

STATYSTYKA I PRAWDOPODOBIEŃSTWO

SZCZEGÓŁOWY OPIS:

W pierwszych czterech rozdziałach skryptu zaczynamy od przeglądu klasycznych narzędzi probabilistycznych, takich jak zasady kombinatoryki, koncepcja prawdopodobieństwa, definicje i twierdzenia takie jak prawdopodobieństwo warunkowe, wzór Bayesa, niezależność zdarzeń.

W rozdziałach 5-8 wprowadzamy pojęcie dyskretnych i ciągłych zmiennych losowych oraz ich rozkłady. Przedstawiamy najważniejsze przykłady jak rozkład dwumianowy, rozkład Poissona i rozkład normalny. Prezentując liczne przykłady pokazujemy ich zastosowania.

Ostatnie rozdziały poświęcone są głównie statystyce. Podajemy podstawowe definicje przedstawiamy główne idee. Na koniec wykonujemy testy statystyczne i przedstawiamy metodę analizy regresji.

Na koniec pokazujemy zastosowanie rozkładu Poissona w teorii niezawodności. Dostępny jest również zestaw ćwiczeń do indywidualnych zajęć i nauki.

CEL: Zdobycie umiejętności rozwiązywania problemów w czasie rzeczywistym, zwłaszcza związanych z morzem i jednostkami pływającymi, za pomocą metod probabilistycznych i statystycznych. Okazuje się, że wiele problemów w dziedzinie morskiej nie jest deterministycznych i potrzebne są narzędzia probabilistyczne.

Efekty kształcenia:

1. Obliczenia z wykorzystaniem klasycznej definicji prawdopodobieństwa i przy użyciu podstawowych zasad kombinatoryki.
2. Obliczanie prawdopodobieństwa całkowitego i warunkowego
3. Sprawdzanie niezależności zdarzeń, poznanie ich zastosowania
4. Poznanie i stosowanie dyskretnych rozkładów zmiennych losowych
5. Poznanie, jak ważne są rozkłady normalne
6. Wykonywanie testów statystycznych
7. Obliczanie regresji liniowej
8. Poznanie zastosowania prawdopodobieństwa i statystyki w dziedzinach morskich

Wymagana wiedza:

Matematyka elementarna, arytmetyka, teoria mnogości, elementy rachunku różniczkowego, w szczególności całki oznaczone.

Związek z rzeczywistymi zagadnieniami morskimi:

Powodem, dla którego stosujemy metodę probabilistyczną i statystyczną w problemach morskich, jest to, że problemy te są często nieprzewidywalne. Ruch statków i okrętów na

morzu jest częściowo losowy. Częściowo oznacza to, że oczywiście możemy sterować jednostką i reagować na rzeczywista sytuacja, ale w wielu szczególnie trudnych momentach to nie wystarczy. Nie możemy wystarczająco wcześnie przewidzieć wielu potencjalnie ryzykownych zdarzeń. Wynika to z faktu, że nie możemy opisać morza w sposób deterministyczny. Fale morskie nie są deterministyczne. Nie ma "równań morskich", które mogłyby opisać, jak fale będą zachowywać się w określonym miejscu i czasie, nawet jeśli znamy tam obecną sytuację.

Drugim powodem, dla którego prawdopodobieństwo i statystyka są tak ważne w dziedzinie morskiej, jest teoria niezawodności, którą stosujemy do większości elementów wyposażenia statków. Jednostki spędzają zazwyczaj długie dni w morzu, dlatego musimy mieć pewność, że wszystkie urządzenia mechaniczne, elektryczne i elektroniczne będą działać prawidłowo.

Spis treści:

1. Podstawowe reguły kombinatoryki
- 2 Zdarzenia. Prawdopodobieństwo
- 3 Prawdopodobieństwo geometryczne
- 4 Wzór Bayesa. Prawdopodobieństwo całkowite
- 5 Dyskretne rozkłady prawdopodobieństwa
- 6 Rozkład dwumianowy i rozkład Poissona
- 7 Ciągłe rozkłady prawdopodobieństwa
- 8 Rozkład normalny
- 9 Statystyka: podstawowa definicje i notacje 10
- 10 Testowanie hipotez statystycznych. Błędy wnioskowań statystycznych
- 11 Regresja, korelacja
- 12 Zastosowania
- 13 Ćwiczenia

STATYSTYKA I PRAWDOPODOBIENSTWO

1. Podstawowe reguły kombinatoryki

Reguła mnożenia (reguła iloczynu)

Jeśli zbiór A zawiera n elementów, a zbiór B zawiera m elementów wtedy jest $n \cdot m$ różnych par (a, b) gdzie a należy do A a b należy do B .

Przykład 1.1

Ile jest dwucyfrowych parzystych liczb naturalnych?

Rozwiązanie:

Dwucyfrowe parzyste liczby naturalne mogą być reprezentowane jako uporządkowane pary (a, b) , gdzie $a \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\} = A$ i $b \in \{0, 2, 4, 6, 8\} = B$. Ilość elementów w zbiorze A jest $n = 9$ oraz ilość elementów w zbiorze B jest $m = 5$. Stąd całkowita liczba dwucyfrowych parzystych liczb naturalnych wynosi

$$9 \cdot 5 = 45.$$

Zauważmy, że ponieważ uporządkowane trójki liczbowe (a, b, c) mogą być reprezentowane jako uporządkowane pary $((a, b), c)$, możemy rozszerzyć regułę mnożenia na przypadek trzech zbiorów.

Reguła mnożenia (trzy i więcej zbiorów)

Jeśli zbiór A zawiera n elementów, a zbiór B zawiera m elementów a zbiór C zawiera p elementów, wtedy są $n \cdot m \cdot p$ różne uporządkowane trójki (a, b, c) , gdzie a należy do A , b należy do B i c należy do C . Podobna zasada dotyczy czterech, pięciu i więcej zbiorów.

Przykład 1.2

Ile jest trzycyfrowych nieparzystych liczb naturalnych?

Rozwiązanie:

Trzycyfrowe nieparzyste liczby naturalne mogą być reprezentowane jako uporządkowane trójki (a, b, c) , gdzie $a \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\} = A$, $b \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\} = B$ a $c \in \{1, 3, 5, 7, 9\} = C$. Ilość elementów w zbiorze A wynosi $n = 9$, liczba elementów w zbiorze B jest $m = 10$ oraz ilość elementów w zbiorze C wynosi $p = 5$. Tak więc całkowita liczba trzycyfrowych nieparzystych liczb naturalnych wynosi

$$9 \cdot 10 \cdot 5 = 450.$$

Reguła dodawania (reguła sumy)

Jeśli zbiór A składa się z n elementów, a zbiór B składa się z m elementów oraz $A \cap B$ składa się z l elementów, wtedy zbiór $A \cup B$ składa się z $n + m - l$ elementów.

Przykład 1.3

Alicja kupiła w sklepie spożywczym: jabłka, śliwki, banany, jagody i pomarańcze. Jej mąż Jan kupił w innym sklepie jeżyny, agrest, winogrona, śliwki, gruszki, grejpfruty i banany. Ile różnych rodzajów owoców zjedzą w domu?

Rozwiązanie:

Mamy

$$n = 5, m = 7, l = 2.$$

Stąd Alicja i jej mąż mają

$$m + n - l = 5 + 7 - 2 = 10$$

różnych owoców.

Definicja: Permutacje

Permutacja to dowolny układ elementów skończonego zbioru w określonym porządku.

Liczba permutacji

Liczba wszystkich permutacji n – elementowego zbioru wynosi

$$P_n = n! = n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1$$

gdzie $n!$ nazywa się silnią n .

Przykład 1.4

Na ile sposobów możemy ułożyć pięć różnych książek na półce.

Rozwiązanie:

Musimy obliczyć liczbę wszystkich permutacji 5 – elementowego zbioru. Liczba ta jest równa

$$P_5 = 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120.$$

Ćwiczenie:

Na ile sposobów siedem różnych statków może zacumować w porcie?

Rozwiązanie:

5040.

Wariacje bez powtórzeń

Ciągi długości k bez powtórzeń, których elementy są zaczerpnięte z n – elementowego zbioru ($k \leq n$) są nazywane k -wyrazowymi wariacjami bez powtórzeń zbioru n -elementowego. (*wariacjami bez powtórzeń*).

Liczba takich wariacji bez powtórzeń wynosi

$$V_n^k = n \cdot (n - 1) \cdot \dots \cdot (n - k + 1) = \frac{n!}{(n - k)!}$$

Przykład 1.5

Na ile sposobów pięć osób można usiąść w rzędzie złożonym z jedenastu krzeseł.

Rozwiązanie:

Poszukiwana liczba N to liczba 5- cio wyrazowych wariacji bez powtórzeń ze zbioru 11- to elementowego.

Zatem, ponieważ

$$n = 11, k = 5$$

$$N = V_n^k = \frac{11!}{(11 - 5)!} = \frac{11!}{6!} = 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 55440.$$

Przykład 1.6.

Istnieje dziewięć wolnych miejsc do dokowania w porcie. Na ile sposobów może tam zacumować sześć różnych statków?

Rozwiązanie:

Liczba N 6- cio wyrazowych wariacji bez powtórzeń ze zbioru 9- cio elementowego powinna być policzona.

Mamy $n = 9$, $k = 6$ zatem

$$N = V_9^6 = \frac{9!}{(9 - 6)!} = \frac{9!}{3!} = 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 = 60480.$$

Wariacje z powtórzeniami

Liczba różnych ciągów długości k , które można utworzyć z elementów zbioru n -elementowego, gdy dozwolone są powtórzenia, wynosi

$$W_n^k = n^k.$$

Różne ciągi długości k które można utworzyć ze zbioru n – elementowego, gdy dopuszczone są powtórzenia nazywamy wariacjami z powtórzeniami.

W dalszej części przyjmiemy, że każdy układ liter tworzy słowo.

Przykład 1.7

Ile trzyliterowych słów można ułożyć z liter a, b, c, d ?

Rozwiązanie:

Liczba N różnych ciągów (słów) o długości 3 wziętych z 4 – elementowego zbioru $\{a, b, c, d\}$, gdy powtórzenia są dozwolone, powinna być obliczona. Wyrażenie "powtórzenia są dozwolone" oznacza, że każda litera może być użyta więcej niż raz, aby utworzyć słowo. Stąd

$$n = 4, k = 3 \quad \text{oraz} \quad N = W_4^3 = 4^3 = 64.$$

Przykład 1.8

Ile czteroliterowych słów można ułożyć z liter a, b, c ?

Rozwiązanie:

Liczbę N różnych ciągów (słów) o długości 4 wziętych ze zbioru 3 – elementowego $\{a, b, c\}$, gdy powtórzenia są dopuszczalne powinna być policzona. Stąd

$$n = 3, k = 4 \quad \text{i} \quad N = W_3^4 = 3^4 = 81.$$

Uwaga:

W niektórych problemach wzór na liczbę wszystkich permutacji z powtórzeniami nie jest łatwy do zastosowania. Łatwiej jest stosować zasadę mnożenia.

Przykład 1.9

Na ile sposobów pięć statków z kraju A może zadokować w trzech różnych portach w kraju B?

Rozwiązanie:

1-szy sposób:

Niech porty w kraju B będą oznaczone symbolami h_1, h_2, h_3 .

Każdy statek ma trzy możliwości wyboru: zadokować w h_1 , w h_2 lub w h_3 .

Stosując regułę mnożenia widzimy, że:

dwa statki mają $3 \cdot 3 = 9$ możliwości wyboru,

trzy statki mają $3 \cdot 3 \cdot 3 = 27$ możliwości wyboru,

cztery statki mają $3 \cdot 3 \cdot 3 \cdot 3 = 81$ możliwości wyboru.

I ostatecznie widzimy, że pięć statków ma

$$N = 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^5 = 243$$

możliwości wyboru.

2- gi sposób:

Mamy ciągi długości $k = 5$ (pięć statków), które składają się z liczb z 3 – elementowego zbioru $\{1,2,3\}$ (porty) ($n = 3$). Powtórzenia są dozwolone, ponieważ dwa lub więcej statków może wybrać ten sam port. Więc

$$N = 3^5 = 243.$$

Definicja: Kombinacje

Każdy k – elementowy podzbiór zbioru n elementowego ($k \leq n$) nazywamy kombinacją k -elementową zbioru n -elementowego.

Liczba kombinacji

Liczba wszystkich możliwości wyboru podzbioru of k elementowego ze zbioru n elementowego (porządek jest nieistotny) wynosi

$$C(n, k) = \binom{n}{k} = \frac{n!}{k!(n-k)!},$$

gdzie $\binom{n}{k}$ jest współczynnikiem dwumianowym.

Przykład 1.10

Anne, Beth, Charlie i Donald próbują wygrać dwa bilety na Bahamy.

- Jakie są wszystkie kombinacje zwycięzców?
- Jaka jest liczba wszystkich możliwości wyboru zwycięzców?
- Jeśli Anne i Charlie zostaną wybrani, czy ma znaczenie, kto dostał pierwszy, a kto drugi bilet?

Rozwiązanie:

- a) Wszystkie kombinacje zwycięzców są następujące:

$$\{A, B\}, \{A, C\}, \{A, D\}, \{B, C\}, \{B, D\}, \{C, D\},$$

gdzie A oznacza Anne, B oznacza Beth itp.

- b) Liczba wszystkich możliwości wyboru zwycięzców wynosi sześć, ponieważ jest to liczba wszystkich kombinacji zwycięzców (por. a)). Z drugiej strony, stosując formułę otrzymujemy tę samą liczbę.

$$N = C(4,2) = \binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{4 \cdot 3}{2 \cdot 1} = 6.$$

- c) Nie, nie ma to znaczenia, ponieważ w zbiorach (podzbiorach) nie mają zadanego porządku. Mamy $\{A, C\} = \{C, A\}$.

Uwaga:

Następujące związki dla współczynników dwumianowych są przydatne w obliczeniach:

- $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n(n-1)\cdots(k+1)}{(n-k)!}$
- $\binom{n}{k} = \binom{n}{n-k}$
- $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{1} = \binom{n}{n-1} = n$

Ćwiczenie:

1. Obliczyć $\binom{4000}{3998}$.
2. Ile podzbiorów zbioru z czterema tysiącami elementów ma dwa elementy?

Rozwiązanie:

1. 7 998 000
2. 7 998 000

Przykład 1.11

Ile różnych komitetów składających się z czterech członków można utworzyć z klubu z dziesięcioma członkami?

Rozwiązanie:

Komitet jest 4 – elementowym podzbiorem 10 – elementowego zbioru członków klubu. Tak więc liczba N różnych komitetów jest równa $C(10,4)$. Mamy

$$N = C(10,4) = \binom{10}{4} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = 210.$$

2. Zdarzenia. Prawdopodobieństwo

Przykład 2.1

Załóżmy, że przeprowadzamy eksperyment, rzucając dwiema różnymi uczciwymi monetami. Niech O reprezentuje orła i R reprezentuje reszkę. Wtedy każdy możliwy wynik eksperymentu jest elementem zbioru

$$\Omega = \{(O, O), (O, R), (R, O), (R, R)\}.$$

Definicja: przestrzeń zdarzeń elementarnych

Przestrzenią zdarzeń elementarnych (*przestrzenią próbek losowych*) nazywamy zbiór Ω złożony z elementów, które odpowiadają jeden do jednego wynikom eksperymentu. Każdy z elementów zbioru Ω jest nazywany *zdarzeniem elementarnym*.

Definicja: zdarzenie

Zdarzeniem nazywamy podzbiór przestrzeni zdarzeń elementarnych.

Uwaga:

Przestrzeń zdarzeń elementarnych Ω sama jest zdarzeniem (*zdarzenie pewne*), każde elementarne zdarzenie jest zdarzeniem, zbiór pusty \emptyset jest zdarzeniem (*zdarzenie niemożliwe*).

Przykład 2.2

Każda litera słowa *lampart* jest zapisana na osobnej karcie, a karty są tasowane. Eksperyment polega na losowym wybieraniu karty. Podaj przestrzeń zdarzeń elementarnych.

Odpowiedź:

$\Omega = \{l, e, o, p, a, r, d\}$ (również $\Omega = \{e, p, l, a, d, o, r\}$ ponieważ kolejność nie jest ważna). Załóżmy teraz, że interesuje nas, czy litera wybrana losowo jest samogłoską. Nazwijmy zdarzeniem A zdarzenie polegające na wyciągnięciu samogłoski. Zdarzenie A jest zatem zdarzeniem polegającym na wybraniu jednej z liter e, o, a . Możemy je reprezentować jako zbiór $A = \{e, o, a\}$ który jest podzbiorem przestrzeni zdarzeń elementarnych Ω .

Definicja: Klasyczna definicja prawdopodobieństwa

Niech Ω będzie przestrzenią zdarzeń elementarnych pewnego eksperymentu, w którym może zajść n możliwych wyników, tak samo możliwych. Jeśli zdarzenie A jest podzbiorem Ω takim, że A zawiera k elementów, wtedy prawdopodobieństwo zdarzenia A , oznaczane $P(A)$, jest dane wzorem

$$P(A) = \frac{k}{n}.$$

Przykład 2.3

Założmy, że przeprowadzamy eksperyment, rzucając dwiema różnymi sześciostronnymi kośćmi do gry. Jakie jest prawdopodobieństwo, że w sumie wypadnie dziesięć?

Rozwiązanie:

Liczba zdarzeń elementarnych wynosi $n = 6 \cdot 6 = 36$. Oznaczmy przez A zdarzenie, że otrzymamy w sumie dziesięć. Mamy

$$A = \{(6, 4), (5, 5), (4, 6)\}$$

więc $k = 3$ i

$$P(A) = \frac{3}{36} = \frac{1}{12}.$$

Niech Ω będzie teraz skończonym lub nieskończonym zbiorem. Zbiór Ω nazywamy przestrzenią zdarzeń elementarnych. Niech \mathcal{F} oznacza pewną klasę podzbiorów Ω . Elementy \mathcal{F} będą nazywane zdarzeniami. Jeśli zbiór Ω jest skończony, wszystkie jego podzbiory są zdarzeniami. Jeśli Ω nie jest zbiorem skończonym, za zdarzenia przyjmujemy dużą nieskończoną klasę podzbiorów Ω .

Definicja: Prawdopodobieństwo (aksjomaty)

Prawdopodobieństwo zdarzenia A jest liczbą rzeczywistą $P(A)$, która spełnia następujące trzy warunki:

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. Dla każdego ciągu, A_1, A_2, \dots wzajemnie wykluczających się zdarzeń zachodzi

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots.$$

Właściwości prawdopodobieństwa

1. $P(\emptyset) = 0$
2. $0 \leq P(A) \leq 1$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $A \subset B \Rightarrow P(A) \leq P(B)$
5. $P(\bar{A}) = 1 - P(A)$
6. $P(A \setminus B) = P(A) - P(A \cap B)$
7. $B \subset A \Rightarrow P(A \setminus B) = P(A) - P(B)$
8. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

3. Prawdopodobieństwo geometryczne

Jako przestrzeń zdarzeń elementarnych rozważamy 1D, 2D lub 3D zbiory o skończonej mierze $m(\Omega)$ którą jest odpowiednio: długość, powierzchnia lub objętość. Jako zdarzenia przyjmujemy podzbiory Ω mające skończoną miarę.

Definicja: prawdopodobieństwo geometryczne

Prawdopodobieństwo, że punkt x leżący w zbiorze Ω leży również w jego podzbiorze A z miarą skończoną $m(A)$ jest zdefiniowany jako stosunek

$$P(A) = \frac{m(A)}{m(\Omega)}.$$

Przykład 3.1

Okrąg jest wpisany w kwadrat. Punkt jest wybierany losowo z obszaru kwadratu. Oblicz prawdopodobieństwo, że leży on wewnątrz okręgu.

Rozwiązanie:

Niech Ω będzie kwadratem i A będzie okręgiem o promieniu $r > 0$ który jest wpisany w Ω . Mamy

$$m(A) = \pi r^2, \quad m(\Omega) = 4r^2.$$

$$P(A) = \frac{\pi r^2}{4r^2} = \frac{\pi}{4}$$

Przykład 3.2

Dwa statki mają czas przybycia do portu między **2** a **3** po południu. Oba statki muszą zacumować na tym samym nabrzeżu. Po zadokowaniu każdy statek potrzebuje **10** minut na uzupełnienie zapasów i opuszczenie doku. Jakie jest prawdopodobieństwo, że statki nie będą musiały czekać na zacumowanie?

Rozwiązanie:

Jako jednostkę czasu przyjmujemy **10** minut. Przedziałem czasu pomiędzy **2** i **3** popołudniu jest zatem $[0, 6]$. Niech x będzie czasem, kiedy pierwsza łódź przybędzie, i y będzie czasem, kiedy druga łódź przybędzie. Mamy $x, y \in [0, 6]$, $|x - y| \geq 1$.

Niech $\Omega = [0, 6] \times [0, 6]$ będzie przestrzenią zdarzeń elementarnych. Pole powierzchni zbioru Ω wynosi $\mu(\Omega) = 36$.

Wydarzenie, że oba statki nie będą musiały czekać, może być reprezentowane przez zbiór

$$A = \{(x, y) \in \Omega: |x - y| \geq 1\}.$$

Ponieważ

$$A = \{(x, y) \in \Omega: y \leq x - 1 \text{ lub } y \geq x + 1\},$$

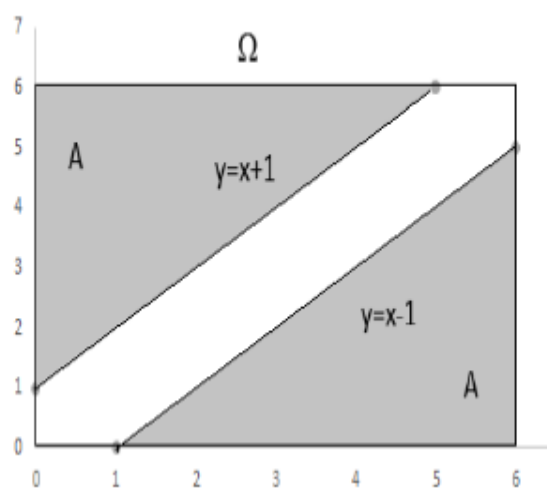
pole zbioru A wynosi

$$\mu(A) = \frac{1}{2} \cdot 5 \cdot 5 + \frac{1}{2} \cdot 5 \cdot 5 = 25.$$

Ze wzoru na prawdopodobieństwo geometryczne mamy

$$P(A) = \frac{m(A)}{m(\Omega)} = \frac{25}{36} \approx 0.6944.$$

Stąd wynika, że mamy **69.44%** szansy, że statki nie będą musiały czekać na zacumowanie.



Rys.: Ilustracja do Przykładu 3.2

4. Wzór Bayesa. Prawdopodobieństwo całkowite.

Definicja: zdarzenia niezależne

Dwa zdarzenia A i B są nazywane niezależnymi (statystycznie niezależnymi), jeżeli

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

Uwaga:

Fizycznie niezależne eksperymenty są statystycznie niezależne, np. kolejne rzuty monetą, kostką itp.

Przykład 4.1

Założmy, że dwie kule są losowane po kolei z torby zawierającej 4 czerwone i 3 czarne kule. Po wylosowaniu pierwszej kuli jest ona zwracana do torby przed wylosowaniem drugiej kuli. Jakie jest prawdopodobieństwo, że obie wybrane kule są czerwone?

Rozwiązanie:

Niech A będzie zdarzeniem polegającym na wylosowaniu czerwonej kuli za pierwszym razem oraz B zdarzenie wylosowania czerwonej kuli za drugim razem. Przestrzeń zdarzeń elementarnych Ω składa się ze wszystkich uporządkowanych par (x, y) , gdzie zarówno x i y oznaczają elementy zbioru 7 wyników (4 czerwone, 3 czarne kule).

Reguła mnożenia oznacza, że Ω składa się $7 \cdot 7 = 49$ elementów. Ponieważ $A \cap B$ składa się z uporządkowanych par postaci (red, red), mamy $4 \cdot 4 = 16$ elementów w $A \cap B$. Stąd,

$$P(A \cap B) = \frac{16}{49} = \frac{4}{7} \cdot \frac{4}{7} = P(A)P(B).$$

Widzimy więc, że zdarzenia A and B są niezależne.

Przykład 4.2

Założmy, że dwie kule są losowane bez zwracania (pierwsza kula nie jest zwracana przed losowaniem drugiej) losowo z torby zawierającej 4 czerwone i 3 czarne kule.

Dowolna z 7 kul może być wylosowana za pierwszym razem. Jedna z 6 kul może być wybrana podczas drugiego losowania (losujemy bez zwracania).

Stąd jest $7 \cdot 6 = 42$ elementów w zbiorze Ω (ordered pairs without repetitions).

Niech A będzie zdarzeniem polegającym na wylosowaniu kuli czerwonej za pierwszym razem i B zdarzeniem polegającym na wylosowaniu kuli czerwonej za drugim razem.

Jest oczywiście $4 \cdot 6 = 24$ elementów w zbiorze A i $6 \cdot 4 = 24$ elementów w zbiorze B . To daje

$$P(A) = P(B) = \frac{24}{42} = \frac{4}{7}.$$

Mamy $4 \cdot 3 = 12$ elementów zbioru $A \cap B$. Stąd, mamy

$$P(A \cap B) = \frac{12}{42} = \frac{2}{7} \neq \frac{4}{7} \cdot \frac{4}{7} = P(A)P(B).$$

Stąd zdarzenia A i B są zależne.

Uwaga:

Ponieważ w [Przykładzie 4.2](#) $P(A \cap B) = \frac{12}{42} = \frac{4}{7} \cdot \frac{3}{6}$ i $P(A) = \frac{4}{7}$ możemy zinterpretować drugi czynnik $\frac{3}{6}$ jako prawdopodobieństwo, że druga wyciągnięta kula jest czerwona pod warunkiem, że pierwsza wyciągnięta kula była czerwona. Oznaczamy to przez $P(A|B)$.

Ostatnia uwaga sugeruje ogólną zależność

Wzór na prawdopodobieństwo całkowite

Niech $P(A)$ oznacza prawdopodobieństwo zdarzenia A , i $P(B|A)$ oznacza prawdopodobieństwo zdarzenia B przy założeniu, że zaszło zdarzenie A . Jeśli $P(A \cap B)$ jest prawdopodobieństwem, że zaszły zdarzenia A i B , wtedy

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

Jeśli $P(A) > 0$ otrzymujemy wzór na *prawdopodobieństwo warunkowe*

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

oraz gdy $P(B) > 0$

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

Uwaga:

W **Przykładzie 4.2** $P(A|B) = \frac{3}{6} = \frac{1}{2}$ oznacza prawdopodobieństwo zdarzenia A pod warunkiem, że zaszło zdarzenie B tzn., że pierwsza kula wylosowana była czerwona pod warunkiem, że druga wylosowana kula jest również czerwona.

$P(A|B)$ nazywa się prawdopodobieństwem warunkowym A pod warunkiem B , lub prawdopodobieństwem A pod warunkiem B .

Uwaga:

Jeśli A i B są zdarzeniami niezależnymi, to $P(A|B) = P(A)$ i $P(B) > 0$.

Ćwiczenie:

Pokaż, że jeśli A i B są zdarzeniami niezależnymi, wtedy

A i \bar{B} są zdarzeniami niezależnymi.

Rozwiązanie:

$$\begin{aligned} P(A \cap \bar{B}) &= P(A \cap (\Omega \setminus B)) = P(A \setminus B) = P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) = P(A)P(\bar{B}). \end{aligned}$$

Ćwiczenie:

Wykazać, że jeśli A i B są zdarzeniami niezależnymi, to również \bar{A} i \bar{B} są zdarzeniami niezależnymi.

Rozwiązanie:

$$\begin{aligned} P(\bar{A} \cap \bar{B}) &= P((\Omega \setminus A) \cap (\Omega \setminus B)) = P(\Omega \setminus (A \cup B)) = 1 - P(A \cup B) \\ &= 1 - P(A) - P(B) + P(A \cap B) = 1 - P(A) - P(B) + P(A)P(B) \\ &= (1 - P(A))(1 - P(B)) = P(\bar{A})P(\bar{B}). \end{aligned}$$

Wzór Bayesa

Przypuśćmy, że $P(B) > 0$. Wtedy

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

Mówimy, że zdarzenia E_1, \dots, E_n tworzą rozbitcie przestrzeni zdarzeń elementarnych Ω jeśli zdarzenia E_1, \dots, E_n są wzajemnie rozłączne (wykluczające się) oraz

$$\Omega = E_1 \cup \dots \cup E_n.$$

Wzór na prawdopodobieństwo całkowite

Załóżmy, że przestrzeń Ω jest rozbita na zdarzenia E_1, \dots, E_n .

Wtedy dla każdego zdarzenia A

$$P(A) = P(A|E_1)P(E_1) + \dots + P(A|E_n)P(E_n).$$

Przykład 4.3

Trzy identyczne urny są oznaczone 1, 2, 3. Pierwsza urna zawiera 3 czerwone i 3 niebieskie kule. Druga urna zawiera 4 czerwone i 2 niebieskie kule. Trzecia urna zawiera 1 czerwoną i 5 niebieskich kul. Najpierw losowo wybierana jest urna, a następnie kula jest losowo wybierana z urny.

1) Jaka jest prawdopodobieństwo, że wybrana kula jest niebieska?

Rozwiązanie:

E_i – zdarzenie: urna nr i została wybrana. Mamy $P(E_i) = \frac{1}{3}$.

A – zdarzenie: niebieska kula została wybrana. Mamy

$$P(A|E_1) = \frac{1}{2}, \quad P(A|E_2) = \frac{1}{3}, \quad P(A|E_3) = \frac{5}{6}$$

oraz

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + P(A|E_3)P(E_3) = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} + \frac{5}{6} \cdot \frac{1}{3} = \frac{5}{9}.$$

2) Wiedząc, że wylosowana została kula niebieska oblicz prawdopodobieństwo, że wcześniej została wybrana urna nr 2?

Rozwiązanie:

Za pomocą wzoru Bayesa otrzymujemy

$$P(E_2|A) = \frac{P(A|E_2)P(E_2)}{P(A)} = \frac{\frac{1}{3} \cdot \frac{1}{3}}{\frac{5}{9}} = \frac{1}{5}.$$

Przykład 4.4

Pewna choroba ma częstość występowania w populacji 1%. Załóżmy, że dla pewnego testu diagnostycznego wskaźnik fałszywie ujemny wynosi 10% a wskaźnik fałszywie dodatni wynosi 2%. Oblicz prawdopodobieństwo, że osoba, wybrana losowo z populacji, która ma pozytywny wynik testu, faktycznie ma tę chorobę.

Rozwiązanie:

Zdefiniujmy zdarzenia dla danej osoby:

D : ma chorobę, T : test wskazuje na chorobę (test jest pozytywny)

Zdarzeniami dopełniającymi są:

\bar{D} : nie ma choroby, \bar{T} : test nie wskazuje na chorobę (test jest negatywny).

Ponieważ częstość występowania choroby wynosi 1%, mamy

$$P(D) = 0.01$$

oraz dla zdarzenia dopełniającego

$$P(\bar{D}) = 1 - P(D) = 0.99.$$

Ponieważ wskaźnik fałszywie ujemny wynosi 10%, mamy

$$P(\bar{T}|D) = 0.1.$$

Ponieważ wskaźnik fałszywie dodatni wynosi 2%, mamy

$$P(T|D) = 0.02.$$

Musimy obliczyć $P(D|T)$.

Stosujemy wzór Bayesa

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)},$$

gdzie

$$P(T|D) = 1 - P(\bar{T}|D) = 1 - 0.1 = 0.9.$$

Ze wzoru na prawdopodobieństwo całkowite

$$P(T) = P(T|D)P(D) + P(T|\bar{D})P(\bar{D}) = 0.9 \cdot 0.01 + 0.02 \cdot 0.99 = 0.0288.$$

Ostatecznie,

$$P(D|T) = \frac{0.9 \cdot 0.01}{0.0288} = 0.3125.$$

Tylko 31.25% losowo wybranych osób, których wynik testu jest pozytywny, ma w rzeczywistości chorobę. Wynik wydaje się zaskakujący. Niska częstość występowania choroby w populacji jest tutaj zmienną która o tym decyduje.

Ćwiczenie:

W powyższym przykładzie obliczyć prawdopodobieństwo, że osoba, wybrana losowo z populacji, która ma wynik negatywny testu, faktycznie nie ma choroby.

Rozwiązanie:

0.99897

5. Dyskretne rozkłady prawdopodobieństwa

Definicja: zmienna losowa

Zmienną losową X nazywamy funkcję o wartościach rzeczywistych zdefiniowaną na przestrzeni zdarzeń elementarnych Ω .

Uwaga:

Piszemy $X: \Omega \rightarrow \mathbb{R}$ i $X(\omega)$ jest wartością X dla zdarzenia $\omega \in \Omega$.

Przykład 5.1

Przypuśćmy, że rzucamy monetą. Wtedy, przestrzenią zdarzeń elementarnych jest $\Omega = \{O, R\}$, gdzie O reprezentuje orła R reprezentuje reszkę. Definiujemy zmienną losową $X: \Omega \rightarrow \mathbb{R}$ wzorem

$$X(O) = 1 \quad \text{oraz} \quad X(R) = 0.$$

Przykład 5.2

Założmy, że rzucamy kostką do gry. Przestrzenią zdarzeń elementarnych jest $\Omega = \{1, 2, 3, 4, 5, 6\}$, gdzie liczby reprezentują ilość oczek na ściankach kostki. Definiujemy zmienną losową $X: \Omega \rightarrow \mathbb{R}$ następująco

$$X(1) = 1, \quad X(2) = 2, \quad X(3) = 3, \quad X(4) = 4, \quad X(5) = 5, \quad X(6) = 6.$$

Mówimy po prostu, że X jest to liczba oczek, które wypadły.

Uwaga:

Jeśli moneta, którą rzucamy w Przykład 5.1 jest uczciwa prawdopodobieństwo, że wypadnie orzeł wynosi $\frac{1}{2}$ i prawdopodobieństwo, że wypadnie reszka wynosi $\frac{1}{2}$. Zapisujemy to następująco:

$$P(O) = P(\{\omega \in \Omega: X(\omega) = 1\}) = \frac{1}{2}$$

oraz

$$P(R) = P(\{\omega \in \Omega: X(\omega) = 0\}) = \frac{1}{2}.$$

Dla uproszczenia piszemy: $P(X = x) = P(\{\omega \in \Omega: X(\omega) = x\})$.

Podobnie, rozumiemy notację: $P(X < x)$, $P(X > x)$, $P(X \leq x)$, $P(X \geq x)$.

Definicja: dystrybuanta

Dystrybuantę zmiennej losowej X definiujemy wzorem,

$$F(x) = P(X \leq x)$$

dla każdej liczby rzeczywistej x .

Właściwości dystrybuanty

1. Dla każdego x : $0 \leq F(x) \leq 1$
2. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$
3. $\lim_{x \rightarrow a^+} F(x) = F(a)$
4. F jest niemalejąca.

Definicja: dyskretna zmienna losowa

Zmienna losowa nazywana jest *dyskretną*, jeśli przyjmuje z dodatnim prawdopodobieństwem skończony lub co najwyżej przeliczalny zbiór wartości, tj. istnieje skończony ciąg x_1, \dots, x_n taki, że

(1)

$$P(X = x_i) = p_i, \quad \sum_{i=1}^n p_i = 1, \quad p_1, \dots, p_n \text{ liczby dodatnie}$$

lub istnieje nieskończony ciąg x_1, x_2, \dots taki, że

(2)

$$P(X = x_i) = p_i, \quad \sum_{i=1}^{\infty} p_i = 1, \quad p_1, p_2, \dots \text{ liczby dodatnie.}$$

Definicja: rozkład prawdopodobieństwa

Wzór (1) lub (2) nazywa się rozkładem prawdopodobieństwa dyskretnej zmiennej losowej.

Ponadto (1) może być zapisany w formie

$X = x_i$	x_1	x_2	...	x_n
$P(X = x_i)$	p_1	p_2	...	p_n

lub prościej

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Uwaga:

W [Przykładzie 5.1](#) jeśli moneta jest uczciwa, mamy

x_i	0	1
p_i	0.5	0.5

Definicja: Wartość oczekiwana

Wartość oczekiwana (wartość przeciętna lub średnia) dyskretnej zmiennej losowej X jest zdefiniowana wzorem

$$E(X) = \sum_i x_i P(X = x_i).$$

Liniowość wartości oczekiwanej

Jeśli $E(X) < \infty$ i $E(Y) < \infty$, wtedy dla dowolnych liczb rzeczywistych a, b

$$E(aX + bY) = aE(X) + bE(Y).$$

Wartość oczekiwana funkcji X .

Niech $Y = g(X)$, gdzie g jest funkcją dyskretnej zmiennej losowej X . Wtedy Y jest dyskretną zmienną losową o wartości oczekiwanej

$$E(Y) = \sum_i g(x_i) P(X = x_i).$$

Definicja: Wariancja, odchylenie standardowe

Wariancja dyskretnej zmiennej losowej X jest zdefiniowana przez

$$V(X) = E(X - E(X))^2 = E(X^2) - E(X)^2.$$

Odchylenie standardowe zmiennej losowej X jest zdefiniowane przez

$$\sigma(X) = \sqrt{V(X)}.$$

Przykład 5.3

Zmienna losowa X ma następujący rozkład

$X = x_i$	-2	0	2	3
$P(X = x_i)$	0.2	0.4	0.1	0.3

Oblicz: $E(X)$, $E(X^2)$, $V(X)$ oraz $\sigma(X)$. Naszkicuj wykres dystrybuanty X .

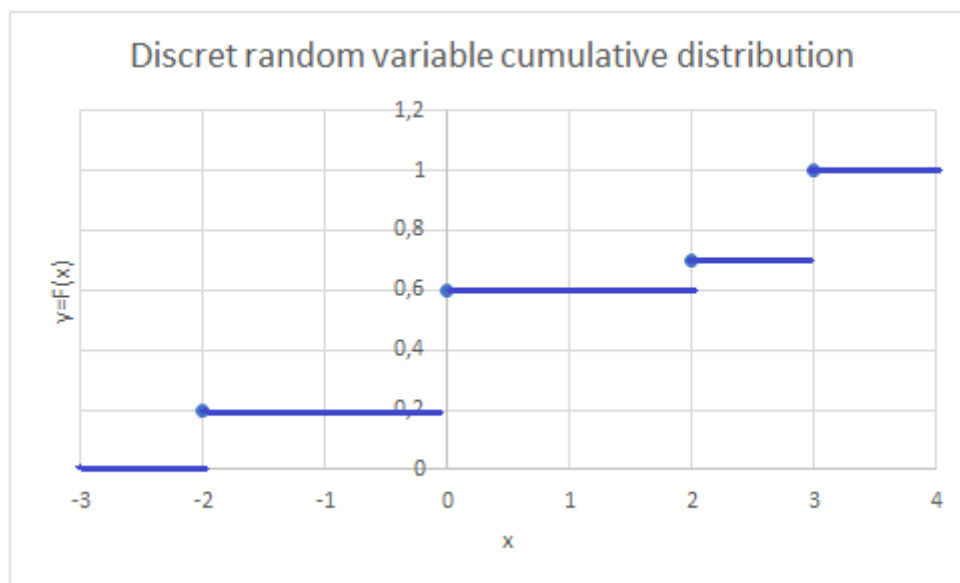
Rozwiązanie:

$$E(X) = (-2) \cdot (0.2) + 0 \cdot (0.4) + 2 \cdot (0.1) + 3 \cdot (0.3) = 0.7$$

$$E(X^2) = (-2)^2 \cdot (0.2) + 0^2 \cdot (0.4) + 2^2 \cdot (0.1) + 3^2 \cdot (0.3) = 3.9$$

$$V(X) = 3.9 - (0.7)^2 = 3.41$$

$$\sigma(X) = \sqrt{3.41} \approx 1.8466.$$



Rys: Ilustracja do Przykładu 5.3

Uwaga:

Zmienna losowa $Y = X^2$ w powyższym przykładzie ma następujący rozkład:

$Y = y_i$	0	4	9
$P(Y = y_i)$	0.4	0.3	0.3

$$P(Y = 0) = P(X^2 = 0) = P(X = 0) = 0.4$$

$$P(Y = 4) = P(X^2 = 4) = P(X = 2 \text{ or } X = -2) = P(X = -2) + P(X = 2) \\ = 0.2 + 0.1 = 0.3.$$

Ponieważ zdarzenia $\{X = -2\}$ i $\{X = 2\}$ są rozłączne, stąd prawdopodobieństwo ich sumy jest równe sumie ich prawdopodobieństw.

$$P(Y = 9) = P(X^2 = 9) = P(X = 3) = 0.3$$

Należy pamiętać, że $P(X = -3) = 0$.

Za pomocą rozkładu zmiennej $Y = X^2$ możemy obliczyć $E(X^2)$ inaczej,

$$E(X^2) = E(Y) = 0 \cdot 0.4 + 4 \cdot 0.3 + 9 \cdot 0.3 = 3.9.$$

Przykład 5.4

Rzucamy dwa razy kostką do gry. Definiujemy zmienną losową X jako sumę uzyskanych liczb. Znajdź rozkład prawdopodobieństwa zmiennej X . Oblicz wartość oczekiwaną, wariancję i odchylenie standardowe zmiennej losowej X .

Rozwiązanie:

Przestrzeń zdarzeń elementarnych dla eksperymentu jest,

$$\Omega = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$$

Liczba elementów zbioru Ω wynosi 36.

W oparciu o klasyczną definicję prawdopodobieństwa mamy:

$$P(X = 2) = P(\{(1,1)\}) = \frac{1}{36}$$

$$P(X = 3) = P(\{(1,2), (2,1)\}) = \frac{2}{36} = \frac{1}{18}$$

$$P(X = 4) = P(\{(1,3), (2,2), (3,1)\}) = \frac{3}{36} = \frac{1}{12}$$

$$P(X = 5) = P(\{(1,4), (2,3), (3,2), (4,1)\}) = \frac{4}{36} = \frac{1}{9}$$

$$P(X = 6) = P(\{(1,5), (2,4), (3,3), (4,2), (5,1)\}) = \frac{5}{36}$$

$$P(X = 7) = P(\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}) = \frac{6}{36} = \frac{1}{6}$$

$$P(X = 8) = P(\{(2,6), (3,5), (4,4), (5,3), (6,2)\}) = \frac{5}{36}$$

$$P(X = 9) = P(\{(3,6), (4,5), (5,4), (6,3)\}) = \frac{4}{36} = \frac{1}{9}$$

$$P(X = 10) = P(\{(4,6), (5,5), (6,4)\}) = \frac{3}{36} = \frac{1}{12}$$

$$P(X = 11) = P(\{(5,6), (6,5)\}) = \frac{2}{36} = \frac{1}{18}$$

$$P(X = 12) = P(\{(6,6)\}) = \frac{1}{36}$$

Dystrybuanta X zapisana w tabeli:

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Wartość oczekiwaną $E(X)$ zmiennej X obliczamy następująco:

$$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + 5 \cdot \frac{1}{9} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{1}{6} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{1}{9} + 10 \cdot \frac{1}{12} + 11 \cdot \frac{1}{18} + 12 \cdot \frac{1}{36} = 7$$

Wariancja X jest obliczana w dwóch krokach. Najpierw obliczamy wartość oczekiwaną X^2 .

$$E(X^2) = 2^2 \cdot \frac{1}{36} + 3^2 \cdot \frac{1}{18} + 4^2 \cdot \frac{1}{12} + 5^2 \cdot \frac{1}{9} + 6^2 \cdot \frac{5}{36} + 7^2 \cdot \frac{1}{6} + 8^2 \cdot \frac{5}{36} + 9^2 \cdot \frac{1}{9} + 10^2 \cdot \frac{1}{12} + 11^2 \cdot \frac{1}{18} + 12^2 \cdot \frac{1}{36} = \frac{329}{6} \approx 54.83$$

Na koniec obliczamy wariancję,

$$V(X) = E(X^2) - E(X)^2 \approx 54.83 - 7^2 = 5.83.$$

Odchylenie standardowe wynosi

$$\sigma(X) = \sqrt{V(X)} \approx 2.42.$$

6. Rozkład dwumianowy i rozkład Poissona

Definicja: n prób Bernoulliego

Zasady eksperymentu losowego *Bernoulliego*:

1. Eksperyment jest powtarzany ustaloną liczbę razy (n prób).
2. Każda próba ma tylko dwa możliwe wyniki, "sukces" i "porażkę". Możliwe wyniki są takie same dla każdej próby.
3. Prawdopodobieństwo sukcesu jest takie same dla każdej próby. Oznaczamy je p . Prawdopodobieństwo porażki dla każdej próby oznaczamy $q = 1 - p$.
4. Próby są niezależne. Wynik poprzedniej próby nie ma wpływu na wynik następczej próby.

Interesuje nas zmienna losowa X gdzie X oznacza liczbę sukcesów. Możliwe wartości zmiennej X wynoszą $0, 1, 2, \dots, n$.

Naszym celem jest wyznaczyć rozkład prawdopodobieństwa zmiennej X , gdzie X oznacza liczbę sukcesów w eksperymencie Bernoulliego o n próbach.

Definicja: rozkład dwumianowy

Jeśli X jest liczbą sukcesów w eksperymencie Bernoulliego o n niezależnych próbach, gdzie prawdopodobieństwo sukcesu wynosi p w każdej próbie a prawdopodobieństwo porażki $q = 1 - p$, to

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

Przykład 6.1

Koszykarz wykonuje 4 niezależne rzuty do kosza. Za każdym razem prawdopodobieństwo zdobycia punktu wynosi 0.6, i każdy zdobyty punkt jest rejestrowany. Tutaj każdy rzut jest próbą, a próby są niezależne. Każda próba ma dwa wyniki: punkt (sukces) lub brak punktu (porażka). Prawdopodobieństwo sukcesu wynosi $p = 0.6$ a prawdopodobieństwo niepowodzenia wynosi $q = 1 - p = 1 - 0.6 = 0.4$. Interesuje nas zmienna X , która liczy liczbę sukcesów w 4 próbach. Jest to przykład eksperymentu Bernoulliego o $n = 4$ próbach.

Rozwiązanie:

$$n = 4, \quad p = 0.6, \quad q = 0.4$$

$$P(X = 0) = \binom{4}{0} \cdot (0.6)^0 \cdot (0.4)^4 = (0.4)^4 = 0.0256$$

$$P(X = 1) = \binom{4}{1} \cdot (0.6)^1 \cdot (0.4)^3 = 4 \cdot 0.6 \cdot (0.4)^3 = 0.1536$$

$$P(X = 2) = \binom{4}{2} \cdot (0.6)^2 \cdot (0.4)^2 = 6 \cdot (0.6)^2 \cdot (0.4)^2 = 0.3456$$

$$P(X = 3) = \binom{4}{3} \cdot (0.6)^3 \cdot (0.4)^1 = 4 \cdot (0.6)^3 \cdot 0.4 = 0.3456$$

$$P(X = 4) = \binom{4}{4} \cdot (0.6)^4 \cdot (0.4)^0 = (0.6)^4 = 0.1296$$

Ostatecznie

x_i	0	1	2	3	4
p_i	0.0256	0.1536	0.3456	0.3456	0.1296

Definicja: Rozkład Poissona

Rozkład Poissona z parametrem $\lambda > 0$ jest dany wzorem

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, 3, \dots$$

Podstawowa własność rozkładu Poissona polega na tym, że jeżeli zmienna losowa X ma rozkład Poissona z parametrem $\lambda > 0$ to $E(X) = V(X) = \lambda$.

Uwaga:

Rozkład Poissona jest rozkładem prawdopodobieństwa liczby zdarzeń, które występują losowo i niezależnie w ustalonym przedziale czasu (lub przestrzeni). Jeśli wartość średnia liczby zdarzeń w przedziale wynosi λ to prawdopodobieństwo zaobserwowania k zdarzeń w tym czasie jest zadane za pomocą rozkładu Poissona.

Przykład 6.2

Założmy, że średnia liczba pewnych zdarzeń w ustalonym przedziale czasu wynosi 2. Oblicz prawdopodobieństwo, że zaobserwujemy dzisiaj pięć zdarzeń we wspomnianym przedziale czasu.

Rozwiązanie:

Zmienna losowa X ma rozkład Poissona ze średnią 2. Należy obliczyć $P(X = 5)$. Ponieważ $\lambda = 2$, mamy

$$P(X = 5) = e^{-2} \frac{2^5}{5!} \approx 0.03609.$$

7. Ciągły rozkład prawdopodobieństwa

Definicja: zmienna losowa ciągła

Zmienną losową X nazywamy ciągłą, jeśli istnieje ciągła (prawie wszędzie) funkcja nieujemna f taka, że dla dowolnego a

$$P(X \leq a) = \int_{-\infty}^a f(x) dx.$$

Funkcję f nazywamy gęstością prawdopodobieństwa (*funkcja gęstości*) zmiennej losowej X .

Uwaga:

Jeśli funkcja jest ciągła poza skończoną liczbą punktów to jest ciągła prawie wszędzie.

Jeżeli f jest funkcją gęstości prawdopodobieństwa, to

$$\int_{-\infty}^{\infty} f(t) dt = 1.$$

Uwaga:

Dystrybuanta ciągłej zmiennej losowej X jest zdefiniowana wzorem:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Dla dowolnej liczby rzeczywistej x .

Związek dystrybuanty z prawdopodobieństwem:

- 1 $P(X \geq a) = P(X > a) = 1 - F(a)$
- 2 $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = F(b) - F(a)$
- 3 $P(X = a) = 0$

Przykład 7.1

Dana jest funkcja prawdopodobieństwa zmiennej losowej X ,

$$f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

Znajdź dystrybuantę zmiennej losowej X .

Rozwiązanie:

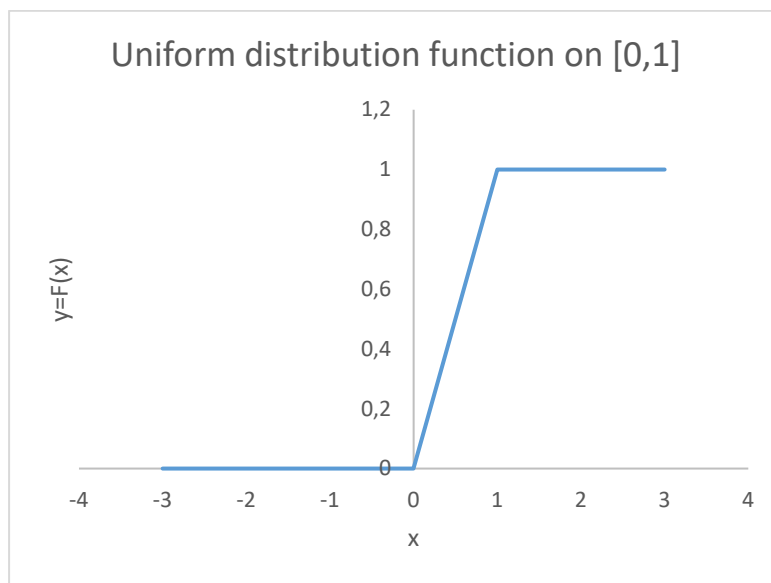
Zastosujemy wzór na dystrybuantę. Musimy rozważyć trzy przypadki:

1. $x < 0$: $F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^x 0dt = 0$
2. $0 \leq x \leq 1$: $F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 0 dt + \int_0^x 1 dt = 0 + x = x$
3. $x > 1$: $F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^0 0dt + \int_0^1 1dt + \int_1^x 0dt = 0 + 1 + 0 = 1$

Ostatecznie otrzymujemy wzór na dystrybuantę

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & x \geq 1. \end{cases}$$

Jej wykres znajduje się poniżej.



Rys: Ilustracja do Przykładu 7.1

Definicja: zmienna losowa o rozkładzie jednostajnym (rozkład jednostajny)

Rozkład jednostajny (prostokątny) zmiennej losowej X na odcinku $[a, b]$ jest dany za pomocą funkcji gęstości,

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{if } x \notin [a, b] \end{cases}.$$

Dystrybuanta rozkładu jednostajnego

Dystrybuanta jednostajnej zmiennej losowej X dana jest

$$F(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b. \\ 1, & x > b \end{cases}$$

Ćwiczenie:

Znajdź dystrybuantę rozkładu jednostajnego zmiennej losowej X na przedziale $[1,4]$.

Rozwiązanie:

Ponieważ $a = 1$, $b = 4$, mamy $b - a = 4 - 1 = 3$ i

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{3}, & x \in [1,4] \\ 1, & x > 4. \end{cases}$$

Definicja: Wartość oczekiwana (wartość średnia)

Wartość oczekiwana (*wartość średnia*) ciągłej zmiennej losowej X jest zdefiniowana wzorem

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

gdzie f jest funkcją gęstości prawdopodobieństwa zmiennej X .

Podobnie jak dla zmiennej dyskretnej, zachodzi liniowość $E(X)$ tzn.

Jeśli $E(X) < \infty$ i $E(y) < \infty$, to dla dowolnych liczb rzeczywistych a, b

$$E(aX + bY) = aE(X) + bE(Y).$$

Expected value of a function of X

Niech $Y = g(X)$, gdzie g jest ciągłą funkcją ciągłej zmiennej losowej X z funkcją gęstości prawdopodobieństwa f . Wtedy, Y jest ciągłą zmienną losową i zachodzi wzór,

$$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Podobnie, jak dla dyskretnej zmiennej losowej definiujemy:

Wariancja zmiennej X

$$V(X) = E(X - E(X))^2 = E(X^2) - E(X)^2.$$

Odchylenie standardowe zmiennej losowej X zdefiniowane jest wzorem

$$\sigma(X) = \sqrt{V(X)}.$$

Odchylenie standardowe zmiennej X jest również oznaczane σ a wariancja σ^2 .

Przykład 7.2

Obliczyć $E(X)$, $E(X^2)$, $V(X)$, $\sigma(X)$ dla zmiennej losowej o rozkładzie jednostajnym na przedziale $[0, 1]$.

Rozwiązanie:

Funkcja gęstości prawdopodobieństwa jest dana wzorem

$$f(x) = 1 \quad \text{dla} \quad x \in [0, 1] \quad \text{oraz} \quad f(x) = 0 \quad \text{dla pozostałych } x.$$

Stąd

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 xdx = \frac{1}{2}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2f(x)dx = \int_0^1 x^2dx = \frac{1}{3}$$

$$V(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

$$\sigma(X) = \sqrt{\frac{1}{12}} \approx 0.2887.$$

8. Rozkład Normalny

Definicja: Rozkład Normalny

Rozkładem normalnym (rozkładem Gaussa) nazywamy rozkład ciągły o gęstości

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

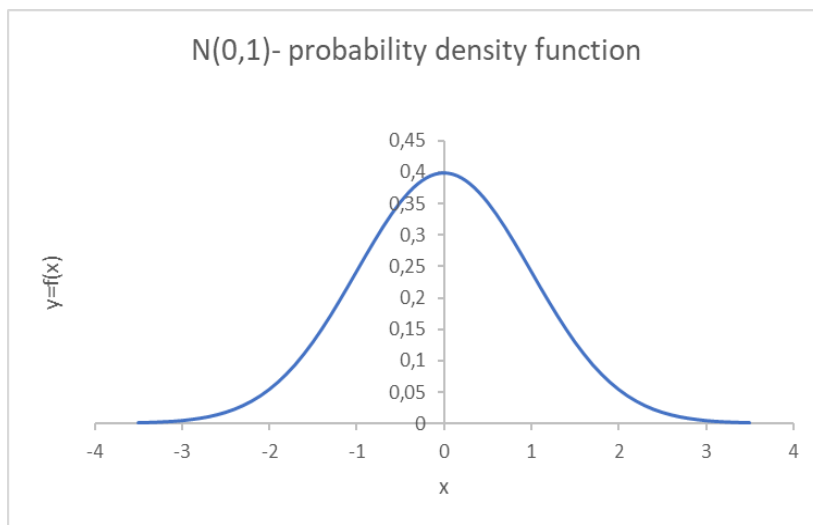
Oznaczamy go $N(\mu, \sigma)$, gdzie μ jest średnią a σ jest odchyleniem standardowym.

Definicja: Standardowy rozkład normalny

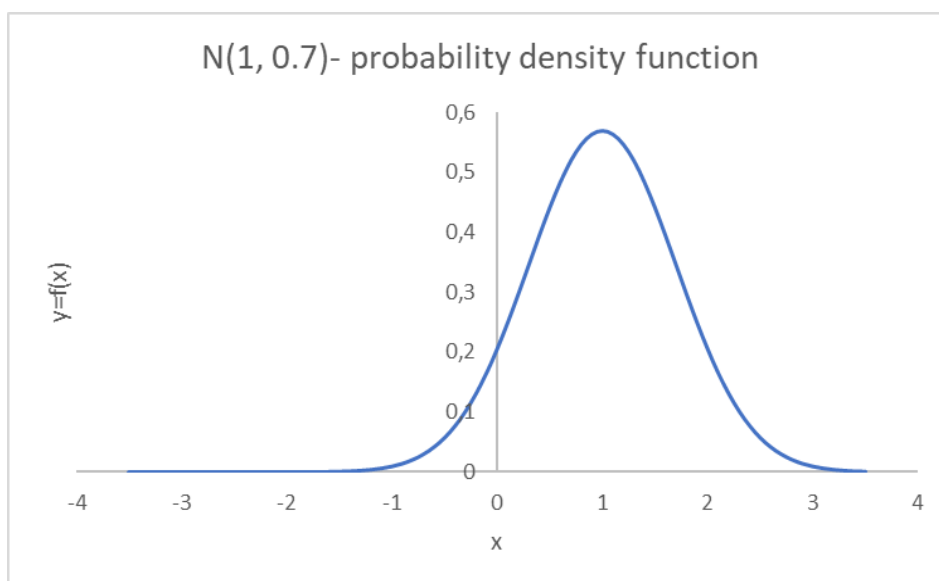
Standardowym rozkładem normalnym (*standardowym rozkładem Gaussa*) nazywamy rozkład normalny ze średnią $\mu = 0$ i wariancją $\sigma^2 = 1$. Funkcja gęstości prawdopodobieństwa jest dana wzorem

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty.$$

Standardowy rozkład normalny oznaczamy $N(0,1)$.



Rys.: Standardowy rozkład normalny,



Rys.: Rozkład normalny z $\mu = 1, \sigma = 0.7$.

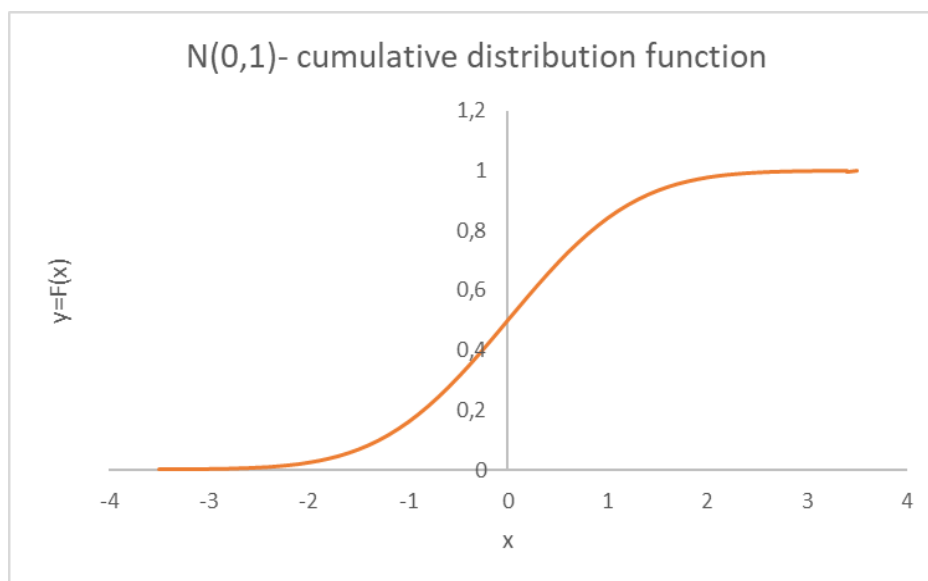
Uwaga:

Wykres funkcji gęstości rozkładu normalnego nazywamy krzywą Gaussa.

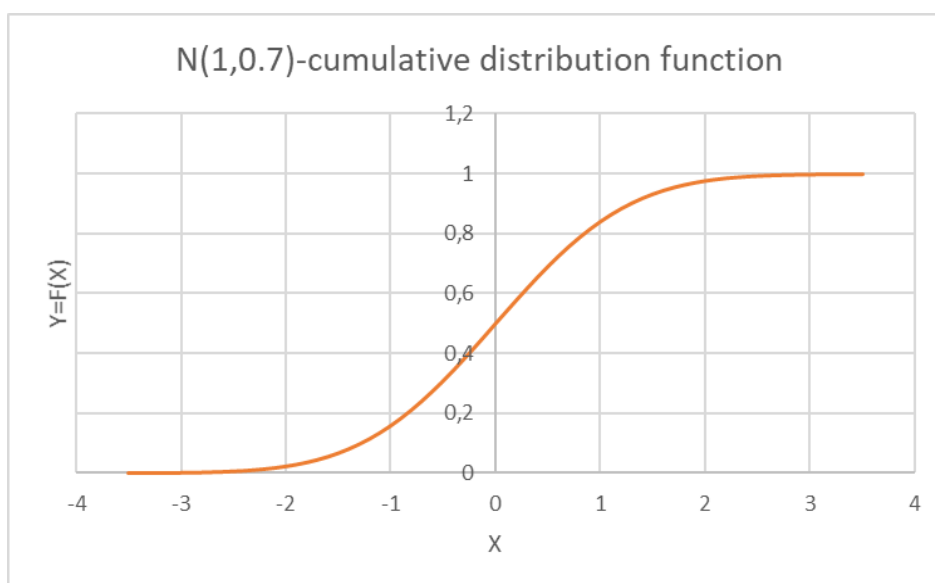
Dystrybuanta rozkładu normalnego dana jest wzorem

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad -\infty < x < \infty.$$

Występująca we wzorze całka nie jest całką elementarną.



Rys.: Dystrybuanta standardowego rozkładu normalnego.



Rys.: Dystrybuanta rozkładu normalnego z $\mu = 1, \sigma = 0.7$

Przykład 7.3

Zmienna losowa X ma rozkład normalny ze średnią 20 i z odchyleniem standardowym 3. Wybieramy jeden wynik losowo.

1. Jakie jest prawdopodobieństwo, że jest powyżej 15?
2. Jakie jest prawdopodobieństwo, że jest w przedziale $[18,22]$?

Rozwiązanie:

1. Zmienna losowa X ma rozkład $N(20,3)$. Musimy policzyć $P(X > 15)$. Niech F będzie dystrybuantą zmiennej X .

$$P(X > 15) = 1 - P(X \leq 15) = 1 - F(15) \approx 1 - 0.0478 = 0.9522$$

Obliczamy $F(15)$ za pomocą `NORM.DIST(15;20;3;1)` w Excelu lub `NORMDIST(15;20;3;1)` w OpenOffice Calc.

2. $P(18 \leq X \leq 22) = F(22) - F(18) \approx 0.7475 - 0.2525 = 0.4950$.

Ćwiczenie:

Przypuśćmy, że długość ciała śledzia jest zmienną losową o rozkładzie normalnym o średniej 8 cali i odchyleniu standardowym 1.5 cala. Statek rybacki jest w stanie złowić 10 ton śledzi dziennie. Oszacować:

1. Jaka ilość śledzi ma spodziewaną długość większą niż 11 cali?
2. Jak jest spodziewana ilość śledzi o długości mniejszej niż 6 cali?
3. Jaka przypuszczalnie ilość śledzi ma długość pomiędzy 8 i 10 cali?

Rozwiązanie:

1. $\approx 0.0228 \cdot 10 t = 0.228 t$

2. $\approx 0.0912 \cdot 10 t = 0.912 t$

3. $\approx 0.4088 \cdot 10 t = 4.088 t$

9. Statystyka, podstawowe definicje i notacje

Statystyka jest nauką zajmującą się zbieraniem, analizą, prezentowaniem i interpretacją danych.

Populacja zawiera wszystkie elementy z zestawu danych.

Próba składa się z jednej obserwacji pobranej z populacji (niewielka część populacji).

Badając populację, jesteśmy zainteresowani zbieraniem informacji o różnych cechach podmiotu (takich jak ich długość, waga lub wiek) w próbie. Te cechy nazywane są zmiennymi.

Istnieją dwie główne gałęzie statystyki: opisowa i reprezentatywna. Statystyka opisowa służy do powiedzenia czegoś o zestawie informacji, które zostały zebrane z niewielkiej części.

Statystyka reprezentatywna jest wykorzystywana do przewidywania lub porównywania większej grupy (populacji) przy użyciu informacji zebranych na temat niewielkiej części tej populacji.

Inne rozróżnienia są dokonywane między typami danych. Dane dyskretne składają się z liczb całkowitych i zwykle opisują wiele obiektów. Przykładowe badanie może opisywać, ile dzieci posiadają różne rodziny.

Dane pomiarowe, w przeciwieństwie do danych dyskretnych, są ciągłe, a zatem mogą przyjąć dowolną wartość rzeczywistą. Na przykład ilość czasu spędzonego przez grupę dzieci na komputerze stanowi dane pomiarowe ciągłe, ponieważ dzieci mogą spędzić dowolną ilość czasu na używaniu komputera.

Dla danej próbki

$$x: x_1, x_2, \dots, x_n.$$

Zdefiniujemy kilka liczb, które ją charakteryzują (charakterystyk, statystyk).

Def: Średnia

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

Inaczej

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n).$$

Przykład 9.1

x_i : 1, 4, 7, 2, 4, 3.

Średnia

$$\bar{x} = \frac{1}{6}(1 + 4 + 7 + 2 + 4 + 3) = \frac{21}{6} = 3.5$$

Średnia: OpenOffice Calc: **AVERAGE()**, **ŚREDNIA()**

Excel: **AVERAGE()**

Def: Wariancja,

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Równoważna formuła

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

Przykład 9.2

Dla próbki

x_i : 1, 4, 7, 2, 4, 3.

Średnia $\bar{x} = 3.5$

Wariancja

$$\begin{aligned} s^2 &= \\ &= \frac{1}{6} [(1 - 3.5)^2 + (4 - 3.5)^2 + (7 - 3.5)^2 + (2 - 3.5)^2 + (4 - 3.5)^2 + (3 - 3.5)^2] = \\ &= \frac{1}{6} [(2.5)^2 + (0.5)^2 + (3.5)^2 + (1.5)^2 + (0.5)^2 + (0.5)^2] = \frac{1}{6} (21.5) \cong 3.58 \end{aligned}$$

Z formuły równoważnej

$$s^2 = \frac{1}{6} [1^2 + 4^2 + 7^2 + 2^2 + 4^2 + 3^2] - (3.5)^2 = \frac{95}{6} - 12.25 \cong 3.58$$

Def: Odchylenie standardowe

$$s = \sqrt{s^2}.$$

Uwaga: Wariancja i odchylenie standardowe mierzą jak bardzo elementy w próbkę są rozproszone wokół średniej (miary rozproszenia)

Przykład 9.3

$$x_i: 3, 3, 3, 3, 3.$$

Średnia

$$\bar{x} = \frac{1}{5}(3 + 3 + 3 + 3 + 3) = 3$$

Wariancja

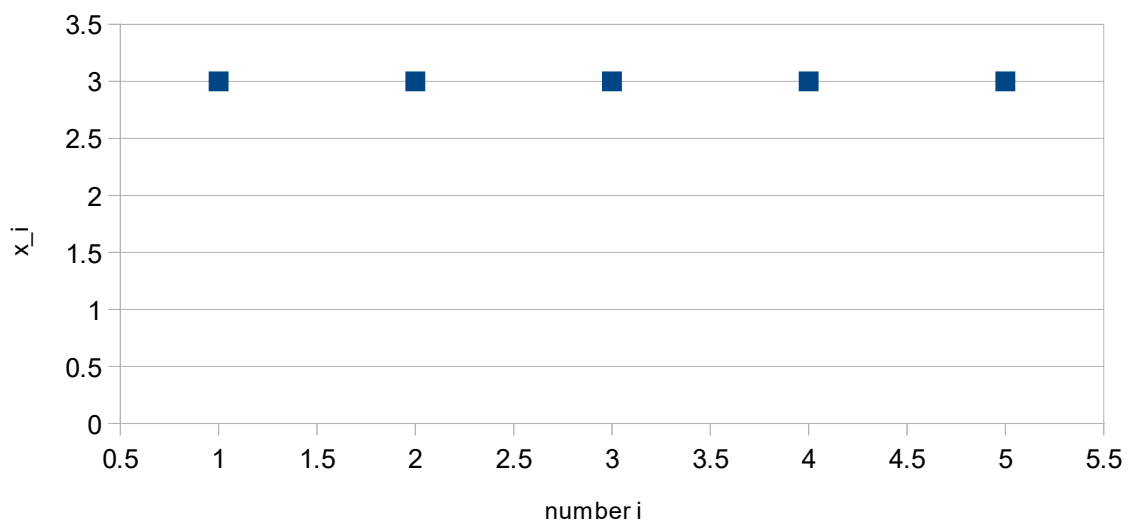
$$s^2 = \frac{1}{5}[(3 - 3)^2 + (3 - 3)^2 + (3 - 3)^2 + (3 - 3)^2 + (3 - 3)^2] = 0$$

Odchylenie standardowe

$$s = \sqrt{0^2} = 0.$$

Nie ma rozproszenia danych wokół średniej.

$$X_{\{i\}}: 3, 3, 3, 3, 3$$



Rys.: Ilustracja do Przykładu 9.3

Przykład 9.4

$$x_i: 3, 2, 4, 3, 3.$$

Średnia

$$\bar{x} = \frac{1}{5}(3 + 2 + 4 + 3 + 3) = 3$$

Wariancja

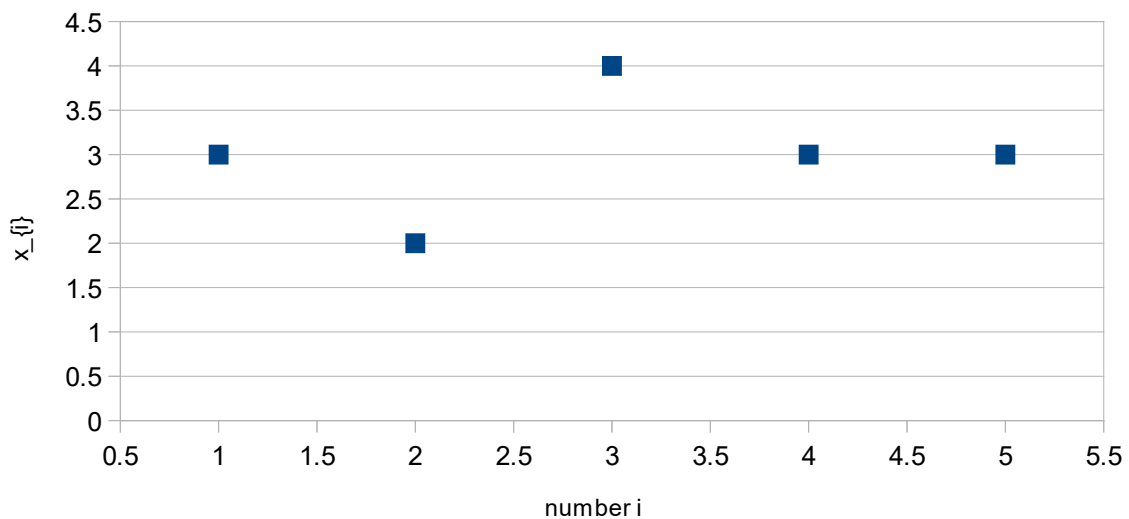
$$s^2 = \frac{1}{5}[(3 - 3)^2 + (2 - 3)^2 + (4 - 3)^2 + (3 - 3)^2 + (3 - 3)^2] = \frac{2}{5} = 0,4.$$

Odchylenie standardowe

$$s = \sqrt{\frac{2}{5}} = 0.63.$$

Występuje pewne rozproszenie danych wokół średniej.

$$x_{\{i\}}: 3, 2, 4, 3, 3,$$



Rys.:1 Ilustracja do Przykładu 9.4

Przykład 9.5

$$x_i: 5, 2, 4, 2, 2.$$

Średnia

$$\bar{x} = \frac{1}{5}(5 + 2 + 4 + 2 + 2) = 3.$$

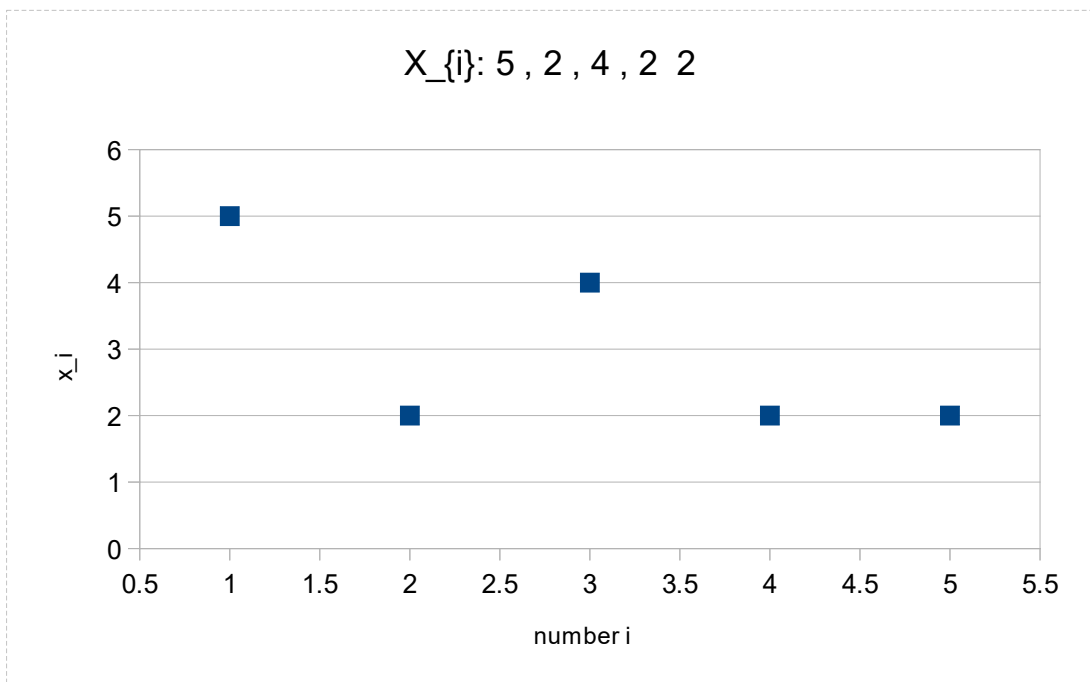
Wariancja

$$s^2 = \frac{1}{5}[(5 - 3)^2 + (2 - 3)^2 + (4 - 3)^2 + (2 - 3)^2 + (2 - 3)^2] = \frac{8}{5} = 1,6.$$

Odchylenie standardowe

$$s = \sqrt{\frac{8}{5}} = 1,26.$$

Występuje wyraźne rozproszenie danych wokół średniej.



Rys.: Ilustracja do Przykładu 9.5

Definicja: Estymator wariancji

Estymator wariancji jest zdefiniowany wzorem,

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Odchylenie standardowe estymatora wariancji jest wyrażone wzorem

$$\hat{s} = \sqrt{\hat{s}^2}.$$

Przykład 9.6

próbka:

5, 2, 4, 2, 2

średnia:

$$\bar{x} = \frac{1}{5}(5 + 2 + 4 + 2 + 2) = 3$$

estymator wariancji:

$$\hat{s}^2 = \frac{1}{4}[(5 - 3)^2 + (2 - 3)^2 + (4 - 3)^2 + (2 - 3)^2 + (2 - 3)^2] = \frac{8}{4} = 2$$

odchylenie standardowe estymatora wariancji:

$$\hat{s} = \sqrt{2} = 1.41.$$

Excel:

Wariancja: WARIANCJA.POP()

Odchylenie standardowe ODCH.STAND.POPUL()

Estymator wariancji: WARIANCJA.PRÓBK()

OpenOffice Calc:

Wariancja: WARIANCJA.POPUL()

Odchylenie standardowe ODCH.STANDARD.POPUL()

Estymator wariancji: WARIANCJA() !!!

10. Badanie hipotez statystycznych. Błędy wnioskowania statystycznego.

Definicja: Hipoteza zerowa i alternatywna

- *Hipoteza zerowa* jest stwierdzeniem, które należy przetestować w eksperymencie. Jest oznaczana H_0 .
- *Alternatywna hipoteza* jest stwierdzeniem przeciwnym do H_0 . Jest oznaczana H_a .

Istnieją dwie opcje dla naszej decyzji: odrzuć H_0 lub nie odrzucać H_0 .

Definicja: Błędy typu I i błędy typu II

- Błąd **typu I** występuje, gdy odrzucamy H_0 ale w rzeczywistości H_0 jest prawdziwa (błędna decyzja).
- Błąd **typu II** występuje, gdy nie odrzucamy H_0 ale w rzeczywistości H_0 jest fałszywa.

Prawdopodobieństwo błędu **typu I** jest oznaczone symbolem α . Prawdopodobieństwo błędu **typu II** jest oznaczone symbolem β .

Uwaga:

Błąd **typu I** i **typ II** są oczywiście **błędnymi** decyzjami.

Pozostałe dwie decyzje są **prawidłowe**:

- Nie odrzucać H_0 kiedy w rzeczywistości, H_0 jest prawdziwa (prawidłowa decyzja).
- Odrzucić H_0 i w rzeczywistości, H_0 jest fałszywa (prawidłowa decyzja).

Wprowadzimy dwa testy, w których chcemy zweryfikować hipotezę, że średnia z populacji wynosi μ_0 . W pierwszym (Model I) zakładamy, że odchylenie standardowe populacji σ jest znane. W drugim teście (Model II) nie znamy wartości σ . W obu przypadkach zakładamy, że populacja ma rozkład normalny $N(\mu, \sigma)$.

Model I testu jest następujący.

Założmy, że populacja ma rozkład normalny $N(\mu, \sigma)$ oraz że odchylenie standardowe populacji σ jest znane. Hipotezą zerową jest $H_0: \mu = \mu_0$, gdzie μ_0 jest przewidywaną średnią wartością dla całej populacji. Z losowej próbki x chcemy sprawdzić hipotezę $H_0: \mu = \mu_0$ przeciwko hipotezie alternatywnej $H_1: \mu \neq \mu_0$. Zakładamy, pewien z góry założony lub wstępnie ustawiony "poziom istotności" α testu. Wstępnie ustawione α to prawdopodobieństwo, że odrzucimy H_0 ale w rzeczywistości H_0 jest prawdziwe, tj. prawdopodobieństwo błędu typu I. Jeśli α nie jest podany, przyjętym standardem jest ustalenie $\alpha = 0.05$.

Weryfikujemy hipotezę zerową H_0 w następujący sposób:

Najpierw obliczamy średnią \bar{x} próbki x . Następnie obliczamy wartość zmiennej losowej u ze standardowym rozkładem normalnym $N(0,1)$ za pomocą wzoru:

$$u = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}.$$

Z tabeli standardowego rozkładu normalnego lub za pomocą kalkulatora odwrotnego rozkładu normalnego (patrz na przykład $N(0,1)$):

http://onlinestatbook.com/2/calculators/inverse_normal_dist.html),

lub używając popularnego arkusza kalkulacyjnego, takiego jak Excel, OpenOffice Calc, znajdujemy taką wartość krytyczną, u_α że

$$P(|u| \geq u_\alpha) = \alpha.$$

Zbiór wartości zdefiniowanych przez nierówność $|u| \geq u_\alpha$ jest krytycznym obszarem tego testu, t.j. jeśli otrzymamy taką wartość u , że $|u| \geq u_\alpha$ hipoteza H_0 musi zostać odrzucona. Jeśli $|u| < u_\alpha$, nie możemy odrzucić hipotezy H_0 .

Test opisany powyżej jest czasami nazywany testem dwustronnym, ponieważ mamy dwustronny obszar krytyczny, tj.

$$|u| \geq u_\alpha \quad \text{oznacza to samo, co } u \in (-\infty, -u_\alpha] \cup [u_\alpha, \infty).$$

Dzieje się tak za każdym razem, gdy mamy alternatywną hipotezę formy $H_a: \mu \neq \mu_0$.

W niektórych przypadkach musimy jednak wykonać testy jednostronne.

Gdy hipoteza alternatywna ma postać $H_a: \mu < \mu_0$ mamy lewostronny obszar krytyczny, tj. $u \leq u_\alpha$. W tym przypadku znajdujemy u_α takie, że

$$P(u \leq u_\alpha) = \alpha.$$

Gdy hipoteza alternatywna ma postać $H_a: \mu > \mu_0$ mamy prawostronny obszar krytyczny i.e. $u \geq u_\alpha$. W tym przypadku znajdujemy u_α takie, że

$$P(u \geq u_\alpha) = \alpha.$$

Uwaga:

Test zgodnie z Modelem I, przeprowadzamy również, jeśli populacja ma rozkład normalny $N(\mu, \sigma)$ a odchylenie standardowe jest nieznane, ale próbka jest duża, tj. zawiera kilkadziesiąt lub więcej elementów. W tym przypadku przyjmujemy $s = \sigma$.

Przykład 10.1

W 81 losowo wybranych firmach produkcyjnych, zmierzono koszt materiału wykorzystywanego do produkcji jednego konkretnego produktu. Średnia z tych wartości wynosi $\bar{x} = 540$ € i odchylenie standardowe $s = 150$ €. Na poziomie istotności $\alpha = 0.05$, zweryfikować hipotezę, że średni koszt materiałów do wytworzenia tego produktu wynosi 600 €.

Rozwiązanie:

Hipotezą zerową H_0 jest to, że średni koszt materiałowy wytworzenia produktu wynosi 600 € i.e.,

$$H_0: \mu_0 = 600.$$

Ponieważ próbka jest duża, możemy założyć, że $s = \sigma = 150$. Stosujemy Model I.

Liczymy

$$u = \frac{540 - 600}{150} \sqrt{81} = -3.6.$$

Mamy

$$P(|u| \geq u_\alpha) = 0.05$$

dla $u_\alpha = 1.96$ (zobacz: tabela, arkusz kalkulacyjny lub kalkulator odwrotnego rozkładu normalnego i.e., http://onlinestatbook.com/2/calculators/inverse_normal_dist.html (opcja: 'outside' ponieważ test jest dwustronny).

Widzimy, że

$$|-3.6| = 3.6 \geq 1.96 = u_\alpha.$$

Stąd na poziomie istotności $\alpha = 0.05$ hipoteza H_0 należy odrzucić na rzecz alternatywnej hipotezy, zgodnie z którą średni koszt materiałów do wytworzenia danego produktu nie jest równy 600 €.

Model II testu jest następujący.

Założmy, że populacja podlega rozkładowi normalnemu $N(\mu, \sigma)$ oraz że średnia μ i odchylenie standardowe populacji σ są nieznanne. Mała próba została wybrana losowo z populacji.

Hipotezą zerową jest $H_0: \mu = \mu_0$. Z losowej próbki x chcemy sprawdzić H_0 przeciwko hipotezie alternatywnej $H_1: \mu \neq \mu_0$. Zakładamy, pewien "poziom istotności" α testu.

Weryfikujemy hipotezę zerową H_0 jak następuje:

Najpierw obliczamy średnią próbkę \bar{x} oraz odchylenie standardowe próbki s lub estymator odchylenia standardowego \hat{s} . Następnie obliczamy wartość zmiennej losowej t za pomocą wzoru:

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{n - 1} = \frac{\bar{x} - \mu_0}{\hat{s}} \sqrt{n}.$$

Pod warunkiem, że H_0 zachodzi zmienna losowa t spełnia rozkład t – studenta z $n - 1$ stopniami swobody (rozkład t).

Z tabeli rozkładu t –lub za pomocą kalkulatora odwrotnego rozkładu normalnego (patrz na przykład: <https://www.statology.org/inverse-t-distribution-calculator>) lub za pomocą popularnego arkusza kalkulacyjnego, takiego jak Excel, OpenOffice Calc, znajdujemy taką wartość krytyczną t_α , taką że

$$P(|t| \geq t_\alpha) = \alpha.$$

Zakres wartości t zdefiniowany przez nierówności $|t| \geq t_\alpha$ jest krytycznym obszarem tego testu tj. jeśli otrzymamy ze wzoru taką wartość t że $|t| \geq t_\alpha$ hipoteza H_0 musi zostać odrzucona na rzecz hipotezy H_1 . Jeśli $|t| < t_\alpha$ nie możemy odrzucić hipotezy H_0 .

Test opisany powyżej jest oczywiście dwustronny (patrz Model I), ponieważ mamy dwustronny obszar krytyczny, tj.

$$|t| \geq t_\alpha$$

Dzieje się tak za każdym razem, gdy mamy alternatywną hipotezę postaci $H_1: \mu \neq \mu_0$.

W niektórych przypadkach musimy jednak wykonać testy jednostronne.

Gdy hipoteza alternatywna ma postać $H_1: \mu < \mu_0$ mamy lewostronny obszar krytyczny, tj.: $t \leq t_\alpha$. W tym przypadku wyznaczamy t_α takie, że

$$P(t \leq t_\alpha) = \alpha.$$

Gdy hipoteza alternatywna ma postać $H_1: \mu > \mu_0$ mamy prawostronny obszar krytyczny, tj., $t \geq t_\alpha$. W tym przypadku wyznaczamy t_α takie, że

$$P(t \geq t_\alpha) = \alpha.$$

Przykład 10.2

Maszyna wytwarza metalowe płyty o określonych wymiarach, przy czym ich nominalna grubość wynosi 0.04 mm . Po losowym wybraniu i zmierzeniu 25 płyt, ich średnia grubość wynosiła $\bar{x} = 0.037 \text{ mm}$ i $\hat{s} = 0.005 \text{ mm}$. Czy można stwierdzić, że produkowane płyty są cieńsze niż 0.04 mm ? Załóżmy poziom istotności $\alpha = 0.01$.

Rozwiązanie:

Hipoteza zerowa H_0 oznacza teraz, że średnia grubość blaszek jest nie mniejsza niż 0.04 mm . Jeżeli w wyniku testu odrzucimy tę hipotezę to uda nam się stwierdzić, że blaszki są cieńsze niż 0.04 mm

$$H_0: \mu_0 \geq 0.04 \text{ and } H_1: \mu_0 < 0.04.$$

Obliczamy

$$t = \frac{0.037 - 0.04}{0.005} \sqrt{25} = -3$$

Używamy rozkładu t z 24 stopniami swobody i test jest lewostronny.

Z tablicy rozkładu t lub za pomocą kalkulatora odwrotnego rozkładu t (patrz na przykład: <https://www.statology.org/inverse-t-distribution-calculator>) lub za pomocą niektórych popularnych arkusz kalkulacyjny, taki jak Excel, OpenOffice Calc,

dostajemy $t_\alpha = -2.492$.

Mamy $t = -3 < -2.492 = t_\alpha$.

Stąd t należy do obszaru krytycznego. Musimy odrzucić H_0 na rzecz H_1 . Produkowane płyty są cieńsze niż 0.04 mm .

11. Regresja i korelacja

Przypuśćmy, że mamy dwie próbki różnych **cech statystycznych pewnej populacji**,

$$x: x_1, x_2, \dots, x_n \quad y: y_1, y_2, \dots, y_n$$

Jesteśmy zainteresowani odpowiedzią na następujące pytanie: Czy jest jakiś związek, zależność, między tymi dwoma cechami **badanej populacji**?

Definicja: Kowariancja

Kowariancję próbek x, y definiujemy wzorem

$$cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Definicja: Współczynnik korelacji Pearsona

Współczynnik korelacji Pearsona definiujemy wzorem

$$r = \frac{cov(x, y)}{s(x)s(y)}$$

gdzie $s(x)$ jest odchyleniem standardowym próbki x a $s(y)$ jest odchyleniem standardowym próbki y .

Uwaga:

Współczynnik korelacji r przyjmuje wartości od -1 do 1 .

Wartość współczynnika korelacji $r = 1$ oznacza, że występuje całkowita dodatnia liniowa korelacja pomiędzy próbkami x i y . Wartość $r = 0$ oznacza, że nie występuje liniowa korelacja pomiędzy x i y . Wartość $r = -1$ oznacza, że występuje całkowita ujemna liniowa korelacja między x i y .

Dodatnia korelacja informuje, że wzrostowi wartości jednej cechy towarzyszy wzrost średnich wartości drugiej cechy. Ujemna korelacja informuje, że wzrostowi wartości jednej cechy towarzyszy spadek średnich wartości drugiej cechy.

0,00-0,3 słaba zależność

0,31-0,6 umiarkowana zależność

0,61-1,0 silna zależność.

Przykład 11.1

Dla

$$x: 3, 3, 4, 5, 5 \quad y: 5, 7, 6, 4, 8$$

$$\bar{x} = 4, \quad \bar{y} = 6$$

$$\text{cov}(x, y) =$$

$$\begin{aligned} & \frac{1}{5}((3-4)(5-6) + (3-4)(7-6) + (4-4)(6-6) + (5-4)(4-6) + (5-4)(8-6)) \\ &= \frac{1}{5}(1-1+0-2+2) = 0 \end{aligned}$$

$$r = \frac{0}{s(x)s(y)} = 0.$$

Cechy nieskorelowane.

Przykład 11.2

Dla

$$x: 3, 3, 4, 5, 5 \quad y: 5, 8, 6, 6, 10$$

$$\bar{x} = 4, \quad \bar{y} = 7$$

$$\text{cov}(x, y) =$$

$$\begin{aligned} & \frac{1}{5}((3-4)(5-7) + (3-4)(8-7) + (4-4)(6-7) + (5-4)(6-7) \\ &+ (5-4)(10-7)) = \frac{1}{5}(2-1+0-1+3) = \frac{3}{5} \end{aligned}$$

$$s^2(x) =$$

$$\begin{aligned} & \frac{1}{5}((3-4)^2 + (3-4)^2 + (4-4)^2 + (5-4)^2 + (5-4)^2) \\ &= \frac{1}{5}(1+1+0+1+1) = \frac{4}{5} \end{aligned}$$

$$s(x) = \sqrt{\frac{4}{5}} = \frac{2}{\sqrt{5}}$$

$$s^2(y) =$$

$$\begin{aligned} & \frac{1}{5}((5-7)^2 + (8-7)^2 + (6-7)^2 + (6-7)^2 + (10-7)^2) \\ &= \frac{1}{5}(4+1+1+1+9) = \frac{16}{5} \end{aligned}$$

$$s(y) = \sqrt{\frac{16}{5}} = \frac{4}{\sqrt{5}}$$

$$r = \frac{\frac{3}{5}}{\frac{2}{\sqrt{5}} \cdot \frac{4}{\sqrt{5}}} = \frac{3}{8} = 0.375.$$

Umiarkowana zależność.

Excel:

KOWARIANCJA.POPUL(A1:A10;B1:B10),

WSP.KORELACJI(A1:A10;B1:B10)

OpenOffice Calc:

KOWARIANCJA(A1:A10;B1:B10),

WSP.KORELACJI(A1:A10;B1:B10),

Regresja liniowa

Przypuśćmy, że mamy dwie próbki:

$$x: x_1, x_2, \dots, x_n, \quad y: y_1, y_2, \dots, y_n.$$

Za pomocą *regresji liniowej*, możemy modelować zależność między dwiema zmiennymi (cechami) x and y .

Pierwsza ze zmiennych (x) jest nazywana zmienną niezależną (objaśniającą). Druga zmienna (y) jest nazywana zmienną zależną (objaśnianą).

Związek pomiędzy zmiennymi x i y jest modelowany za pomocą funkcji liniowej

$$y = ax + b,$$

której nieznane parametry a, b (nachylenie i odcięta) są szacowane na podstawie danych.

Aby znaleźć a, b używamy metody najmniejszych kwadratów. Metoda polega na utworzeniu linii, która minimalizuje sumę kwadratów odległości wszystkich punktów od tej linii. Taką linię (prostą) nazywamy linią najlepszego dopasowania.

Aby znaleźć linię najlepszego dopasowania musimy rozwiązać następujący problem minimalizacyjny:

$$\sum_{i=1}^n (y_i - ax_i - b)^2 \rightarrow \min, \quad a, b = ?$$

Można udowodnić, że taki problem minimalizacyjny zawsze ma rozwiązanie a, b dane wzorami

$$a = \frac{\text{cov}(x, y)}{s^2(x)}, \quad b = \bar{y} - a\bar{x},$$

gdzie

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

kowariancja x, y ,

$$s^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Jest wariancją z próbki x oraz

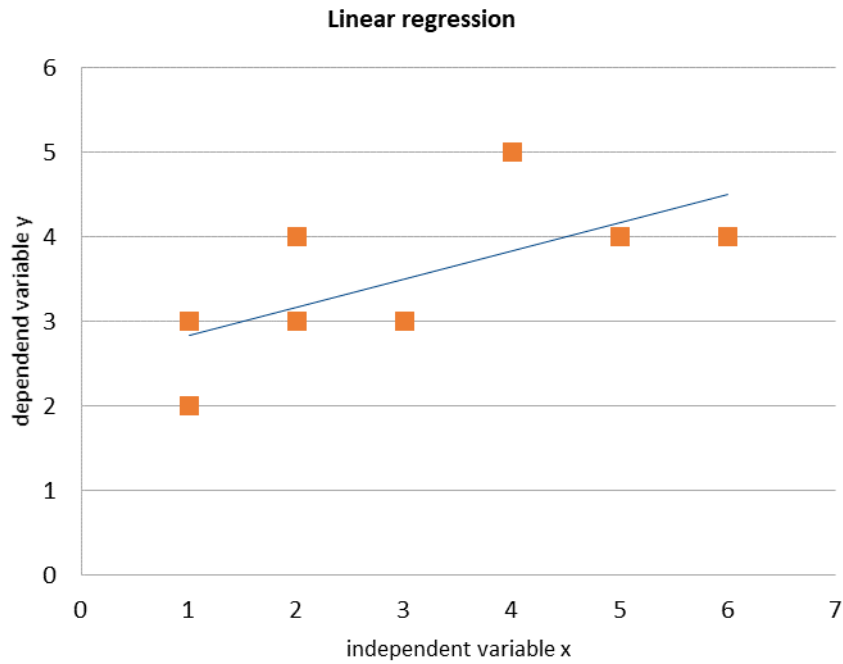
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Są średnimi odpowiednio z próbek x i y .

Przykład 11.3

Mamy dane

$$x: 3, 5, 2, 2, 1, 4, 6, 1 \quad y: 3, 4, 3, 4, 2, 5, 4, 3.$$



Rys.: Funkcja regresji: $y = \frac{1}{3}x - \frac{5}{2}$.

Mamy

$$\text{cov}(x, y) = 1, \quad s^2(x) = 3, \quad \bar{x} = 3, \quad \bar{y} = \frac{7}{2}$$

$$a = \frac{\text{cov}(x, y)}{s^2(x)} = \frac{1}{3}, \quad b = \bar{y} - a\bar{x} = \frac{7}{2} - \frac{1}{3} \cdot 3 = \frac{5}{2}$$

Liniową regresją (linią najlepszego dopasowania) jest $y = \frac{1}{3}x + \frac{5}{2}$.

Jej wykres widzimy na rysunku.

Ćwiczenie:

Znajdź funkcję regresji liniowej:

$$x: 1, 2, 4 \quad y: 0, 2, 1.$$

Rozwiązanie:

$$y = \frac{3}{14}x + \frac{1}{2}.$$

a

Excel: **SLOPE**(B1:B5,A1:A5), **NACHYLENIE**(B1:B5,A1:A5)

OpenOffice Calc: **SLOPE**(B1:B5,A1:A5), **NACHYLENIE**(B1:B5,A1:A5)

b

Excel: **INTERCEPT**(B1:B5,A1:A5), **ODCIĘTA**(B1:B5,A1:A5)

OpenOffice Calc: **INTERCEPT**(B1:B5,A1:A5) , **ODCIĘTA**(B1:B5,A1:A5)

12. Zastosowania

Niezawodność – zastosowanie rozkładu Poissona.

Prawdopodobieństwo k usterek w jednostce czasu dane jest wzorem

$$f(k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

gdzie λ jest średnią liczbą usterek (wartością oczekiwaną liczby usterek) w jednostce czasu.

Funkcja

$$f(k; \lambda, t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

wyraża prawdopodobieństwo k usterek w czasie t .

Zauważmy, że prawdopodobieństwo 0 usterek w czasie t jest dane przez,

$$f(0; \lambda, t) = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Definicja: Funkcja niezawodności

Funkcja niezawodności $R(t)$ wyraża prawdopodobieństwo zero liczby usterek w przedziale czasu o długości t . Jest dana wzorem

$$R(t) = \begin{cases} 1, & t < 0 \\ e^{-\lambda t}, & t \geq 0 \end{cases}.$$

Prawdopodobieństwo wystąpienia nie więcej niż m usterek w przedziale czasu o długości $t > 0$ jest dane przez

$$R(t, m) = \sum_{k=0}^m e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Przykład:

Pewien złożony system ma średni współczynnik awaryjności $\lambda = 0.002$ awarii tranzystora na godzinę. Oblicz jaka jest niezawodności systemu w przeciągu 30 dni, jeśli liczba awarii tranzystorów nie może przekroczyć 1?

Rozwiązanie:

$$\begin{aligned}\lambda &= 0.002 \\ t &= 30 \cdot 24 = 720 \\ m &\leq 1 \\ \lambda t &= 1.44\end{aligned}$$

Musimy obliczyć $R(720,1)$.

$$R(720,1) = e^{-1.44} + e^{-1.44} \cdot 1.44 \approx 0.58$$

12. Ćwiczenia

Ćwiczenies:

1. Ile jest dwucyfrowych nieparzystych liczb naturalnych większych niż 30?
2. Ile jest trzycyfrowych parzystych liczb naturalnych mniejszych niż 500?
3. Na ile sposobów możemy ułożyć litery słowa *MONTANA*?
4. W kraju A znajduje się siedem portów. Na ile sposobów mogą tam dokować cztery różne statki?
5. Na ile sposobów można ukryć sześć monet w czterech pudełkach.
6. Sześć punktów leży na obwodzie okręgu. Ile wpisanych trójkątów można narysować, mając te punkty jako wierzchołki?
7. Jeśli litery wyrazu *algebra* są umieszczone losowo w rzędzie, jakie jest prawdopodobieństwo, że dwie kolejne litery będą *a*.

8. Jeśli litery słowa *about* są umieszczane losowo w rzędzie, jakie jest prawdopodobieństwo, że trzy kolejne litery będą samogłoskami.
9. Para chce mieć troje dzieci. Przypuśćmy, że każda płeć dziecka jest równie prawdopodobna. Podaj prawdopodobieństwo, że:
- para będzie miała co najmniej jednego chłopca,
 - nie będzie miała dziewczyny starszej niż chłopiec,
 - para będzie mieć dokładnie jedną dziewczynę.
10. Dwa statki A i B mają czas przybycia między godz. **13** a **17**. Oba statki muszą zacumować na tym samym nabrzeżu. Po zadokowaniu każdy z nich potrzebuje **30** minut na uzupełnienie zapasów i opuszczenie doku. Jakie jest prawdopodobieństwo
- że statki nie będą musiały czekać na miejsce do zacumowania?
 - że statek A nie będzie musiał czekać na miejsce do zacumowania?
 - że statek B będzie musiał czekać na miejsce do zacumowania?
11. Załóżmy, że dwie kule są losowane bez wymiany (pierwsza kula nie jest zastępowana przed losowaniem drugiej) z torby zawierającej 4 czerwone i 3 czarne kule. Niech *A* będzie zdarzeniem polegającym na wylosowaniu czerwonej kuli za pierwszym razem a zdarzenie *B* wylosowania czerwonej kuli za drugim razem. Czy zdarzenia *A* and *B* są niezależne.
12. Trzy identyczne miski są oznaczone 1, 2, 3. Pierwsza miska zawiera 3 czerwone i 3 niebieskie kulki. Druga miska zawiera 4 czerwone i 2 niebieskie kulki. Trzecia miska zawiera 1 czerwoną i 5 niebieskich kulek. Miska jest wybierana losowo, a następnie jest losowo wybierana kulka z miski. a) Jakie jest prawdopodobieństwo, że wylosowana kulka jest niebieska? b) Biorąc pod uwagę, że wylosowana kulka jest czerwona, jakie jest prawdopodobieństwo, że została wybrana miska 2.
13. Załóżmy, że prawdopodobieństwo, że Jan rozwiąże pewien problem, wynosi $\frac{2}{3}$, że Maria go rozwiąże, jest $\frac{3}{4}$ a że Lech go rozwiąże jest $\frac{1}{2}$. Jakie jest prawdopodobieństwo
- że przynajmniej jedna osoba go rozwiąże?
 - że Maria i Lech rozwiążą go, ale Jan nie?
 - że Jan i Maria rozwiążą go, ale Lech nie?
 - że co najmniej dwie osoby go rozwiążą?
14. Pudełko zawiera trzy monety: dwie zwykłe monety i jedną fałszywą monetę z dwoma orłami ($P(O) = 1$). Wybieramy losowo monetę i ją rzucamy.
- Jakie jest prawdopodobieństwo, że wyląduje orłem do góry?
 - Wybieramy losowo monetę, rzucamy nią i wypada orzeł. Jakie jest prawdopodobieństwo, że jest to moneta z dwoma orłami?

15. Rzucamy uczciwą monetą trzy razy:
- Jakie jest prawdopodobieństwo wypadnięcia trzech orłów?
 - Jakie jest prawdopodobieństwo, że wypadnie dokładnie jeden orzeł?
 - Biorąc pod uwagę, że wypadł co najmniej jeden orzeł, jakie jest prawdopodobieństwo, że wypadną co najmniej dwa orły?
16. Artykuły przechodzące przez linię inspekcyjną są kontrolowane wzrokowo przez dwóch kolejnych inspektorów. Kiedy wadliwy artykuł przechodzi przez linię kontrolną, prawdopodobieństwo, że dostanie się do pierwszego inspektora, wynosi 0.1 . Drugi inspektor przegapi pięć na dziesięć wadliwych artykułów, które przejdą przez pierwszego inspektora. Jakie jest prawdopodobieństwo, że wadliwy przedmiot przedostanie się przez obu inspektorów?
17. Na egzaminie dwa problemy 1 and 2, są zadane. 35% studentów rozwiązało problem 1 a 15% studentów rozwiązało oba problemy. Ilu uczniów, którzy rozwiązali pierwszy problem, rozwiązało również drugi?
18. Z grupy 50 osób 35 pali papierosy. Wśród nich znajduje się 20 mężczyzn. Wśród osób niepalących znajduje się 10 kobiet. Jakie jest prawdopodobieństwo, że jeśli osoba wybrana losowo jest mężczyzną, to jest palaczem?
19. Pewna choroba ma częstość występowania 2% . Załóżmy, że dla pewnego testu diagnostycznego wskaźnik fałszywie ujemny wynosi 1% a wskaźnik fałszywie dodatni wynosi 1% . Oblicz prawdopodobieństwo, że osoba wybrana losowo z populacji:
- która ma pozytywny wynik testu, faktycznie ma chorobę.
 - która ma negatywny wynik testu, w rzeczywistości nie ma choroby
20. Rozkład prawdopodobieństwa dla zmiennej losowej X jest podany w tabeli.

x_i	-3	-1	0	2	3
p_i	0.1	0.4	0.2	0.2	0.1

Znajdź średnią, wariancję i odchylenie standardowe X .

21. Pięciu na stu mężczyzn i dwie na 1000 kobiet są daltonistami. Z grupy o tej samej liczbie mężczyzn i kobiet wybrano losowo jedną osobę. Osoba ta okazała się nie rozróżniać kolorów. Jakie jest prawdopodobieństwo, że był to mężczyzna?
22. Trzy identyczne miski są oznaczone 1, 2, 3. Pierwsza miska zawiera 3 czerwone, 4 niebieskie i 3 czarne kule. Druga miska zawiera 6 czerwonych, 2 niebieskie i 2 czarne kule. Trzecia miska zawiera 2 czerwone, 5 niebieskich i 3 czarne kule. Miska jest wybierana losowo, a następnie kula jest losowo wybierana z tej miski.
- Jakie jest prawdopodobieństwo, że wybrana kula jest czarna?

- b) Wiedząc, że wybrana kula jest czarna, obliczyć jakie jest prawdopodobieństwo, że miska 3 została wybrana?
- c) Wiedząc, że wybrana kula jest niebieska, jakie jest prawdopodobieństwo, że miska 1 została wybrana?
23. Trzy identyczne miski są oznaczone 1, 2, 3. Pierwsza miska zawiera 3 czerwone, 4 niebieskie i 3 czarne kule. Druga miska zawiera 6 czerwonych, 2 niebieskie i 2 czarne kule. Trzecia miska zawiera 2 czerwone, 5 niebieskich i 3 czarne kule. Miska jest wybierana losowo, a następnie dwie kule są losowo wybierane bez zwracania z tej miski.
- a) Jakie jest prawdopodobieństwo, że obie wybrane kulki są niebieskie?
- b) Wiedząc, że obie wybrane kule są niebieskie, jakie jest prawdopodobieństwo, że została wybrana miska 2?
- c) Wiedząc, że wybrano kule niebieską i czarną, jakie jest prawdopodobieństwo, że miska 2 została wybrana?
24. Kierownik w fabryce ma trzech mężczyzn i trzy kobiety pracujących w jego dziale. Chce wybrać dwóch pracowników do specjalnej pracy. Postanawia ich wybrać losowo. Niech Y oznacza liczbę kobiet w jego wyborze. Znajdź rozkład prawdopodobieństwa dla Y .
25. Każda z trzech kulek jest losowo umieszczana w jednej z trzech misek. Znajdź rozkład prawdopodobieństwa dla $Y =$ liczba pustych misek.
26. Zbalansowana moneta jest rzuca trzy razy. Niech Y równa się liczbie orłów które wypadły.
- a) Oblicz prawdopodobieństwa związane z $Y = 0, 1, 2, \text{ i } 3$.
- b) Utwórz tabelę rozkładu prawdopodobieństwa.
- c) Znajdź wartość oczekiwaną i odchylenie standardowe Y .
27. Firma ubezpieczeniowa proponuje roczną \$2000 polisę ubezpieczeniową od zdarzenia A , które historycznie zdarza się 1 raz na każdym 100 posiadaczy polisy. Ile firma powinna pobierać za polisę, jeśli zakłada, aby oczekiwany zysk na każdej polisie wynosił \$75?
28. Koszykarz wykonuje 4 niezależne rzuty osobiste z prawdopodobieństwem 0.7 trafienia do kosza (punkt) w każdym rzucie. Niech $Y =$ liczba punktów, które uzyskuje. Znajdź rozkład prawdopodobieństwa dla zmiennej losowej Y . Znajdź prawdopodobieństwo, że uzyska co najmniej 3 trafienia.
29. Załóżmy, że radio zawiera sześć tranzystorów, z których dwa są wadliwe. Trzy tranzystory są wybierane losowo, usuwane z radia i sprawdzane. Niech X jest równe liczbie zaobserwowanych wad. Znajdź rozkład prawdopodobieństwa dla X .
30. Załóżmy, że dwie kulki są losowane z wymianą (pierwsza kula jest zastępowana przed losowaniem drugiej) z torby zawierającej 5 czerwonych i 3 czarne kule. Niech X jest

równe liczbie wylosowanych czarnych kulek. Znajdź rozkład prawdopodobieństwa dla zmiennej losowej X .

31. Załóżmy, że dwie kulki są losowane bez wymiany z torby zawierającej 5 czerwonych i 3 czarne kule. Niech X jest równa liczbie wylosowanych czarnych kulek. Znajdowanie rozkładu prawdopodobieństwa dla zmiennej losowej X .
32. Rozkład prawdopodobieństwa dla zmiennej losowej X jest podana w tabeli.

x_i	0	2	4	5	6
p_i	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{12}$

- a) Znajdź dystrybuantę rozkładu zmiennej X . Naszkicuj wykres.
- b) Oblicz: i) $P(2 \leq X \leq 4)$ ii) $P(0 < X < 4)$ iii) $P(X > 1)$.
33. Piłkarz wykonuje trzy niezależne rzuty karne z prawdopodobieństwem $\frac{4}{5}$ strzelenia gola przy każdym strzale. Niech X będzie liczbą bramek, które zdobywa. Znajdź rozkład prawdopodobieństwa dla zmiennej losowej X . Znajdź oczekiwaną liczbę bramek.
34. Zmienna X ma rozkład normalny ze średnią 10 oraz odchyleniem standardowym wynoszącym 2. Jeden wynik jest losowo wybierany. Jakie jest prawdopodobieństwo, że jest między 11 a 12?
35. Głębokość morza zmierzono w 5 niezależnych pomiarach, a wyniki (w metrach) były następujące: 862, 870, 876, 866, 871. Wiedząc, że rozkład pomiarów jest rozkładem normalnym z odchyleniem standardowym 5 m, a poziom istotności jest $\alpha = 0,05$, zweryfikuj hipotezę, że średnia głębokość morza na tym obszarze wynosi 870m.
36. W eksperymencie biochemicznym zmierzono długość życia pewnego organizmu. Rozkład czasu życia tego organizmu można przyjąć jako rozkład normalny. Wykonano 8 pomiarów, a wyniki (w godzinach) wynoszą: 4.7; 5.3; 4.0; 3.8; 6.2; 5.5; 4.5; 6.0. Przyjmując poziomu istotności $\alpha = 0,05$, zweryfikować hipotezę, że średnia długość życia tych organizmów wynosi 4.0 godziny.
37. Znajdź funkcję liniowej regresji dla danych: $x = -1, 1, 2$, $y = 1, -2, 2$.
38. Pewien złożony układ ma średni wskaźnik awaryjności $\lambda = 0.005$ awarii lamp na godzinę. Jaka jest niezawodność systemu przez 60 dni, jeżeli liczba uszkodzeń lamp nie może przekroczyć 2?

Rozwiązania:

1. 35
2. 200
3. 1260
4. 720
5. 4096
6. 20
7. $\frac{2}{7}$
8. $\frac{1}{10}$
9. a) $\frac{7}{8}$ b) $\frac{1}{2}$ c) $\frac{3}{8}$
10. a) ≈ 0.7656 b) ≈ 0.8828 c) ≈ 0.1172
11. not
12. a) $\frac{5}{9}$ b) $\frac{1}{2}$
13. a) $\frac{23}{24}$ b) $\frac{1}{8}$ c) $\frac{1}{4}$ d) $\frac{17}{24}$
14. a) $\frac{2}{3}$ b) $\frac{1}{2}$
15. a) $\frac{1}{8}$ b) $\frac{3}{8}$ c) $\frac{4}{7}$
16. 0.05
17. $\approx 42.8\%$
18. $\frac{4}{5}$
19. a) ≈ 0.6689 b) ≈ 0.9998
20. 0, 3, $\sqrt{3}$
21. $\frac{25}{26}$
22. a) $\frac{4}{15}$ b) $\frac{3}{8}$ c) $\frac{4}{11}$
23. a) $\frac{17}{135}$ b) $\frac{1}{17}$ c) $\frac{4}{393}$
24. $P(Y = 0) = \frac{1}{5}$, $P(Y = 1) = \frac{3}{5}$, $P(Y = 2) = \frac{1}{5}$.
25. $P(Y = 0) = \frac{2}{9}$, $P(Y = 1) = \frac{2}{3}$, $P(Y = 2) = \frac{1}{9}$.
26. a) $P(Y = 0) = \frac{1}{8}$, $P(Y = 1) = \frac{3}{8}$, $P(Y = 2) = \frac{3}{8}$, $P(Y = 3) = \frac{1}{8}$ c) $E(Y) = 1.5$,
 $V(Y) = 0.75$.
27. \$95
28. $P(Y = 0) = 0.0081$, $P(Y = 1) = 0.0756$, $P(Y = 2) = 0.2646$, $P(Y = 3) = 0.4116$,
 $P(Y = 4) = 0.2401$, $P(Y \geq 3) = 0.6517$.
29. $P(X = 0) = \frac{1}{5}$, $P(X = 1) = \frac{3}{5}$, $P(X = 2) = \frac{1}{5}$.
30. $P(X = 0) = \frac{25}{64}$, $P(X = 1) = \frac{15}{32}$, $P(X = 2) = \frac{9}{64}$.
31. $P(X = 0) = \frac{5}{14}$, $P(X = 1) = \frac{15}{28}$, $P(X = 2) = \frac{3}{28}$.

32. a) $F(x) = 0$ $x < 0$, $F(x) = \frac{1}{3}$ $0 \leq x < 2$, $F(x) = \frac{5}{12}$ $2 \leq x < 4$, $F(x) = \frac{3}{4}$ $4 \leq x < 5$, $F(x) = \frac{11}{12}$ $5 \leq x < 6$, $F(x) = 1$ $x \geq 6$. b) i) $\frac{5}{12}$ ii) $\frac{1}{12}$ iii) $\frac{2}{3}$.

33. $P(X = 0) = \frac{1}{125}$, $P(X = 1) = \frac{12}{125}$, $P(X = 2) = \frac{48}{125}$, $P(X = 3) = \frac{64}{125}$.
 $E(X) = 2.4$.

34. 0.1499

35. $|u| = 0.447 < 1.96 = u_{\alpha}$, H_0 nie można odrzucić

36. $|t| = 3.17 > 2.365 = t_{\alpha}$, H_0 musi zostać odrzucona

37. $y \cong 0.0714x + 0.2857$

38. ≈ 0.0255