



# WYCENA.COM.PL

**Podstawowe pojęcia statystycznej  
analizy danych wykorzystywane  
w analizie rynku  
i wycenie nieruchomości  
(Kurs REx01)**

**Arkusze rex\_histogram  
Krótka Instrukcja Obsługi  
wersja: 1.0.4 z 20.05.2014 r.**

**Autor:**

*Tomasz Kotrański*

[kotrasinski@realexperts.pl](mailto:kotrasinski@realexperts.pl)

[www.realexperts.pl](http://www.realexperts.pl)

**Warszawa, 20.05.2014 r.**

## **Arkusz rex\_histogram**

**wersja: 1.0.4 z 20.05.2014 r.**

### **Możliwości arkusza**

Arkusz umożliwia:

- ✓ wizualizację rozkładu zmiennej (np. cen lub cech) przy pomocy histogramu;
- ✓ wizualizację podobieństwa rozkładu empirycznego badanej zmiennej do rozkładu normalnego;
- ✓ wizualizację położenia średniej arytmetycznej, mediany, dominanty i średniej geometrycznej;
- ✓ określanie parametrów statystycznych badanej zmiennej w zakresie: tendencji centralnej, zmienności i symetryczności/asymetrii rozkładu;

### **Ograniczenia ilościowe arkusza**

Maksymalna liczba przypadków możliwych do analizy: 1000.

### **Środowisko uruchomieniowe**

Arkusz przeznaczony jest zasadniczo dla Excela 2007/2010, ale pracuje też w Excelu 2003, choć z utratą zaawansowanych formatowań i estetyki wyglądu.

### **Makra**

Do prawidłowego działania arkusza w zakresie skalowania wykresu histogramu niezbędne jest włączenie makr.

---

### **Proszę pamiętać!**

Proszę pamiętać, że ten arkusz jest jedynie narzędziem analityczno-obliczeniowym.

Sam nie zbiera ani nie opisuje i nie przygotowuje danych mogących być podstawą efektywnej analizy rynku. Nie szuka, nie zastanawia się ani myśli za analityka.

Jednak osobom, które posiadają wartości do analizy dane rynkowe i chcą znaleźć w nich porządek, ten może, jak sądzę, oddać nieocenione usługi.



WYCENA.COM.PL

## Wprowadzanie danych

Dane przeznaczone do analizy należy przekopiować do arkusza „**dane**” do kolumny oznaczonej „Dane”. Najlepiej zrobić to przez funkcję „wklej specjalnie”->„wklej wartości”.

Lp.	Dane
1	1715,93
2	1837,67
3	1951,63
4	1971,39
5	2217,32
6	2230,32
7	2290,76
8	2312,51
9	2349,62
10	2354,75
11	2358,85
12	2365,88

Zalecane jest aby przed wprowadzeniem nowych danych do analizy zawsze wyczyścić istniejące dane za pomocą przycisku „**Czyść dane**”. Pozwoli to uniknąć kłopotów z możliwym niezauważeniem przy wprowadzaniu danych w dalszej części arkusza, których pozostawienie zakłóci analizę. Najczęściej może się to zdarzyć przy wprowadzeniu mniejszego zbioru danych niż już istniejący w arkuszu.

Czyść dane

Zaleca się zawsze, dla każdej analizy, robić kopię arkusza, na której się pracuje. Uniknie się wtedy sytuacji, w której zniszczy się przez nieuwagę arkusz.

Dla pracy arkusza nie jest konieczne uruchamianie żadnego makra uruchamiającego pobranie danych do analizy. Wyniki analizy są od razu widoczne w arkuszu „**histogram**”

## Analiza danych

W arkuszu **histogram**, znajdują się wyniki analizy w zakresie podstawowego opisu statystycznego danych w zakresie tendencji centralnej, zmienności i symetrii oraz wykres histogramu obrazującego rozkład wartości analizowanej.

## Miary tendencji centralnej (skupienia)

W tabeli pierwszej znajdują się parametry określające miary skupienia (tendencji centralnej) charakteryzujące analizowaną zmienną, określające średnią arytmetyczną wraz z przedziałem ufności dla określenia tej średniej, mediany (wartości środkowej), dominanty (wartości typowej) oraz średniej geometrycznej.

Lp.	Miary tendencji centralnej	Wartość
1	Średnia arytmetyczna ( $\mu$ )	2.947,60 zł
2	Poziom istotności dla średniej [ $\alpha$ ]	5%
3	Przedział ufności średniej [ $u(\alpha)$ ] $\pm$	102,38 zł
4	Przedział ufności średniej $\pm$ %	3,5%
5	Przedział średniej od... do...	2.845,22 zł 3.049,98 zł
6	Mediana (Me)	2.917,12 zł
7	Dominanta (D)	2.888,89 zł
8	Średnia geometryczna (G)	2.903,99 zł

**średnia arytmetyczna:** Średnia przydatna jest dla danych, dla których stwierdza się wpływ addytywny (dodający) oraz rozkład symetryczny. Jest bardzo wrażliwa na wpływ wartości odstających.

**poziom istotności:** maksymalne dopuszczalne prawdopodobieństwo popełnienia błędu I rodzaju (zazwyczaj oznaczane symbolem  $\alpha$  - alfa). Określa maksymalne ryzyko błędu, jakie jesteśmy skłonni zaakceptować przyjmując hipotezę pozytywną. W przypadku arkusza rex\_histogram hipotezę, że z prawdopodobieństwem określonym przez poziom ufności ( $1 - \alpha$ ) wartość parametru mieści się w obliczonym przedziale ufności.

**poziom ufności** – dopełnienie poziomu istotności, wynoszące  $1 - \alpha$ . Przy poziomie istotności 5%, poziom istotności wynosi  $1 - 0,05 = 0,95$ , czyli 95%.

**przedział ufności:** przedział, w którym z zadaniem prawdopodobieństwem (poziomem ufności) mieści się rzeczywista wartość badanego parametru. Zazwyczaj określana jako  $X \pm x$ . W przypadku arkusza rex\_histogram jest obliczany przedział ufności dla średniej arytmetycznej, pokazujący z jaką dokładnością możemy znać wartość średniej dla pewnej liczby danych (próba) cechujących się właściwą im wariancją (zmiennością).

**mediana:** typ średniej pozycyjnej obrazujący wartość środkową, powyżej i poniżej której znajduje się połowa liczebności danych. Mediana ma zastosowanie uniwersalne. Jest bardzo odporna na wpływ wartości odstających.

**dominanta:** (inaczej moda, modalna, wartość najczęstsza, typowa). Wartość, która występuje najczęściej lub najliczniejszy przedział wartości. Uwaga: dominanta dla wartości ciągłych (w przeciwieństwie do średnich i mediany) nie jest miarą stałą i zależy od wybranego podziału

**średnia geometryczna:** rodzaj średniej obliczany dla  $n$  danych jako pierwiastek stopnia  $n$  z iloczynu wartości tych danych. Przydatna jest dla danych, dla których stwierdza się wpływ multiplikatywny (mnożnikowy) oraz rozkład niesymetryczny (prawoskośny). Jest dość odporna na wpływ wartości odsta-



jących. Średnia geometryczna stanowi odwróconą średnią arytmetyczną logarytmów danej wartości. Uwaga: średnia geometryczna może być obliczona jedynie dla wartości ściśle pozytywnych. W wypadku wystąpienia w analizowanych danych jakichkolwiek wartości ujemnych średnia geometryczna nie zostaje obliczona (N/D).

## Miary zmienności

W drugiej tabeli znajdują się parametry dotyczące zmienności: odchylenie standardowe, współczynnik zmienności, rozstęp empiryczny i rozstęp estymowany określany przez odchylenia standardowe.

**Odchylenie standardowe:** bezwzględna miara zmienności określająca średnie (pierwiastkowe) odchylenie danych od wartości średniej arytmetycznej. Określana w jednostkach mierzonych, czyli np. w zł/m<sup>2</sup> (dla cen) lub w m<sup>2</sup> (dla powierzchni).

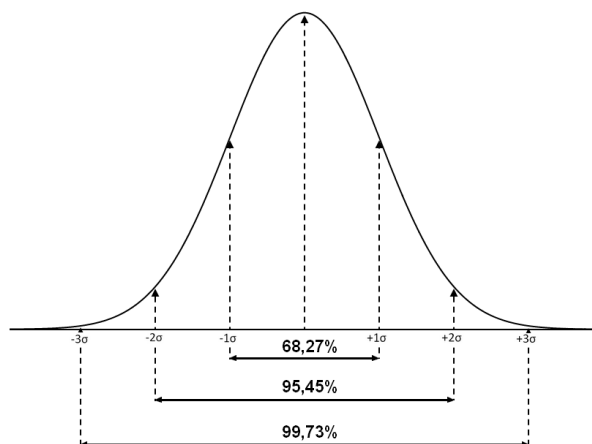
**Współczynnik zmienności:** względna miara zmienności.

Oblicza się ją przez podzielenie odchylenia standardowego przez średnią arytmetyczną. Określana jest w %. Dzięki temu możliwe jest porównywanie zróżnicowania różnych rynków. Dzięki temu możemy stwierdzić, że rynek cechujący się odchyleniem standardowym cen jednostkowych wynoszącym 50 zł/m<sup>2</sup> i średnią arytmetyczną wynoszącą 100 zł/m<sup>2</sup> jest rynkiem bardziej zróżnicowanym cenowo niż rynek cechujący się odchyleniem standardowym cen jednostkowych wynoszącym 500 zł/m<sup>2</sup> i średnią arytmetyczną wynoszącą 4000 zł/m<sup>2</sup>. W pierwszym przypadku współczynnik zmienności wynosi bowiem  $50/100 = 50\%$  a w drugim przypadku wynosi  $500/4000 = 12,5\%$ . Pierwszy z rynków jest  $50/12,5 = 4$ -razy bardziej zróżnicowany niż drugi.

Lp.	Miary zmienności	Wartość
8	Odchylenie standardowe ( $\sigma$ )	503,77 zł
9	Wsp. zmienności ( $\sigma/\mu$ )%	17,1%
10	Wartość minimalna	1.715,93 zł
11	Wartość maksymalna	4.181,93 zł
12	Rozstęp empiryczny	2.466,00 zł
13	Liczba odchyłeń standardowych $\pm$	3,0
14	Średnia - 3 * odchylenie standardowe	1.436,29 zł
15	Średnia + 3 * odchylenie standardowe	4.458,91 zł
16	Rozstęp $\pm 3$ * odchylenie standardowe	3.022,62 zł

W arkuszu znajduje się także określenie empirycznego rozstępu (różnicy między zanotowaną wartością maksymalną i minimalną) oraz rozstępu estymowanego dla rozkładu normalnego poprzez odchylenia standardowe.

Zgodnie z właściwościami rozkładu normalnego:



- ✓ w przedziale  $\pm 1 \cdot \sigma$  mieści się ok. 68,27% przypadków,
- ✓ w przedziale  $\pm 1,96 \cdot \sigma$  mieści się ok. 95,0% przypadków,
- ✓ w przedziale  $\pm 2 \cdot \sigma$  mieści się ok. 95,45% przypadków,
- ✓ w przedziale  $\pm 3 \cdot \sigma$  mieści się ok. 99,73% przypadków.

Uwzględniając powyższe, znając średnią i odchylenie standardowe, można określić rozstęp cenowy obejmujący np. 99,73% przypadków (czyli praktycznie wszystkie) za pomocą ustawienia  $\pm 3 \cdot \sigma$ . Domyślnie w arkuszu ustawiono właśnie taki rozstęp estymowany.

Nawet jeżeli rozkład nie jest normalny, to dla dowolnego rozkładu, dla którego możemy obliczyć średnią arytmetyczną ( $\mu$ ) i odchylenie standardowe ( $\sigma$ ), zgodnie z nierównością Czebyszewa zachodzi prawdziwość:

- ✓ w przedziale  $\pm 2 \cdot \sigma$  mieści się ok. 75% przypadków,
- ✓ w przedziale  $\pm 3 \cdot \sigma$  mieści się ok. 89% przypadków.



## Miary symetrii/asymetrii rozkładu

W trzeciej tabeli znajdują się parametry związane głównie z symetrycznością/asymetrią rozkładu analizowanej zmiennej, takie jak skośność, różnica między średnią arytmetyczną, medianą, dominantą i średnią geometryczną oraz współczynniki asymetrii dla mediany, dominanty i średniej geometrycznej.

Dzięki tym miarom możemy stwierdzić jak blisko/daleko rozkład zmiennej badanej odbiega od rozkładu symetrycznego i czy nie jest niezbędne np. przekształcenie zmiennej (zwłaszcza zmiennej objaśnianej, jak np. ceny) aby uzyskać rozkład bardziej symetryczny.

Należy pamiętać, że badanie symetryczności rozkładu jest głównym kryterium wyboru typu modelu. W wypadku silnie niesymetrycznego rozkładu cen (np. zbliżonego do log-normalnego) posłużenie się poprawkami symetrycznymi względem przyjętej miary tendencji centralnej (średniej arytmetycznej, mediany, dominanty lub średniej geometrycznej) musi z konieczności, ze względów czysto matematycznych, prowadzić do kłopotów z adekwatnością modelu.

Zastosowane w modelu poprawki, czyli charakter modelu musi kształtowi rozkładu zmiennej objaśnianej (cen itp.), inaczej wynik analizy może być błędny a model źle skonstruowany.

W tabeli widać, że rozkład badanej zmiennej cechuje się znaczną symetrycznością.

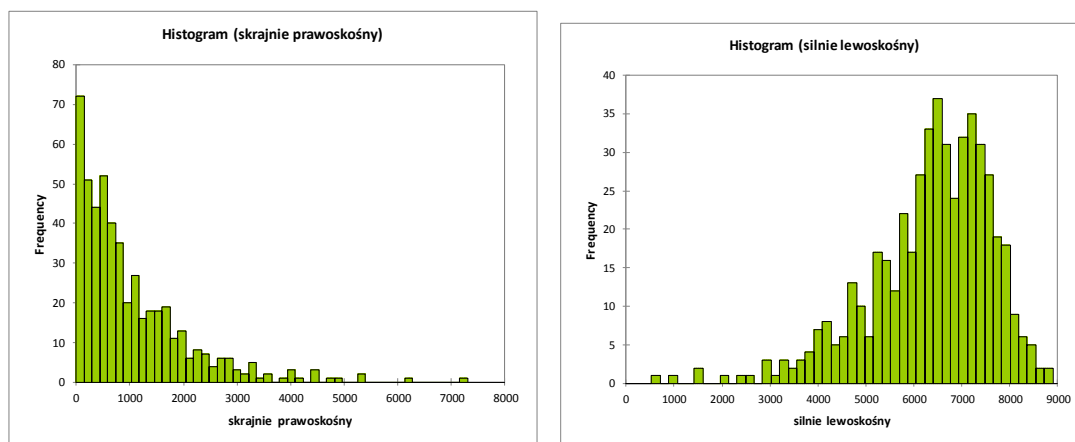
Skośność obliczana jest na podstawie tzw. trzeciego momentu centralnego. Dla rozkładów symetrycznych przyjmuje wartość 0, wartości ujemne dla rozkładów lewoskośnych i dodatnie dla rozkładów prawoskośnych.

L.p.	Miary symetrii/asymetrii	Wartość
17	Skośność	0,106
18	Średnia - mediana (w %)	1,0%
19	Średnia - dominanta (w %)	2,0%
20	Średnia - śr. geometryczna (w %)	1,5%
21	Wsp. asymetrii mediany $[(\mu-Me)/\sigma]$	0,061
22	Wsp. asymetrii dominanty $[(\mu-D)/\sigma]$	0,117
23	Wsp. asymetrii śr. geometrycznej $[(\mu-G)/\sigma]$	0,087
24	Wsp. asymetrii kwartyli	0,160

**Skośność:** miara asymetrii rozkładu. Istnieje wiele miar asymetrii, takie jak współczynnik skośności i miary asymetrii mediany i dominanty. Skośność (asymetrię) można określać na różnych poziomach dokładności, np. dla kwartyli, decyli i centyli.

Najpopularniejsze są: współczynnik skośności oraz współczynnik asymetrii mediany.

Rozkłady o długim prawym ramieniu nazywamy „prawoskośnymi” a rozkłady o długim lewym ramieniu nazywamy „lewoskośnymi”.



Uwzględniając fakt, że dla rozkładu idealnie symetrycznego wszystkie jego miary tendencji centralnej (średnia arytmetyczna, średnia geometryczna, mediana i dominanta) pokrywają się rozmięczenie się między średnią arytmetyczną a innymi miarami stają się miarami symetryczności/asymetrii rozkładu. W tabeli podane są względne (w % średniej arytmetycznej) różnice między średnią geometryczną, medianą i dominantą. W analizowanym przypadku widać, że różnice te są bardzo małe i wynoszą w granicach 1,0%-2,0% czyli są w ogólności niższe nawet niż 95% przedział ufności dla obliczonej średniej arytmetycznej.

Dodatkowo w tabeli obliczone są współczynniki asymetrii dla mediany, dominanty i średniej geometrycznej wyrażone w stosunku do odchylenia standardowego, co pozwala porównywać asymetrię różnych rozkładów niezależnie od ich wariancji.

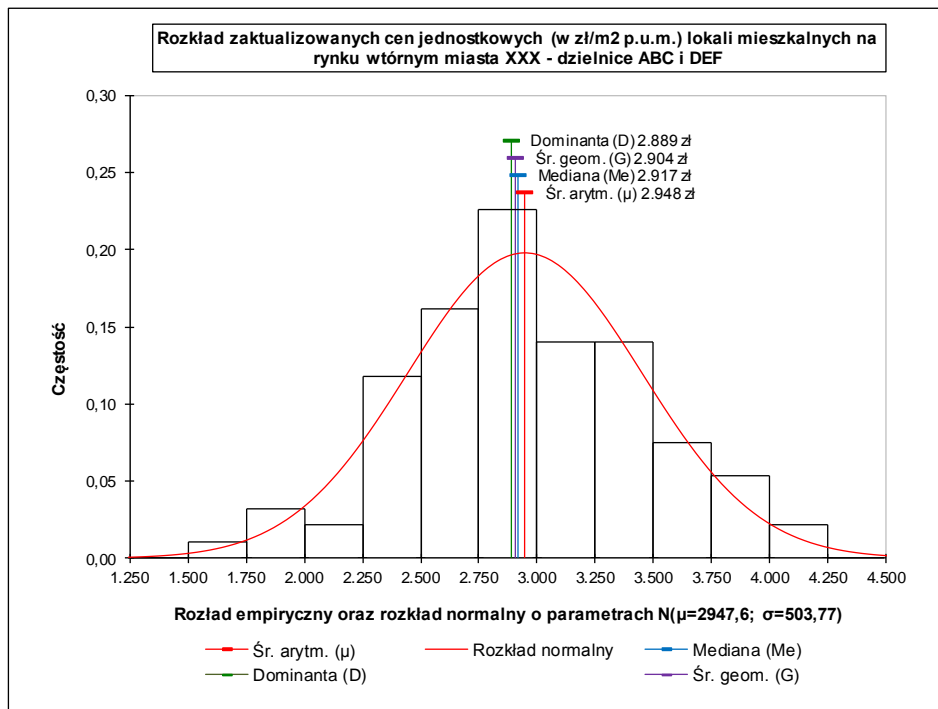


## Histogram rozkładu

Arkusz zawiera także histogram przedstawiający rozkład licznosci analizowanej zmiennej dla poszczególnych przedziałów klasowych. Na histogramie, oprócz słupków obrazujących licznosc poszczególnych przedziałów klasowych zaznaczono także linię rozkładu normalnego o parametrach takich samych (tj. średniej i odchylenia standardowego) jak parametry cechujące analizowaną zmienną.

**Histogram:** wykres obrazujący rozkład wartości danej cechy w postaci licznosci dla poszczególnych przedziałów. Np. Ceny w przedziałach o szerokości 100 zł/m<sup>2</sup> albo powierzchni w przedziałach co 250 m<sup>2</sup>.

W efekcie łatwo jest, poprzez porównanie rozkładu teoretycznego i empirycznego, zaobserwować, czy rozkład empiryczny jest zbliżony do normalnego, czy też bardzo mocno od niego odbiega (np. w przypadku rozkładu log-normalnego, cechującego rozleglejsze rynki przy dużych rozpiętościach cenowych), co ułatwia np. analizę dotyczącą zastosowania wartości zlogarytmowanych w analizie wartości (cechujących się rozkładem bardziej zbliżonym do rozkładu normalnego).



Dla skonstruowania histogramu niezbędne jest wprowadzenie parametrów początku histogramu oraz szerokości przedziałów. Szerokość przedziałów należy dobrać eksperymentalnie tak aby uzyskać najbardziej reprezentatywną demonstrację rozkładu. Szerokość przedziałów może być „zaokrąglona” np. do 100 zł/m<sup>2</sup>, 500 zł/m<sup>2</sup> itd. ale może także być niezaokrąglona i wynosić np. 258 zł/m<sup>2</sup>. Szerokości przedziałów nie mają wpływu na żadne obliczenia poza tymi, które związane są z wartością dominanty oraz z wyglądem histogramu.

Początek przedziałów	<b>1.500</b>
Szerokość przedziału	<b>250</b>

Powyżej histogramu znajduje się pole w którym można wpisać tytuł histogramu.

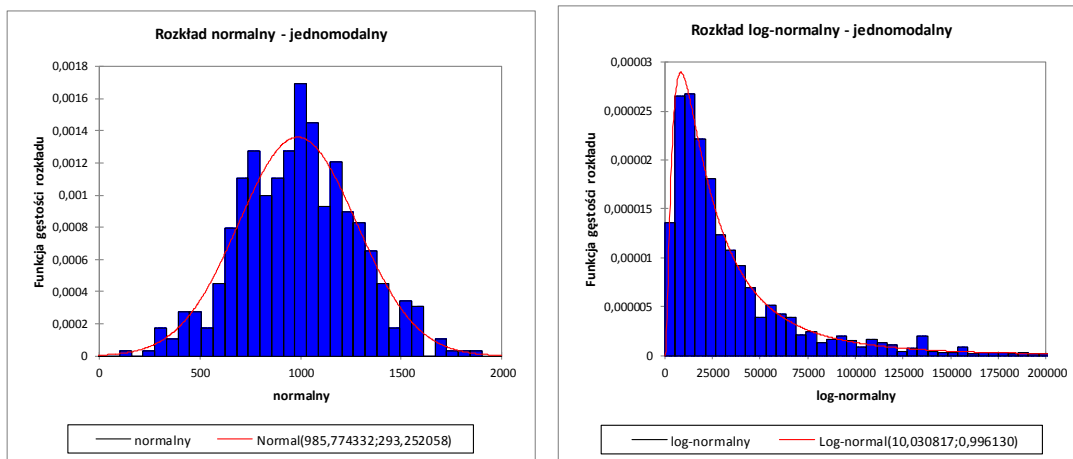
Po wprowadzeniu parametrów szerokości przedziałów histogram jest skalowany automatycznie za pomocą przycisku „Skaluj histogram”. W przypadku „trudnego” rozkładu danych na wykresie zawsze można wykres skalować ręcznie, bezpośrednio na osiach zarówno co do skali lub użytych jednostek.

Skaluj histogram

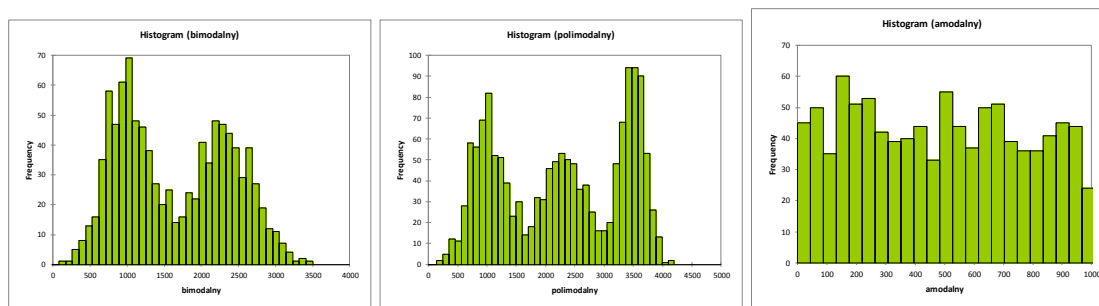
Analiza „optyczna” histogramu rozkładu analizowanej zmiennej pozwala także m.in. na przybliżone określenie charakterystyki rozkładu co do modalności, co może być pomocne w stwierdzeniu faktu, że np. dla cen jednostkowych (lub całkowitych) mamy do czynienia z kilkoma rynkami.



**Rozkład jednomodalny:** Rozkład zmiennej, który posiada jedną modę (dominantę), czyli wyraźną wartość dominującą w rozkładzie (niezależnie od tego czy rozkład jest symetryczny czy niesymetryczny).



**Rozkład wielomodalny:** Rozkład zmiennej, który posiada więcej niż jedną modę (dominantę) Bimodalny: dwie, polimodalny: wiele, amodalny: brak wyraźnej dominanty.



Tomasz Kotrański  
[kotrasinski@realexperts.pl](mailto:kotrasinski@realexperts.pl)  
[www.realexperts.pl](http://www.realexperts.pl)



## Postanowienia licencyjne

### Arkusz rex\_histogram

#### Wersja: 1.0.4

Wszystkie prawa do niniejszego arkusza posiada RealExperts.pl sp. z o.o.

[www.realexperts.pl](http://www.realexperts.pl)

Arkusz jest licencjonowany.

Licencja obejmuje wykorzystanie arkusza zarówno dla celów niekomercyjnych, jak i komercyjnych.

Licencja obejmuje prawo do aktualizacji arkusza do nowych wersji w okresie 6 miesięcy od daty zakupu niniejszej wersji oraz bezpłatnej pomocy technicznej przez okres 12 miesięcy.

Licencjobiorca ma prawo wykorzystywać arkusz w dowolnej liczbie kopii na dowolnej liczbie komputerów ale wyłącznie dla wykorzystania przez licencjobiorcę.

Licencjobiorca ma prawo dokonywania na własny użytek i dla własnej wygody zmian i dostosowań niniejszego arkusza w zakresie jaki uzna za stosowne.

Zabroniona jest sprzedaż lub przekazywanie arkusza lub jego kopii albo modyfikacji osobom trzecim zarówno dla celów komercyjnych, jak i niekomercyjnych.

W wypadku wykorzystania niniejszego arkusza wykorzystujący powinien umieścić informację o treści: „*wykorzystano arkusz rex\_histogram v. 1.0.4 RealExperts.pl sp. z o.o., autor: Tomasz Kotrański*” lub innej treściowo i funkcjonalnie tożsamej (szczegółowo uzależnione od kontekstu i stylistyki).

RealExperts.pl sp. z o.o. nie bierze żadnej odpowiedzialności za działanie arkusza ani jego wykorzystanie względem wykorzystującego niniejszy arkusz ani wobec osób trzecich.

RealExperts.pl sp. z o.o. udziela pomocy technicznej związanej z działaniem i wykorzystaniem arkusza jedynie licencjonowanym użytkownikom.

*Tomasz Kotrański*

[kotrasinski@realexperts.pl](mailto:kotrasinski@realexperts.pl)

[www.realexperts.pl](http://www.realexperts.pl)



WYCENA.COM.PL