

Zmienne losowe

Zmienne losowe oznaczają się dużymi literami alfabetu łacińskiego, na przykład X , Y , Z . Natomiast wartości jakie one przyjmują odpowiednio małymi literami x , y , z i nazywają się je realizacjami.

Zmienną losową nazywamy zbiór wartości, które mogą być realizowane i którym przyporządkowane są jednoznacznie prawdopodobieństwa poszczególnych realizacji.

Zmienne losowe

Rozróżnia się dwa rodzaje zmiennych losowych a mianowicie: skokowe (dyskretne) i ciągłe.

Zmienne losowe, które przyjmują skończoną lub co najwyżej przeliczalną liczbę wartości nazywamy *skokowymi*. Przykładem może być tutaj liczba zakupionych lokali mieszkalnych w danej miejscowości, w danym okresie, albo liczba zajmowanych pokoi w mieszkaniu.

Przyporządkowanie wszystkim wartościom zmiennej losowej X prawdopodobieństwa określa rozkład zmiennej losowej skokowej.

Przyporządkowanie to nazywamy funkcją prawdopodobieństwa zmiennej losowej skokowej. Jeżeli oznaczymy przez x_i ($i = 1, 2, \dots, n$) wszystkie możliwe wartości zmiennej losowej X , to wówczas funkcję prawdopodobieństwa zapisujemy w następujący sposób:

$$P(X = x_i) = p(x_i) = p_i \quad i = 1, 2, \dots, n$$

przy czym:

$$\sum_{i=1}^n p_i = 1$$

W przypadku, gdy zbiór wartości zmiennej losowej X jest nieskończony, wtedy spełniona jest równość: $\sum_{i=1}^{\infty} p_i = 1$.

Rozkład zmiennej losowej X można także zapisać za pomocą funkcji zwanej dystrybuantą:

$$F(x) = P(X < x).$$

Wartościami dystrybuanty są więc prawdopodobieństwa zdarzeń losowych polegających na tym, że zmienna losowa X przyjmuje wartości mniejsze niż liczba rzeczywista x .

Zmienna losowa, której wartością może być każda liczba rzeczywista z określonego przedziału tych liczb nazywamy zmienną losową ciągłą.

Rozkład zmiennej losowej ciągłej opisuje się za pomocą nieujemnej ciągłej funkcji $f(x)$, zwanej funkcją gęstości prawdopodobieństwa, mającą następującą własność:

$$\int_a^b f(x)dx = P(a < X < b)$$

dla każdego $a \leq b$.

Dystrybuanta dowolnej zmiennej losowej ma następujące własności:

$$0 \leq F(x) \leq 1, \quad \text{dla wszystkich } x$$

$$\text{Jeśli, } x_2 > x_1 \quad \text{to, } F(x_2) \geq F(x_1)$$

$$\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$$

$$\lim_{x \rightarrow +\infty} F(x) = F(+\infty) = 1$$

ROZKŁAD DWUMIANOWY

W rozkładzie dwumianowym, nazywanym również od nazwiska jego twórcy rozkładem Bernoulliego, można wyróżnić dwa zdarzenia: sukces (występujący z prawdopodobieństwem p) i porażkę (występującą z prawdopodobieństwem q).

Prawdopodobieństwo p i q dla wszystkich losowań jest stałe, a losowania są niezależne.

Zmienną losową X jest liczba sukcesów. Prawdopodobieństwo uzyskania k sukcesów oblicza się według wzoru:

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

gdzie:

$$q = 1 - p; \quad 0 \leq p \leq 1;$$

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

n – liczba losowań,

p - prawdopodobieństwo sukcesu w pojedynczym losowaniu,

q - prawdopodobieństwo porażki w pojedynczym losowaniu.

Obliczane prawdopodobieństwa są składnikami rozwinięcia dwumianu Newtona $(p+q)^n$. Dlatego też, omawiany rozkład nazywany jest rozkładem dwumianowym (binomialnym). Fakt, że zmienna losowa X ma rozkład dwumianowy o parametrach n i p zapisujemy następująco: $X:B(n,p)$.

Przykład

Właściciel pewnej firmy zajmującej się pośrednictwem nieruchomości na podstawie swoich obserwacji ocenił, że 30% klientów, którzy oglądali przynajmniej dwukrotnie daną nieruchomość, zdecydowali się w przyszłości na jej zakup.

W ostatnim miesiącu powyższy warunek spełniło 10 klientów. Obliczyć prawdopodobieństwo sprzedaży przynajmniej dwóch nieruchomości.

Rozwiązanie

$$p = 0,3$$

$$q = 0,7$$

$$n = 10$$

Należy obliczyć prawdopodobieństwo $P(X \geq 2)$.

W tym przypadku łatwiej będzie obliczyć zdarzenie przeciwne $P(X < 2)$, wiedząc że suma zdarzeń przeciwnych równa jest 1.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - (P(X = 1) + P(X = 0))$$

$$P(X = 1) = \binom{10}{1} (0,3)^1 \cdot (0,7)^9 = \frac{10!}{9!1!} \cdot (0,3)^1 \cdot (0,7)^9 = 10 \cdot 0,3 \cdot 0,040354 \approx 0,121$$

=ROZKŁAD.DWUM(1;10;0,3;FAŁSZ)

$$P(X = 0) = \binom{10}{0} (0,3)^0 \cdot (0,7)^{10} = 1 \cdot 1 \cdot (0,7)^{10} \approx 0,0282$$

=ROZKŁAD.DWUM(0;10;0,3;FAŁSZ)

$$P(X \geq 2) = 1 - (0,121 + 0,0282) = 0,8508$$

Prawdopodobieństwo zdarzenia polegającego na tym, że przynajmniej dwóch z 10 klientów, którzy dwukrotnie oglądali tę samą nieruchomość w poprzednim miesiącu, dokona jej zakupu, wynosi 85,08%.

Rozkład Poissona

Rozkład Poissona jest rozkładem granicznym dla rozkładu dwumianowego.

Jeśli w rozkładzie Bernoulliego liczba doświadczeń n dąży do nieskończoności

a prawdopodobieństwo sukcesu k dąży do zera, wówczas prawdopodobieństwo $P(X=k)$ można obliczyć korzystając z następującego wzoru:

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

gdzie:

$$np = \lambda > 0 \quad ,$$

$e = 2,7182$ (e jest podstawą logarytmu naturalnego).

Rozkład Poissona wykorzystywany jest jako przybliżenie rozkładu dwumianowego. Przybliżenie to jest wystarczająco dobre wtedy, gdy $n \geq 20$ i jednocześnie $p \leq 0,2$.

Rozkład ten nazywany jest rozkładem zdarzeń rzadkich. Często wykorzystuje się go w statystycznej kontroli jakości, szczególnie wtedy, gdy - na przykład - kontroli podlega liczba wad w jednostce produktu.

Parametr λ jest zarówno wartością oczekiwaną jak i odchyleniem standardowym tego rozkładu. Fakt, że zmienna losowa X ma rozkład Poissona o parametrze λ zapisujemy następująco: $X: P(\lambda)$.

Przykład

Właściciel pewnej firmy pośredniczącej w sprzedaży nowych mieszkań stwierdził (na podstawie swoich doświadczeń), że 2% klientów reklamuje zakupione mieszkania z powodu występujących w nich wad budowlanych. Firma ta w zeszłym roku sprzedała 100 mieszkań. Jakie jest prawdopodobieństwo tego, że nie więcej niż jeden klient będzie reklamował zakupione mieszkanie?

Rozwiązanie

$$n = 100$$

$$p = 0,02$$

$$\lambda = 100 \cdot 0,02 = 2$$

$$P(X \leq 1) = P(X = 1) + P(X = 0)$$

$$P(X = 1) = \frac{2^1}{1} e^{-2} = 2 \cdot (2,7183)^{-2} = 0,2706$$

=ROZKŁAD.POISSON(1;2;FAŁSZ)

$$P(X = 0) = \frac{2^0}{1} e^{-2} = 1 \cdot (2,7183)^{-2} = 0,1353$$

=ROZKŁAD.POISSON(0;2;FAŁSZ)

$$P(X \leq 1) = 0,2706 + 0,1353 = 0,4059$$

Prawdopodobieństwo zdarzenia polegającego na tym, że wystąpi nie więcej niż jedna reklamacja wynosi 40,59%.

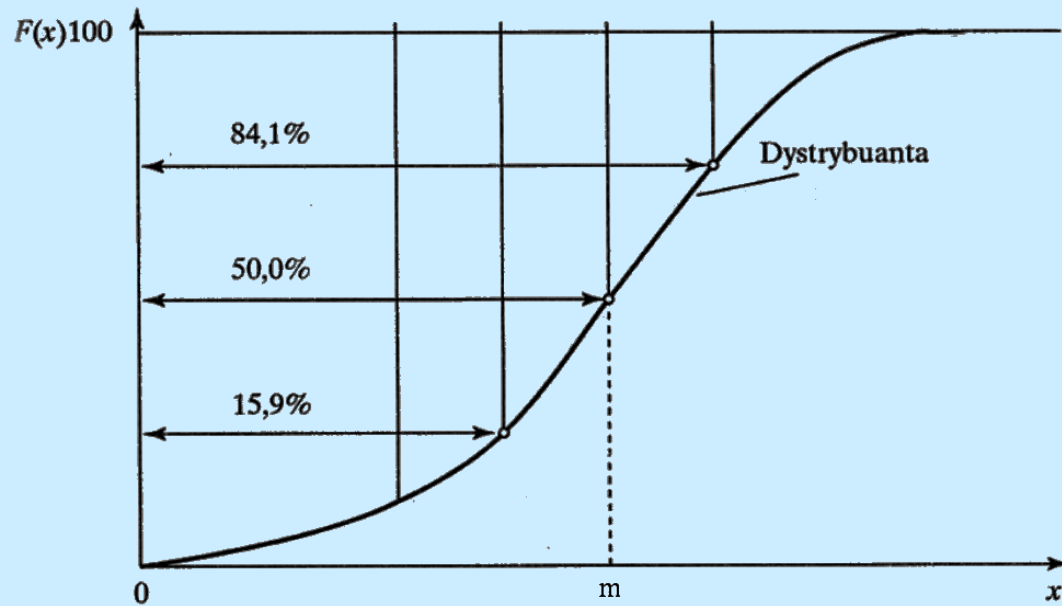
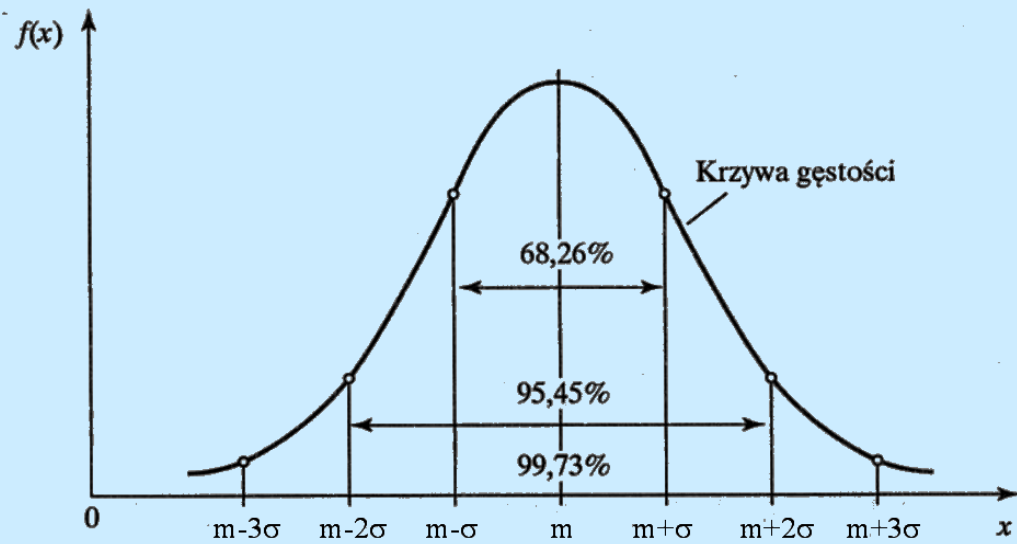
Rozkład normalny

Zmienna losowa X ma rozkład normalny $X:N(m,\sigma)$, jeżeli jej funkcja gęstości prawdopodobieństwa ma postać:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right], \quad -\infty < x < \infty$$

Własności rozkładu normalnego:

- jest rozkładem symetrycznym,
- do jego określenia potrzebne są dwa parametry:
wartość oczekiwana m i odchylenie standardowe σ ,
- wartość oczekiwana m jest jednocześnie medianą i modalną,
- wartości zmiennej losowej należą do zbioru liczb rzeczywistych, co jest charakterystyczne dla tego rozkładu.



Funkcja gęstości prawdopodobieństwa i dystrybuanta normalnej zmiennej losowej

Dystrybuanta zmiennej losowej $X:N(m,\sigma)$ określona jest wzorem:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx,$$

Rysunek przedstawia własności funkcji gęstości prawdopodobieństwa oraz dystrybuanty normalnej zmiennej losowej X .

Oprócz wyżej podanych własności na uwagę zasługuje reguła trzech sigm. Polega ona na tym, że niemal wszystkie możliwe realizacje zmiennej losowej $X:N(m,\sigma)$ należą do przedziału $(m-3\sigma ; m+3\sigma)$:

- 68,26% wartości zmiennej należy do przedziału $(m-\sigma ; m+\sigma)$,
- 95,45% wartości zmiennej należy do przedziału $(m-2\sigma ; m+2\sigma)$,
- 99,73% wartości zmiennej należy do przedziału $(m-3\sigma ; m+3\sigma)$.

W rozważaniach teoretycznych i zastosowaniach istotną rolę odgrywa tzw. standaryzowana (znormalizowana lub unormowana) normalna zmienna losowa:

$$U = \frac{X - E(X)}{D(X)},$$

o wartościach:

$$u = \frac{x_0 - m}{\sigma}.$$

Po dokonaniu zabiegu standaryzacji wartość oczekiwana znormalizowanej zmiennej losowej jest równa 0, natomiast odchylenie standardowe przyjmuje wartość 1.

W tablicach można odczytać wartości funkcji gęstości prawdopodobieństwa:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

lub dystrybuanty:

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^u \exp\left(-\frac{u^2}{2}\right) du.$$

Obliczanie prawdopodobieństwa

- a) Obliczanie prawdopodobieństwa zdarzenia losowego A , polegającego na tym, że zmienna losowa $X:N(m,\sigma)$ przyjmie wartość mniejszą od x_0 . Wykorzystując definicję dystrybuanty oraz dokonując normalizacji zmiennej losowej X zapisujemy

$$P(A) = P(X < x_0) = F(x_0) = \Phi\left(\frac{x_0 - m}{\sigma}\right) = \Phi(u_0).$$

Tablice dystrybuanty rozkładu normalnego opisane są z reguły dla wartości u z przedziału $[0;4]$. Tak więc:

dla $u_0 \geq 0$ wartość $\Phi(u_0)$ można odczytać bezpośrednio z tablicy

dla $u_0 < 0$ wartość $\Phi(u_0)$ obliczamy korzystając z zależności

$$\Phi(-u_0) = 1 - \Phi(u_0).$$

b) Obliczanie prawdopodobieństwa zdarzenia losowego B , polegającego na tym, że zmienna losowa $X: N(m, \sigma)$ przyjmie wartość większą niż x_0 .

$$P(B) = P(X > x_0) = 1 - F(x_0) = 1 - \Phi\left(\frac{x_0 - m}{\sigma}\right) = 1 - \Phi(u_0).$$

Dalsze postępowanie jest analogiczne do podpunktu a)

c) Obliczanie prawdopodobieństwa zdarzenia losowego C , polegającego na tym, że zmienna losowa $X: N(m, \sigma)$ przyjmie wartość z przedziału (x_1, x_2) :

$$P(C) = P(x_1 < X < x_2) = F(x_2) - F(x_1) = \Phi\left(\frac{x_2 - m}{\sigma}\right) - \Phi\left(\frac{x_1 - m}{\sigma}\right) = \Phi(u_2) - \Phi(u_1).$$

Przykład

Zakładając, że ceny jednostkowe lokali mieszkalnych z danego przykładu w miejscowości A, w pierwszym kwartale 2017 r., mają rozkład zbliżony do rozkładu normalnego $N(3460;241)$, oszacować:

- Jaki procent lokali mieszkalnych jest tańszych niż 3300 zł/m²?
- Jaki procent lokali mieszkalnych jest droższych niż 3600 zł/m²?
- Jaki procent lokali mieszkalnych znajduje się przedziale cenowym od 3400 do 3600 zł/m²?

Rozwiązanie

a) Korzystając z określonych wzorów obliczamy prawdopodobieństwo:

$$P(X < 3300) = F(3300) = \Phi\left(\frac{3300 - 3460}{241}\right) = \Phi(-0,664) = 1 - \Phi(0,664) = 1 - 0,745 = 25,5\%$$

=ROZKŁAD.NORMALNY(3300;3460;241;PRAWDA)

W celu uzyskania wartości dystrybuanty rozkładu normalnego $U: N(0;1)$ należy wpisać formułę:

=ROZKŁAD.NORMALNY.S (,wartość zmiennej standaryzowanej")

Ceny jednostkowe 25,5% lokali mieszkalnych są niższe niż 3300 zł/m².

b) Należy obliczyć prawdopodobieństwo:

$$P(X > 3600) = 1 - F(3600) = 1 - \Phi\left(\frac{3600 - 3460}{241}\right) = 1 - \Phi(0,581) = 1 - 0,719 = 28,1\%$$

=1 - ROZKŁAD.NORMALNY(3600;3460;241;PRAWDA)

Ceny jednostkowe 28,1% lokali mieszkalnych są wyższe niż 3600 zł/m².

Należy obliczyć prawdopodobieństwo :

$$\begin{aligned} P(3400 < X < 3600) &= F(3600) - F(3400) = \Phi\left(\frac{3600 - 3460}{241}\right) - \Phi\left(\frac{3400 - 3460}{241}\right) = \Phi(0,581) - \Phi(-0,249) = \\ &= \Phi(0,581) - (1 - \Phi(0,249)) = 0,719 - 1 + 0,599 = 31,8\% \end{aligned}$$

=ROZKŁAD.NORMALNY(3600;3460;241;PRAWDA)-
ROZKŁAD.NORMALNY (3400;3460;241;PRAWDA)

Do zadanego przedziału cenowego należy 31,8% lokali mieszkalnych.

Pamiętając o właściwościach rozkładu normalnego można stwierdzić, że w poszczególnych przedziałach cenowych:

3460 ± 241 [zł/m²] znajduje się 68,26% lokali mieszkalnych,

3460 ± 482 [zł/m²] znajduje się 95,45% lokali mieszkalnych,

3460 ± 723 [zł/m²] znajduje się 99,73% lokali mieszkalnych.

Reguła 3 σ

$$P(|X - m| > \sigma) = 0,3174$$

$$P(m - \sigma < X < m + \sigma) = 0,6826$$

$$P(|X - m| > 2\sigma) = 0,0455$$

$$P(m - 2\sigma < X < m + 2\sigma) = 0,9545$$

$$P(|X - m| > 3\sigma) = 0,0027$$

$$P(m - 3\sigma < X < m + 3\sigma) = 0,9973$$

$$P(|X - m| > u_\alpha \sigma) = 2\alpha$$

$$P(m - u_\alpha \sigma < X < m + u_\alpha \sigma) = 1 - 2\alpha$$