

ĆWICZENIE 2

POPULACJE I PRÓBY DANYCH

Cel

Wprowadzenie podstawowych pojęć statystycznych, wskazanie różnic pomiędzy próbą danych a populacją.

Wprowadzenie teoretyczne

Celem badań statystycznych jest opis zjawisk zachodzących w zbiorowisku zwanym **populacją generalną** lub krócej – **populacją**. W skład populacji wchodzi elementy (osobniki) jednego rodzaju, np. populację zdefiniowaną „każdy element jest zwierzęciem” będą stanowiły wszystkie zwierzęta. W celu uzyskania wniosków dotyczących badanej populacji, należy zmierzyć jej cechy będące przedmiotem zainteresowania. Jest to czynność niezwykle pracochłonna, a często wręcz niemożliwa do wykonania. W związku z tym, wnioskowanie przeprowadza się na podstawie części populacji, czyli na próbie.

Populacją próbną (próbą) nazywamy skończony zbiór elementów pobranych z populacji generalnej. Pobrana próba powinna być **reprezentatywna**, czyli powinna w jak najlepszy sposób oddawać strukturę badanej populacji. Najprostszym typem próby reprezentatywnej jest **próba losowa prosta**. Otrzymuje się ją wtedy, gdy każdy element populacji ma taką samą szansę trafienia do próby, a wybór elementów jest losowy. Drugim, często stosowanym typem próby jest **próba pobierana systematycznie**. W tym przypadku elementy nie są wybierane losowo, lecz z góry określony sposób, np. co dziesiąty element.

Każda populacja może być charakteryzowana odpowiednimi parametrami, takimi jak **wartość oczekiwana** i **wariancja**. Wartości parametrów charakteryzujących populację z reguły nie są znane (z powodu wielkości populacji i niemożności zmierzenia wszystkich jej elementów). Na podstawie próby wylosowanej z badanej populacji można oszacować wartości parametrów, czyli wyznaczyć ich **estymatory**. Estymatorem wartości oczekiwanej badanej cechy jest **średnia arytmetyczna** elementów próby, estymatorem wariancji jest tzw. **wariancja próbkowa**. Estymatory te dane są następującymi wzorami:

- średnia arytmetyczna $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, gdzie n oznacza liczbę elementów próby, x_i są elementami próby,
- wariancja próbkowa $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Zadania do wykonania

1. Podać trzy przykłady populacji oraz prób z tych populacji pobranych.
2. Wykonać następujące czynności:
 - 2.1. wczytać do Excela pliki *średnia0_10.txt*, *średnia0_100.txt*, *średnia0_1000.txt*,
 - 2.2. wyznaczyć liczbę elementów prób,
 - 2.3. wyznaczyć estymatory wartości oczekiwanej i wariancji dla każdej z prób,
 - 2.4. porównać otrzymane wartości estymatorów z rzeczywistymi wartościami parametrów,
 - 2.5. obliczyć medianę dla każdej próby i zinterpretować jej wartość,
 - 2.6. wyznaczyć maksymalną i minimalną wartość każdej próby,
 - 2.7. czynności powtórzyć dla plików *średnia5_10.txt*, *średnia5_100.txt*, *średnia5_1000.txt*.
3. Odpowiedzieć na następujące pytania.
 - 3.1. Jak rozmiar próby wpływa na jakość estymatora?
 - 3.2. Jak wartość wariancji wpływa na wartości elementów próby?

Źródła:

- Łomnicki A. „Wprowadzenie do statystyki dla przyrodników”, PWN, Warszawa 2007
- Żuk B. „Biometria stosowana”, PWN, Warszawa 1989