Elementy statystyki opisowej

Wojciech Zieliński

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.

STATYSTYCZNA ANALIZA DANYCH: etap badania statystycznego polegający na wykrywaniu - przy użyciu odpowiednich metod - prawidłowości kształtowania się zjawisk statystycznych oraz związków i zależności między nimi, a także na interpretacji wyników badań i formułowaniu wniosków

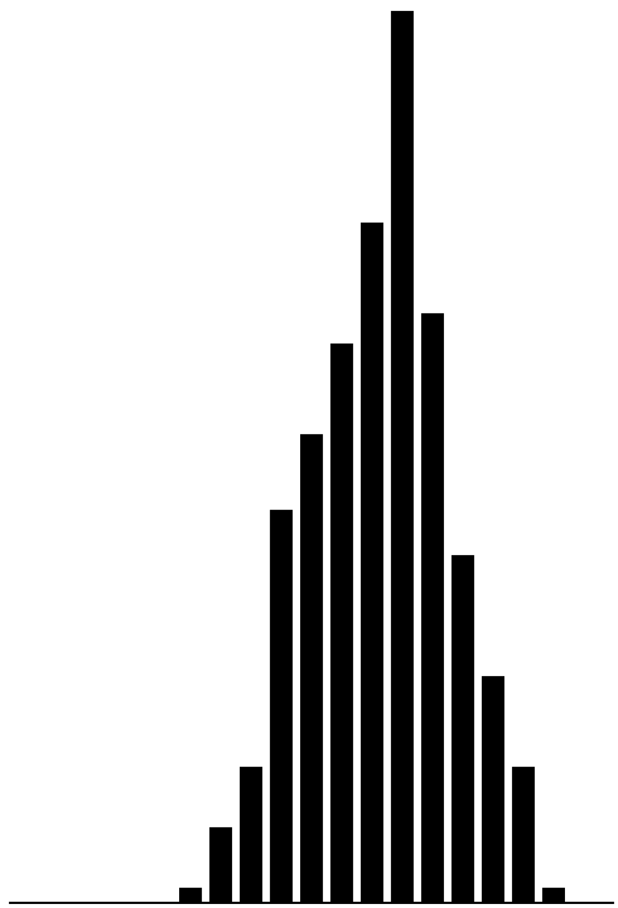
Literatura

Tukey J. W. 1977: Exploratory data analysis, Addison Wesley Publishing Company
Rao C. R. 1994: Statystyka i prawda, PWN
Domański C. 2001: Metody statystyczne, teoria i zadania, wyd. IV, Wydawnictwo UŁ
Zieliński W. 1999, Tablice Statystyczne, Wyd. III poprawione i uzupełnione, Fundacja Rozwój SGGW (wyd. IV - 2000; wyd. V - 2001)
. . . .
Gupta C. B. 1982, An Introduction to Statistical Methods, Wyd. IX, Vikas Publishing House PVT Ltd., New Delhi, India
Reichmann W. J. 1968, Drogi i bezdroża statystyki, PWN, Warszawa
Kassyk–Rokicka H. 1986, Statystyka nie jest trudna: mierniki statystyczne, PWE, Warszawa
Michalski T. 1994, Statystyka, WSiP, Warszawa

Punkty z klasówki

0.00:	0
0.05:	0
0.10:	0
0.15:	0
0.20:	0
0.25:	0
0.30:	1
0.35:	5
0.40:	9
0.45:	26
0.50:	31
0.55:	37
0.60:	45
0.65:	59
0.70:	39
0.75:	23
0.80:	15
0.85:	9
0.90:	1
0.95:	0
1.00:	0

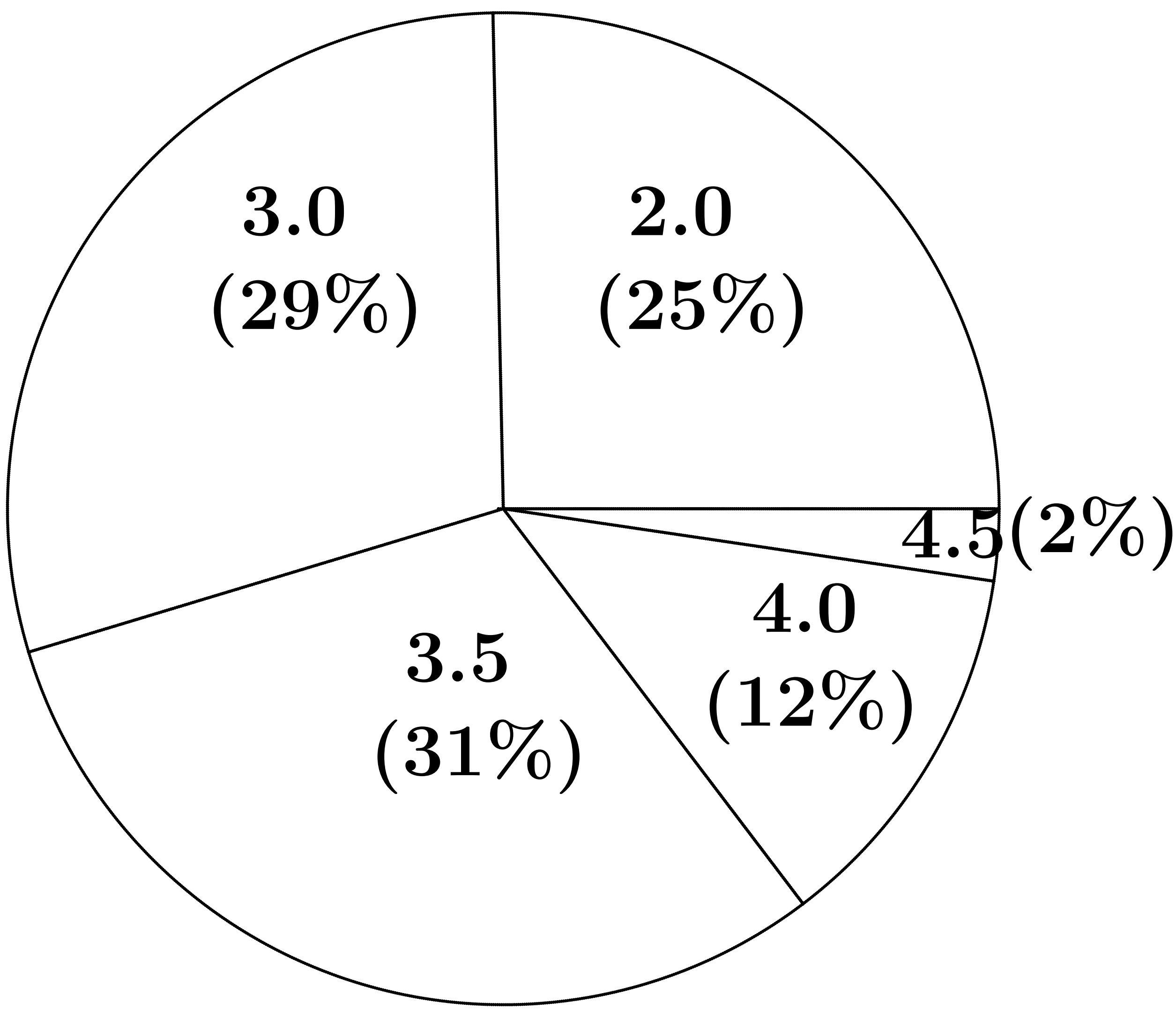
Średnia	0.589
Kwartyl dolny	0.508
Mediana	0.595
Kwartyl górny	0.672



Stopnie z klasówki

2.0:	76
3.0:	88
3.5:	92
4.0:	37
4.5:	7
5.0:	0

Średnia	3.06
Kwartyl dolny	2.00
Mediana	3.00
Kwartyl górny	3.50



Oceny z klasówki

	Negatywne	Pozytywne	Razem
Ogółem	76 (25%)	224 (75%)	300
Kobiety	46 (28%)	116 (72%)	162
Mężczyźni	30 (22%)	108 (78%)	138
Grupa 1	3 (10%)	27 (90%)	30
Grupa 2	11 (37%)	19 (63%)	30
Grupa 3	17 (57%)	13 (43%)	30
Grupa 4	0 (0%)	30 (100%)	30
Grupa 5	6 (20%)	24 (80%)	30
Grupa 6	4 (13%)	26 (87%)	30
Grupa 7	12 (40%)	18 (60%)	30
Grupa 8	3 (10%)	27 (90%)	30
Grupa 9	7 (23%)	23 (77%)	30
Grupa 10	13 (43%)	17 (57%)	30

Pytania

Jaka jest struktura ocen?

Jaki jest odsetek ocen pozytywnych?

Czy są różnice w ocenach między grupami?

Czy są różnice w ocenach między płciami?

$\Big\uparrow\Big\uparrow\Big\uparrow\Big\uparrow\Big\uparrow$ *Analiza danych* $\Big\uparrow\Big\uparrow\Big\uparrow\Big\uparrow\Big\uparrow$

. . . .

$\Big\downarrow\Big\downarrow\Big\downarrow\Big\downarrow\Big\downarrow$ *Wnioskowanie statystyczne* $\Big\downarrow\Big\downarrow\Big\downarrow\Big\downarrow\Big\downarrow$

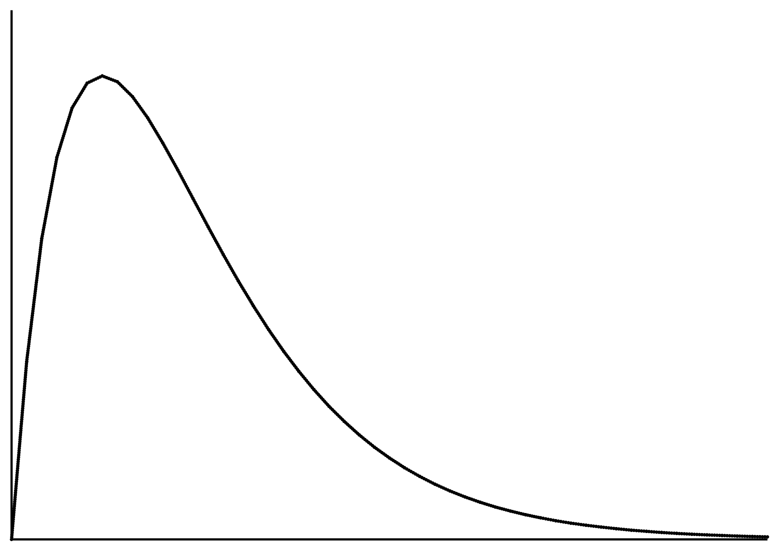
Jakich wyników można oczekiwać na następnej klasówce?

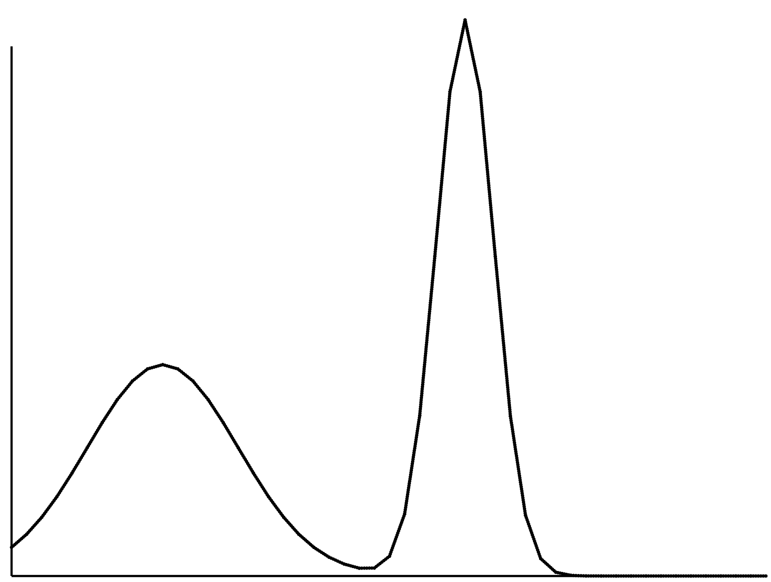
Jakich wyników można oczekiwać na egzaminie?

Jaki jest wpływ prowadzącego zajęcia?

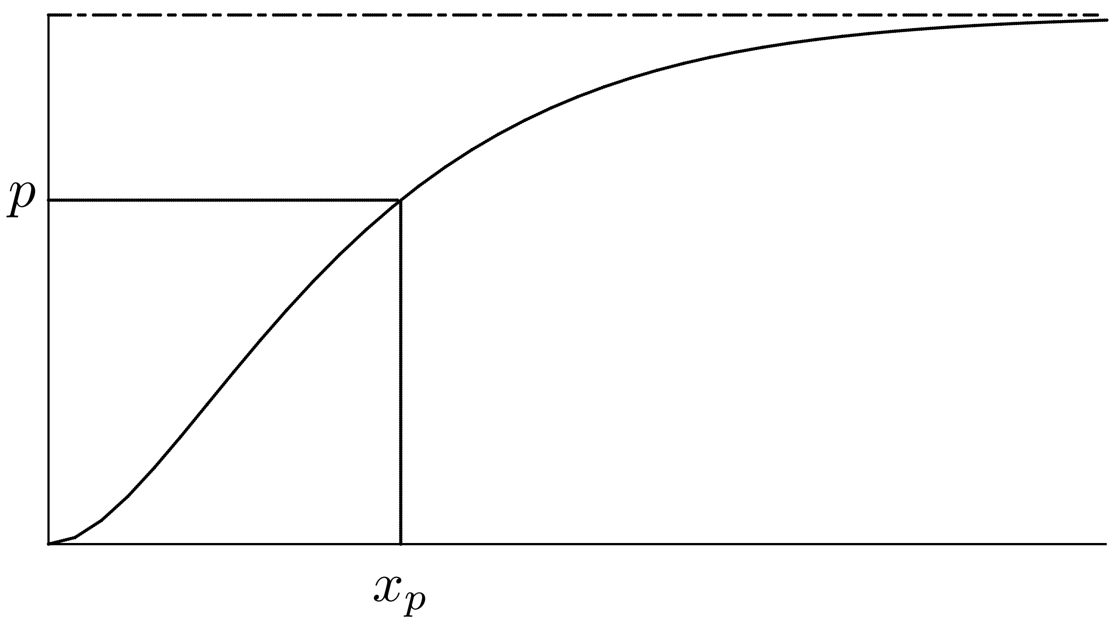
Niech ξ będzie zmienną losową o rozkładzie z dystrybuantą F i gęstością f.

Moda (dominanta):
punkt, w którym gęstość osiąga maksimum.





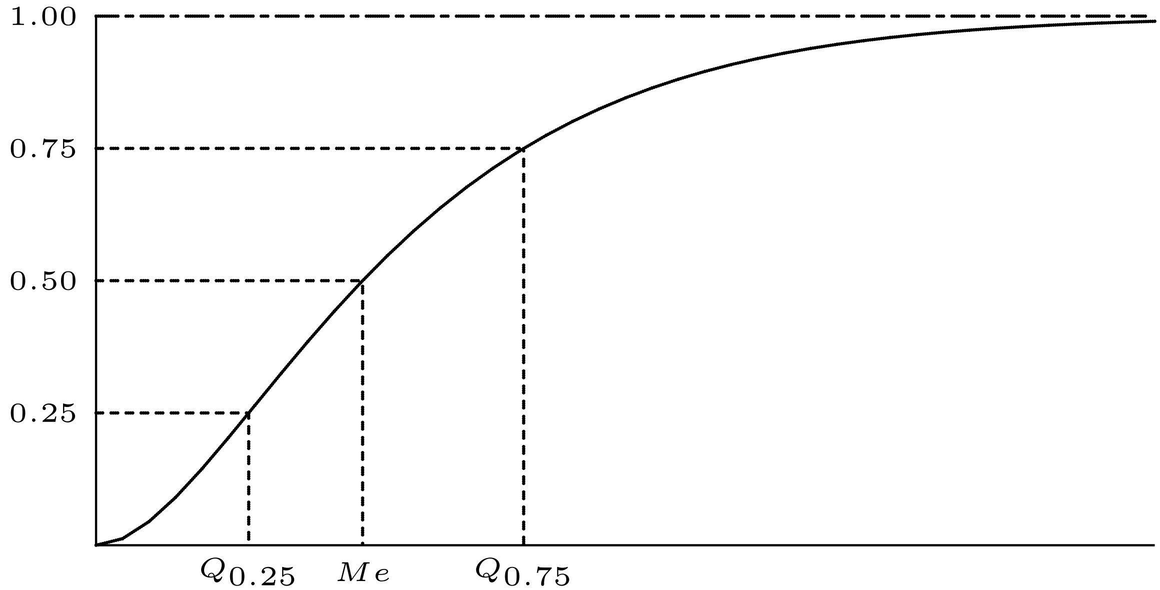
Kwantyl rzędu p: $x_p = \inf\{x : F(x) \geq p\}$

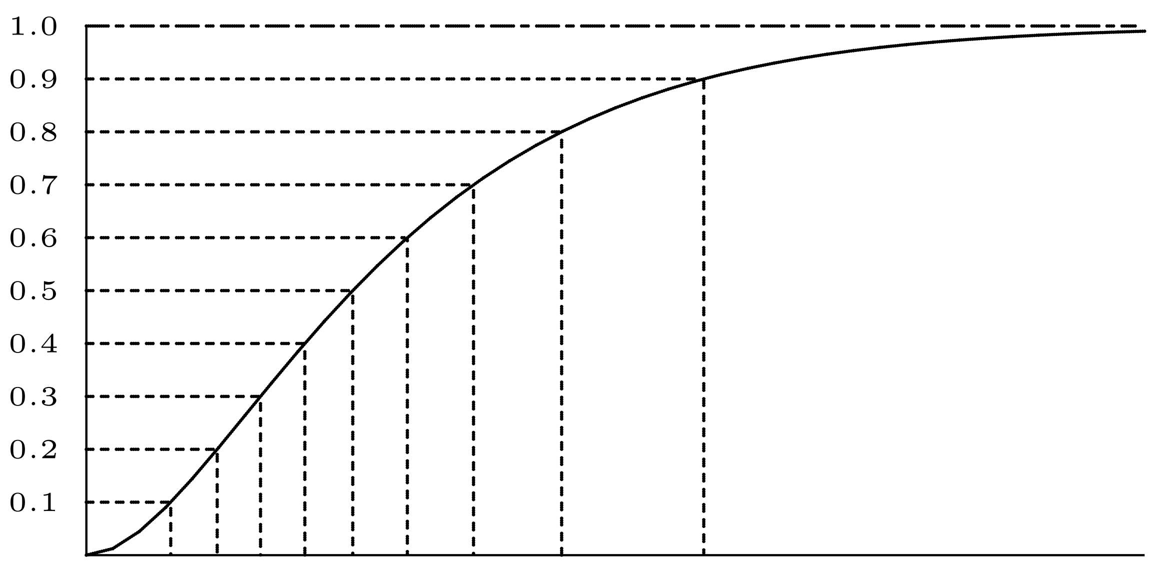


Mediana: $Me = x_{0.5}$

Dolny kwartyl: $Q_{0.25} = x_{0.25}$

Górny kwartyl: $Q_{0.75} = x_{0.75}$


1.00
0.75
0.50
0.25
$Q_{0.25}$
Me
$Q_{0.75}$


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

Moment rzędu k:

$$E\xi^k = \int_{-\infty}^{\infty} x^k dF(x)$$

Wartość oczekiwana:

$$E\xi = \int_{-\infty}^{\infty} x dF(x)$$

Moment centralny rzędu k:

$$E(\xi - E\xi)^k = \int_{-\infty}^{\infty} (x - E\xi)^k dF(x)$$

Wariancja:

$$D^2\xi = E(\xi - E\xi)^2 = \int_{-\infty}^{\infty} (x - E\xi)^2 dF(x)$$

Odchylenie standardowe:

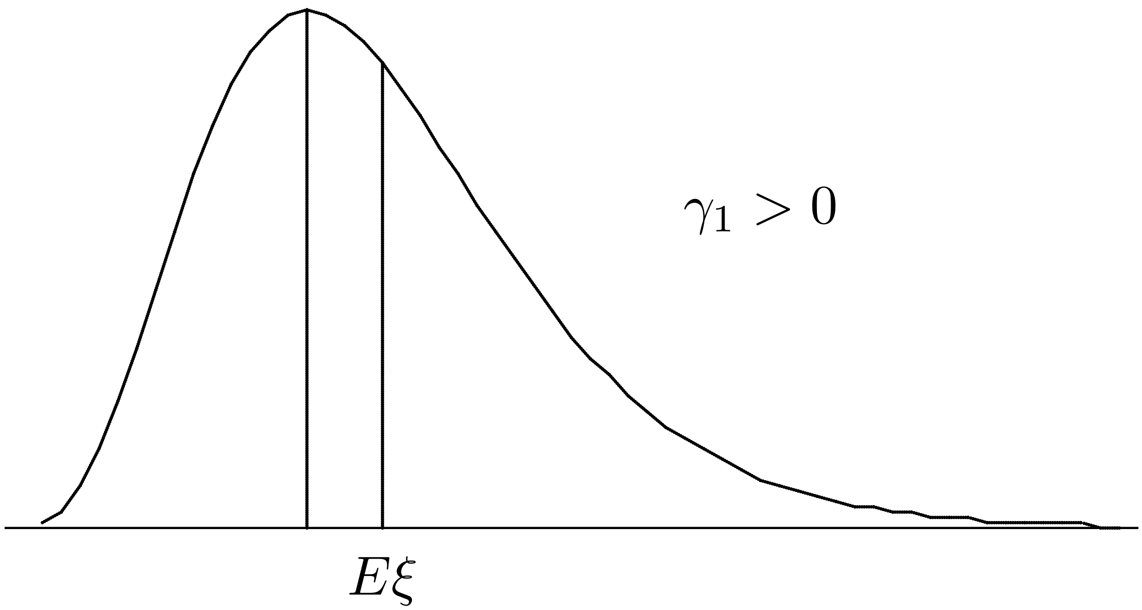
$$D\xi = \sqrt{D^2\xi}$$

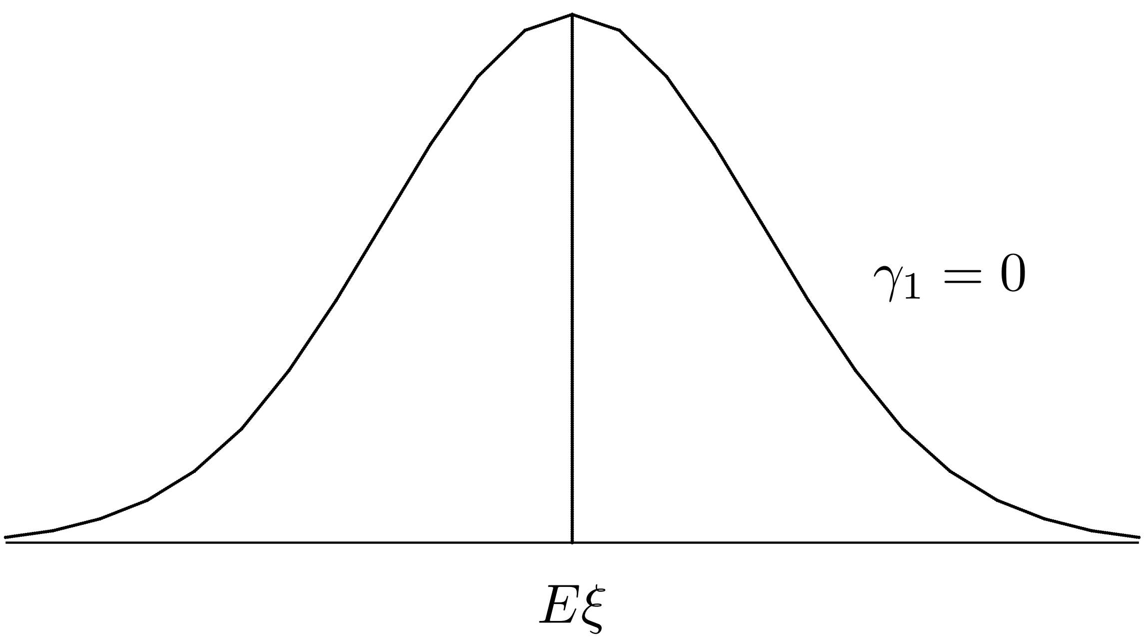
Moment absolutny rzędu k:

$$E|\xi - E\xi|^k = \int_{-\infty}^{\infty} |x - E\xi|^k dF(x)$$

Współczynnik asymetrii (skośności)

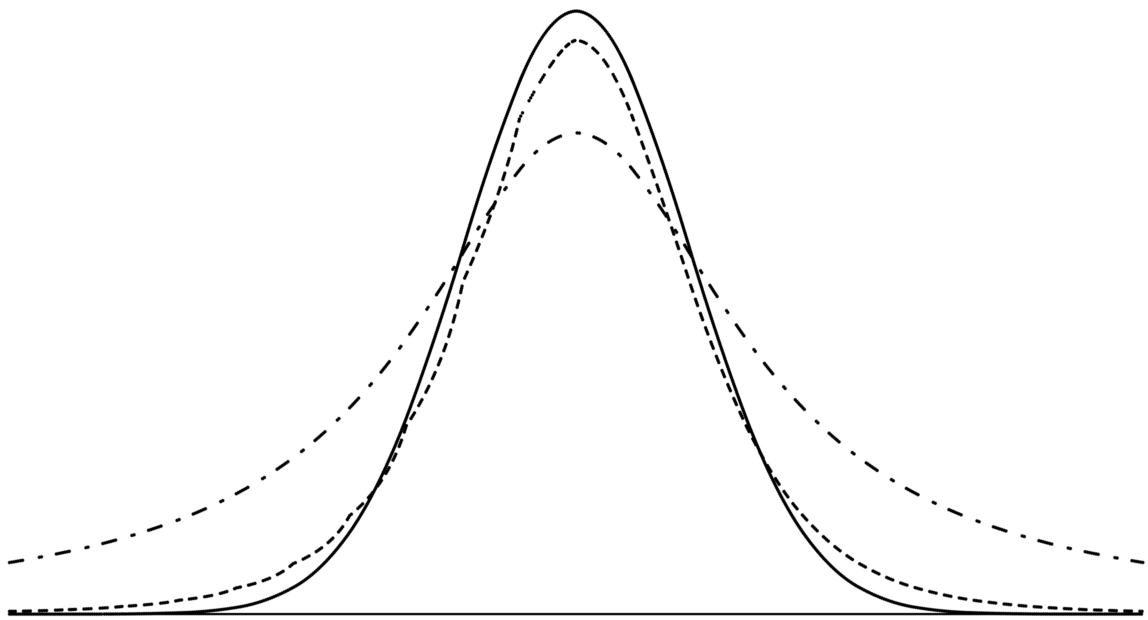
$$\gamma_1 = \frac{E(\xi - E\xi)^3}{(D\xi)^3}$$





Kurtoza (spłaszczenie, eksces)

$$\gamma_2 = \frac{E(\xi - E\xi)^4}{(D\xi)^4}$$



—— rozkład normalny ($\gamma_2 = 3$)

-------- rozkład Studenta $t(5)$ ($\gamma_2 = 6$)

- · - · - rozkład Cauchy'ego ($\gamma_2 = +\infty$)

Niech ξ_n, $n = 1, 2, \ldots$, będzie ciągiem zmiennych losowych oraz niech $S_n = \sum_{i=1}^{n} \xi_i$.

Jeżeli
$$\frac{S_n - ES_n}{n} \to 0 \ p.n.$$
to ciąg spełnia **mocne prawo wielkich liczb**.

Jeżeli
$$\frac{S_n - ES_n}{n} \to 0 \text{ według prawdopodobieństwa}$$
to ciąg spełnia **słabe prawo wielkich liczb**.

. . . .

Jeżeli $\xi_1, \xi_2, \ldots$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie oraz $E|\xi_1| < \infty$, to
$$\frac{S_n}{n} \to E\xi_1 \ p.n.$$

Centralne twierdzenie graniczne (CTG)

Jeżeli $\xi_1, \xi_2, \ldots$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie takim, że

$$E\xi_1 = \mu \quad \text{oraz} \quad D^2\xi_1 = \sigma^2 < \infty,$$

to

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1) \text{ według rozkładu}$$

. . . .

$$\sup_{x\in\mathbb{R}} \left| P\left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \le x \right\} - \Phi(x) \right| \to 0$$

gdzie $\Phi(x)$ jest dystrybuantą rozkładu $N(0,1)$

. . . .

$$\sup_{x\in\mathbb{R}} \left| P\left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \le x \right\} - \Phi(x) \right| \le C\,\frac{E|\xi_1 - E\xi_1|^3}{\sigma^2\sqrt{n}}$$

gdzie $1/\sqrt{2\pi} \le C < 0.8$

Rozkłady Pearsona

Rozkłady o gęstości $f(x)$ takiej, że

$$\frac{1}{f(x)}\frac{df(x)}{dx} = \frac{x-b}{c_0 + c_1 x + c_2 x^2}$$

Niech ξ będzie zmienną losową o rozkładzie Pearsona

$E\xi = \mu,\ D^2\xi = \sigma^2,\ E(\xi-\mu)^3 = \mu_3,\ E(\xi-\mu)^4 = \mu_4$

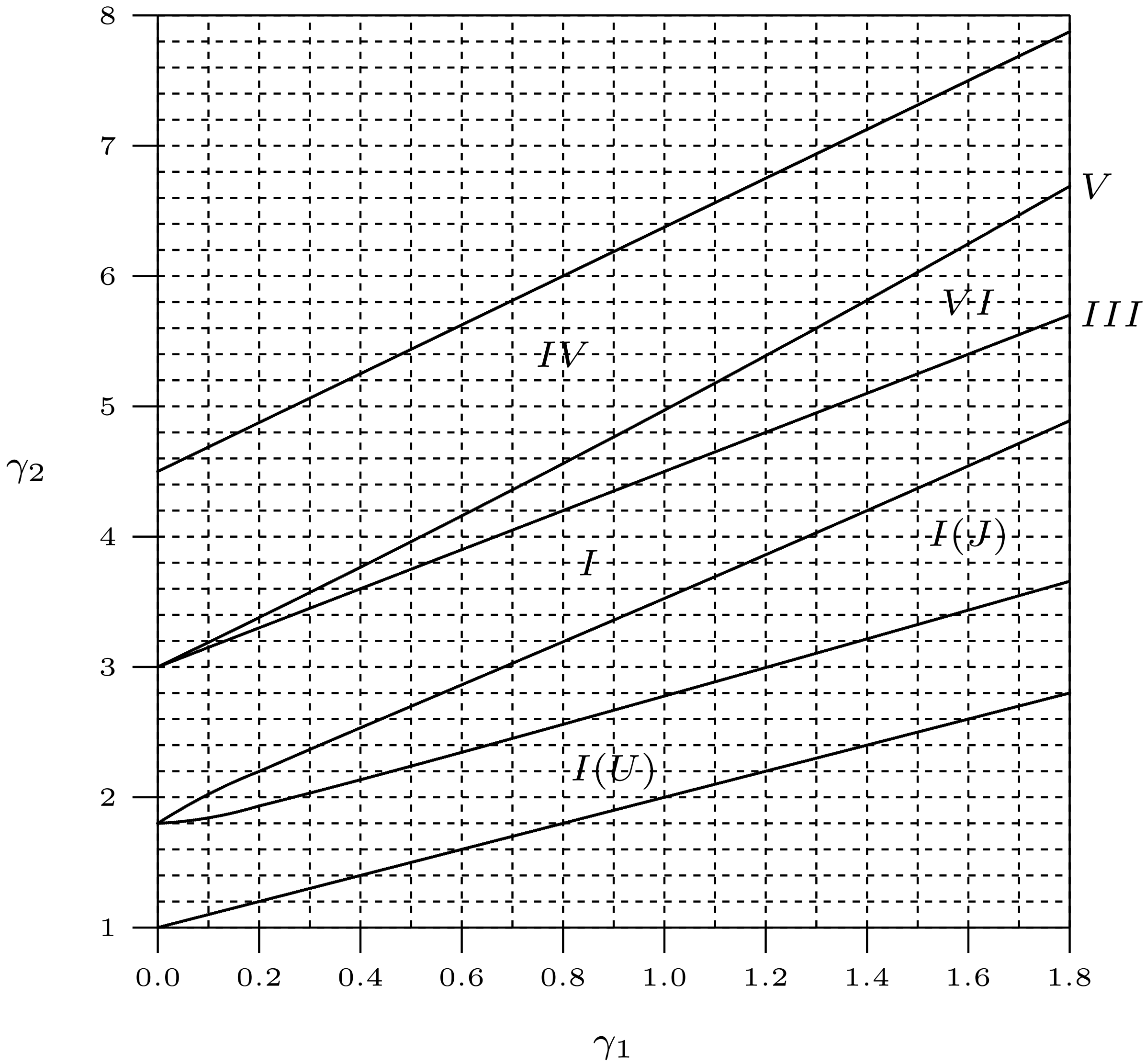
Zmienna losowa $(\xi-\mu)/\sigma$ ma rozkład Pearsona o gęstości spełniającej równanie

$$\frac{1}{f(x)}\frac{df(x)}{dx} = -\frac{x+b_1}{b_0 + b_1 x + b_2 x^2}$$

$$b_0 = \frac{4\gamma_2 - 3\gamma_1}{2(5\gamma_2 - 6\gamma_1 - 9)}$$

$$b_1 = \frac{\sqrt{\gamma_1}(\gamma_2 + 3)}{2(5\gamma_2 - 6\gamma_1 - 9)}$$

$$b_2 = \frac{2\gamma_2 - 3\gamma_1 - 6}{2(5\gamma_2 - 6\gamma_1 - 9)}$$



$$
\begin{array}{lcl}
II(U) & : & \gamma_1 = 0;\ 1 \leq \gamma_2 < 1.8 \\
II & : & \gamma_1 = 0;\ 1.8 < \gamma_2 < 3 \\
N & : & \gamma_1 = 0;\ \gamma_2 = 3 \\
VII & : & \gamma_1 = 0;\ 3 < \gamma_2 \leq 4.5
\end{array}
$$