

## Analiza statystyczna przy użyciu pakietu STATISTICA

1. Zaprojektować badanie statystyczne: określić cel badania, zdefiniować populację, co do której będą wyciągane wnioski, dokonać wyboru cech statystycznych (co najmniej dwie), określić źródło pozyskiwania danych (Internet, aktualna prasa, publikacje, dane z laboratorium, inne źródła), podać rozmiar próby (co najmniej 30) i sposób próbkowania (losowe, systematyczne, metodą najłatwiejszego pozyskiwania, czy warstwowe);
2. Scharakteryzować badane cechy (mieralne ciągłe, quasi-ciągłe lub dyskretne);
3. Zbudować szeregi rozdzielcze:
  - a) dla cechy ciągłej lub quasi-ciągłej dobierając odpowiednią ilość klas ( $k \approx \sqrt{n}$ ):  
[Statystyka] – [Statystyki podstawowe i tabele] – [Tabele licznosci] – [Więcej] – [Dokładna liczba przedziałów lub [Przybliżona liczba przedziałów];
  - b) dla cechy dyskretnej:  
[Statystyka] – [Statystyki podstawowe i tabele] – [Tabele licznosci] – [Więcej] – [Wszystkie różne wartości];
4. Wykreślić
  - a) histogramy częstości (oś Y %):  
[Wykresy] – [Histogramy] – [Więcej] – [Standardowe] – [Kategorie] (dla cechy ciągłej wybierając taką samą liczbę kategorii jak w poprzednim punkcie) lub [Tryb całkowity] (dla cechy dyskretnej);
  - b) histogramy częstości skumulowanej (oś Y %):  
zmieniając [Standardowe] na [Skumulowany] ;
5. Skomentować uzyskane wykresy, co do symetrii i kształtu. Wysunąć hipotezy, dotyczące zgodności z rozkładem Poissona (cechy dyskretne) lub normalnym (cechy ciągłe i quasi-ciągłe);
6. Obliczyć i skomentować parametry statystyczne: średnia, mediana, moda, odchylenie standardowe, wariancja, skośność, kurioza, eksces, dolny i górny kwartył, rozstęp, rozstęp kwartyłowy:  
[Statystyka] – [Statystyki podstawowe i tabele] – [Statystyki opisowe] – [Więcej]
7. Zbudować i opisać wykresy „ramka-wąsy”:  
[Statystyka] – [Statystyki podstawowe i tabele] – [Statystyki opisowe] – [Podstawowe] – [Wykres ramka wąsy];

8. Zweryfikować hipotezę, że badane cechy mają rozkład normalny (ewentualnie Poissona, gdy cecha dyskretna) używając testów  $\chi^2$  i Kołmogorowa-Smirnowa.

[Statystyka] – [Dopasowanie rozkładu] – [Rozkłady ciągłe] (lub [Rozkłady dyskretny]) – [Opcje] – [Test Kołmogorowa-Smirnowa] – [Tak], [Test chi-kwadrat] – bez połączonych kategorii;

Dla cech ciągłych lub quasi-ciągłych dodatkowo przeprowadzić badanie normalności używając testów Shapiro-Wilka i Lillieforsa. Wyniki testów porównać oraz podjąć decyzję o odrzuceniu ( $p < 0,05$ ) lub nie odrzuceniu ( $p > 0,05$ ) hipotezy zerowej.

[Statystyka] – [Statystyki podstawowe i tabele] – [Statystyki opisowe] – [Normalność] – [Testy normalności K-S i Lillieforsa] oraz [Test W Shapiro-Wilka];

9. Wzmocnić podjętą decyzję wykreślając i komentując wykresy normalności:  
[Statystyka] – [Statystyki podstawowe i tabele] – [Statystyki opisowe] – [W. prawdopodob. i rozrzutu] – [Wykres normalności];

10. Zbadać zależność pomiędzy cechami:

Zweryfikować hipotezę o istotności współczynnika korelacji, a w przypadku istotności tego współczynnika ( $p < 0,05$ ) zdecydować, która cecha jest zmienną objaśnianą (zależną), a która objaśniającą (niezależną) i na podstawie tej decyzji skonstruować odpowiednią prostą regresji. Wyjaśnić znaczenie współczynników  $a$  i  $b$  w tym równaniu, a także wielkość współczynnika korelacji  $r$  oraz współczynnika determinacji  $R^2$ :

[Wykresy] – [Wykresy rozrzutu] – [Więcej] – [R kwadrat], [Wsp. korel. i  $p$ ], [Równanie regresji];

11. Skonstruować histogramy trójwymiarowe i wytłumaczyć położenie słupków na podstawie wniosków z punktu 10:

[Wykresy] – [Wykresy sekwencyjne 3W] – [Histogramy dwóch zmiennych] ([Tryb całkowity] dla cechy dyskretny, [Kategorie] dla cechy ciągłej lub quasi-ciągłej).