

Zmienna losowa

Zmienna losowa to zmienna liczbowa, której wartość zależy od wyniku eksperymentu.

Przykłady:

- liczba orłów uzyskanych w rzucie monetą,
- liczba orłów uzyskanych w 10 rzutach monetą,
- liczba oczek uzyskanych w rzucie kostką.

Zmienna losowa

Zmienna losowa dyskretna

Zmienna losowa dyskretna może przyjmować wartości ze zbioru skończonego lub przeliczalnego: $\{x_1, x_2, \dots\}$. Rozkład prawdopodobieństwa zmiennej losowej to przyporządkowanie prawdopodobieństwa (liczby z przedziału $[0, 1]$) poszczególnym wartościom.

Zmienna losowa ciągła

Zmienna losowa ciągła może przyjmować wszystkie wartości rzeczywiste z pewnego odcinka albo całej prostej. Rozkład prawdopodobieństwa zmiennej losowej ciągłej opisujemy podając (nieujemną) funkcję gęstości $f(x)$, która przyporządkowuje prawdopodobieństwo $\int_a^b f(x)dx$ (liczbę z przedziału $[0, 1]$) każdemu przedziałowi wartości $[a, b]$.

Zmienna losowa dyskretna

- $E(X) = \sum x_i \cdot P(X = x_i) = \mu_X$
- $Var(X) = \sum (x_i - \mu_X)^2 \cdot P(X = x_i) = \sum x_i^2 \cdot P(X = x_i) - \mu_X^2$

Zmienna losowa ciągła

- $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) = \mu_X$
- $Var(X) = \int_{-\infty}^{\infty} (x - EX)^2 \cdot f(x) = \int_{-\infty}^{\infty} x^2 f(x) - (EX)^2$

Własności wartości oczekiwanej i wariancji

$$E(aX + b) = aEX + b$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Dla dwóch zmiennych losowych X i Y :

$$E(X + Y) = EX + EY$$

$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$, o ile X i Y są niezależne

Z – standardowe oznaczenie na zmienną losową mającą standardowy rozkład normalny $N(0, 1)$

Przykład 1

Założmy, że poziom cholesterolu w pewnej populacji ma rozkład normalny o średniej $\mu = 220$ i odchyleniu standardowym $\sigma = 40$, tzn. zmienna Y ma rozkład $N(220, 40)$. Jaka część populacji ma poziom cholesterolu powyżej 240?

Zmienna $Z = (Y - \mu)/\sigma$ ma rozkład normalny o średniej 0 i odchyleniu standardowym 1

$$P(Y > 240) = P(Z > \frac{240-220}{40}) = P(Z > 0.5) = 1 - 0.6915 = 0.3085$$

Przykład 2

Poziom serum CK

$$n = 36, \quad \bar{x} = 98.28, \quad s = 40.38$$

$$\frac{26}{36} = 72\% \text{ obserwacji jest w przedziale } \bar{x} \pm s$$

$$\frac{34}{36} = 94\% \text{ obserwacji jest w przedziale } \bar{x} \pm 2s$$

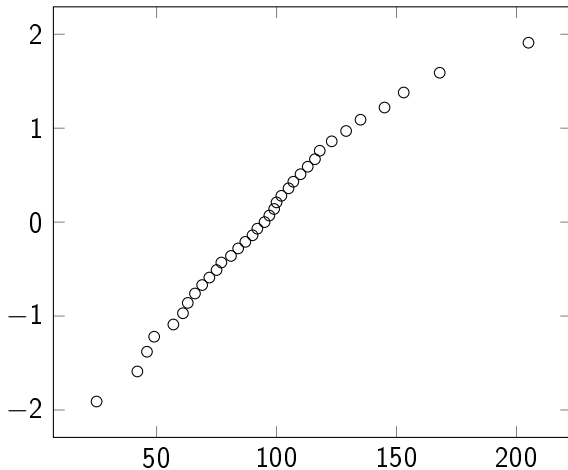
$$\frac{36}{36} = 100\% \text{ obserwacji jest w przedziale } \bar{x} \pm 3s$$

To w przybliżeniu odpowiada wartościom dla rozkładu normalnego.

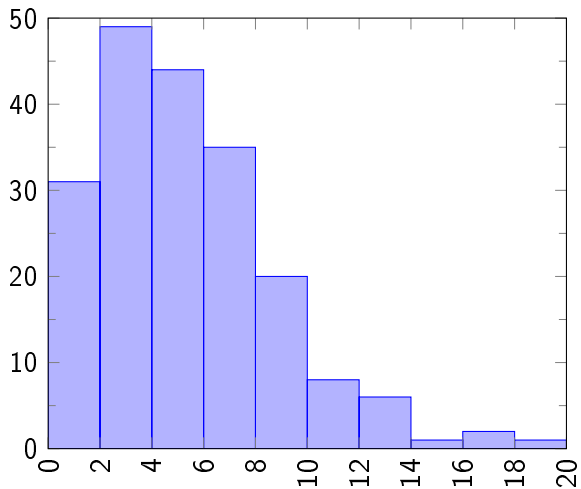
Obserwacja wyznaczająca:

- 2.5 centyl powinna odpowiadać wartości $z = -2$
 - 16 centyl powinna odpowiadać wartości $z = -1$
 - 50 centyl powinna odpowiadać wartości $z = 0$
 - 84 centyl powinna odpowiadać wartości $z = 1$
 - 97.5 centyl powinna odpowiadać wartości $z = 2$
- 1 Porządkujemy obserwacje od najmniejszej do największej i dla każdej ustalamy centyl, który ona wyznacza.
 - 2 Wyliczamy wartość z (standardowego rozkładu normalnego $N(0, 1)$) dla każdego z centyli.
 - 3 Rysujemy wykres wartości z dla każdej obserwacji. Jeśli dane mają rozkład zbliżony do normalnego, punkty na wykresie powinny leżeć blisko prostej.

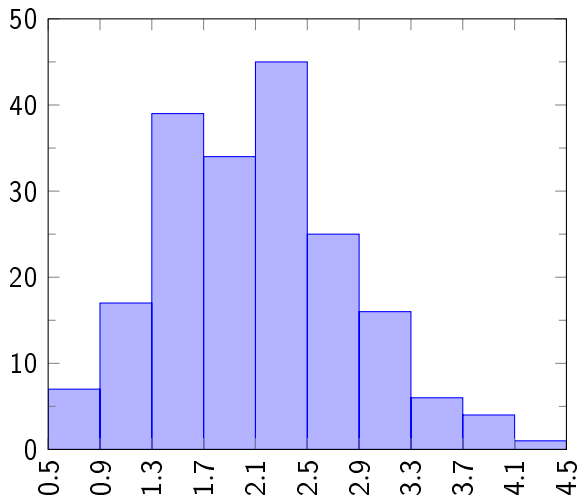
Przykład 2 (raz jeszcze)



Histogram zmiennej X



Histogram zmiennej $\sqrt{X}$



Histogram zmiennej $\ln X$

