



WIESŁAWA MAŁSKA

Opisowe charakterystyki rozkładów cech statystycznych

Descriptive Characteristics of Distributions of Statistical Variables

Doktor inżynier, Politechnika Rzeszowska, Wydział Elektrotechniki i Informatyki, Katedra Energoelektroniki i Elektroenergetyki, Polska

Streszczenie

W dziale statystyki opisowej często wyznacza się pewne charakterystyki opisujące właściwości rozkładu badanej cechy zarówno ilościowej (mierzalnej) i jakościowej (niemierzalnej). Dla zmiennych (danych) ilościowych, jeżeli dane pochodzą z próby losowej (małej o liczebności $n \leq 30$ lub dużej o liczebności $n > 30$) charakterystyki te często nazywa się statystykami (miarami z próby losowej), w przypadku gdy dane dotyczą całej populacji generalnej wówczas charakterystyki te nazywamy parametrami. W dydaktyce przedmiotów związanych ze statystyką opisową i statystyką matematyczną ważna jest świadomość różnicy między poszczególnymi kategoriami miar. Przedstawienie własności rozkładu cech statystycznych za pomocą miar określa się także jako parametryczny opis rozkładu w próbie. Ocena rozkładu za pomocą parametrów lub statystyk (miar z próby) może być związana z położeniem cechy, jej zróżnicowaniem, symetrią (lub niesymetrią), a także koncentracją (skupieniem). W artykule zaprezentowano możliwość wykorzystania programu komputerowego STATISTICA do obliczeń i prezentacji opisowych charakterystyk rozkładów cech statystycznych. Wykorzystanie dedykowanych modułów do obliczeń i graficznej prezentacji rozkładów powinno być pomocne w procesie dydaktycznym.

Słowa kluczowe: miary położenia, miary zmienności, miary skośności, miary koncentracji, rozkłady zmiennych, statystyka opisowa

Abstract

In the section of descriptive statistics, some characteristics are often defined describing the distribution properties of the tested feature both quantitative (measurable) and qualitative (non-measurable). For quantitative variables (data), if the data comes from a random sample (small size $n \leq 30$ or large number $n > 30$), these characteristics are often referred to as statistics (random sample measures), when the data concern the whole general population then the characteristics these are called parameters. In the didactics of subjects related to descriptive statistics and mathematical statistics, it is important to be aware of the difference between particular categories of measures. The presentation of the distribution properties of statistical features by means of measures is also referred to as a parametric description of the distribution in the sample. Evaluation of the distribution using parameters or statistics (measurements from the sample) may be related to the location of the feature, its differentiation, symmetry (or asymmetry), as well as con-

centration (focus). The article presents the possibility of using the STATISTICA computer program for the calculation and presentation of descriptive characteristics of distributions of statistical features. The use of dedicated modules for calculations and graphical presentation of distributions should be helpful in the didactic process.

Keywords: location measures, measures of variability, measures of skewness, measures of concentration, distributions of variables, descriptive statistics

Wstęp

Metody mające na celu przedstawienie charakterystyk danej cechy nazywane są metodami opisu statystycznego. Opis statystyczny sprowadza się do wyznaczania pewnych liczbowych parametrów, określających badany zbiór danych (Rabiej, 2012; Józwiak, 2009). Opis statystyczny jest badaniem zamkniętym, jeżeli analizujemy dane dla populacji generalnej (dla wszystkich wartości) i wówczas otrzymane wyniki są parametrami. W przypadku, gdy zbiór danych stanowi próbę losową, opis statystyczny może być podstawą do wnioskowania w działach statystyki matematycznej (Luszniewicz, Słaby, 2001). Pełna analiza danych statystycznych powinna doprowadzić do zwięzłego zaprezentowania wyników badań za pomocą odpowiednich charakterystyk liczbowych.

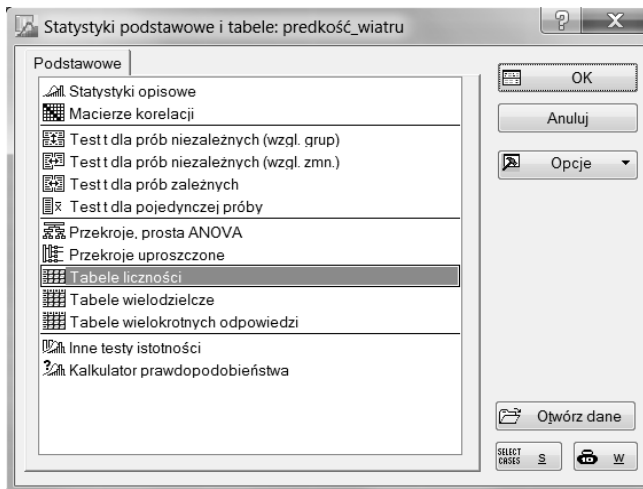
Syntetyczna charakterystyka danej zbiorowości nie oznacza określenia tylko jednego parametru. Im mniej parametrów wykorzystujemy do opisu statystycznego, tym więcej tracimy informacji o strukturze zbiorowości. Często znajomość rozkładu cechy statystycznej (jej obraz graficzny) pozwala na dobór odpowiednich miar.

Opisowe charakterystyki rozkładów

Pierwszą kategorię stanowią miary położenia, które służą do określenia tej wartości zmiennej opisanej przez rozkład, wokół której skupiają się pozostałe wartości zmiennej (Ostasiewicz, Rusnak, Siedlecka, 2011; Sobczyk, 2007). Wśród miar położenia najczęściej rozpatrywane są tzw. miary tendencji centralnej. Jest to wartość średnia, wartość modalna i mediana (wartość środkowa). Miary rozproszenia, często zwane miarami zmienności, służą do badania stopnia zróżnicowania wartości zmiennej. Najczęściej wyznacza się wariancję, odchylenie standardowe, odchylenie przeciętne, klasyczne współczynniki zmienności, a także rozstęp. Miary niesymetrii (skośności) służą do badania kierunku zróżnicowania wartości zmiennej. Ostatnią grupę stanowią miary koncentracji (skupienia) służące do analizy stopnia skupienia wartości zmiennej wokół średniej. Współczynnik skośności służy do badania symetrii rozkładu cechy statystycznej. Jeżeli współczynnik ten jest równy zero, to rozkład jest symetryczny. Przy ujemnej wartości współczynnika mówimy, że skośność jest lewostronna, a przy wartości dodatniej mówimy, że skośność jest prawostronna. Analogicznie interpretujemy współczynnik niesymetrii (asymetrii). W ostatnich latach najczęściej

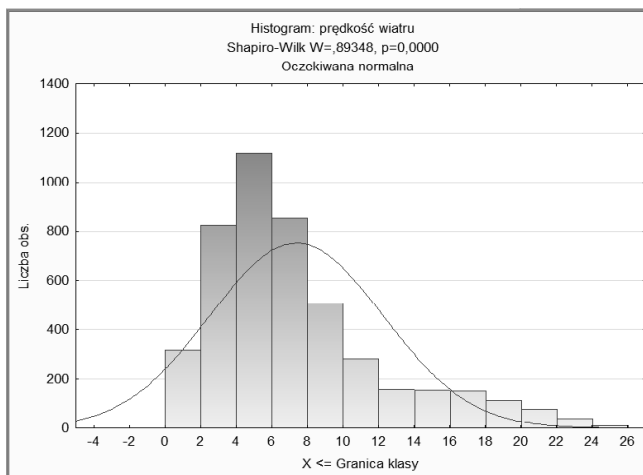
wyznaczaną miarą koncentracji jest kurtoza i eksces. Kurtoza jest miarą tzw. smukłości rozkładu.

Dane wykorzystane w przykładzie obliczeniowym dotyczą prędkości wiatru. Prędkość wiatru na danym terenie jest istotnym parametrem, od którego zależy produkcja energii elektrycznej w farmie wiatrowej. W programie STATISTICA można też sprawdzić rozkład zmiennej, wykorzystując przykładowo moduł *Statystyka/Statystyki podstawowe i tabele/Tabele licznosci* (rys. 1).



Rysunek 1. Widok modułu *Statystyka/Statystyki podstawowe i tabele/Tabele licznosci*

Źródło: opracowanie własne w programie STATISTICA.

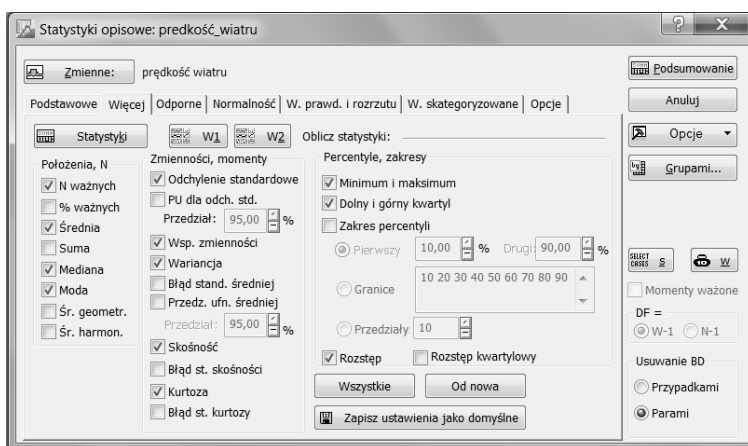


Rysunek 2. Histogram analizowanej zmiennej – prędkość wiatru

Źródło: opracowanie własne w programie STATISTICA.

Korzystając z modułu *Tabele licznosci* i wykresu w postaci histogramu można w przejrzysty sposób przedstawić rozkład empiryczny rozważanej cechy, w tym przypadku prędkości wiatru. Przedstawiony na rys. 2 histogram prędkości wiatru świadczy o niesymetrycznym rozkładzie, z niesymetrią prawostronną.

Dodatkowo na histogramie naniesiony jest wykres funkcji gęstości rozkładu normalnego (tzw. oczekiwana normalna) wraz z wynikiem testu normalności Shapiro-Wilka. Wartość p (p – prawdopodobieństwo testowe w programie STATISTICA) jest mniejsza od współczynnika istotności ($\alpha = 0,05$), stąd możemy uznać, że rozkład prędkości wiatru nie jest zgodny z rozkładem normalnym (z rozkładem Gaussa). Do pełnego opisu wykorzystuje się miary, które można obliczyć, korzystając z modułu *Statystyka/Statystyki podstawowe i tabele/Statystyki opisowe* i kartę *Więcej* (rys. 3). Korzystając z tego modułu, można wyznaczyć potrzebne miary do pełnego opisu statystycznego.



Rysunek 3. Widok okna modułu *Statystyka/Statystyki podstawowe i tabele/Statystyki opisowe* – karta *Więcej*

Źródło: opracowanie własne w programie STATISTICA.

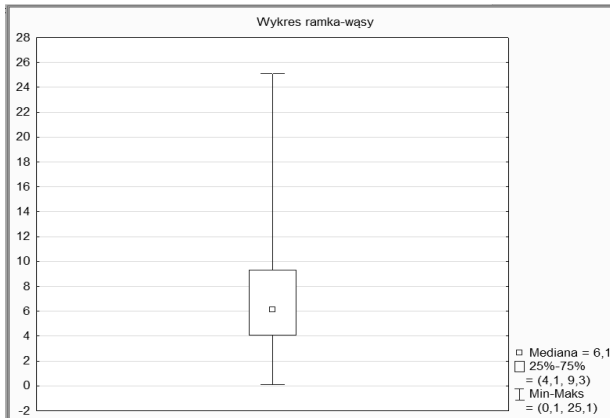
Zmienna	Statystyki opisowe (predkośc_wiatru)									
	Nważnych	Srednia	Mediana	Moda	Liczność Mody	Minimum	Maksimum	Odch.std	Skośność	Kurtoza
prędkość wiatru	4601	7,389459	6,100000	5,100000	74	0,100000	25,10000	4,879666	1,232913	1,135702

Rysunek 4. Wyniki miar dla zmiennej prędkość wiatru – fragment arkusza

Źródło: opracowanie własne w programie STATISTICA.

Na rys. 4 zaprezentowano fragment arkusza wynikowego dla analizowanej zmiennej prędkość wiatru. Otrzymane wartości tzw. miar tendencji centralnej, tj. średnia arytmetyczna, moda i mediana dla zmiennej prędkość wiatru, mają różne wartości, a to świadczyć może o niesymetrycznym rozkładzie. Do prezentacji

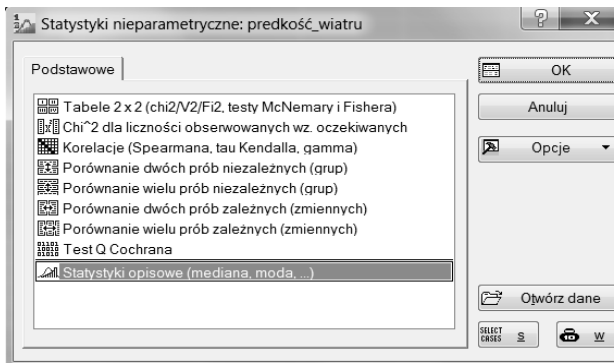
pozycyjnych miar położenia tzw. kwartyli (z podziałem danej zbiorowości na 4 równe części) w statystyce dedykowany jest odpowiedni typ wykresu, tzw. wykres ramka-wąsy. Dla rozważanej zmiennej wykres ramka-wąsy przedstawiono na rys. 5. Wykres ten oprócz prezentacji poszczególnych kwartyli świadczy również o niesymetrycznym rozkładzie analizowanej zmiennej. W tym przypadku świadczy o niesymetrii prawostronnej. Wartość współczynnika skośności ma wartość dodatnią, co również wskazuje na niesymetrię prawostronną (wartość ujemna może świadczyć o niesymetrii lewostronnej).



Rysunek 5. Wykres ramka-wąsy dla zmiennej prędkość wiatru

Źródło: opracowanie własne w programie STATISTICA.

W przypadku konieczności obliczenia wartości odchylenia przeciętnego należy wykorzystać moduł dostępny w *Statystyka/Statystyki nieparametryczne/Statystyki opisowe* (rys. 6).



Rysunek 6. Widok okna modułu Statystyka/Statystyki nieparametryczne/Statystyki opisowe

Źródło: opracowanie własne w programie STATISTICA.

W module przedstawionym na rys. 6 można obliczyć wszystkie miary, które mogą służyć do pełnego opisu statystycznego struktury danej populacji generalnej (w tym module nie ma możliwości wyboru poszczególnych interesujących nas miar, obliczane są wszystkie miary). Zaprezentowane wybrane moduły dostępne w programie STATISTICA nie są jedynymi modułami, które można wykorzystać do wyznaczania charakterystyk i pełnego opisu statystycznego. Można w tym celu wykorzystać także inne moduły dedykowane do tej tematyki (np. moduł analizy przekrojowej – *Tabele wielodzielcze*). Aby we właściwy sposób dokonać interpretacji, potrzebna jest także wiedza teoretyczna dotycząca zagadnień związanych z opisem i charakterystykami cech rozkładów empirycznych analizowanych cech statystycznych.

Podsumowanie

Wykorzystując wspomaganie komputerowe w procesie dydaktycznym oraz w nauczaniu przedmiotów związanych ze statystyką opisową, z prezentacją danych itp., należy w opisie statystycznym pamiętać, czy dane dotyczą całej zbiorowości statystycznej czy dotyczą tylko wyników badań częściowych dla małej lub dla dużej próby losowej. Dedykowane wykresy i moduły w programie STATISTICA do wyznaczania charakterystyk rozkładów cech mogą być pomocne w procesie dydaktycznym. Interpretacja otrzymanych wyników analiz i obliczeń rozważanych cech statystycznych powinna przyczynić się do właściwej prezentacji rozkładów i właściwej charakterystyki rozkładów.

Literatura

[http:// www.statsoft.pl](http://www.statsoft.pl) (4.05.2018).

Józwiak, J., Podgórski, J. (2009). *Statystyka od podstaw*. Warszawa: PWE.

Luszniewicz, A., Słaby, T. (2001). *Statystyka z pakietem komputerowym STATISTICA*. Warszawa: C.H. Beck.

Ostasiewicz, S., Rusnak, Z., Siedlecka, U. (2011). *Statystyka. Elementy teorii i zadania*. Wrocław: Wyd. AE we Wrocławiu.

Rabiej, M. (2012). *Statystyka z programem STATISTICA*. Gliwice: Helion.

Sobczyk, M. (2007). *Statystyka*. Warszawa: PWN.