \in (3, 3.25)\}$,
- $P\{X \in (2.5, 3.5)\}$.

Rozkład normalny

Rozwiązanie

$$P\{X \in (3, 3.25)\} = \texttt{ROZKŁAD.NORMALNY}(3.25; 3; 0.5; 1) - \texttt{ROZKŁAD.NORMALNY}(3; 3; 0.5; 1) = 0.1915.$$

Odpowiedź: nieco ponad 19% noworodków ma wagę między 3 a 3.25 kilograma

Rozkład normalny

Rozwiązanie

$$P\{X \in (2.5, 3.5)\} = P\{X \in (\text{średnia} - \text{odchylenie}, \text{średnia} + \text{odchylenie})\} = 0.6827.$$

Odpowiedź: prawie 68.3% noworodków ma wagę między 2.5 a 3.5 kilograma

5 Estymacja

5.1 Estymatory punktowe

Estymatory punktowe

Określenie
Estymujemy parametr θ rozkładu cechy X
Próba: $X_1, X_2, \ldots, X_n$
Estymator (punktowy) jest funkcją próby

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$$

przybliżającą wartość parametru θ

Estymatory punktowe

Rozkład normalny $N(\mu, \sigma^2)$
Estymator średniej μ - średnia arytmetyczna

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n}$$

Estymatory punktowe

Rozkład normalny $N(\mu, \sigma^2)$
Estymator wariancji σ^2 - wariancja próbkowa

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Suma kwadratów odchyleń od średniej

$$\mathrm{var}X = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2$$

Estymatory punktowe

Rozkład normalny $N(\mu, \sigma^2)$
Estymator odchylenia standardowego σ

$$S = \sqrt{S^2}$$

Estymatory punktowe

Rozkład dwumianowy $B(n, p)$
k - liczba sukcesów w próbie n elementowej
Estymator punktowy:

$$\hat{p} = \frac{k}{n}$$

5.2 Przedziały ufności

Przedział ufności

Określenie
Przedziałem ufności nazywamy przedział o końcach zależnych od próby, który z pewnym z góry zadanym prawdopodobieństwem pokrywa nieznaną wartość parametru θ

$$P\{\theta \in (\underline{\theta}(X_1, \ldots, X_n),\ \overline{\theta}(X_1, \ldots, X_n))\} = 1 - \alpha$$

Poziom ufności: prawdopodobieństwo $1 - \alpha$

Przedział ufności

Długość przedziału d

- Liczność próby ($n \nearrow \Longrightarrow d \searrow$)
- Poziom ufności ($1 - \alpha \nearrow \Longrightarrow d \nearrow$)
- Wariancja cechy ($\sigma^2 \searrow \Longrightarrow d \searrow$)

5.3 Przedział ufności dla średniej

Przedział ufności dla wartości średniej

Rozkład normalny $N(\mu, \sigma^2)$
Wariancja σ^2 jest nieznana
Poziom ufności: $1 - \alpha$

$$\left(\bar{X} - t(\alpha; n-1)\frac{S}{\sqrt{n}}\ ,\ \bar{X} + t(\alpha; n-1)\frac{S}{\sqrt{n}}\right)$$

$t(\alpha; \nu)$: wartość krytyczna rozkładu t (Studenta) z ν stopniami swobody

Długość przedziału: $d = 2t(\alpha; n-1)\dfrac{S}{\sqrt{n}}$

Przedział ufności dla wartości średniej

Przykład - treść
Oszacować przeciętną ilość punktów uzyskiwanych na klasówce.

$$n = 300 \qquad \sum x_i = 176.566 \qquad \sum x_i^2 = 107.845302$$

Przedział ufności dla wartości średniej

Przykład - rozwiązanie
Populacja: Słuchacze podstawowego kursu statystyki
Cecha X: ilość punktów zdobytych na klasówce
Założenie: cecha X ma rozkład normalny $N(\mu, \sigma^2)$
Zadanie: oszacować parametr μ
Technika statystyczna: przedział ufności dla średniej
poziom ufności $1 - \alpha = 0.95$

Przedział ufności dla wartości średniej

Przykład - rozwiązanie
Obliczenia

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{176.566}{300} = 0.589$$

$$\mathrm{var}X = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 107.845302 - \frac{176.566^2}{300} = 3.92679$$

$$s^2 = \frac{3.92679}{300-1} = 0.01313, \qquad s = \sqrt{s^2} = 0.11460$$

Przedział ufności dla wartości średniej

Przykład - rozwiązanie
Obliczenia cd

$$t(0.05; 299) \approx 1.96$$

$$t(0.05; 299)\frac{s}{\sqrt{n}} = 1.96\frac{0.11460}{\sqrt{300}} = 0.01297$$

$$(0.589 - 0.013, 0.589 + 0.013) = (0.576, 0.602)$$

Przedział ufności dla wartości średniej

Przykład - rozwiązanie
Odpowiedź: $\mu \in (0.576, 0.602)$

Wniosek. Przeciętna ilość punktów zdobywana na klasówce jest liczbą z przedziału $(0.576, 0.602)$. Zaufanie do tego wniosku wynosi 95%.

Jednostronny przedział ufności dla wartości średniej

Rozkład normalny $N(\mu, \sigma^2)$
Wariancja σ^2 jest nieznana
Poziom ufności: $1 - \alpha$

$$(-\infty, \quad \bar{X} + t(2\alpha; n-1)\frac{S}{\sqrt{n}})$$

$$(\bar{X} - t(2\alpha; n-1)\frac{S}{\sqrt{n}}, \quad \infty)$$

5.4 Przedział ufności dla wariancji

Przedział ufności dla wariancji

Rozkład normalny $N(\mu, \sigma^2)$
Wartość średnia μ jest nieznana
Poziom ufności: $1 - \alpha$

$$\left(\frac{\text{var}X}{\chi^2\left(\frac{\alpha}{2}; n-1\right)}, \quad \frac{\text{var}X}{\chi^2\left(1-\frac{\alpha}{2}; n-1\right)}\right)$$

$\chi^2(\alpha; \nu)$ jest stablicowaną wartością krytyczną rozkładu chi–kwadrat z ν stopniami swobody

5.5 Przedział ufności dla frakcji

Przedział ufności dla prawdopodobieństwa sukcesu

Rozkład dwumianowy $B(n, p)$
k - liczba sukcesów w próbie n elementowej
Poziom ufności: $1 - \alpha$

Dokładny przedział ufności

$$\left(p_1\left(1 - \frac{\alpha}{2}; k, n-k\right)\ ,\quad 1 - p_1\left(1 - \frac{\alpha}{2}; n-k, k\right)\right)$$

Przybliżony przedział ufności

$$\left(\hat{p} - u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

u_α jest kwantylem rzędu α rozkładu $N(0, 1)$.

Przedział ufności dla prawdopodobieństwa sukcesu

Przykład - treść
Oszacować odsetek ocen dostatecznych otrzymywanych na klasówce.

$$n = 300 \quad k = 88$$

Przedział ufności dla prawdopodobieństwa sukcesu

Przykład - rozwiązanie
Populacja: Słuchacze podstawowego kursu statystyki
Cecha X**:** ocena dostateczna/inna z klasówki
Założenie: cecha X ma rozkład dwupunktowy $D(p)$
Zadanie: oszacować parametr p
Technika statystyczna: przybliżony przedział ufności dla prawdopodobieństwa p
poziom ufności $1 - \alpha = 0.95$

Przedział ufności dla prawdopodobieństwa sukcesu

Przykład - rozwiązanie
Obliczenia

$$\widehat{p} = \frac{88}{300} = 0.29$$

$$u_{1-\alpha/2} = u_{0.975} = 1.96$$

$$0.29 - 1.96\sqrt{\frac{0.29(1-0.29)}{300}} = 0.2387$$

$$0.29 + 1.96\sqrt{\frac{0.29(1-0.29)}{300}} = 0.3413$$

Przedział ufności dla prawdopodobieństwa sukcesu

Przykład - rozwiązanie
Odpowiedź: $p \in (0.2387, 0.3413)$

Wniosek. Odsetek ocen dostatecznych zdobywanych na klasówce jest liczbą z przedziału $(23.87\%, 34.13\%)$. Zaufanie do tego wniosku wynosi 95%.

Jednostronny przedział ufności dla prawdopodobieństwa sukcesu

Rozkład dwumianowy $B(n,p)$
k - liczba sukcesów w próbie n elementowej
Poziom ufności: $1-\alpha$

Dokładne jednostronne przedziały ufności

$$(0\ , 1 - p_1\left(1-\alpha; n-k, k\right))$$

$$(p_1\left(1-\alpha; k, n-k\right)\ , 1)$$

Jednostronny przedział ufności dla prawdopodobieństwa sukcesu

Rozkład dwumianowy $B(n,p)$
k - liczba sukcesów w próbie n elementowej
Poziom ufności: $1-\alpha$

Przybliżone jednostronne przedziały ufności

$$\left(0,\ \hat{p} + u_{1-\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$\left(\hat{p} - u_{1-\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ 1\right)$$