



Dofinansowane przez Unię Europejską



Ćwiczenie nr 4 Obliczenia statystyczne w Excelu





PRACOWNIA KOMPUTEROWA



Rzeczpospolita Polska Dofinansowane przez Unię Europejską



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. Kwartyli są trzy:

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

Jak obliczyć kwartyle?

- 1. Posortuj dane w rosnącej kolejności.
- 2. Znajdź drugi kwartyl (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 kwartyl (Q1): Pierwszy kwartyl to mediana dolnej połowy danych, czyli tych, które są mniejsze niż Q2.
- 4. Znajdź trzeci kwartyl (Q3): Trzeci kwartyl 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).

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



Rzeczpospolita Polska Dofinansowane przez Unię Europejską



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 wyrażeń w Excelu: ODCH.STANDARD.PRÓBKI(dane), WARIANCJA.PRÓBKI(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:



Narysuj wykres kolumnowy (2D) gdzie etykietami będą granicie 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

poje programa Excer					
Ogólne Formuły	Wyświetlanie dodatków pakietu Microsoft Offic	e i zarządzanie nimi			
Sprawdzanie	Dodatki				
Zapisywanie	Nazwa ~	Lokalizacja	Тур	Dodatki	7 >
Język Zaawansowane	Aktywne dodatki aplikacji Brak aktywnych dodatków aplikacji			Destance de dethi	
Dostosowywanie Wstążki Parak naczedni Snibli docten	Nieaktywne dodatki aplikacji Analysis ToolPak Analysis ToolPak	C:\fice15\Library\Analysis\ANALYS32.XLL	Dodatek programu Excel	Analysis ToolPak	АОК
Dodatki	Data (XML) Dodatek Solver	C:\ES(Eb)ary\analysis(ar)+vBacht,Calin C:\icrosoft shared\Smart Tag\MOFL.DLL C:\ffice15\Library\SOLVER\SOLVER.XLAM	Akcja Dodatek programu Excel	Analysis ToolPak - VBA Dodatek Solver Euro Currency Tools	Anuluj
Centrum zaufania	IMS Sales Analyzer Microsoft Actions Pane 3	C:\th\IMS Sales Analyzer\Sales Analyzer.xla	Dodatek COM Dodatek programu Excel Pakiet rozszerzenia XML Dodatek programu pol		<u>P</u> rzeglądaj
	Dodatki związane z dokumentami Brak dodatkiw związanych z dokumentami	C. (CE(OTTCE) (EDTBIJ)(EDTOTOCE)(EDT	bouates program a cer		<u>A</u> utomatyzacja.
	Wyłączone dodatki aplikacji Brak wyłączonych dodatkiw aplikacji				
	Dodatek: Analysis ToolPak Wydawca: Microsoft Corporation Zgodność: Brak dostępnych informacji o zgodności Lokalizacja: C\Program Files\Microsoft Office\OfficeL	5\Library\Analysis\Ata=LYS32.XLL			•
	Opis: Udostępnia narzędzia do analizowania dan	ych statysty mych i inżynierskich		Analysis ToolPak	
	Zarządzaj: Dodatki programu Excel 🔹 📴 rzejdź			Udostępnia narzędzia analizy danych danych statystycznych i inżynierskic	n do analizowania h
			OK Anuluj		

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.:

Zakres wyjściowy:	\$F\$32	1	
O Nowy arkusz:	Analiza		

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:
 - Przeciągnij "Płeć" oraz "BMI" do obszaru Wiersze.

Pola tabeli przesta	iwnej	\sim ×
Wybierz pola, które chcesz doo	dać do raportu:	ବ୍ୟୁ •
Wyszukaj		ر ر
Nr pacjenta		i
Masa ciała		
BMI		I
Więcej tabel		
Przeciągnij pola między obsza	rami poniżej:	
T Filtry	III Kolumny	
Wiersze	Σ Wartości	
Płeć 🗸		
BMI ~		

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

К	М								
21,67125803	16,62049861			0		-	_		
22,30814991	18,58789213	ĭ	Tytu	wykresu		ĭ + <	Elen	anty wykresu	
24,8357635	20,86111966		Tytta	wykrosu				iency wykresu	
25,07619305	21,20099137	40				1		Osie	
26,5625	22,5981405	35		34.36791383				Tytuły osi	
27,12031558	24,25127831	20	33,141	0.39553271				Tytuł wykresu	
27,19178983	24,39105884	50	×28,1400	28,68651305				Etykiety danych	
28,90625	25,88057064	25	2	5.447769724,96312869105884				Linie siatki	
28,95900015	27,45865421	O ₂₀		21,03105552	— К	þ		Legenda	
28,98114344	27,76465028	45		16,62049861	M				_
29,17488494	29,60837581	15							
30,35925114	30,93043808	10							
30,40762656	34,36791383	5							
32,04614645									
32,89329415		0							
33,74770387		L		0		5			

Dodaj do wykresu legendę oraz etykiety danych.

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

> Opcje etyk	iet	
✓ Liczby		
<u>K</u> ategoria		
Liczbowe		•
Miejsca <u>d</u>	ziesiętne: 2	

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):
 - Umieść (przeciągnij i upuść) "Oddział" w wierszach tabeli przestawnej.
 - Umieść "Pacjent ID" w wartościach tabeli, licząc pacjentów (zmień z sumowania na "Liczba z Pacjent ID").

Wyszukaj			
Pacjent ID Wiek			
Płeć		Przenieś w górę	
Wynik badania krwi	Przenieś w <u>d</u> ół		
Wynik badania moczu		Przenieś na <u>p</u> oczątek	
□ Diagnoza		Przenieś na <u>k</u> oniec	
Przeciągnij pola między o	т	Przenieś do filtru raportu	
		Przenieś do etykiet wiersza	
T Filtry	=	Przenieś do etykiet kolumi	
	Σ	Przenieś do wartości	
	×	Usuń pole	
Wiersze	6	Ustawie <u>n</u> ia pola wartości	
Oddział	~	Liczba z Pacient 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:
 - Umieść (przeciągnij i upuść) "Płeć" w wierszach tabeli przestawnej.
 - Umieść "Wynik badania krwi" w wartościach tabeli przestawnej, ustawiając jako średnią (nie sumę).

Pola tabeli prze Wybierz pola, które chcesz	sta dod	wnej v X ać do raportu:		
Wyszukaj		م		
Wiek		A		
Wynik badania krwi		Przenieś w górę		
🗆 Wynik badania moczu		Przenieś w <u>d</u> ół		
Diagnoza		Przenieś na początek		
Oddział		Przenieś na <u>k</u> oniec		
Przeciagnii pola miedzy o	T	Przenieś do filtru raportu		
_		Przenieś do etykiet wiersza		
▼ Filtry	=	Przenieś do etykiet kolumn		
	Σ	Przenieś do wartości		
	×	Usuń pole		
Wiersze	6	Ustawienia pola wartości		
Płeć	~	Średnia z Wynik badani 🛩		

Odkopiuj tabelę obok, żeby móc pracować nad kolejnym zadaniem.

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



Rzeczpospolita Polska Dofinansowane przez Unię Europejską



Odkopiuj tabelę obok, żeby móc pracować nad kolejnym zadaniem.

- e) Rozkład diagnoz pacjentów w zależności od oddziału:
 - Umieść "Oddział" w wierszach, a "Diagnoza" w kolumnach tabeli przestawnej.
 - Umieść "Pacjent ID" w wartościach, licząc pacjentów.

Zaproponuj reprezentację graficzną: wykres, aby pokazać powyższy rozkład. f) Filtr danych:

 Skorzystaj z filtrów tabeli przestawnej, aby wyświetlić dane tylko dla kobiet lub mężczyzn. Do tego celu wstaw pole Płeć w obszar "Filtry".

-ola labell przesta	awnej 🗸 🗸	Ý X
Wybierz pola, które chcesz do	dać do raportu:	⊗ ~
Wyszukaj		Q
U Wiek		
Y Płeć		7
🗌 Wynik badania krwi		- 1
🗌 Wynik badania moczu		
Diagnoza		~
✓ Oddział		
Przeciągnij pola między obsz	arami poniżej:	
Przeciągnij pola między obsz T Filtry Płeć *	arami poniżej: III Kolumny Diagnoza	v
Przeciągnij pola między obsz. T Filtry Płeć Wiersze	arami poniżej: III Kolumny Diagnoza Σ Wartości	v

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	В	С	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)?



Dofinansowane przez Unię Europejską



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.



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*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 μ =25 kg i odchyleniem standardowym σ =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 odpowiedz 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)