

STATYSTYKA OPISOWA (lab. 10)

GRAFICZNA ANALIZA DANYCH – część 2.

Na dzisiejszych zajęciach zapoznamy się z wybranymi **wykresami statystycznymi** – jest to nieformalne określenie, które dotyczyć będzie wykresów, które pokazują wyniki pewnych analiz statystycznych, a nie dane surowe. Najczęściej są to proste analizy, dotyczące grupowania danych (odpowiadające im wykresy to: **histogramy** i **wykresy kołowe**) albo wyliczania statystyk opisowych (**wykresy ramka-wąsy**).

Wykresy „statystyczne” nie są „wrażliwe” na zmiany struktury zbioru danych – w szczególności sortowanie.

UWAGA: Wszystkie wykorzystywane wykresy są dostępne po wybraniu polecenia *Wykresy / Wykresy 2W*.

Histogram w wersji wielokrotnej Dane: Ankieta studencka 2013-2016																																																																																					
<p>Histogramy są to wykresy pozwalające przedstawić częstość występowania wartości pewnej cechy w badanej zbiorowości. Oczywiście, w zależności od potrzeb użytkownika, program może zliczać wszystkie różne wartości bądź grupować je w pewne przedziały.</p> <p>Celem analizy będzie jednoczesna wizualizacja odpowiedzi ankietowanych na trzy pytania, dotyczące ważności trzech problemów społecznych w opinii respondentów: <i>Polityki prorodzinnej</i>, <i>Legalizacji związków partnerskich</i> i <i>Przeciwdziałania migracji zarobkowej Polaków</i>.</p> <p>W tym celu ustalamy rodzaj wykresu na Wielokrotny. Następnie w zakładce Więcej włączamy pewne dodatkowe opcje: Pokaż procenty, Odstępy między kolumnami i wyłączamy opcję: Dopasuj, aby zwiększyć czytelność wykresu. Wykonujemy wykres i starannie go formatujemy, wykorzystując zdobyte wcześniej umiejętności.</p>	<table border="1" style="margin: auto; border-collapse: collapse;"> <caption>Liczba ankietowanych osób w zależności od ważności problemu</caption> <thead> <tr> <th>Ważność</th> <th>Polityka prorodzinna</th> <th>Legalizacja związków partnerskich</th> <th>Przeciwdziałanie emigracji zarobkowej Polaków</th> </tr> </thead> <tbody> <tr> <td>nieważny</td> <td>1%</td> <td>38%</td> <td>1%</td> </tr> <tr> <td>mało ważny</td> <td>9%</td> <td>44%</td> <td>8%</td> </tr> <tr> <td>dość ważny</td> <td>32%</td> <td>12%</td> <td>21%</td> </tr> <tr> <td>ważny</td> <td>35%</td> <td>4%</td> <td>32%</td> </tr> <tr> <td>bardzo ważny</td> <td>23%</td> <td>2%</td> <td>38%</td> </tr> </tbody> </table>	Ważność	Polityka prorodzinna	Legalizacja związków partnerskich	Przeciwdziałanie emigracji zarobkowej Polaków	nieważny	1%	38%	1%	mało ważny	9%	44%	8%	dość ważny	32%	12%	21%	ważny	35%	4%	32%	bardzo ważny	23%	2%	38%																																																												
Ważność	Polityka prorodzinna	Legalizacja związków partnerskich	Przeciwdziałanie emigracji zarobkowej Polaków																																																																																		
nieważny	1%	38%	1%																																																																																		
mało ważny	9%	44%	8%																																																																																		
dość ważny	32%	12%	21%																																																																																		
ważny	35%	4%	32%																																																																																		
bardzo ważny	23%	2%	38%																																																																																		
Wykres ramka-wąsy w wersji wielokrotnej Dane: Bezrobocie w gminach 2003-2022																																																																																					
<p>Wykres ramka-wąsy służy do wizualizacji statystyk opisowych, a więc miar przeciętnego poziomu i zmienności danej cech liczbowych. Celem analizy będzie prezentacja bezrobocia ogółem w 160 gminach Podkarpacia w latach 2003-2022:</p> <ol style="list-style-type: none"> 1) znajdujemy w grupie <i>Wykresy 2W / Wykres ramka-wąsy</i>; 2) na liście zmiennych wskazujemy wskaźniki bezrobocia ogółem z lat 2003-2022 i zmieniamy rodzaj wykresu na Wielokrotny; 3) włączamy warunek selekcji Województwo = „podkarpackie” (prościej $v3=9$); 4) w zakładce Więcej ustawiamy Punkt środkowy jako Medianę, Ramkę jako Centyl 25 i Wąs jako Minimum-maksimum oraz [wartości] Odstające jako Wyłączone; 5) na wykresie dokonujemy formatowania – zmieniamy położenie legendy, dodajemy tytuł osi i główny tytuł; 6) opisy lat na osi poziomej są niewygodne (program wykorzystał nazwy zmiennych „nie wiedząc”, że w zupełności wystarczą same lata), aby je zmienić klikamy dwa razy etykiety osi poziomej i w zakładce Wartości skali//Edytuj etykiety użytkownika wprowadzamy własne etykiety (w polu <i>Pozycja</i> należy podać 1, a w polu <i>Tekst</i> – 2003; następnie 2 i 2004, 3 i 2005 i tak dalej). Można tu wykorzystać arkusz <i>Excelsa</i> i tam przygotować tabelę, a potem wkleić ją do programu <i>STATISTICA</i>. 	<table border="1" style="margin: auto; border-collapse: collapse;"> <caption>Wskaznik bezrobocia w gminach woj. podkarpackiego w latach 2003-2022</caption> <thead> <tr> <th>Rok</th> <th>Mediana</th> <th>25%-75%</th> <th>Min-Maks</th> </tr> </thead> <tbody> <tr><td>2003</td><td>15%</td><td>12-18%</td><td>8-26%</td></tr> <tr><td>2004</td><td>14%</td><td>11-17%</td><td>7-25%</td></tr> <tr><td>2005</td><td>13%</td><td>10-16%</td><td>6-24%</td></tr> <tr><td>2006</td><td>12%</td><td>9-15%</td><td>5-23%</td></tr> <tr><td>2007</td><td>11%</td><td>8-14%</td><td>4-22%</td></tr> <tr><td>2008</td><td>10%</td><td>7-13%</td><td>3-21%</td></tr> <tr><td>2009</td><td>9%</td><td>6-12%</td><td>2-20%</td></tr> <tr><td>2010</td><td>8%</td><td>5-11%</td><td>1-19%</td></tr> <tr><td>2011</td><td>7%</td><td>4-10%</td><td>0-18%</td></tr> <tr><td>2012</td><td>6%</td><td>3-9%</td><td>0-17%</td></tr> <tr><td>2013</td><td>5%</td><td>2-8%</td><td>0-16%</td></tr> <tr><td>2014</td><td>4%</td><td>1-7%</td><td>0-15%</td></tr> <tr><td>2015</td><td>3%</td><td>0-6%</td><td>0-14%</td></tr> <tr><td>2016</td><td>2%</td><td>0-5%</td><td>0-13%</td></tr> <tr><td>2017</td><td>1%</td><td>0-4%</td><td>0-12%</td></tr> <tr><td>2018</td><td>0%</td><td>0-3%</td><td>0-11%</td></tr> <tr><td>2019</td><td>0%</td><td>0-2%</td><td>0-10%</td></tr> <tr><td>2020</td><td>0%</td><td>0-1%</td><td>0-9%</td></tr> <tr><td>2021</td><td>0%</td><td>0-0%</td><td>0-8%</td></tr> <tr><td>2022</td><td>0%</td><td>0-0%</td><td>0-7%</td></tr> </tbody> </table>	Rok	Mediana	25%-75%	Min-Maks	2003	15%	12-18%	8-26%	2004	14%	11-17%	7-25%	2005	13%	10-16%	6-24%	2006	12%	9-15%	5-23%	2007	11%	8-14%	4-22%	2008	10%	7-13%	3-21%	2009	9%	6-12%	2-20%	2010	8%	5-11%	1-19%	2011	7%	4-10%	0-18%	2012	6%	3-9%	0-17%	2013	5%	2-8%	0-16%	2014	4%	1-7%	0-15%	2015	3%	0-6%	0-14%	2016	2%	0-5%	0-13%	2017	1%	0-4%	0-12%	2018	0%	0-3%	0-11%	2019	0%	0-2%	0-10%	2020	0%	0-1%	0-9%	2021	0%	0-0%	0-8%	2022	0%	0-0%	0-7%
Rok	Mediana	25%-75%	Min-Maks																																																																																		
2003	15%	12-18%	8-26%																																																																																		
2004	14%	11-17%	7-25%																																																																																		
2005	13%	10-16%	6-24%																																																																																		
2006	12%	9-15%	5-23%																																																																																		
2007	11%	8-14%	4-22%																																																																																		
2008	10%	7-13%	3-21%																																																																																		
2009	9%	6-12%	2-20%																																																																																		
2010	8%	5-11%	1-19%																																																																																		
2011	7%	4-10%	0-18%																																																																																		
2012	6%	3-9%	0-17%																																																																																		
2013	5%	2-8%	0-16%																																																																																		
2014	4%	1-7%	0-15%																																																																																		
2015	3%	0-6%	0-14%																																																																																		
2016	2%	0-5%	0-13%																																																																																		
2017	1%	0-4%	0-12%																																																																																		
2018	0%	0-3%	0-11%																																																																																		
2019	0%	0-2%	0-10%																																																																																		
2020	0%	0-1%	0-9%																																																																																		
2021	0%	0-0%	0-8%																																																																																		
2022	0%	0-0%	0-7%																																																																																		

STATYSTYKA OPISOWA (lab. 10)

GRAFICZNA ANALIZA DANYCH – część 2.

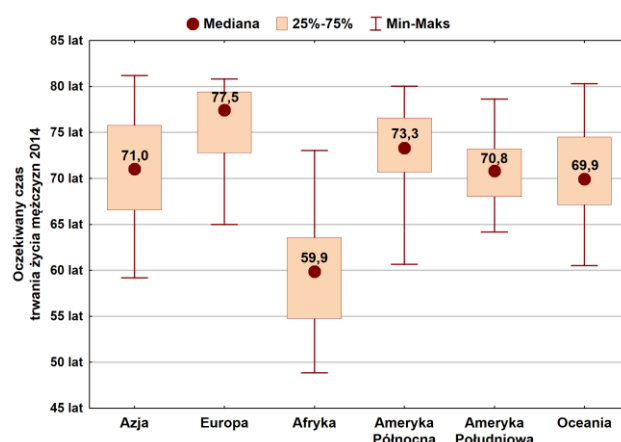
Wykres ramka wąsy ze zmienną grupującą Dane: Czas trwania życia na świecie 1960-2022

Celem analizy będzie porównanie poziomu oczekiwanego czasu trwania życia mężczyzn w państwach świata w roku 2014 według kontynentów. W tym celu ponownie wykorzystamy **wykres ramka-wąsy**. W oknie wyboru zmiennych na pierwszej liście wskazujemy *Oczekiwany czas trwania życia mężczyzn 2014*, a na drugiej *Kontynent*.

W zakładce *Więcej* ustawiamy *Punkt środkowy* jako *Mediana*, *Wąs* jako *Minimum-maksimum* i *Odstające* jako *Wyłączone*. Wykonujemy wykres. Dokonujemy formatowania wykresu, tak by uzyskać jego wygląd zgodny z załączonym rysunkiem.

Szczegółowe wskazówki:

- w zakładce *Ramka-wąsy* ustalamy szerokość ramki na 60%;
- w zakładce *Etykiety punktów* włączamy *Etykiety punktów* i zaznaczamy *Współrzędne Y*, odpowiednio je formatując;
- formatujemy tytuły (pogrubiamy i powiększamy czcionkę), przenosimy legendę do tytułu, etc.



Wykres kołowy ze zmienną grupującą Dane: Opinie o integracji z UE 2004

Celem analizy będzie pokazanie sposobu głosowania w referendum akcesyjnym ankietowanych osób, w zależności od płci i miejsca zamieszkania.

W oknie wyboru zmiennych **Wykresu kołowego** wskazujemy *Sposób głosowania w referendum europejskim*, a następnie w zakładce *Skategoryzowane* włączamy *Kategorię X i Y*, wskazując na listach odpowiednio zmienne: *Płeć* i *Miejsce zamieszkania*. Tworzymy wykres.

Na gotowym wykresie włączamy wyświetlanie procentów (pod dwukrotnym kliknięciu etykiet przy wykresie), w zakładce *Kategoryzacja* ustawiamy *Nazwy zmiennych i klas* na *Tylko nazwy klas*.

Formatujemy tytuł wykresu, usuwamy zbędne opisy.

Pozostaje problem **sposobu wyliczania** procentów – jak widać, na wykresie pokazane są procenty względem całej zbiorowości, a powinny się one sumować do 100 na każdym wykresie. Aby zmienić sposób wyświetlania procentów, znajdujemy zakładkę *Histogram* i ustawiamy *Procent na wykresach skategoryzowanych* jako *Kategoria*.

