

STATYSTYKA OPISOWA (lab. 11)

INTERNETOWE ŹRÓDŁA DANYCH • PRZYGOTOWANIE ARKUSZA • INDEKSY DYNAMIKI

Przykład 1 (Ludność państwa świata 1950-2050)

Celem analizy będzie porównanie zmian w poziomie ludności w wybranych krajach świata, przy czym do porównania wybrano kraje rywalizujące ze sobą na płaszczyźnie politycznej i gospodarczej. Przed rozpoczęciem analizy, danym należy nadać układ czasowy (tak, żeby lata stanowiły przypadki arkusza danych), co należy wykonać za pomocą poleceń *Dane / Transponuj / Plik*. Następnie proszę wykonać *Wykres liniowy (zmiennych)* w *wersji wielokrotnej*, który umożliwi porównanie zmian ludności w następujących krajach – dla każdego punktu oddzielny wykres:

- 1) Niemcy, Turcja, Egipt i Iran;
- 2) Indie i Chiny;
- 3) Polska, Ukraina i Niemcy;
- 4) Japonia i Meksyk.

Na wszystkich wykresach proszę uwzględnić następujące elementy formatowania:

- na osi pionowej ustawić początek skali od 0;
- na osi poziomej wyświetlić etykiety co 5 lat – w położeniu prostokątnym do osi;
- osi pionowej nadać tytuł „Ludność (w mln)”.

Analizując zjawiska przedstawione na wykresach proszę zwrócić szczególną uwagę na prognozy dotyczące zmian ludnościowych w kolejnych latach (aż do 2050 r.).

Przykład 2 (Dane ze strony GUS – stat.gov.pl)

1) POZYSKANIE DANYCH O RUCHU TURYSTYCZNYM W WOJEWÓDZTWACH

Na głównej GUS stronie proszę znaleźć *Bank Danych Lokalnych*, a następnie *Dane według dziedzin*:

- znajdujemy dziedzinę *Turystyka / Turystyczne obiekty noclegowe / Wykorzystanie turystycznych obiektów noclegowych według miesięcy* po czym naciskamy przycisk *Dalej*;
- zaznaczamy *lata* – **wszystkie po czym odznaczamy rok 2026** (z uwagi na niepełne dane) i *miesiące* – **wszystkie, rodzaje obiektów** – **turystyczne obiekty noclegowe ogółem, zakres przedmiotowy** – **turyści (korzystający) ogółem** i ponownie naciskamy przycisk *Dalej*;
- w kolejnym oknie zaznaczamy **Polskę i wszystkie województwa** i dodajemy zaznaczone regiony do *Wybranych* po czym znowu naciskamy przycisk *Dalej*;
- uzyskujemy podgląd danych;
- teraz trzeba je pobrać w odpowiedniej formie – w tym celu znajdujemy polecenie i wskazujemy opcję *XLS – tablica relacyjna (zip)* po czym zapisujemy ten plik na dysku.

2) PRZYGOTOWANIE DANYCH DO ANALIZY

Zakończyliśmy pobieranie danych, zamykamy stronę internetową i przystępujemy do przygotowania danych do analizy, naszym celem będzie utworzenie dwóch plików: z danymi **miesięcznymi** dla Polski i województw oraz z danymi **rocznymi**.

- otwieramy ściągnięty ze strony GUS arkusz *Excels* i wykonujemy następujące czynności: zostawiamy tylko kolumny *Nazwa, Miesiące, Rok* oraz *Wartość* – pozostałe kolumny **usuwamy**;
- zaznaczamy cały arkusz i kopiujemy, otwieramy program *STATISTICA* i tworzymy nowy arkusz, a następnie wklejamy dane poleceniem *Wklej z nagłówkami / Z nazwami zmiennych* i zapisujemy arkusz pod nazwą *Turystyka*;
- przekształcimy teraz arkusz do danych rocznych – w tym celu wykorzystamy polecenie *Dane / Rozrzucić po zmiennych...*, gdzie jako zmienną kodującą kolumny wybieramy *Nazwa*, zmienną kodującą wiersze – *Rok*, a zmienną z wartościami – *Wartość*. Po wyborze zmiennych ustawiamy jeszcze opcję *Jeśli wiele wierszy ma ten sam kod, to jako Oblicz sumę* i wywołujemy arkusz wynikowy, w którym kopiujemy wartości ze zmiennej *Rok* do nazw przypadków, usuwamy zmienną *Rok* i zapisujemy pod nazwą *Turystyka (dane roczne)* i na razie go zamykamy;
- wracamy do arkusza *Turystyka* i w podobny sposób jak opisano powyżej przekształcimy go do układu czasowego dla danych miesięcznych, jednak wcześniej proszę **usunąć etykiety tekstowe ze zmiennej Miesiące** – teraz zastosujemy polecenie *Dane / Rozrzucić po zmiennych...*, tak samo jak w poprzednim punkcie, tylko że na liście zmiennych kodujących wiersze wybierzemy *Rok* i *Miesiące* (w tej kolejności!) i nie będziemy zmieniać ustawień w opcji *Jeśli wiele wierszy ma ten sam kod*. Wygenerowany arkusz zapisujemy pod nazwą *Turystyka (dane miesięczne)*;
- dokonamy jeszcze pewnych modyfikacji w tym arkuszu: proszę dodać nową zmienną po zmiennej *Rok* i nazwać ją *DATA*, a następnie zaznaczyć ją i za pomocą przycisku *Zmienne* znaleźć polecenie *Operacje na danych*, gdzie należy wskazać numery zmiennej *Miesiąc* i *Rok*, po czym przyciskiem *OK* utworzyć wartości *DATY*;
- zmienną *DATA* proszę sformatować stosując format użytkownika *mmmm yyyy*;
- zapisujemy ostateczną wersję arkusza

3) PODSUMOWANIE

W tym momencie zakończone zostało przygotowywanie danych do analizy – w następnych przykładach odwoływać będziemy się do arkuszy *Turystyka (dane roczne)* oraz *Turystyka (dane miesięczne)*.

Widać, iż proces przygotowania danych do analizy jest dość złożony, choć składa się z sekwencji prostych czynności.

STATYSTYKA OPISOWA (lab. 11)

INTERNETOWE ŹRÓDŁA DANYCH • PRZYGOTOWANIE ARKUSZA • INDEKSY DYNAMIKI

Przykład 3 (Turystyka (dane roczne))

Celem analizy będzie opisanie dynamiki ruchu turystycznego w całej Polsce oraz w woj. podkarpackim, mazowieckim i pomorskim w latach 2009-2025. Szczególnie ważne będzie porównanie spadku ruchu turystycznego wskutek tak zwanych „lockdownów” w trakcie kryzysu w latach 2020-21 (w 2022 r. sektor turystyczny działał już w miarę normalnie). Wyznamy indeksy dynamiki o stałej i zmiennej podstawie – obliczenia będą wykonane w programie Excel.

- W programie STATISTICA za pomocą polecenia **Dane / Podzbiór** tworzymy arkusz zawierający dane o ruchu turystycznym w całej Polsce oraz w woj. podkarpackim, mazowieckim i pomorskim.
- Zaznaczamy ten arkusz i kopiujemy za pomocą polecenia **Edycja / Kopiuj z nagłówkami**.
- Otwieramy nowy arkusz w Excelu i wklejamy dane, a następnie usuwamy pierwszy, pusty wiersz z arkusza.
- Kopiujemy nazwy regionów i wklejamy dwukrotnie obok, tworząc miejsce na wyliczenie indeksów o stałej i zmiennej podstawie.
- Aby wyznaczyć indeksy dynamiki **o stałej podstawie** dzielimy wartość dla Polski z roku 2009 przez tę samą wartość i mnożymy przez 100, a następnie blokujemy drugi wiersz w mianowniku tego wyrażenia ($=B2/B\$2*100$), formułę tę przeciągamy w dół i w prawo, a następnie formatujemy wyniki z dokładnością do 1-go miejsca po przecinku, a w celu ułatwienia interpretacji stosujemy **Skale kolorów** dostępne za pomocą **Formatowania warunkowego**.
- Indeksy dynamiki o zmiennej podstawie wyznaczamy pomijając pierwszy rok, zaś sama formuła będzie identyczna, tylko bez znaku „\$”.
- Wyniki przedstawiono poniżej.

ROK					Indeksy dynamiki o stałej podstawie				Indeksy dynamiki o zmiennej podstawie			
	POLSKA	PODKARPACKIE	MAZOWIECKIE	POMORSKIE	POLSKA	PODKARPACKIE	MAZOWIECKIE	POMORSKIE	POLSKA	PODKARPACKIE	MAZOWIECKIE	POMORSKIE
2009	19 353 712	654 908	2 786 608	1 610 730	100,0	100,0	100,0	100,0	105,7	x	x	x
2010	20 461 496	697 000	3 187 278	1 628 830	105,7	106,4	114,4	101,1	105,7	106,4	114,4	101,1
2011	21 476 616	718 911	3 311 910	1 661 540	111,0	109,8	118,9	103,2	105,0	103,1	103,9	102,0
2012	22 635 388	730 177	3 375 976	1 872 589	117,0	111,5	121,2	116,3	105,4	101,6	101,9	112,7
2013	23 401 138	782 127	3 648 184	1 966 831	120,9	119,4	130,9	122,1	103,4	107,1	108,1	105,0
2014	25 083 978	873 470	3 867 197	2 197 688	129,6	133,4	138,8	136,4	107,2	111,7	106,0	111,7
2015	26 942 056	957 687	4 106 961	2 439 185	139,2	146,2	147,4	151,4	107,4	109,6	106,2	111,0
2016	30 108 308	1 110 605	4 737 844	2 672 714	155,6	169,6	170,0	165,9	111,8	116,0	115,4	109,6
2017	31 989 344	1 226 038	5 091 968	2 863 071	165,3	187,2	182,7	177,7	106,2	110,4	107,5	107,1
2018	33 895 930	1 263 445	5 336 963	3 047 790	175,1	192,9	191,5	189,2	106,0	103,1	104,8	106,5
2019	35 668 091	1 297 039	5 536 187	3 238 675	184,3	198,0	198,7	201,1	105,2	102,7	103,7	106,3
2020	17 878 969	710 332	2 116 493	1 972 559	92,4	108,5	76,0	122,5	50,1	54,8	38,2	60,9
2021	22 198 972	803 938	3 027 315	2 370 937	114,7	122,8	108,6	147,2	124,2	113,2	143,0	120,2
2022	34 249 004	1 209 430	5 893 738	3 289 820	177,0	184,7	211,5	204,2	154,3	150,4	194,7	138,8

Indeksy dynamiki – ważne uwagi:

- 1) Za pomocą wskaźników dynamiki można także porównać dane czasowe dotyczące zmiennych o zupełnie odmiennym charakterze – na przykład można na jednym wykresie porównać zmiany w czasie w Polsce następujących wielkości: poziomu wynagrodzeń, cen benzyny, stopy bezrobocia, inflacji, przestępczości, etc.
- 2) Istnieje możliwość wykorzystania formuł programu STATISTICA do obliczenia wskaźników dynamiki, bez konieczności przenoszenia danych do arkusza programu EXCEL, ale jest to bardziej pracochłonne.
- 3) Z wartością indeksów dynamiki wiąże się ściśle pojęcie zmiany procentowej – na przykład:

Indeks dynamiki = 123 to wzrost o 23% Indeks dynamiki = 84 oznacza spadek o 16%

Przy zmianach procentowych należy **koniecznie** podać nie tylko wartość, ale także kierunek (wzrost czy spadek?).

Arkusz Excela, w którym wyznaczono indeksy dynamiki proszę zapisać pod nazwą *Turystyka (dynamika)*, gdyż będzie on jeszcze wykorzystywany na następnych zajęciach.

Na podstawie wyliczonych indeksów dynamiki proszę odpowiedzieć na następujące pytania:

- W którym z województw pomiędzy 2009 i 2019 rokiem nastąpił największy wzrost ruchu turystycznego – o ile procent wzrósł ruch turystyczny w tym województwie
- W którym z województw zanotowano największy kryzys w branży turystycznej w 2020 r. w stosunku do roku 2019 i o ile procent spadł wtedy ruch turystyczny.....?
- W którym województwie spadek ruchu turystycznego w 2020 r. względem 2019 r. był najmniejszy i ile wyniósł procent
- Analiza indeksów dynamiki o zmiennej podstawie, wyznaczanych w układzie **rok do roku** może prowadzić do mylnych wniosków, że rok 2021 czy 2022 r. był pozytywny dla branży turystycznej. Proszę wyznaczyć indeksy dynamiki dla lat 2020-25 względem stałej podstawy, którą będzie rok 2019 i określić procentowy spadek ruchu turystycznego w 2021 r. względem roku 2019: *podkarpackie*:, *mazowieckie*:, *pomorskie*:
- Czy w którymś województwie ruch turystyczny w 2022 r. nie przekroczył poziomu z 2019 r.? Jeżeli tak, to proszę podać to województwo i procent, o jaki ruch turystyczny spadł w 2022 względem 2019 r.?