

# STATYSTYKA OPISOWA (lab. 4)

## GRUPOWANIE DANYCH STATYSTYCZNYCH

### Wprowadzenie

Sposób prezentacji danych przekrojowych, polegający na *sortowaniu* czy *rangowaniu* wartości i prezentacji graficznej w postaci *wykresu słupkowego* nie zdaje egzaminu w przypadku danych ankietowych. Ten typ danych należy pogrupować i podać licznosc oraz czestość występowania poszczególnych wariantów odpowiedzi (jest to tworzenie tak zwanego **szeregu rozdzielczego**). Również dane typu przekrojowego czy czasowego chcemy czasem jedynie „policzyć”, a nie przedstawiać w dokładny sposób jak to było robione na poprzednich zajęciach. Na przykład celem analizy może być wskazanie liczby (i procentowego udziału) powiatów, w których poziom przeciętnych wynagrodzeń zawiera się w przedziale 3000-3500 zł bez wyszczególniania, jakie to powiaty.

W programie *STATISTICA* do grupowania danych służą następujące narzędzia: **Tabele licznosci** oraz dwa typy wykresów: **Histogramy** i **Wykresy kołowe**. Analizy te można wykonywać niezależnie – informację o strukturze procentowej można uzyskać za pomocą samych wykresów, bez wcześniejszego wykonywania tabel.

### Przykład 1 (Opinie studentów o UE 2010-2012)

#### TABELE LICZNOŚCI DLA CECH NOMINALNYCH + HISTOGRAMY

Celem analizy jest poznanie opinii studentów na temat zysków Polski w wyniku integracji z UE (*Czy Polska zyskała na integracji z UE*) oraz wpływu członkostwa w UE na zamożność Polaków (*Wpływ członkostwa w UE na zamożność Polaków*). Aby przedstawić wyniki w formie tabel wybieramy polecenie *STATYSTYKA / STATYSTYKI PODSTAWOWE I TABELE / TABELE LICZNOŚCI* po czym na liście zmiennych wskazujemy obie wyżej wymienione cechy. W zakładce **WIĘCEJ** sprawdzamy *Metodę kategoryzacji dla tabel i wykresów*, która powinna być ustawiona jako *Wszystkie różne wartości (z etykietami tekstowymi)*. Za pomocą przycisku *Podsumowanie* wywołujemy skoroszyt wyników, w którym umieszczone zostały dwie tabele, zawierające rozkład odpowiedzi na interesujące nas pytania.

Na podstawie uzyskanych wyników, proszę odpowiedzieć na następujące pytania:

- czy studenci uważają, że Polska zyskała na integracji – ile procent osób jest takiego zdania .....
- jaka jest najbardziej popularna opinia jeśli chodzi o ocenę wpływu integracji na zamożność Polaków i przez jaki procent ankietowanych jest wygłaszana .....

Aby przedstawić wyniki w postaci *Histogramu* wystarczy przywołać okno analizy *Tabele licznosci* i nacisnąć stosowny przycisk (UWAGA! Nie trzeba powtarzać procedury wyboru zmiennych i innych żmudnych czynności okna analiz po wywołaniu wyników są minimalizowane w dolnej części ekranu – wystarczy przywołać odpowiednią analizę lub wykorzystać skrót **Ctrl + R**).

Więcej możliwości daje jednak wykonanie wykresów za pomocą polecenia *WYKRESY / WYKRESY 2W / HISTOGRAMY* lub *WYKRESY KOŁOWE*. Proszę sporządzić wizualizację rozkładu odpowiedzi na oba rozważane pytania w formie wykresu kołowego.

### Przykład 2 (Opinie o integracji z UE 2004)

#### GRUPOWANIE DANYCH W PRZEDZIAŁY

W analogiczny sposób przedstawimy rozkład wieku ankietowanych 300 osób w badaniu ankietowym na temat integracji z UE przeprowadzonym w roku 2004. Proszę wykorzystać *Tabele licznosci* i przedstawić wynik w postaci tabeli, a następnie histogramu. Czy przedstawione wyniki są czytelne, czy potrzebny jest taki poziom szczegółowości prezentacji danych?

W przypadku danych mających charakter liczbowy (a zwłaszcza, gdy cecha przyjmuje szeroki zakres wartości) sensowne wydaje się stworzenie **szeregu rozdzielczego przedziałowego**, co wymaga określenia początku pierwszego przedziału i ich rozpiętości. Przyjmijmy, iż chcemy pogrupować wiek w następujących przedziałach 15-24, 25-34, ..., itd.. W tym celu wykorzystujemy opcję *Krok* według wskazówek zamieszczonych obok tabeli (zakładka *Więcej* w oknie *Tabele licznosci*).

Wiek	<i>n</i>	<i>n<sub>SK</sub></i>	%	% <sub>SK</sub>
[15; 25)				
[25; 35)				
[35; 45)				
[45; 55)				
[55; 65)				
[65; 75)				
[75; 85)				

Krok:

lub  od minimum

*n* – liczba osób, *n<sub>SK</sub>* – liczebność skumulowana, % – procent, %<sub>SK</sub> – procent skumulowany

Na podstawie uzyskanych wyników proszę udzielić odpowiedzi na pytanie, ile ankietowanych osób (i jaki procent) miało mniej niż 45 lat .....

Proszę przedstawić wyniki w formie *Histogramu*, wykorzystując domyślny wykres wbudowany w okno analizy *Tabele licznosci*.

## STATYSTYKA OPISOWA (lab. 4)

### GRUPOWANIE DANYCH STATYSTYCZNYCH

#### Przykład 3 (Wynagrodzenia w powiatach 2002-2017)

Zanim przejdziemy do nowej analizy, krótka powtórka z poprzednich zajęć...

Proszę wpisać do tabeli nazwę powiatu miejsca swojego stałego zamieszkania, a następnie za pomocą procedury rangowania wyznaczyć miejsce tego powiatu na tle wszystkich powiatów w Polsce pod względem wynagrodzenia dla poszczególnych lat. Wyniki proszę wpisać do poniższej tabeli.

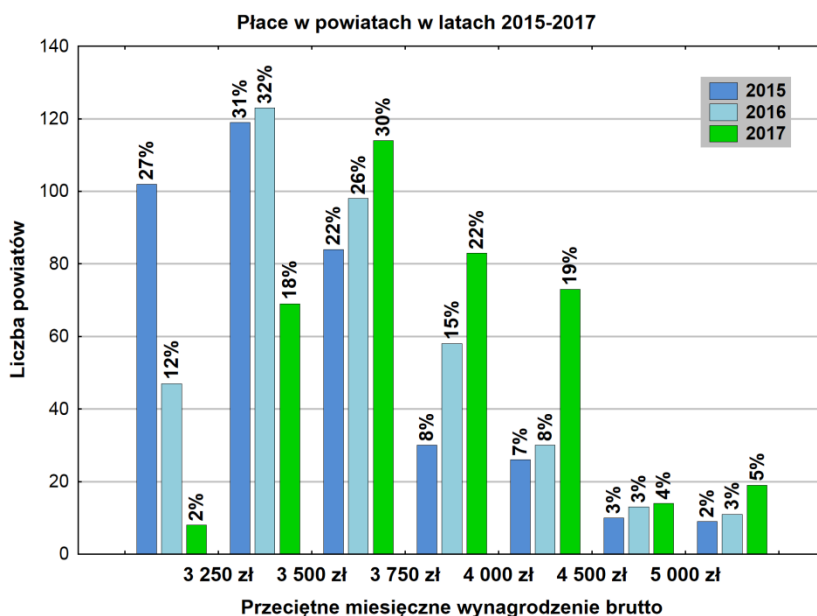
Powiat miejsca zamieszkania studenta	Pozycja w Polsce pod względem płac					
	2002	2005	2008	2011	2014	2017
.....						

### HISTOGRAMY

Celem analizy jest dokonanie opisu poziomu wynagrodzeń w Polsce w przekroju powiatów w roku 2015, 2016 i 2017. Oczywiście do analizy można wykorzystać procedury poznane na wcześniejszych zajęciach: sortowania, rangowania, wizualizacji w postaci wykresów słupkowych. Tym razem jednak nie będziemy analizować „imiennie” powiatów, lecz zbadamy ogólny rozkład wynagrodzeń, tworząc szereg przedziałowy o przedziałach różnej rozpiętości – wykorzystamy w tym celu możliwość wizualizacji danych w postaci histogramu.

Oto kolejne etapy tworzenia odpowiedniego wykresu:

- Za pomocą poleceń **Wykresy / Wykresy 2W / Histogramy** wywołujemy okno analizy **Histogramy 2W** po czym wybieramy jako zmienne **2015, 2016 i 2017**;
- W zakładce **Więcej** wskazujemy **Rodzaj wykresu** jako **Wielokrotny** oraz wyłączamy opcję **Dopasuj**;
- Aby ustalić sposób grupowania płac, znajdujemy opcję **Granice**, po czym wprowadzamy następujące liczby **3250 3500 3750 4000 4500 5000** – w ten sposób dane zostaną pogrupowane w 7 przedziałów, przy czym skrajne przedziały są otwarte. Jak widać, przedziały nie muszą być równej długości – w przypadku płac, które wykazują się dość dużą asymetrią, wygodnie jest wybrać „krótsze” przedziały dla niższych przeciętnych wynagrodzeń.
- Zaznaczamy jeszcze opcję **Pokaż procenty** i **Odstępy między kolumnami**, po czym sporządzamy wykres.
- Wykres formatujemy, tak by był jak najbardziej zgodny ze wzorcem zamieszczonym obok.
- Wykres proszę zapisać pod nazwą **Płace w powiatach 2015-2017**. Wykres zapisujemy wywołując menu podręczne prawym przyciskiem myszki, a następnie wybierając polecenie **Zapisz wykres**.



#### Przykład 4 (Ankieta studencka 2013-2016)

### WYKRESY KOŁOWE

Za pomocą wykresu kołowego proszę wyznaczyć procentową strukturę odpowiedzi na pytanie: *Czy Polska powinna jak najszybciej przystąpić do strefy EURO?* Na wykresie proszę włączyć informację o procentowym udziale poszczególnych wariantów odpowiedzi (w tym celu wystarczy dwukrotnie kliknąć którąkolwiek z etykiet na wykresie).

### Umiejętności do sprawdzenia na kolejnych zajęciach / kolokwium:

1. Grupowanie danych ankietowych za pomocą analizy **Tabele licznosci** – dla danych nominalnych w postaci **szeregu szczegółowego**, dla danych liczbowych w postaci **szeregu przedziałowego** o określonym początku i długości przedziału.
2. Tworzenie prezentacji graficznej danych ankietowych za pomocą **histogramów** i **wykresów kołowych**. Wraz z podaniem struktury procentowej.