Iwona Mróz,

Instytut Fizyki Doświadczalnej,

Uniwersytet Wrocławski

**„Podstawy statystyki i analizy danych” – materiały do wykładu nr 1 z dnia 8.10.2021 r.**

**Zalecana literatura:**

1. Hogg, R.V., Tanis, E.A., *Probability and statistical inference,*  Upper Saddle River, New Jersey 07458, 18th edition – bardzo dobry podręcznik przedstawiający statystykę na poziomie podstawowym i średnim, gorąco polecam.
2. Freund, J.E., Perles B.M., *Modern elementary statistics*, Upper Saddle River, New Jersey 07458, 12th edition – podręcznik zawiera wiele ciekawych przykładów i wyczerpujące wyjaśnienia zagadnień podstawowych.
3. Aczel, A.D., *Statystyka w zarządzaniu*, Warszawa 2000 – podręcznik jest adresowany do kadry zarządzającej, która charakteryzuje się ponadprzeciętną inteligencją, ale nie zawsze posiada przygotowanie matematyczne. Niewątpliwą zaletą podręcznika jest ciekawy wykład i bardzo interesujące przykłady. Wadą jest brak opisu matematycznego właściwego dla studentów nauk ścisłych.
4. Ostasiewicz, S., Rusnak, Z., Siedlecka, U., *Statystyka. Elementy teorii i zadania*, Wrocław 1999 – książka przyda się na początku kursu. Jest to podręcznik na poziomie podstawowym, napisany w sposób bardzo rzetelny. Zawiera wstępy teoretyczne i bardzo dużo ciekawych przykładów, rozwiązanych i do samodzielnego rozwiązania. Będziemy z niego korzystać także jak ze zbioru zadań. Podobnie jak w przypadku podręcznika Aczela, wadą jest brak pogłębionego opisu matematycznego omawianych zagadnień.
5. Plucińska, A., Pluciński.E, *Zadania z rachunku prawdopodobieństwa i statystyki matematycznej dla studentów politechnik*, Państwowe Wydawnictwo Naukowe, Warszawa 1976 – wartościowy zbiór zadań.
6. Bobrowski, D., *Probabilistyka w zastosowaniach technicznych*, Wydawnictwa Naukowo-techniczne, Warszawa 1986 – książka zawiera krótkie wprowadzenia teoretyczne i wiele rozwiązanych przykładów. Będziemy z niej korzystać jak ze zbioru zadań.
7. Stanisz, A., *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, t. 1-3*, Kraków 2006 – książka w bardzo przystępny sposób opisuje jak wykorzystywać w praktyce różne, także zaawansowane, metody statystyczne. Warto do niej sięgać, chociaż brakuje wyczerpującego opisu matematycznego.

W trakcie zajęć będziemy korzystać także z innych źródeł, pokazywanych na bieżąco.

Niniejsze materiały mają charakter roboczy. Bardzo proszę o zgłaszanie zauważonych błędów, braków, niedociągnięć i niejasności. Prośba dotyczy też przypisów. Z góry dziękuję za pomoc☺.

**Czym zajmuje się statystyka**

„*Statystyka* jest to nauka traktująca o metodach ilościowych badania prawidłowości zjawisk (procesów) masowych.”[[1]](#footnote-1)

„Przez badanie statystyczne rozumie się ogół prac mających na celu poznanie struktury określonej zbiorowości statystycznej. ”[[2]](#footnote-2)

„Zbiory dowolnych elementów (osób, przedmiotów, faktów) podobnych pod względem określonych cech (ale nie identycznych) i poddanych badaniom statystycznym nazywa się *zbiorowościami statystycznymi*  lub *populacjami statystycznymi*.”[[3]](#footnote-3)

„*Zbiorowość (populację) generalną* stanowią wszystkie elementy, będące przedmiotem badania, co do których chcemy formułować wnioski ogólne.”[[4]](#footnote-4)

Jeżeli mamy możliwość zbadania wszystkich elementów należących do populacji generalnej, to żadnych wniosków ogólnych nie musimy wyciągać. Wystarczy sporządzić opis zbiorowości.

„Zbiorowość próbna (próba) jest to podzbiór populacji generalnej, obejmujący część jej elementów – wybranych w określony sposób. Próba podlega badaniu statystycznemu, a wyniki badania są uogólniane na zbiorowość generalną.”[[5]](#footnote-5)

Jakość wniosków wyciągniętych na podstawie badania statystycznego zależy od wielu czynników. Między innymi, ważne jest, żeby próba była reprezentatywna, co wymaga odpowiednich metod doboru próby oraz pobrania próby o odpowiedniej liczebności.[[6]](#footnote-6) Zagadnienia związane z doborem próby omówimy w przyszłości. Pamiętajmy, że często przyjmowane liczebności prób określanych jako małe (np. n ≤ 30) lub duże (n > 30), chociaż często uzasadnione, nie są uniwersalne.

Uogólnianie wniosków na populację generalną odbywa się na drodze estymacji lub testowania hipotez statystycznych.

**Cechy statystyczne**

Badania statystyczne mogą być całkowite (gdy badamy wszystkie jednostki statystyczne tworzące populację generalną) lub częściowe (gdy badamy próbę). Właściwości jednostek statystycznych to *cechy statystyczne*.[[7]](#footnote-7)

Według klasyfikacji przytoczonej przez A. Stanisza,[[8]](#footnote-8) wyróżniamy dwa zasadnicze typy cech statystycznych: cechy jakościowe (niemierzalne) oraz cechy ilościowe (mierzalne).

Cechy jakościowe (niemierzalne) opisujemy inaczej niż przy pomocy liczby, podając jakąś charakterystykę, np. płeć (kobieta-mężczyzna), stan zdrowia (zdrowy-chory), kolor skóry (czarny-śniady-biały). Wżną grupą cech statystycznych niemierzalnych są cechy opisujące rozkład przestrzenny badanych zjawisk, tzw. cechy geograficzne[[9]](#footnote-9).

Cechy ilościowe (mierzalne) można przedstawić przy pomocy liczby wyrażonej w pewnej skali. Wyróżniamy cechy ciągłe (np. ciśnienie) i dyskretne (skokowe, przyjmujące najczęściej wartości całkowite), np. liczba trombocytów we krwi.

Charakterystyki lub liczby przypisane w wyniku obserwacji cechom statystycznym możemy wyrazić w różnych skalach pomiarowych. Obecnie bardzo często używane są cztery skale pomiarowe opisane m.in. przez A Stanisza:[[10]](#footnote-10)

1. Skala nominalna – identyfikujemy kategorie (np. płeć, palenie, grupa krwi) i przyporządkowujemy obiekty do tych kategorii. W wyniku tego powstają zbiory rozłączne obiektów.
2. Skala porządkowa – można ją zastosować gdy badaną cechę można uporządkować, np. gdy charakteryzuje się ona natężeniem (np. mało, średnio, dużo).
3. Skala równomierna (przedziałowa) – stosowana do badań ilościowych. Podczas stosowania zakłada się. że zbiór wyników można opisać liczbami rzeczywistymi. Zakłada się stałą jednostkę miary, a zero na skali obiera się dowolnie. Przykładem takiej skali jest skala Celsjusza. Różnica pomiędzy pomiarami ma sens merytoryczny.
4. Skala ilorazowa – oprócz cech skali przedziałowej występuje tu zero bezwzględne oraz stałe stosunki. Przykładem skali ilorazowej jest skala bezwzględna pomiary temperatury.

1. Ostasiewicz, S., Rusnak, Z., Siedlecka, U., Statystyka. Elementy teorii i zadania, Wrocław 1999, s. 7. [↑](#footnote-ref-1)
2. Ibidem, s.7. [↑](#footnote-ref-2)
3. Ibidem, s.7. [↑](#footnote-ref-3)
4. Ibidem,.s. 7. [↑](#footnote-ref-4)
5. Ibidem, s. 7. [↑](#footnote-ref-5)
6. Ibidem, s.8. [↑](#footnote-ref-6)
7. Ibidem, s. 8. [↑](#footnote-ref-7)
8. Stanisz, A. Przystępny kurs statystyki z zastosowaniem STATISTICA.PL na przykładach z medycyny. StatSoft, Kraków 2006, t.1 Statystyki podstawowe, s. 22. [↑](#footnote-ref-8)
9. Ostasiewicz, S., …, op. cit. s. 9. [↑](#footnote-ref-9)
10. Stanisz, A., op. cit. s. 23-24. [↑](#footnote-ref-10)