



Fundusze Europejskie
dla Rozwoju Społecznego



Rzeczpospolita
Polska

Dofinansowane przez
Unię Europejską



Ćwiczenie nr 4

Obliczenia statystyczne w Excelu



**POLITECHNIKA
BYDGOSKA**
im. Jana i Jędrzeja Śniadeckich



**POLITECHNIKA
BYDGOSKA**
Wydział Technologii
i Inżynierii Chemicznej



**POLITECHNIKA
BYDGOSKA**
Wydział Medyczny

PRACOWNIA KOMPUTEROWA



Wstęp

Statystyka odgrywa kluczową rolę w medycynie, umożliwiając lekarzom oraz badaczom wyciąganie wniosków na podstawie zebranych danych. Współczesna praktyka medyczna oraz badania kliniczne opierają się na analizie wyników dużej liczby pacjentów, a umiejętność poprawnej interpretacji tych danych jest niezbędna dla podejmowania właściwych decyzji klinicznych. Program Microsoft Excel, ze swoimi narzędziami do analizy danych, jest często wykorzystywany w medycynie do przeprowadzania podstawowych analiz statystycznych.

Przypomnienie podstawowych miar:

Moda (w Excel nazwa Tryb) jest to wartość pojawiająca się w zbiorze danych najczęściej.

Mediana: jeśli liczba danych jest nieparzysta, mediana to środkowa wartość zbioru. Jeśli liczba danych jest parzysta, mediana to średnia arytmetyczna dwóch środkowych wartości.

Kwartyle są statystycznymi miarami pozycyjnymi, które dzielą uporządkowany zbiór danych na cztery równe części, co pozwala lepiej zrozumieć rozkład danych. Kwartyle są trzy:

1. Pierwszy kwartył (Q1) - 25% danych jest mniejszych lub równych tej wartości.
2. Drugi kwartył (Q2) - to samo co mediana, 50% danych jest mniejszych lub równych tej wartości.
3. Trzeci kwartył (Q3) - 75% danych jest mniejszych lub równych tej wartości.

Jak obliczyć kwartyle?

1. Posortuj dane w rosnącej kolejności.
2. Znajdź drugi kwartył (Q2):
 - Jeśli liczba danych jest nieparzysta, mediana to środkowa wartość zbioru.
 - Jeśli liczba danych jest parzysta, mediana to średnia arytmetyczna dwóch środkowych wartości.
3. Znajdź pierwszy kwartył (Q1): Pierwszy kwartył to mediana dolnej połowy danych, czyli tych, które są mniejsze niż Q2.
4. Znajdź trzeci kwartył (Q3): Trzeci kwartył to mediana górnej połowy danych, czyli tych, które są większe niż Q2.

Wariancja określa rozrzut wyników wokół wartości średniej.

Odchylenie standardowe: miara niepewności pojedynczego pomiaru w serii. To pierwiastek z wariancji:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Odchylenia standardowego średniej (błąd standardowy średniej) to odchylenie standardowe (s) podzielone przez pierwiastek z liczby pomiarów (n).

$$SE = \frac{s}{\sqrt{n}}$$



Cel

Celem tego ćwiczenia jest nabycie umiejętności przeprowadzania podstawowych analiz statystycznych z wykorzystaniem programu Microsoft Excel. Studenci nauczą się:

1. Tworzenia zestawień danych medycznych w arkuszu kalkulacyjnym.
2. Obliczania podstawowych miar statystycznych, takich jak średnia, mediana, odchylenie standardowe czy kwartyle.
3. Wizualizacji danych medycznych przy użyciu wykresów, takich jak histogramy, wykresy liniowe i pudełkowe (boxploty).
4. Interpretacji wyników analizy statystycznej w kontekście badań medycznych.

Przebieg ćwiczenia

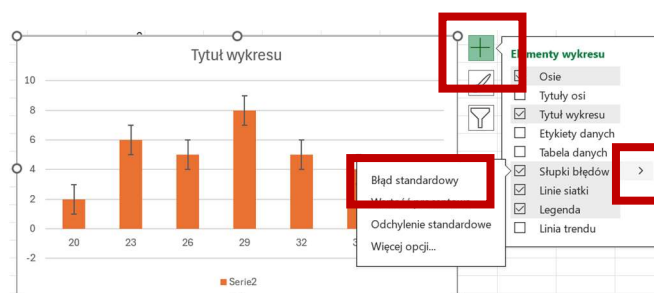
1. Statystyka opisowa

Odszukaj zedytowany przez Ciebie plik BMI_cw3.xlsx z poprzedniego zadania (Zadanie 1 Ćwiczenia 3), gdzie policzone zostały wskaźniki BMI.

- a) Wykonaj statystykę opisową policzonych w poprzednim zadaniu wskaźników BMI.
- i. Policz średnią, medianę i modę. Poszczególne wyrażenia w Excelu to: ŚREDNIA(dane), MEDIANA(dane), WYST.NAJCZĘŚCIEJ.WART(dane).
 - ii. Policz błąd standardowy, odchylenie standardowe, wariancję. Użyj wyrażen w Excelu: ODCZ.STANDARD.PRÓBK(dane), WARIANCJA.PRÓBK(dane). Błąd standardowy to odchylenie standardowe podzielone przez pierwiastek z liczby pomiarów (patrz Wstęp). Wariancja to odchylenie standardowe podniesione do kwadratu.
 - iii. Policz wszystkie trzy kwartyle: KWARTYL.PRZEDZ.ZAMK(dane;numer), gdzie numery 1-3 to numery kwartyli natomiast 0 oraz 4 to wartości min i max.
 - iv. Podobnie jak w poprzednim zadaniu (Zad. 1 w ćwiczeniu 3) stwórz histogram (użyj wykresu kolumnowego 2d) ale tym razem z większą liczbą kategorii od 20 do 38 co 3. Przepisz więc następującą kolumnę (zaznaczoną na czerwono) i użyj funkcji CZĘSTOŚĆ:

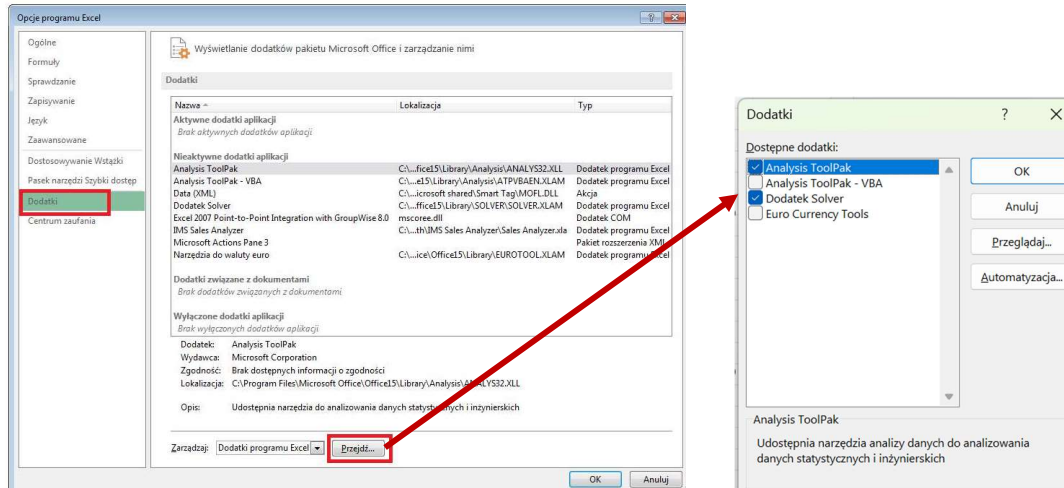
| | |
|----|--|
| 20 | =CZĘSTOŚĆ(E2:E31;H2:H8) |
| 23 | CZĘSTOŚĆ(tablica_dane; tablica_przedziały) |
| 26 | 5 |
| 29 | 8 |
| 32 | 5 |
| 35 | 4 |
| 38 | 0 |

Narysuj wykres kolumnowy (2D) gdzie etykietami będą granice podziałów (20-38) i dodaj na wykresie słupki błędów pokazujące błąd standardowy:

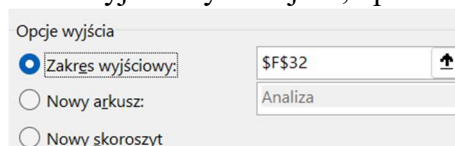




- v. Sprawdź zgodność policzonych przez Ciebie statystyk z wynikami uzyskanymi przez użycie Analizy danych z dodatku Analysis ToolPak. Sprawdź czy dodatek jest włączony (podobnie jak Solver) klikając **Plik->Opcje->Dodatki->Przejdź**

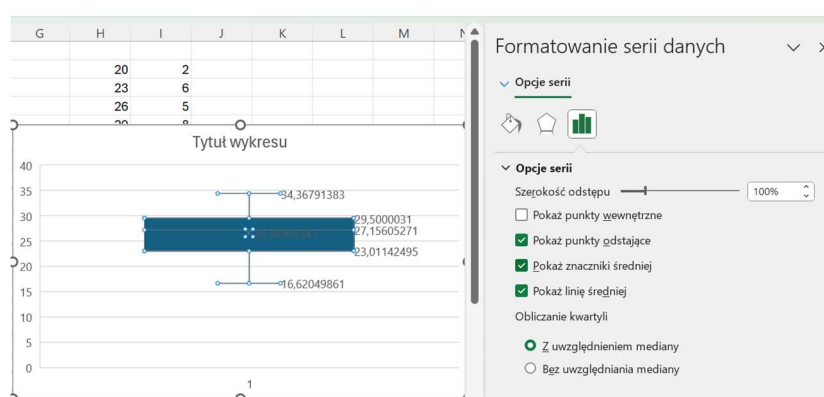


Następnie odszukaj na wstążce opcję: **Dane->Analiza danych** i wybierz „Statystyka opisowa” dla kolumny BMI. Wyniki analizy możesz umieścić na osobnym arkuszu lub we wskazanym w polu „Zakres wyjściowy” miejscu, np.:



Zaznacz checkbox „Statystyka podsumowująca” i dopiero wtedy kliknij „OK”.

- b) Stwórz wykres pudełkowy z uwzględnieniem mediany dla wartości BMI. W tym celu zaznacz dane (kolumna z BMI) i wstaw wykres „Skrzynka i wąsy”. Dodaj etykiety danych (klikając + koło wykresu jak pokazano powyżej), oraz sformatuj wykres jak poniżej:

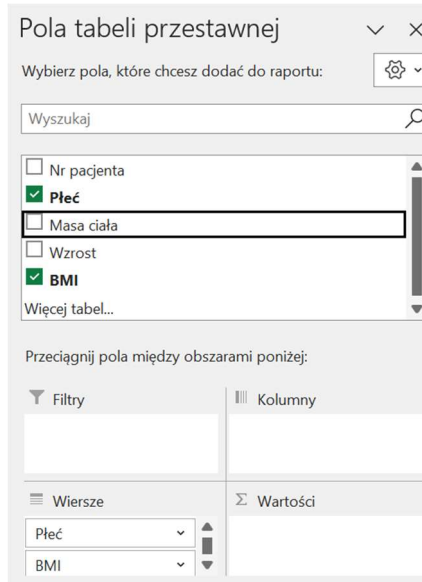


- c) Rozsegreguj powyższe dane na 2 tabele: kobiety i mężczyźni i stwórz wykres jak powyżej pokazując wykres pudełkowy dla kobiet i mężczyzn aby porównać ich BMI. Do rozsegregowania użyjemy tabeli przestawnej:

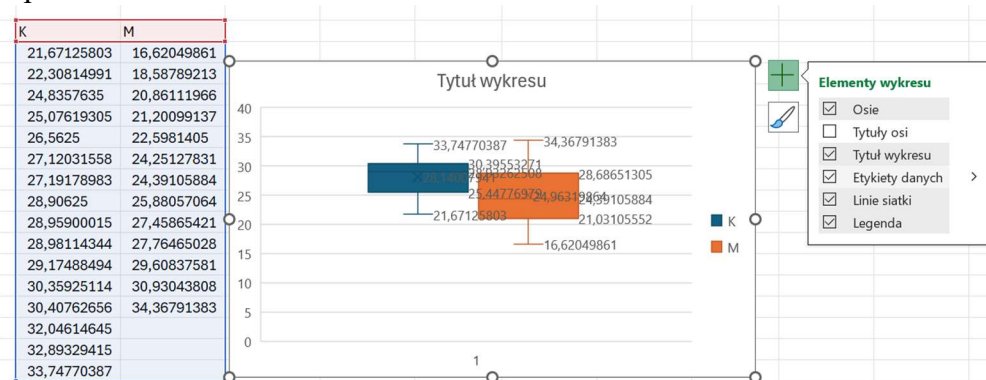
1. Zaznacz całą tabelę danych (włącznie z nagłówkami). Użyj: **Wstawianie->Tabela przestawna**.



2. W oknie dialogowym wybierz, gdzie chcesz umieścić tabelę przestawną (np. na nowym arkuszu) i kliknij **OK**.
3. W polach tabeli przestawnej po prawej stronie:
 - o Przeciągnij **"Płeć"** oraz „**BMI**” do obszaru **Wiersze**.

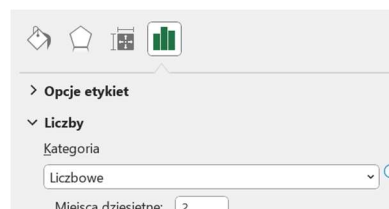


4. Skopiuj dane poza tabelę przestawną, tak, aby dane utworzyły dwie kolumny (jedna dla kobiet i druga dla mężczyzn) a następnie stwórz wykres pudełkowy jak poprzednio:



Dodaj do wykresu legendę oraz etykiety danych.

Zaznacz etykiety danych na wykresie i zmień format liczb na liczbowe z 2 miejscami po przecinku:



Przeanalizuj jakie różnice widzisz.

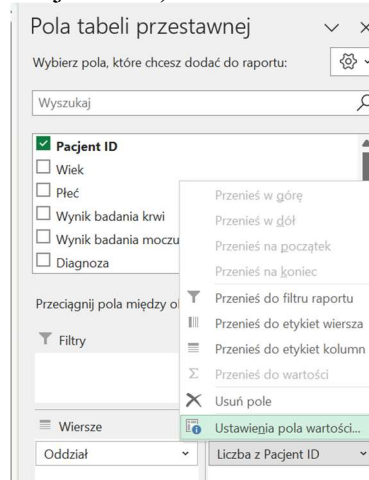
- d) Wykonaj statystykę opisową (**Dane->Analiza danych**) osobno dla kobiet i mężczyzn na wspólnym Arkuszu również w celu porównania wyników.



2. Tabele przestawne

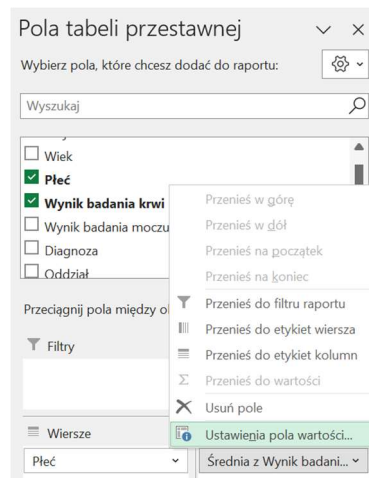
Pobierz plik dane_cw4.xlsx (link do pliku poda prowadzący) i zapisz go na własną nazwę np. NAZWISKO_IMIĘ_WM_ĆW_4.xlsx

- a) Zaznacz dane (włącznie z nagłówkami) i wstaw tabelę przestawną (**Wstawianie->Tabela przestawna**).
- b) Podsumowanie liczby pacjentów na oddziałach (uwaga: nie klikaj nazw nagłówków tylko przeciągnij i upuść):
 - o Umieść (przeciągnij i upuść) "Oddział" w wierszach tabeli przestawnej.
 - o Umieść "Pacjent ID" w wartościach tabeli, licząc pacjentów (zmień z sumowania na „Liczba z Pacjent ID”).



Odkopiuj rozsortowaną tabelę obok, żeby móc pracować nad kolejnym zadaniem na tabeli przestawnej. Usuń wybrane wcześniej pola z wierszy i wartości, żeby otrzymać pustą listę.

- c) Średni wynik badania krwi w zależności od płci:
 - o Umieść (przeciągnij i upuść) "Płeć" w wierszach tabeli przestawnej.
 - o Umieść "Wynik badania krwi" w wartościach tabeli przestawnej, ustawiając jako średnią (nie sumę).



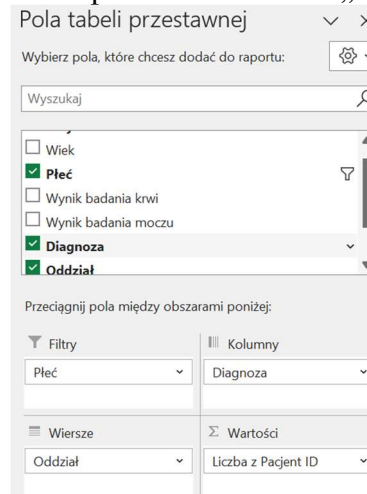
Odkopiuj tabelę obok, żeby móc pracować nad kolejnym zadaniem.

- d) Średni czas hospitalizacji dla każdej diagnozy:
 - o Umieść "Diagnoza" w wierszach tabeli przestawnej.
 - o Umieść "Czas hospitalizacji (dni)" w wartościach, ustawiając jako średnią.



Odkopiuj tabelę obok, żeby móc pracować nad kolejnym zadaniem.

- e) Rozkład diagnoz pacjentów w zależności od oddziału:
- o Umieść "Oddział" w wierszach, a "Diagnoza" w kolumnach tabeli przestawnej.
 - o Umieść "Pacjent ID" w wartościach, licząc pacjentów.
- Zaproponuj reprezentację graficzną: wykres, aby pokazać powyższy rozkład.
- f) Filtr danych:
- o Skorzystaj z filtrów tabeli przestawnej, aby wyświetlić dane tylko dla kobiet lub mężczyzn. Do tego celu wstaw pole Płeć w obszar „Filtr”.



3. Elementy statystyki: częstość i kumulacja

Na pewnej uczelni postanowiono zbadać, z jak licznych rodzin pochodzą studenci. Przeprowadzono zatem badanie polegające na zapytaniu każdego studenta, ile ma rodzeństwa. Po dodaniu do liczby podanej przez studenta jedności otrzymano, ile dzieci jest w danej rodzinie. Okazało się, że największa liczba rodzeństwa wynosiła cztery, co oznacza, że rozpiętość wartości badanej cechy (liczba dzieci w rodzinie) wynosi od jeden do pięciu. Po zliczeniu, ile osób znajduje się w poszczególnych klasach cechy, otrzymane dane przedstawiono w formie tabeli prezentującej w pierwszych dwóch kolumnach szereg statystyczny rozkładu liczby dzieci w rodzinach studentów.

Uzupełnij tabelę i zaprezentuj uzyskane wyniki graficznie (histogramy: znormalizowany i nieznormalizowany (użyj wykresu kolumnowego) , wykres kołowy, wykres dystrybuanty).

| | A | B | C | D | E | F |
|---|----------------------|-------------------|-----------------|---------------------|------------------------------|----------------------------|
| 1 | Wartość cechy | Liczebność | Częstość | Częstość [%] | Kumulacja liczebności | Kumulacja częstości |
| 2 | 1 | 892 | | | | |
| 3 | 2 | 1157 | | | | |
| 4 | 3 | 798 | | | | |
| 5 | 4 | 435 | | | | |
| 6 | 5 | 121 | | | | |
| 7 | Razem | | | | | |

Czy jesteś w stanie powiedzieć jaki to rozkład (tylko wizualnie bez używania testów)?



4. Rozkłady

- a) W pewnym szpitalu oddział ratunkowy przyjmuje średnio 3 pacjentów z zawałem serca na dobę. Załóż, że liczba pacjentów z zawałem serca, którzy trafiają na oddział ratunkowy, jest zmienną losową o rozkładzie Poissona.

Przypomnienie:

Rozkład Poissona opisuje prawdopodobieństwo wystąpienia pewnej liczby zdarzeń w określonym czasie, gdy znana jest średnia liczba tych zdarzeń (w tym przypadku średnia liczba pacjentów).

- a) Jakie jest prawdopodobieństwo, że w ciągu jednego dnia na oddział ratunkowy trafi dokładnie 5 pacjentów z zawałem serca?

Odp: Policz rozkład dla średniej 3 i liczby szukanej 5
=ROZKŁ.POISSON(5;3;0)

Uwaga: zamiast 0 można napisać: Fałsz

- b) Jakie jest prawdopodobieństwo, że w ciągu jednego dnia na oddział ratunkowy trafi 2 lub mniej pacjentów z zawałem serca?

Sposób 1: Policz rozkład dla liczby pacjentów ≤ 2 czyli 0, 1 oraz 2. Następnie zsumuj te prawdopodobieństwa.

| | A | B | C | D | E | F |
|---|---|---|--|---|---|---|
| 1 | | 0 | =ROZKŁ.POISSON(A1;3;0) | | | |
| 2 | | 1 | ROZKŁ.POISSON(x; średnia; skumulowany) | | | |
| 3 | | 2 | | | | |

Sposób 2: Użyj funkcji skumulowanej - dystrybuanty (ostatni parametr ustaw na 1), czyli: =ROZKŁ.POISSON(2;3;1)

Uwaga: zamiast 1 można napisać: Prawda

- c) Jakie jest prawdopodobieństwo, że w ciągu trzech dni na oddział ratunkowy trafi 9 pacjentów z zawałem serca?

Odp: Policz jaka jest średnia liczba pacjentów w ciągu 3 dni i taką średnią wstaw w rozkład. Odp: $3 \cdot 3 = 9$ więc obliczamy: =ROZKŁ.POISSON(9;9;0)

- d) Narysuj rozkład i dystrybuantę (wykresy kolumnowe) w zakresie 0-10 dla tego przykładu.

- b) Załóżmy, że masa ciała 8-letnich dzieci ma rozkład normalny z średnią $\mu=25$ kg i odchyleniem standardowym $\sigma=4$ kg.

- a) Jakie jest prawdopodobieństwo, że losowo wybrane 8-letnie dziecko będzie miało masę ciała mniejszą niż 30 kg?

Odp: Szukamy dystrybuanty rozkładu normalnego:

=ROZKŁ.NORMALNY(30;25;4;1)

Uwaga: Jeśli na końcu parametru podamy 0 zamiast 1 to otrzymamy rozkład a nie dystrybuantę czyli otrzymamy odpowiedź jakie jest prawdopodobieństwo że losowo wybrane dziecko waży dokładnie 30 kg.

- b) Jakie jest prawdopodobieństwo, że losowo wybrane 8-letnie dziecko będzie miało masę ciała większą niż 30 kg?

Odp: odwracamy prawdopodobieństwo, czyli odejmujemy od jedynki wynik poprzedniego zadania: =1-ROZKŁ.NORMALNY(30;25;4;1)