

# Elementy statystyki

Tomasz Przechlewski

3/2019

## 1 Zadanie do wykonania

Na podstawie danych z szeregu szczegółowego: 1. Wyznaczyć szereg rozdzielczy (funkcja CZĘSTOŚĆ/FREQUENCY). 2. wykreślić histogram (na podstawie wartości szeregu rozdzielczego). 3. Scharakteryzować średni poziom cechy, zróżnicowanie i asymetrię (szczegóły niżej)

## 2 Szereg rozdzielczy/histogram

Aby wykreślić histogram należy dane pogrupować w szereg rozdzielczy. Do tego celu służy funkcja CZĘSTOŚĆ, której składnia jest następująca:

CZĘSTOŚĆ (obszar-danych ; obszar-końców-przedziałów)

W omawianym przykładzie obszar-danych, to kolumna zawierająca liczbę hoteli w powiecie.

Argument obszar-końców-przedziałów to obszar górnych końców przedziałów szeregu rozdzielczego. Funkcja CZĘSTOŚĆ zwraca obszar o jedną komórkę większy od obszaru górnych końców przedziałów – liczebność tej komórki, to liczba elementów większych od wartości ostatniego górnego końca przedziału.

Uwaga: funkcja CZĘSTOŚĆ jest specjalna (bo zwraca obszar): po jej wpisaniu do komórki należy nacisnąć Ctrl-Shift-Enter a nie zwyczajne Enter.

Kształt histogramu (a co za tym idzie wnioski dotyczące kształtu rozkładu) zależą w dużym stopniu od sposobu pogrupowania danych w szereg rozdzielczy.

## 3 Średni poziom cechy, zróżnicowanie i asymetria

Do obliczenia średniej, mediany, kwartyli, wariancji, odchylenia standardowego itp. służą odpowiednie funkcje (podano nazwy polskie/angielskie z programu OpenOffice calc):

SUMA (obszar-danych)	SUM(obszar-danych)
ŚREDNIA (obszar-danych)	AVERAGE(obszar-danych)
MEDIANA (obszar-danych)	MEDIAN(obszar-danych)
DOMINANTA (obszar-danych)	MODE(obszar-danych)
WARIANCJA (obszar-danych)	VAR(obszar-danych)
ODCHYL.STD(obszar-danych)	STDEV(obszar-danych)
SKOŚNOŚĆ (obszar-danych)	SKEW(obszar-danych)
KWARTYL(obszar-danych;typ)	QUARTILE(obszar-danych;typ)

jeżeli typ=0 KWARTYL oblicza wartość minimalną, jeżeli typ=4 maksymalną

## 4 Przykład

[0] Plik `hotele.csv` zawiera liczbę całorocznych obiektów hotelowych w Polsce w latach 2012 oraz 2017. Źródłem danych jest <https://bdl.stat.gov.pl/> (kategoria/grupa/podgrupa Turystyka/Turystyczne obiekty noclegowe i ich wykorzystanie/Turystyczne obiekty noclegowe wg rodzajów). Celem analizy jest porównanie struktury obu zbiorowości.

[1] Konstruujemy szereg rozdzielczy/histogram.

[2] Łącznie w Polsce w badanych okresach było odpowiednio 3918 oraz 3326 hoteli całorocznych. W roku 2012 średnio w powiecie było 8,9 hoteli całorocznych, podczas gdy w roku 2017 takich hoteli było 10,3. W ciągu 5 lat nastąpił zatem wzrost o 1,4 hoteli (albo o  $1,4/8,9=15,7\%$ )

[3] W roku 2012 w połowie powiatów było 5 hoteli i mniej, a połowie 5 hoteli i więcej.

[4] W roku 2012 w połowie powiatów było 5 hoteli i mniej, a połowie 5 hoteli i więcej. W roku 2017 w połowie powiatów było 7 hoteli i mniej, a połowie 7 hoteli i więcej.

[5] W roku 2012 w 25% powiatów było 3 hoteli i mniej, a 75% 3 hoteli i więcej. W roku 2017 w 25% powiatów było 4 hoteli i mniej, a 75% 4 hoteli i więcej.

[6] W roku 2012 w 75% powiatów było 10 hoteli i mniej, a 25% 10 hoteli i więcej. W roku 2017 w 75% powiatów było 11 hoteli i mniej, a 25% 11 hoteli i więcej.

[7] Jeżeli chodzi o zróżnicowanie liczby hoteli to przeciętne odchylenie od średniej arytmetycznej wyniosło 13,5 hoteli (2012) oraz 15,7 hoteli (2017). W wartościach bezwzględnych odnotowano zatem niewielki wzrost zmienności.

[8] Zmienność względna (mierzona jako udział odchylenia standardowego w średniej) spadła nieznacznie i wyniosła odpowiednio 151,8% oraz 154,2%. Analiza zmienności wskazuje, że przyrost liczby hoteli był równomierny.

[9] Wartość odchylenia ćwiartkowego była identyczna w obu badanych okresach i wyniosła 3,5 hoteli.

[10] Wartości klasycznego współczynnika asymetrii wyniosły odpowiednio 6,0 oraz 5,9 co wskazuje na znaczącą asymetrię prawostronną. Współczynniki Pearsona z kolei mają wartości znacznie niższe, bo wynoszą one 0,12 oraz 0,29 (odpowiednio). Szczegółowa analiza rozkład liczebności klas w szeregu rozdzielczym (na terenie 75% powiatów jest 0–6 hoteli) wskazuje że rozkład hoteli cechuje się znaczącą skośnością dodatnią (prawostronną) w obu analizowanych okresach. Niska wartość współczynnika Pearsona w omawianym przypadku wskazuje, że dla rozkładu hoteli w Polsce ten współczynnik nie jest dobrą miarą oceny asymetrii.

Objaśnienie:

[0] Populacja, cechy stałe. Cel analizy

[2] Średnia (suma jeżeli ma to sens)

[3] Mediana

[4] Dominanta (jeżeli ma sens – w przypadku cechy ciągłej może nie mieć)

[5,6] Kwartył1, Kwartył3

[7] Odchylenie standardowe (wariancji się nie interpretuje)

[8] Współczynnik zmienności

[9] Odchylenie ćwiartkowe

[10] Skośność (współczynnik klasyczny; funkcja SKOŚNOŚĆ/SKEW)