

# STATYSTYKA OPISOWA (lab. 7)

## STATYSTYKI OPISOWE W GRUPACH

W programie *STATISTICA* obliczenie **statystyk w grupach** nie jest o wiele trudniejsze niż przeprowadzenie takiej analizy dla całego zbioru danych (za pomocą analizy *Statystyki opisowe*). Wystarczy wybrać odpowiednie narzędzie i poza zmiennymi liczbowymi, dla których będą liczone statystyki wskazać dodatkowo zmienną grupującą.

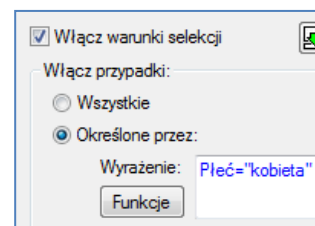
### Przykład 1 (Opinie o integracji z UE 2004)


W roku 2003<sup>1</sup> odbyło się referendum w sprawie wstąpienia Polski do Unii Europejskiej. Baza danych zawiera informacje o sposobie głosowania i innych wybranych cechach grupy respondentów z woj. podkarpackiego. Celem analizy jest zweryfikowanie hipotezy o wpływie wieku na sposób głosowania w referendum akcesyjnym. Do analizy wykorzystana zostanie metoda *Przekroje (prosta ANOVA)* z grupy *Statystyki podstawowe i tabele*. Proszę wyznaczyć wskazane w tabeli statystyki opisowe *Wieku* względem *Sposobu głosowania*.

Sposób głosowania w referendum europejskim	Wiek [w latach]			
	<i>N</i>	$\bar{x}$	Me	<i>s</i>
tak				
nie				

Proszę wrócić do okna analizy (Ctrl + R) i znaleźć w jednej z zakładki *Wykres interakcji*. Korzystając z wyników zamieszczonych w tabeli i z wykresu proszę ocenić, czy uzasadnione jest stwierdzenie, że osoby relatywnie starsze były bardziej sceptycznie w kwestii wstąpienia Polski do UE?

**UWAGA!** W programie *STATISTICA* każdą analizę, którą wykonujemy na całym pliku danych można też w prosty sposób zawęzić do wybranego podzbioru. W tym celu należy zastosować narzędzie *Selekcji przypadków*. Wykonamy jeszcze raz analizę zależności sposobu głosowania od wieku, lecz tylko wśród ankietowanych kobiet. Wznawiamy analizę (Ctrl + R), cofamy się za pomocą *Anuluj* do momentu



wyboru zmiennych i odnajdujemy przycisk: . W oknie selekcji przypadków wpisujemy warunek w sposób pokazany na rysunku obok.

Proszę podać średni wiek kobiet głosujących za wstąpieniem do UE ..... i kobiet głosujących przeciwko wstąpieniu do UE ..... W analogiczny sposób proszę wykonać analizę dla tych samych cech w grupie mężczyzn.

### Przykład 2 (Bezrobocie w gminach Podkarpacia 2003-2015)

Celem analizy będzie porównanie poziomu *Bezrobocia ogółem* w latach 2003-2015 w gminach **miejskich**, **wiejskich** oraz **miejsko-wiejskich**. Najpierw jednak musimy stworzyć zmienną, która określać będzie typ gminy.

- **Wyodrębnianie zmiennej z typem gminy:** Na końcu arkusza proszę dodać nową zmienną i nazwać ją *Rodzaj gminy*. Kod gminy zawarty jest w 1 kolumnie arkusza danych i jest to ostatnia cyfra w 10-cyfrowym kodzie jednostki terytorialnej. Aby wyodrębnić tę wartość posłużymy się formułą: =ToNumber(Mid(v1;10;1)). Proszę wprowadzić tę formułę do nowoutworzonej zmiennej. Po wyliczeniu wartości proszę nadać im *Etykiety tekstowe* według następującej zasady: **1 – miejska, 2 – wiejska, 3 – miejsko-wiejska**. Następnie proszę podać ile jest na Podkarpaciu gmin miejskich \_\_\_\_\_, wiejskich \_\_\_\_\_ i miejsko-wiejskich \_\_\_\_\_.

*Formuła Mid(tekst, z1, z2) wyodrębnia z zadanego tekstu z2 znaków począwszy od znaku z1. Formuła ToNumber() powoduje, iż program traktuje wynik operacji tekstowej jako liczbę.*

- **Obliczenie statystyk opisowych w grupach:** Za pomocą poleceń *Statystyka / Statystyki podstawowe i tabele / Przekroje, prosta ANOVA* otwieramy okno wyboru zmiennych w analizie *Statystyki w grupach* i wybieramy zmienne: na liście zmiennych zależnych wskazujemy wszystkie wskaźniki bezrobocia **ogółem**, zaś na liście zmiennych grupujących **Rodzaj gminy**. Zatwierdzamy wybór i przechodzimy (**OK**) do okna *Wyników statystyk w grupach*, gdzie w zakładce *Statystyki opisowe* określamy, które miary mają być zamieszczone w tabeli wynikowej. Wartość średnia jest wyliczana obligatoryjnie, więc nie ma jej na liście statystyk. Ponieważ w naszej analizie ograniczymy się tylko do wartości średnich proszę odznaczyć wszystkie pozostałe statystyki. Proszę uzupełnić tabelę i ocenić, czy rodzaj gminy różnicuje poziom bezrobocia.

Rodzaj gminy	Średni poziom bezrobocia <b>ogółem</b> w gminach				
	2003	2006	2009	2012	2015
miejska					
wiejska					
miejsko-wiejska					

<sup>1</sup> Dokładnie 7 i 8 czerwca 2003 r., a więc mniej więcej na rok przed późniejszym wejściem Polski do UE. W referendum za akcesją opowiedziało się ok. 77% głosujących.

# STATYSTYKA OPISOWA (lab. 7)

## STATYSTYKI OPISOWE W GRUPACH

### Przykład 3 (Wybrane wskaźniki (Europa))

Celem analizy będzie porównanie poziomu rozwoju społeczeństw **postkomunistycznych** do pozostałych państw europejskich. Analiza będzie miała charakter ogólny, porównamy pewne statystyki opisowe wybranych wskaźników rozwoju społecznego i gospodarczego w grupie państw postkomunistycznych i kapitalistycznych<sup>2</sup>. Ponieważ jedna z porównywanych cech ma charakter nominalny a pozostałe liczbowy, do analizy wykorzystana zostanie metoda **Przekroje (prosta ANOVA)** z grupy **Statystyki podstawowe i tabele**.

Po otwarciu okna tej analizy należy dokonać wyboru zmiennych.

**UWAGA:** Proszę zwrócić uwagę, iż na drugiej liście zmiennych wybieramy cechę nominalną (grupującą) – w naszym przypadku zmienną **Historia**, zaś na pierwszej liście cechę (lub cechy) o charakterze liczbowych – lista tych cech znajduje się w poniższej tabeli.

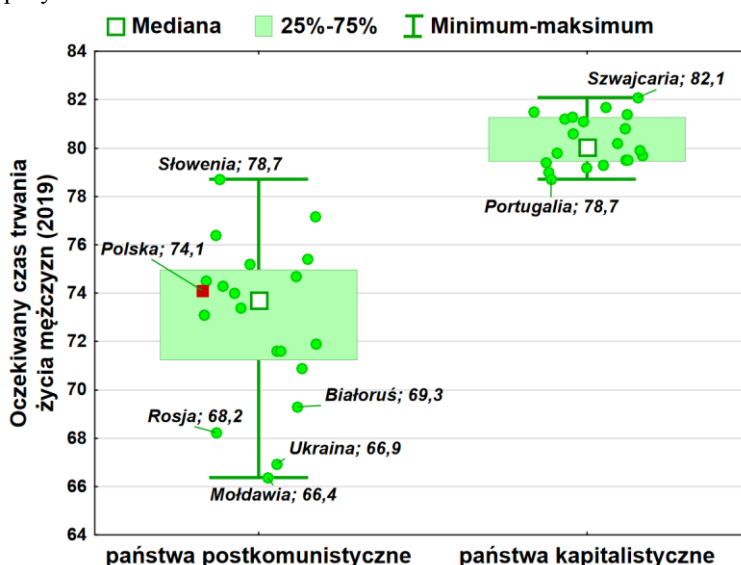
Po zatwierdzeniu wyboru zmiennych, przechodzi do okna **Statystyki w grupach (wyniki)**. W tym oknie w zakładce **Statystyki opisowe** wybieramy takie miary, by móc uzupełnić poniższą tabelę. Wyniki proszę sformatować, a wybrane statystyki umieścić w poniższej tabeli.

Wybrane miary poziomu rozwoju społecznego i gospodarczego	Państwa					
	postkomunistyczne			kapitalistyczne		
	Me	min	max	Me	min	max
Oczekiwany czas trwania życia mężczyzn (2019)						
PKB per capita (2019) \$PPP						
Dzietność (2019)						

Na podstawie uzyskanych wielkości proszę opisać wyniki, zwracając szczególną uwagę na kwestię porównania obu grup krajów .....

Proszę wznowić analizę (Ctrl + R) i wykonać ilustrację graficzną wyników za pomocą wykresu **skategoryzowanego ramka-wąsy**, który można znaleźć w zakładce **Statystyki opisowe** (proszę wykonać wykresy, na których pokazany zostanie poziom **mediany**, **kwartyli** oraz **minimum** i **maksimum** dla każdego wskaźnika w grupie państw postkomunistycznych i pozostałych).

Na poniższym, przykładowym wykresie widzimy doskonale, że poziom przeciętnego czasu trwania życia w „najlepszym” z państw postkomunistycznych w 2019 roku (niemal 30 lat po tzw. transformacji) był taki sam jak w „najgorszym” z państw kapitalistycznych. Poza tym widoczne jest ogromne zróżnicowanie w obrębie grupy państw postkomunistycznych – w „najgorszym” państwie postkomunistycznym mężczyźni żyją, średnio rzecz biorąc, o niemal 12 lat krócej niż w „najlepszym”.



Oczywiście powyższy wykres został przed zaprezentowaniem go „odbiorcy” odpowiednio sformatowany – proszę spróbować sformatować, choć jeden z wykonanych wykresów, według powyższego wzorca.

<sup>2</sup> Oczywiście w tej chwili nie ma w Europie państw komunistycznych, przynajmniej z formalnego punktu widzenia, wszędzie odbywają się wybory i panuje demokracja. Wykorzystywany w zadaniu podział dotyczy więc *de facto*, kwestii historycznych, ale jak się okaże, ciągle mocno różnicujących poziom życia w Europie...