1 Podstawy rachunku prawdopodobieństwa

Dystrybuantą zmiennej losowej X nazywamy prawdopodobieństwo przyjęcia przez zmienną losową X wartości mniejszej od x, tzn. $F(x) = P[X < x]$.

1. dla zmiennej losowej skokowej (dyskretnej)

$$F(x) = P[X < x] = \sum_{x_i < x} P[X = x_i] = \sum_{x_i < x} p_i$$

   $P[X < x]$ oblicza się przez sumowanie wszystkich prawdopodobieństw p_i dla $x_i < x$

2. dla zmiennej losowej ciągłej

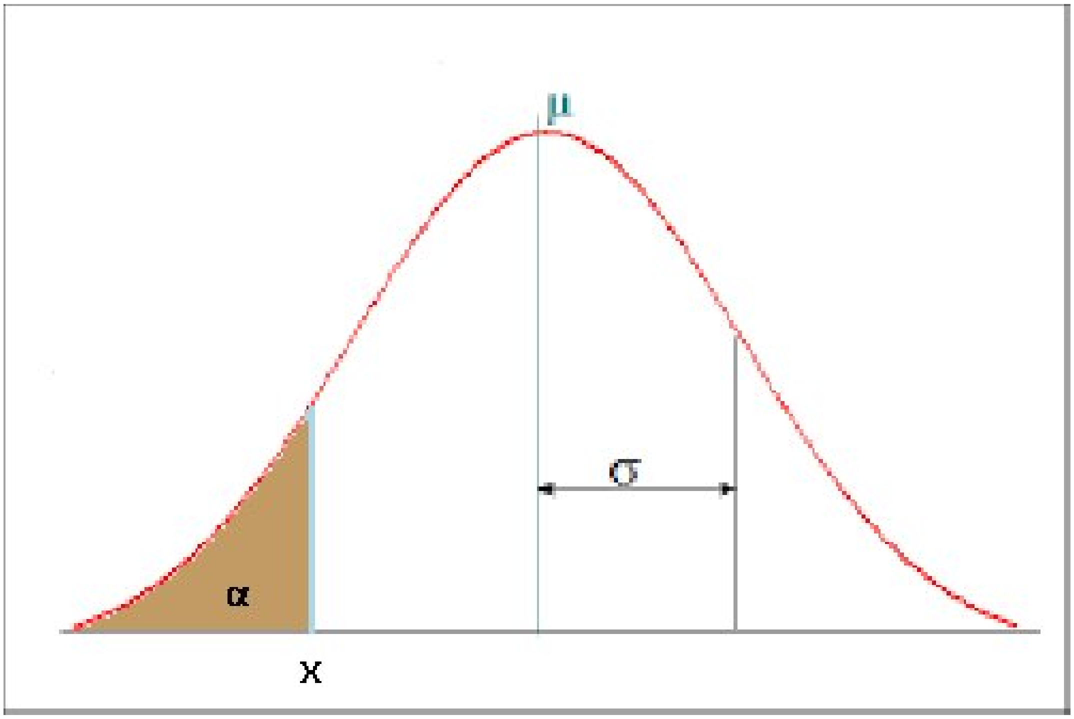
$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \quad \text{gdzie } f \text{ - funkcja gęstości}$$

Charakterystyki liczbowe zmiennej losowej:

Zmienna losowa skokowa (dyskretna)	Zmienna losowa ciągła
$EX = \sum_{i=1}^{n} x_i p_i$ gdzie $p_i = P(X = x_i)$	$EX = \int_{-\infty}^{\infty} x f(x) dx$ gdzie f - funkcja gęstości
$D^2X = E(X - EX)^2 =$ $= \sum_{i=1}^{n} (x_i - EX)^2 p_i$ $D^2X = EX^2 - (EX)^2$ gdzie $EX^2 = \sum_{i=1}^{n} x_i^2 p_i$	$D^2X = E(X - EX)^2 =$ $= \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx$ $D^2X = EX^2 - (EX)^2$ gdzie $EX^2 = \int_{-\infty}^{\infty} x^2 f(x) dx$
$DX = \sqrt{D^2X}$	$DX = \sqrt{D^2X}$
Mo - jest to wartość x_i o największym prawdopodobieństwie wystąpienia lub wartość najczęściej występująca w próbie (różna od wartości skrajnych)	Mo - jest to wartość x, dla której funkcja gęstości $f(x)$ osiąga wartość największą (osiąga maksimum)
Me - jest to liczba spełniającą warunki: $P(X \leqslant Me) \geqslant \frac{1}{2}$ i $P(X \geqslant Me) \geqslant \frac{1}{2}$	Me - jest to liczba x spełniająca równanie $F(x) = \frac{1}{2}$
kwantyl rzędu p, $p \in (0, 1)$ - jest to liczba x_p spełniającą warunki: $P(X \leqslant x_p) \geqslant p$ i $P(X \geqslant x_p) \geqslant 1 - p$	kwantyl rzędu p, $p \in (0, 1)$ - jest to liczba x spełniająca równanie $F(x) = p$

Podstawowe rozkłady zmiennej losowej ciągłej:

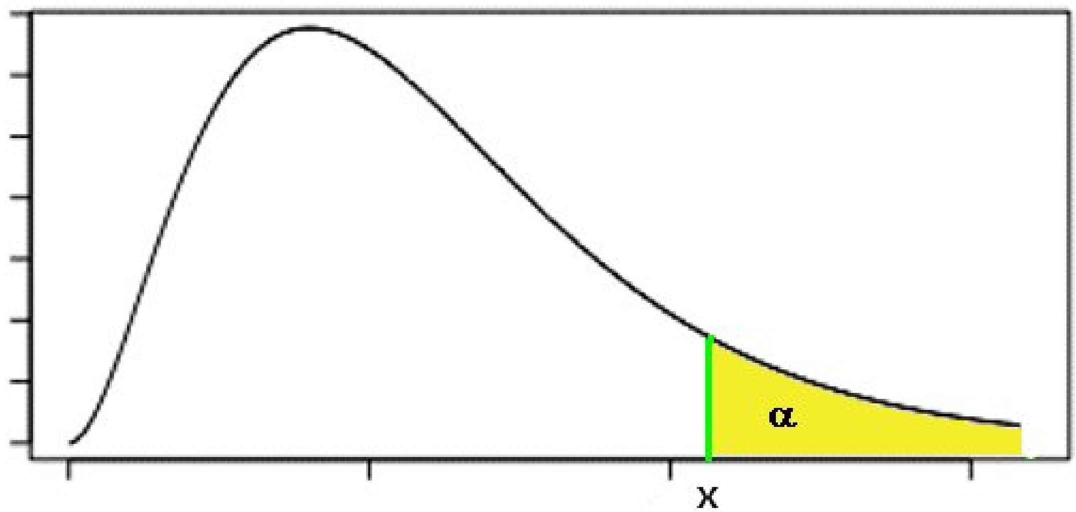
1. Rozkład normalny $N(\mu, \sigma)$



Wartość	Excel 2016, LibreOffice
$\alpha = P[N(\mu, \sigma) < x]$	ROZKŁ.NORMALNY(x;μ;σ;PRAWDA)
$x : P[N(\mu, \sigma) < x] = \alpha$	ROZKŁ.NORMALNY.ODWR(α;μ;σ)

Wartość	Excel, OpenOffice, LibreOffice
$\alpha = P[N(\mu, \sigma) < x]$	ROZKŁAD.NORMALNY(x;μ;σ;PRAWDA)
$x : P[N(\mu, \sigma) < x] = \alpha$	ROZKŁAD.NORMALNY.ODW(α;μ;σ)

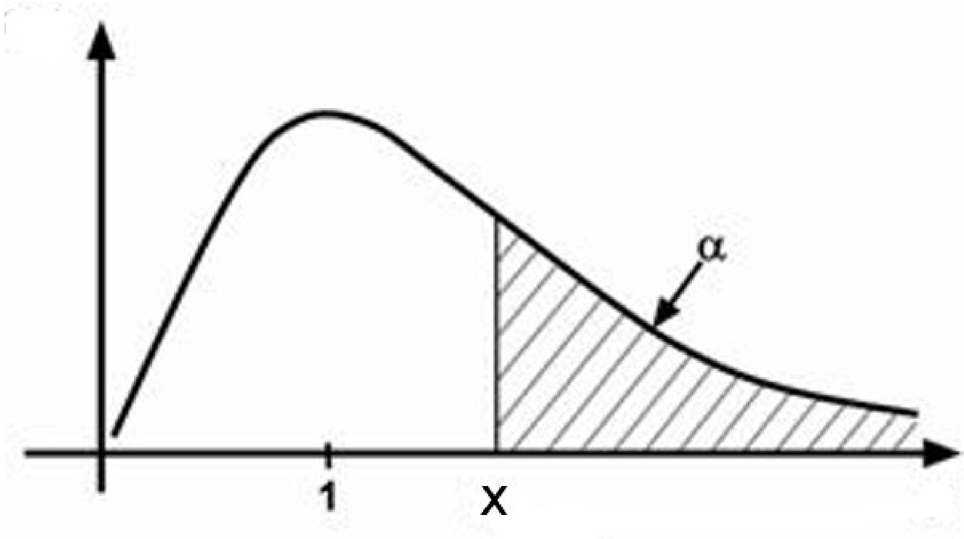
2. Rozkład chi-kwadrat $\chi^2(n)$



Wartość	Excel 2016, LibreOffice
$\alpha = P[\chi^2(n) < x]$	ROZKŁ.CHI(x;n;PRAWDA)
$x : P[\chi^2(n) < x] = \alpha$	ROZKŁ.CHI.ODWR(α;n)

Wartość	Excel, OpenOffice, LibreOffice
$\alpha = P[\chi^2(n) < x]$	1-ROZKŁAD.CHI(x;n)
$\alpha = P[\chi^2(n) > x]$	ROZKŁAD.CHI(x;n)
$x : P[\chi^2(n) < x] = \alpha$	ROZKŁAD.CHI.ODW($1 - \alpha$;n)
$x : P[\chi^2(n) > x] = \alpha$	ROZKŁAD.CHI.ODW(α;n)

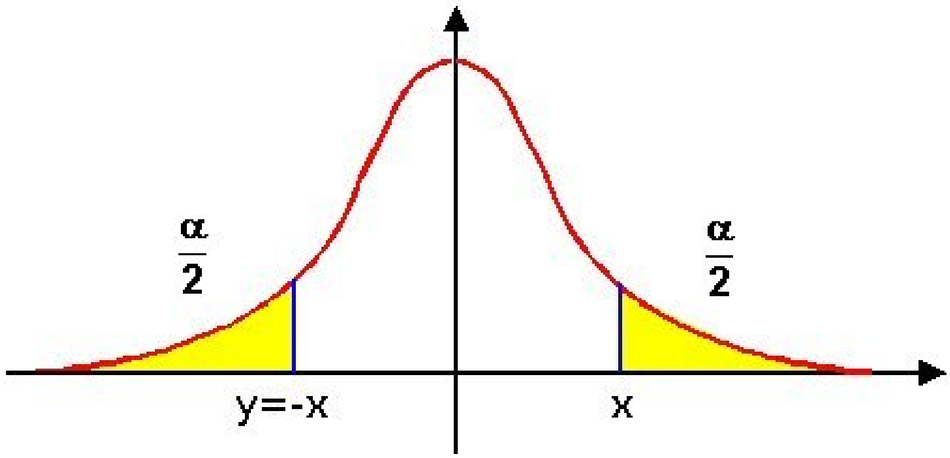
3. Rozkład F Fishera-Snedecora $F(m, n)$



Wartość	Excel 2016, LibreOffice
$\alpha = P[F(m,n) < x]$	ROZKŁ.F(x;m;n;PRAWDA)
$x : P[F(m,n) < x] = \alpha$	ROZKŁ.F.ODWR(α;m;n)

Wartość	Excel, OpenOffice, LibreOffice
$\alpha = P[F(m,n) < x]$	1-ROZKŁAD.F(x;m;n)
$\alpha = P[F(m,n) > x]$	ROZKŁAD.F(x;m;n)
$x : P[F(m,n) < x] = \alpha$	ROZKŁAD.F.ODW($1-\alpha$;m;n)
$x : P[F(m,n) > x] = \alpha$	ROZKŁAD.F.ODW(α;m;n)

4. Rozkład t Studenta $t(n)$



Wartość	Excel 2016, LibreOffice
$\alpha = P[t(n) < y]$	ROZKŁ.T(y;n;PRAWDA)
$y : P[t(n) < y] = \alpha$	ROZKŁ.T.ODWR(α;n)

Wartość	Excel, OpenOffice, LibreOffice
$\alpha = P[t(n) < x]$ dla $x \geqslant 0$	1-ROZKŁAD.T(x;n;1)
$\alpha = P[t(n) > x] = P[t(n) < -x]$ dla $x \geqslant 0$	ROZKŁAD.T(x;n;1)
$\alpha = P[t(n) > x] + P[t(n) < -x]$ dla $x \geqslant 0$	ROZKŁAD.T(x;n;2)
$x : P[t(n) > x] = \alpha$ dla $x \geqslant 0$	ROZKŁAD.T.ODW(2α;n)
$x : P[t(n) < x] = \alpha$ dla $x \geqslant 0$	ROZKŁAD.T.ODW(2-2α;n)
$y : P[t(n) < y] = \alpha$ dla $y < 0$	-ROZKŁAD.T.ODW(2α;n)
$(x > 0) : P[t(n) < -x] + P[t(n) > x] =$ $= 2P[t(n) > x] = 2P[t(n) < -x] = \alpha$	ROZKŁAD.T.ODW(α;n)
$(x > 0) : P[-x < t(n) < x] = \alpha$	ROZKŁAD.T.ODW(1-α;n)

2 Statystyka opisowa

Etapy tworzenia szeregu rozdzielczego przedziałowego:

1. wyznaczamy $x_{\max}$ oraz $x_{\min}$
2. wyznaczamy rozstęp z próby $\quad R = x_{\max} - x_{\min}$
3. wyznaczamy ilość przedziałów klasowych K (w zależności od liczebności próby n)
   - $K = \sqrt{n}$
   - $K = 1 + 3,322 \log n$
   - $K \leqslant 5 \log n$
   - tabela:

liczba pomiarów n	liczba przedziałów klasowych K
30-60	6-8
60-100	7-10
100-200	9-12
200-500	11-17
500-1500	16-25

4. wyznaczamy długość przedziału klasowego h
   minimalna długość przedziału klasowego $h_{\min} \approx \dfrac{R+\alpha}{K}$, gdzie α jest dokładnością pomiaru
5. wyznaczamy lewy koniec pierwszego przedziału klasowego

$$x_0 = x_{\min} - \frac{\alpha}{2}, \qquad \text{gdzie } \alpha \text{ jest dokładnością pomiaru}$$

Miary statystyczne można podzielić na: :

1. **miary położenia** - służą do określenia tej wartości cechy, wokół której skupiają się wszystkie pozostałe wartości; mierzą przeciętny poziom cechy

a) średnia z próby

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ - dane niepogrupowane}$$
$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i \text{ - szereg rozdzielczy punktowy}$$
$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i^* n_i \text{ - szereg rozdzielczy przedziałowy, gdzie}$$

x_i^* - środek przedziału klasowego

b) moda (dominanta) - wartość najczęstsza

- dane niepogrupowane i szereg rozdzielczy punktowy:
  moda to wartość najczęstsza, o ile nie jest to wartość skrajna
- szereg rozdzielczy przedziałowy:

$$Mo = x_m + \frac{(n_m - n_{m-1})\ h}{n_m - n_{m-1} + n_m - n_{m+1}}$$

gdzie x_m - lewy koniec przedziału z modą (czyli przedziału o największej liczebności, ale różnego od przedziału pierwszego i ostatniego), h - długość przedziału z modą, n_m - liczebność przedziału z modą, n_{m-1} - liczebność przedziału poprzedzającego przedział z modą, n_{m+1} - liczebność przedziału następującego po przedziale z modą

c) mediana - wartość środkowa w uporządkowanej próbie

- dane niepogrupowane i szereg rozdzielczy punktowy:

$$Me \ = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & , \quad \text{gdy } \ n \text{ jest nieparzyste} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & , \quad \text{gdy } \ n \text{ jest parzyste} \end{cases}$$

tzn. mediana jest to środkowa liczba, gdy n jest liczbą nieparzystą, albo średnia arytmetyczna dwóch środkowych liczb, gdy n jest liczbą parzystą

- szereg rozdzielczy przedziałowy:

$$Me = x_{Me} + \frac{h}{n_{Me}}\left(\frac{n}{2} - \sum_{i=1}^{k-1} n_i\right)$$

gdzie x_{Me} - lewy koniec przedziału z medianą, h - długość przedziału z medianą, n_{Me} - liczebność przedziału z medianą, k - numer przedziału zawierającego medianę; $\sum_{i=1}^{k-1} n_i$ - sumujemy liczebności przedziałów poprzedzających przedział z medianą

d) kwartyle (dolny Q_1 i górny Q_3) - wartości, które dzielą uporządkowaną próbę w stosunku 1:3 i 3:1

- szereg rozdzielczy przedziałowy:

$$Q_1 = x_{Q_1} + \frac{h}{n_{Q_1}}\left(\frac{n}{4} - \sum_{i=1}^{k-1} n_i\right)$$

$$Q_3 = x_{Q_3} + \frac{h}{n_{Q_3}}\left(\frac{3n}{4} - \sum_{i=1}^{k-1} n_i\right)$$

gdzie x_{Q_1} - lewy koniec przedziału zawierającego Q_1, h - długość przedziału zawierającego Q_1, n_{Q_1} - liczebność przedziału zawierającego Q_1, k - numer przedziału zawierającego Q_1

e) kwantyle rzędu p, $p \in (0,1)$ - wartości, które dzielą uporządkowaną próbę w stosunku $p : (1-p)$

f) inne średnie (geometryczna, harmoniczna)

2. **miary rozproszenia** (zmienności, rozrzutu) - służą do badania stopnia zróżnicowania jednostek populacji ze względu na wartość badanej cechy

a) rozstęp $\quad R = x_{\max} - x_{\min}$

b) wariancja

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \text{ - dane niepogrupowane}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 n_i \text{ - szereg rozdzielczy punktowy}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^* - \overline{x})^2 n_i \text{ - szereg rozdzielczy przedziałowy, gdzie}$$

x_i^* - środek przedziału klasowego

c) odchylenie standardowe $\quad s = \sqrt{s^2}$

typowy przedział zmienności $\quad (\overline{x} - s; \overline{x} + s)$

d) odchylenie przeciętne od średniej

$$d_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}| \text{ - dane niepogrupowane}$$
$$d_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}|n_i \text{ - szereg rozdzielczy punktowy}$$
$$d_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i^* - \overline{x}|n_i \text{ - szereg rozdzielczy przedziałowy, gdzie}$$
$$x_i^* \text{ - środek przedziału klasowego}$$

e) odchylenie przeciętne od mediany

$$d_2 = \frac{1}{n}\sum_{i=1}^{n}|x_i - Me| \text{ - dane niepogrupowane}$$
$$d_2 = \frac{1}{n}\sum_{i=1}^{n}|x_i - Me|n_i \text{ - szereg rozdzielczy punktowy}$$
$$d_2 = \frac{1}{n}\sum_{i=1}^{n}|x_i^* - Me|n_i \text{ - szereg rozdzielczy przedziałowy, gdzie}$$
$$x_i^* \text{ - środek przedziału klasowego}$$

f) odchylenie ćwiartkowe $Q = \frac{1}{2}(Q_3 - Q_1)$
rozstęp kwartylowy $Q = Q_3 - Q_1$

g) współczynnik zmienności $V = \frac{s}{\overline{x}} \cdot 100\%$

h) współczynnik nierównomierności $H = \frac{d_1}{\overline{x}} \cdot 100\%, \quad H_2 = \frac{d_2}{Mo} \cdot 100\%$

3. **miary asymetrii** - służą do badania kierunku zróżnicowania wartości badanej cechy; mierzą czy przeważająca liczba jednostek znajduje się powyżej czy poniżej przeciętnego poziomu badanej cechy

a) wskaźnik asymetrii $ws_1 = \overline{x} - Mo, \quad ws_2 = \overline{x} - Me$

b) współczynnik asymetrii

$$A = \frac{M_3}{s^3}, \quad \text{gdzie}$$

$$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3 \text{ - dane niepogrupowane}$$
$$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3 n_i \text{ - szereg rozdzielczy punktowy}$$
$$M_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i^* - \overline{x})^3 n_i \text{ - szereg rozdzielczy przedziałowy, gdzie}$$

x_i^* - środek przedziału klasowego

c) współczynnik skośności

$$A_1 = \frac{\overline{x} - Mo}{s}, \quad A_2 = \frac{\overline{x} - Mo}{d_1}, \quad A_3 = \frac{Q_1 + Q_3 - 2Me}{2Q}$$

4. **miary spłaszczenia** (koncentracji) - służą do analizy stopnia skupienia poszczególnych jednostek wokół średniej

a) kurtoza (współczynnik skupienia)

$$K = \frac{M_4}{s^4}, \quad \text{gdzie}$$

$$M_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4 \text{ - dane niepogrupowane}$$
$$M_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4 n_i \text{ - szereg rozdzielczy punktowy}$$
$$M_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i^* - \overline{x})^4 n_i \text{ - szereg rozdzielczy przedziałowy, gdzie}$$

x_i^* - środek przedziału klasowego

b) eksces $\quad q = K - 3$

Miary statystyczne można podzielić na:

1. **miary pozycyjne** - są niewrażliwe na obserwacje mocno odstające (moda, mediana, kwartyle, rozstęp kwartylowy)

2. pozostałe

3 Estymacja przedziałowa

1. PRZEDZIAŁY UFNOŚCI DLA ŚREDNIEJ

   (a) **MODEL I**

   Badana cecha ma rozkład normalny $N(\mu, \sigma)$ o **nieznanym parametrze** μ **i znanym** σ. Przedział ufności:

   $$\mu \in \left[\overline{x} - u\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\ ;\ \overline{x} + u\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\right]$$

   gdzie $u(1 - \frac{\alpha}{2})$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ rozkładu normalnego $N(0, 1)$; $\overline{x}$, n - to średnia i liczebność próby.

   (b) **MODEL II**

   Badana cecha ma rozkład normalny $N(\mu, \sigma)$ o **nieznanych parametrach** μ **i** σ. Przedział ufności:

   $$\mu \in \left[\overline{x} - t\left(1 - \frac{\alpha}{2}, n-1\right)\frac{s}{\sqrt{n-1}}\ ;\ \overline{x} + t\left(1 - \frac{\alpha}{2}, n-1\right)\frac{s}{\sqrt{n-1}}\right]$$

   gdzie $t(1 - \frac{\alpha}{2}, n - 1)$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ rozkładu Studenta o $n - 1$ stopniach swobody; $\overline{x}$, s, n - to średnia, odchylenie standardowe i liczebność próby.

   (c) **MODEL III**

   Badana cecha ma dowolny rozkład (niekoniecznie normalny), o nieznanej wartości oczekiwanej μ i nieznanym odchyleniu standardowym σ, zaś liczebność próby jest duża ($\mathbf{n \geqslant 100}$). Przedział ufności:

   $$\mu \in \left[\overline{x} - u\left(1 - \frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}\ ;\ \overline{x} + u\left(1 - \frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}\right]$$

   gdzie $u(1 - \frac{\alpha}{2})$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ rozkładu normalnego $N(0, 1)$; $\overline{x}$, s, n - to średnia, odchylenie standardowe i liczebność próby.

   Jeśli σ jest parametrem znanym, to zamiast s wstawiamy σ.

2. PRZEDZIAŁY UFNOŚCI DLA WARIANCJI

   (a) **MODEL I**

   Badana cecha ma rozkład normalny $N(\mu, \sigma)$ o nieznanych parametrach μ i σ, zaś próba jest mała ($\mathbf{n} \leqslant \mathbf{50}$). Przedział ufności:

$$\sigma^2 \in \left[\frac{ns^2}{\chi^2\left(1-\frac{\alpha}{2}, n-1\right)} \ ; \ \frac{ns^2}{\chi^2\left(\frac{\alpha}{2}, n-1\right)} \right]$$

   gdzie $\chi^2\left(1-\frac{\alpha}{2}, n-1\right)$ i $\chi^2\left(\frac{\alpha}{2}, n-1\right)$ są kwantylami rzędu $1-\frac{\alpha}{2}$ i $\frac{\alpha}{2}$ (odpowiednio) rozkładu chi-kwadrat o $n-1$ stopniach swobody; s^2, n - to wariancja i liczebność próby.

   (b) **MODEL II**

   Badana cecha ma rozkład normalny $N(\mu, \sigma)$ o nieznanych parametrach μ i σ, zaś próba jest duża ($\mathbf{n} > \mathbf{50}$). Przedział ufności:

$$\sigma^2 \in \left[\frac{2ns^2}{\left(\sqrt{2n-3}+u(1-\frac{\alpha}{2})\right)^2} \ ; \ \frac{2ns^2}{\left(\sqrt{2n-3}-u(1-\frac{\alpha}{2})\right)^2} \right]$$

   gdzie $u(1-\frac{\alpha}{2})$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ rozkładu normalnego $N(0,1)$; s^2, n - to wariancja i liczebność próby.

3. PRZEDZIAŁY UFNOŚCI DLA WSKAŹNIKA STRUKTURY

   Jeśli próba jest duża ($\mathbf{n} \geqslant \mathbf{100}$), to przedział ufności dla wskaźnika struktury p jest postaci:

$$p \in \left[q - u\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{q(1-q)}{n}} \ ; \ q + u\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{q(1-q)}{n}} \right]$$

   gdzie $q = \frac{m}{n}$; m jest liczbą elementów w próbie, które posiadają badaną cechę; $u(1-\frac{\alpha}{2})$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ rozkładu normalnego $N(0,1)$.