
METODY ANALIZY DANYCH

Grażyna Niedziela, Laura Slebioda
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu

PROGRAM SZKOLENIA

1. Word jako narzędzie wspomagające pracę

- Tabele
- Kreatywne certyfikaty i dyplomy z wykorzystaniem korespondencji seryjnej
- Plany

2. Excel – podstawowe funkcje

- Formatowanie danych
 - Blokowanie komórek
 - Makra
-

4. Analiza danych – kilka słów wstępu

- Wstęp do metodologii statystycznej
 - Podstawowe pojęcia: populacja, próba, zmienne losowe
 - Dobór próby – zasady zbierania danych
- Statystyka opisowa
 - Podstawowe funkcje i statystyki opisowe: średnia, mediana, moda, wariancja, odchylenie standardowe, rozstęp
 - Częstość, częstość względna i skumulowana, kwantyle

5. Excel jako narzędzie w analizie danych

- Przygotowanie danych do analizy
- Obliczenia
- Tabele przestawne
- Wizualizacja danych za pomocą wykresów
- Analiza trendów przy pomocy wykresów

6. Formułowanie wniosków

WORD jako narzędzie wspomagające pracę:

- Tabele
- Kreatywne certyfikaty i dyplomy z wykorzystaniem korespondencji seryjnej
- Plany

TABELE

- Często używane do prezentacji danych liczbowych.
- Stosowane do prezentowania tekstu w dowolnej formie.
- Każda komórka stanowi oddzielny akapit i może być dowolnie formatowana.

Realizacja zadań obejmujących m.in.:

- Tworzenie tabel z wykorzystaniem różnych metod
 - Formatowanie tabel
 - Wprowadzanie danych do tabeli
-

Kreatywne certyfikaty i dyplomy z wykorzystaniem korespondencji seryjnej

Przygotowanie dyplomu, który będzie zawierał takie elementy jak: kolor tła, obramowanie, napis *WordArt*, kolor czcionki, ułożenie tej czcionki w odpowiednim miejscu oraz dodatkowy element graficzny.

Następnie z wykorzystaniem korespondencji seryjnej dyplomy zostaną spersonalizowane - uzupełnione o zwrot grzecznościowy oraz imię i nazwisko osób z listy zawartej w pliku dane.doc.

EXCEL :

- Formatowanie danych
 - Blokowanie komórek
 - Makra
-

FORMATOWANIE DANYCH

- Przygotowywanie formatów komórek wg odpowiednich wzorców.
- Wprowadzanie i organizacja danych

BLOKOWANIE KOMÓREK

- Wykonanie zadania z wykorzystaniem blokowania komórek

MAKRA

Makra to narzędzia, dzięki którym można zautomatyzować powtarzające się zadania.

- Wykonywanie poniższych poleceń na udostępnionym pliku:
 1. Rejestracja makra Wiadomość, które wyświetla pole tekstowe
 2. Rejestracja makra, które usuwa pole tekstowe.
 3. Dodawanie przycisku "Wyświetl wiadomość" oraz przycisku "Zamknij wiadomość".
 4. Kliknąć przyciski i sprawdzić działanie makra.
-

ANALIZA DANYCH

Analiza danych to ciągły proces polegający na przekształcaniu danych w celu wypracowania wniosków i zdobycia przydatnych informacji do podejmowania decyzji.

Poprzez efektywne gromadzenie i interpretowanie danych można zgłębić wiedzę w danej dziedzinie, identyfikować wzorce i trendy. Dzięki wykorzystaniu mocy danych można podejmować świadome decyzje, dostosowane do potrzeb i preferencji.

Trzeba mieć świadomość, że każdy zbiór danych jest indywidualny. Nie ma uniwersalnego sposobu postępowania dla różnych zbiorów danych.

Rozpoczynając analizę danych trzeba posiadać wiedzę dotyczącą badanego zjawiska.

Proces analizy danych obejmuje trzy główne kroki:

1. Zbieranie, gromadzenie danych.

2. Analiza danych i wnioski.

3. Wdrażanie zmian.

Podstawowe pojęcia:

Populacja - zbiór dowolnych elementów, których mają dotyczyć wnioski.

Próba - podzbiór populacji, obejmujący skończoną część jej elementów.

Próba podlega badaniu statystycznemu.

PODZIAŁ ZMIENNYCH

- Zmienne jakościowe – nie przyjmują wartości liczbowych: płeć, zawód.
 - Zmienne ilościowe – pozwalają na przedstawienie ich wartości w postaci liczb: wiek, wzrost.
 - Zbiór wartości zmiennej może być zakwalifikowany do jednej ze skal pomiarowych:
 - Skala nominalna
 - Skala porządkowa
 - Skala interwałowa
 - Skala ilorazowa
-

OPIS ZMIENNEJ

- Do podstawowego opisu statystycznego zmiennych stosujemy:
 - Miary tendencji centralnej
 - Miary rozproszenia (dyspersji)
- Prezentacja graficzna statystyk opisowych

MIARY TENDENCJI CENTRALNEJ

- **Moda** (dominanta): określa wartość najczęstszą w danym zbiorze. Może być stosowana dla skal nominalnych, porządkowych i ilościowych.
 - **Mediana:** wartość środkowa w uporządkowanym według wielkości zbiorze. Może być stosowana dla skal porządkowych i ilościowych. W przypadku gdy liczba elementów w próbie jest nieparzysta mediana jest wartością zmiennej elementu środkowego. Gdy liczba elementów w próbie jest parzysta, medianę wyznaczamy obliczając średnią z wartości zmiennej dwóch środkowych elementów.
 - **Średnia arytmetyczna:** wartość przeciętna. Ma zastosowanie wyłącznie do skal ilościowych (interwałowych i ilorazowych)
 - **Kwartyle:** dzielą uporządkowaną grupę na cztery równe części.
-

MIARY ROZPROSZENIA

- **Rozstęp:** różnica między maksymalną i minimalną wartością w grupie.
 - **Odchylenie standardowe:** miara rozproszenia danych wokół ich średniej. Im większe odchylenie standardowe, tym większa zmienność danych.
 - **Wariancja:** kwadrat odchylenia standardowego.
-

Omówienie funkcji programu Excel, które są niezwykle przydatne w analizie danych.

Przyjęcie roli dydaktyków w fikcyjnej szkole i analizowanie danych dotyczących wyników nauczania.
Identyfikacja wzorców, szukanie trendów oraz podejmowanie decyzji mających na celu optymalizację nauczania.

Strategie pozwalające zwiększyć efektywność zmian opartych na danych:

1. Ustalanie jasnych celów i założeń.
 2. Wykorzystywanie wielu źródeł danych.
 3. Wykorzystywanie danych w nauce adaptacyjnej.
 4. Wykorzystywanie informacji zwrotnych.
 5. Współpraca ze wszystkimi zainteresowanymi stronami.
-

WSKAŹNIK ŁATWOŚCI ZADANIA

$$p = \frac{\Sigma \text{ punktów za zadanie (wszyscy uczniowie)}}{\text{Liczba punktów za zadanie} * \text{liczba uczniów}}$$

Przedział wartości wskaźnika p	Poziom trudności zadania
0,00 – 0,19	bardzo trudne
0,20 – 0,49	trudne
0,50 – 0,69	umiarkowanie trudne
0,70 - 0,89	łatwe
0,90 – 1,00	bardzo łatwe

q – wskaźnik trudności zadania

$$q = 1 - p$$

RZETELNOŚĆ POMIARU – WSKAŹNIK KUDERA-RICHARDSONA

$$KR_{20} = \frac{k}{k-1} \left(1 - \frac{\sum (p_i * q_i)}{s^2} \right)$$

k – liczba zadań w teście
p_i – łatwość danego zadania
q_i – trudność danego zadania
s² – wariancja wyników testowania

Przedział wskaźnika rzetelności KR ₂₀	Stopień rzetelności testu	Dopuszczalność wnioskowania
0,80 – 1,00	test rzetelny	Test daje rzetelny obraz wiedzy i umiejętności uczniów
0,50 – 0,79	test mało rzetelny	Test pozwala analizować wyniki grup, klas, szkół, ale nie można na jego podstawie wystawiać ocen
0,00 – 0,49	test nierzetelny	test nie nadaje się do dalszych analiz

TABELE PRZESTAWNE

- Zaawansowane narzędzie do analizowania danych, wykonywania obliczeń i tworzenia podsumowań.
 - Możliwość szybkiego wykonywania podsumowania dużej ilości danych.
 - Możliwość wyciągnięcia z dużych i skomplikowanych tabel dokładnie tych informacji, które są potrzebne do przeprowadzanej analizy bądź też tworzonego raportu.
 - Możliwość dowolnego przekształcania jej wierszy oraz kolumn bez bezpośredniej ingerencji w oryginalne zestawy danych.
-

Przygotowanie danych wejściowych:

- Organizowanie danych w kolumnach;
 - Kolumny muszą mieć nagłówki;
 - Nagłówek musi obejmować jedną komórkę i znajdować się w pierwszym wierszu tabeli;
 - W każdej kolumnie powinien znajdować się jeden określony typ danych.
-

1. Otworzyć arkusz "Wyniki_Matur_Szkola".

2. Przy pomocy tabeli przestawnej odpowiedzieć na pytania:

- Ile osób zdawało maturę z języka polskiego w poszczególnych klasach?
- Jaka była średnia wyniku matury z matematyki w poszczególnych klasach?
- W której klasie uzyskano najwyższy wynik z angielskiego?
- W której klasie uzyskano najniższy wynik z biologii?
- W której klasie nie było osób zdawających maturę z chemii?

3. Stworzyć tabelę przestawną, która pokaże wynik z matematyki oraz języka polskiego dla każdego ucznia (filtr).

4. Przy pomocy tabeli przestawnej odpowiedzieć jaka była średnia całej szkoły z historii, a jaka z WOSu.

Różnice między używaniem formuł, a używaniem tabel przestawnych:

Dokonanie porównania na tabeli, która zawiera wykaz sklepów, rodzaj produktów jakie te sklepy sprzedawały oraz na jaką wartość została dokonana sprzedaż.

Na udostępnionym pliku "Dane-sklepow.xls" wykonać obliczenia za pomocą tabeli przestawnej oraz za pomocą innych formuł w celu zobaczenia różnicy.

- Obliczyć ile razy występuje każdy rodzaj produktu.
 - Obliczyć ile razy występuje każdy ze sklepów.
 - Obliczyć ile wynosi wartość sprzedaży dla każdego ze sklepów z podziałem na rodzaj produktu.
-

WYKRESY

- Otworzyć plik Wyniki_Sprawdzianu i w zakładce "Boxplot" wykonać wykres skrzynka i wąsy.
 - Zidentyfikować wszystkie charakterystyczne elementy wykresu:
 - Górna obserwacja ekstremalna
 - Górny outlier (obserwacja odstająca)
 - Najwyższy wynik, który nie jest outlierem
 - Górny czyli trzeci kwartył – q_3
 - Mediana czyli drugi kwartył – q_2
 - Dolny czyli pierwszy kwartył – q_1
 - Najniższy wynik, który nie jest outlierem
 - Dolny outlier (obserwacja odstająca)
 - Dolna obserwacja ekstremalna
-

WYKRESY

1. Otworzyć plik Oceny_klasa_1A_Matematyka i uzupełnić:

- a. W kolumnie G stworzyć funkcję, która obliczy średnią ważoną.
- b. W kolumnie H użyć funkcji ZAOKR().
- c. W wierszu 28 użyć odpowiedniej funkcji do uzupełnienia pól.
- d. W "Oceny – podsumowanie" użyj funkcji LICZ.JEŻELI().

2. Następnie stworzyć wykresy:

- a. Histogram średniej ocen (dodać tytuł wykresu, wypełnienie: gradientowe, obramowanie: linia ciągła; szerokość 2 pkt)
 - b. Wykres kolumnowy "Oceny – podsumowanie" (dodać tytuł wykresu, dodać tytuły osi, wypełnienie: deseń, obramowanie: linia ciągła - 1 pkt, zmień typ kreski, zmienić kolor tła)
 - c. Wykres liniowy przedstawiający średnią sprawdzianów/kartków. Zaznaczyć dane tak, aby etykietami osi poziomej (kategorii) były poszczególne sprawdziany/kartkówki (dodać tytuł wykresu, dodać etykiety danych, zmienić kształt i kolor znaczników)
-

Analiza wyników egzaminów:

Celem przeprowadzenia egzaminu było uzyskanie informacji dotyczącej poziomu opanowania przez uczniów poszczególnych umiejętności, a w efekcie podnoszenie jakości pracy szkoły.

1. Otworzyć plik Egzamin_analiza.xlsx.
 2. Uzupełnić obliczeniami zaznaczone pole dla każdej klasy.
 3. Wykonać wykresy dla każdej klasy przedstawiające:
 - a. wartości współczynnika łatwości poszczególnych zadań (wykres liniowy ze znacznikami)
 - b. wartości współczynnika łatwości poszczególnych standardów (wykres kolumnowy)
 - c. średnie wyniki poszczególnych uczniów (wykres kolumnowy)
 4. Sformułować wnioski dla wyników otrzymanych dla każdej klasy.
-

-
5. Uzupełnić obliczeniami zaznaczone pola arkusza szkoła.
 6. Wykonać wykresy dla obliczeń wykonanych w arkuszu szkoła.
 - a. wartości współczynnika łatwości poszczególnych zadań
 - b. średnie wyniki poszczególnych klas
 7. Sformułować wnioski.
-

DZIĘKUJEMY ZA UWAGĘ!

Materiał powstał w ramach projektu „Najlepsi z natury! Podnoszenie kompetencji osób dorosłych przez UPP”, współfinansowanego ze środków Europejskiego Funduszu Społecznego Plus, w ramach programu Fundusze Europejskie dla Rozwoju Społecznego 2021-2027, na podstawie umowy nr FERS.01.05-IP.08-0469/23



Fundusze Europejskie
dla Rozwoju Społecznego



Rzeczpospolita
Polska

Dofinansowane przez
Unię Europejską

