

# Statystyka opisowa

## Wykład 6:

### Analiza danych czasowych

Wykresy, indeksy dynamiki,  
analiza sezonowości

# Definicja szeregu czasowego

Szeregiem czasowym będziemy nazywać ciąg obserwacji dowolnej cechy statystycznej, dokonanych w kolejnych momentach czasowych.

Wartości szeregu będziemy oznaczać symbolem:

$$Y_t \quad (t = 1, \dots, T)$$

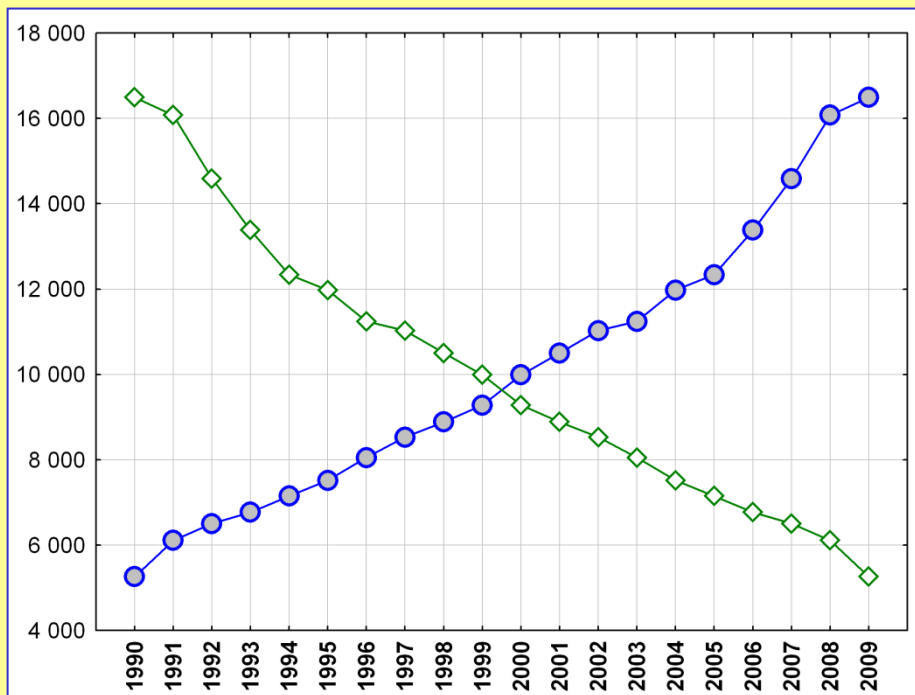
Tak zdefiniowany szereg czasowy ma charakter jednowymiarowy i obecnie zajmujemy się specjalną grupą metod statystycznych, które służą do porównywania a także prognozowania kolejnych jego wartości.

Oczywiście, do badania danych czasowych można też stosować poznane uprzednio metody tabelarycznej i graficznej analizy danych, choć jak pokazano na następnym slajdzie nie zawsze ma to sens.

# Specyfika analizy danych czasowych

Sposób doboru metod analizy statystycznych zawsze powinien być związany z charakterem zbioru danych oraz ich merytorycznym znaczeniem. Ten sam zbiór liczb musi być inaczej traktowany, gdy dotyczy danych czasowych, a inaczej, gdy dotyczy danych przekrojowych lub ankietowych.

Na wcześniejszych wykładach, jako najbardziej popularną metodę opisu danych liczbowych wskazano wyznaczanie statystyk opisowych, poniższy przykład pokazuje, że dla danych czasowych nie zawsze ma to sens.



Wykres przedstawia dynamikę dwóch zjawisk – „nieco” różniących się kierunkiem zmian!

Tymczasem statystyki opisowe dla obu szeregów czasowych są identyczne:

~~Średnia = 10085~~

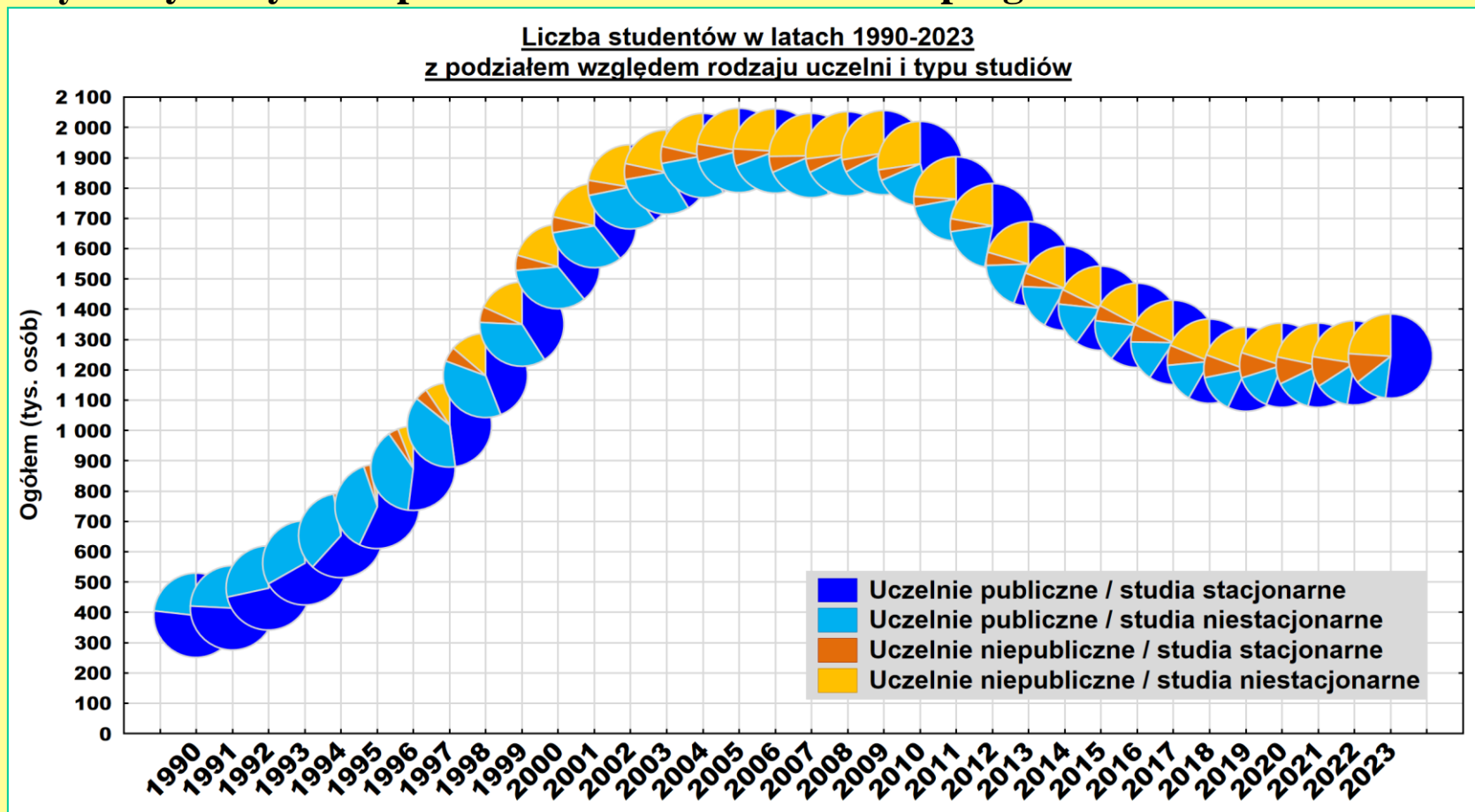
~~Minimum = 5261~~

~~Maksimum = 16495~~

# Wykresy danych czasowych (1)

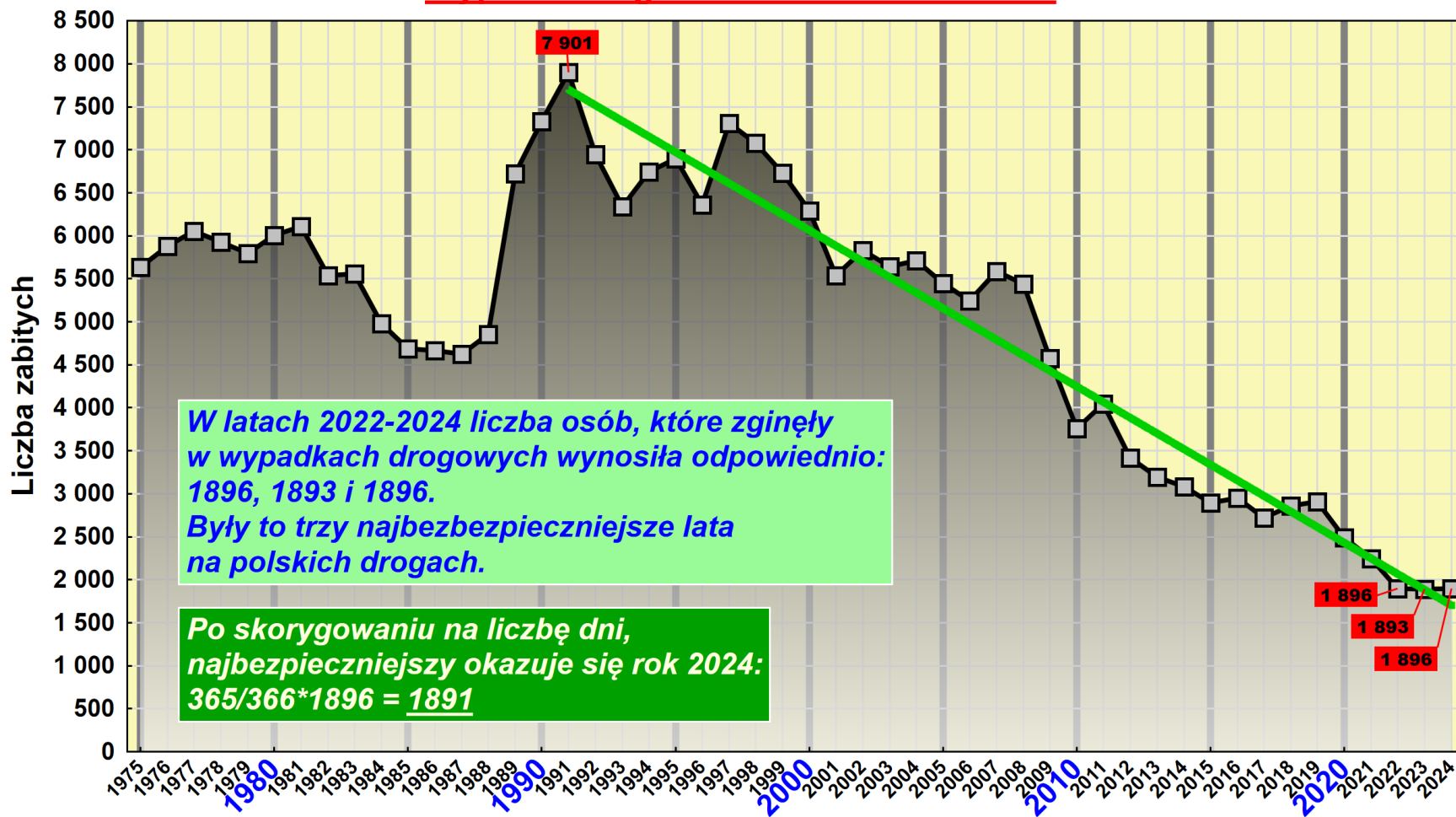
Znaczenie analizy graficznej danych czasowych jest ogromne – ponieważ zostało to już uzasadnione na poprzednim slajdzie. Metody graficznej analizy danych zostały omówione na wykładzie nr 3.

Poniżej zaprezentowano, dla przypomnienia, kilka form graficznej prezentacji danych, w których wykorzystano pewne dodatkowe możliwości programu *STATISTICA*.



# Wykresy danych czasowych (2)

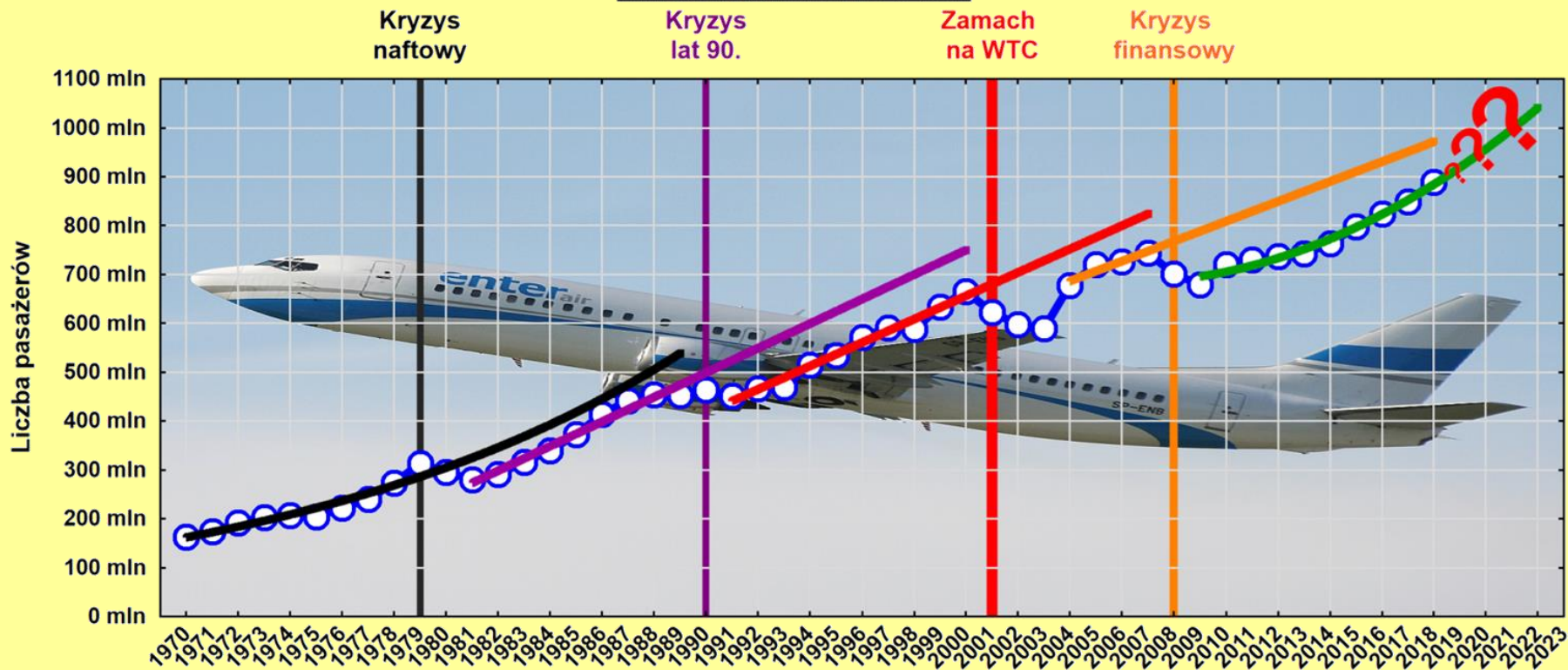
## Wypadki drogowe w Polsce 1975-2024



Źródło: policja.pl (do 2023), obserwatoriumbrd.pl (2024)

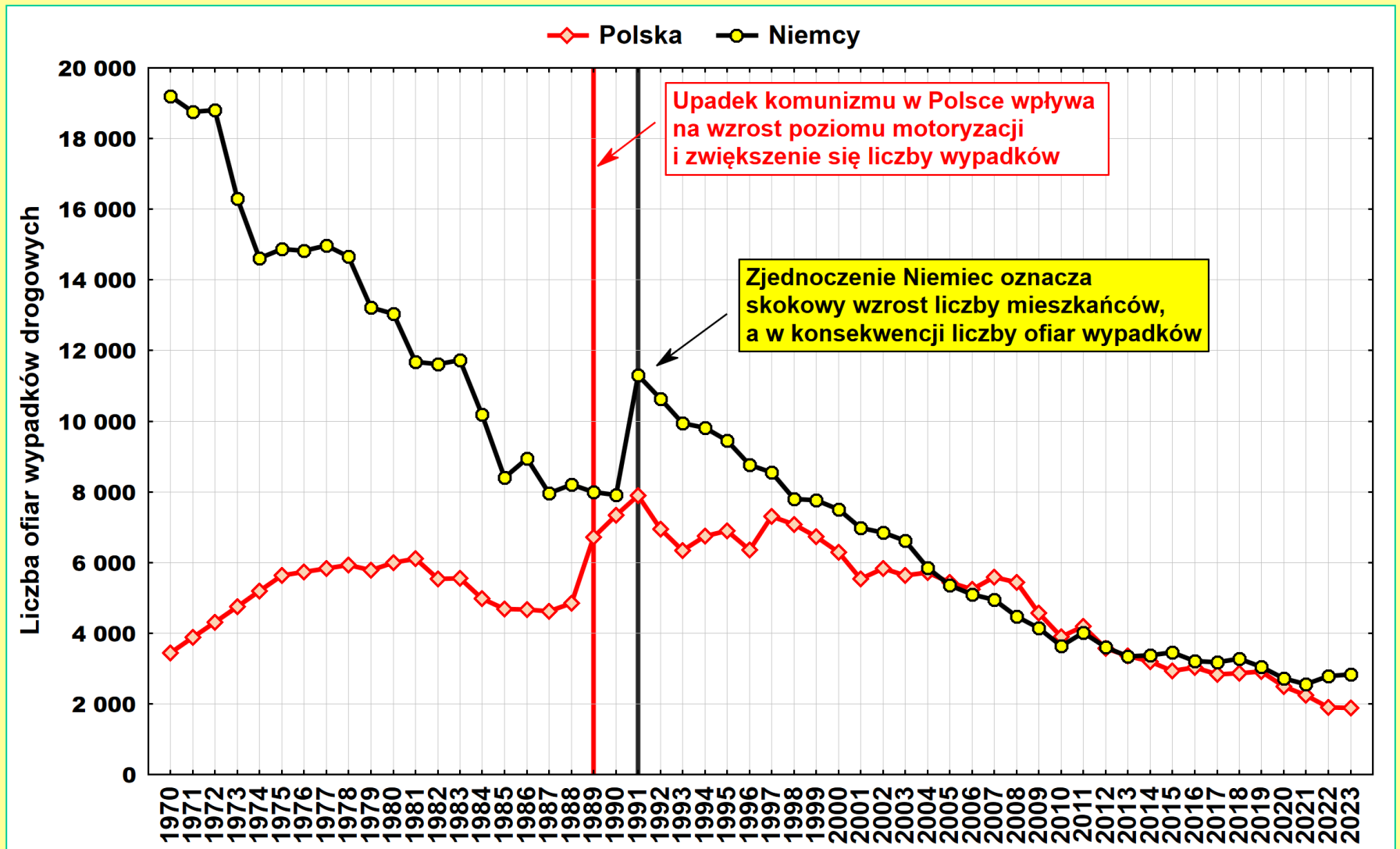
# Wykresy danych czasowych (3)

Wpływ kryzysów na przewozy pasażerskie transportem lotniczym w USA



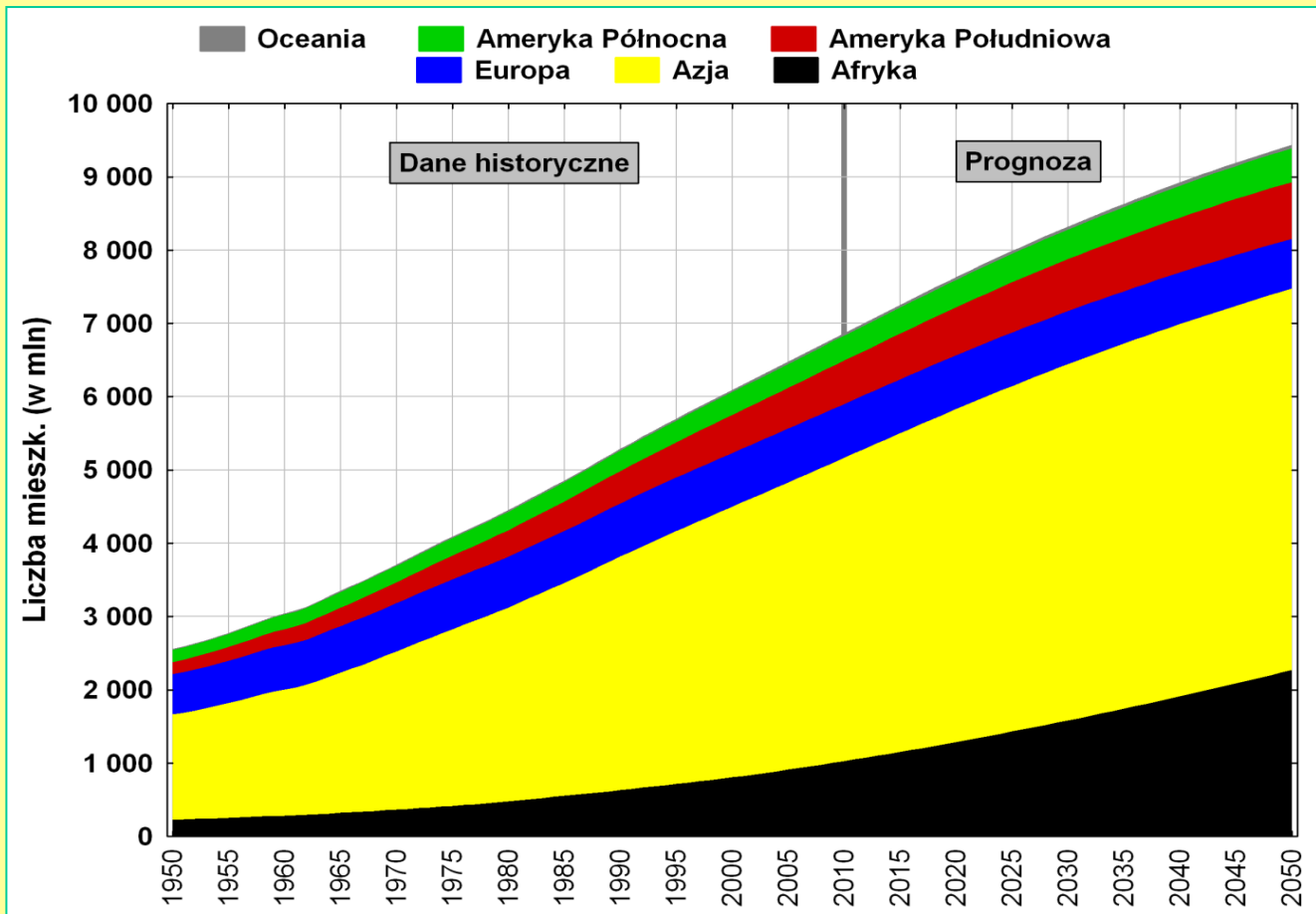
Źródło: [www.transtats.bts.gov](http://www.transtats.bts.gov)

# Wykresy danych czasowych (4)



# Wykresy danych czasowych (5)

Ludność zamieszkująca na poszczególnych kontynentach w latach 1950-2010 i prognoza na lata 2011-2050



Wykres  
warstwowy

Źródło: [www.census.gov](http://www.census.gov)

# Indeksy dynamiki (o stałej i zmiennej podstawie)

Proste indeksy dynamiki określają tempo zmian pojedynczego szeregu czasowego.

Wyodrębnia się dwa podstawowe typy indeksów:

- indeksy o stałej podstawie;
- indeksy o zmiennej podstawie.

Indeksy o stałej podstawie wyznacza się na podstawie wzoru:

Indeks (dla okresu  $t$ ) = Wartość w okresie  $t$  / Wartość w okresie bazowym

$$I_{t/T} = Y_t / Y_T * 100$$

Indeksy o zmiennej podstawie wyznacza się na podstawie wzoru:

Indeks (dla okresu  $t$ ) = Wartość w okresie  $t$  / Wartość w okresie poprzednim

$$I_{t/t-1} = Y_t / Y_{t-1} * 100$$

# Indeksy złożone

Indeksy złożone (agregatowe) pozwalają ocenić tempo i kierunek zmian wartości kombinacji wielu zmiennych jednocześnie.

Do najczęściej stosowanych (ale nie tylko) indeksów złożonych należą formuły wiążące ilość towarów i ich ceny. Takie indeksy stosowane do pewnego „koszyka” dóbr służą między innymi do szacowania poziomu inflacji.

Jako przykład indeksu agregatowego omówiony zostanie indeks Laspeyresa – w indeksie tym odnosi się sumaryczną wartość koszyka dóbr w danym okresie do wartości tego samego koszyka dóbr w okresie bazowym.

$$\text{Indeks Laspeyresa} = \frac{\sum p_i q_0}{\sum p_0 q_0} \cdot 100$$

Oznaczenia:  $p_i$  – ceny poszczególnych dóbr,  $q_i$  – ilości poszczególnych dóbr

# Indeksy dynamiki (zastosowanie)

**Za pomocą indeksów dynamiki można:**

- porównać tempo zmian szeregów czasowych wyrażonych w zupełnie innych jednostkach (np. tempo zmian cen benzyny i zarobków Polaków);
- porównać tempo zmian tej samej cechy dla różnych obiektów (np. liczba ofiar wypadków drogowych w Polsce, Niemczech i we Francji);
- lepiej poznać wewnętrzną strukturę danych – np. stałe indeksy dynamiki w ujęciu rok do roku świadczą o tym, że mamy do czynienia z trendem wykładniczym.

# Przygotowanie arkusza danych w EXCELU

Przykład będzie dotyczył danych o liczbie wypadków drogowych w Polsce, we Francji, Niemczech i w Wielkiej Brytanii. Analizą obejmujemy jednak tylko dane od roku 2000.

## Krok 1 – wybór odpowiednich danych

Za pomocą polecenia *Dane / Podzbiór* tworzymy nowy arkusz zawierający dane z interesujących nas państw

## Krok 2 – kopiowanie danych z programu *STATISTICA* do *Excela*

Obliczenia łatwiej będzie wykonać w programie *Excel*. Zaznaczamy interesujący nas zakres danych (od 2000 roku) i kopiujemy za pomocą polecenie *Edycja / Kopiuj z nagłówkami*. Wklejamy dane w nowym arkuszu programu *Excel*.

# Indeksy dynamiki o stałej podstawie

Wyznaczanie indeksu o stałej podstawie polega na odniesieniu wartości z danego roku do wartości roku bazowego (a następnie pomnożeniu przez 100). Wartości indeksów względem 2000 roku przedstawiono poniżej.

Zastosowanie narzędzia *Formatowanie warunkowe* pozwala przedstawić dane w interesujący sposób.

	A	B	C	D	E	F	G	H	I
1		Francja	Niemcy	Polska	Wielka Brytania	Francja	Niemcy	Polska	Wielka Brytania
2	2000	8 079	7 503	6 294	3 580	100,0	100,0	100,0	100,0
3	2001	8 162	6 977	5 534	3 598	101,0	93,0	87,9	100,5
4	2002	7 655	6 842	5 826	3 581	94,8	91,2	92,6	100,0
5	2003	6 058	6 613	5 642	3 658	75,0	88,1	89,6	102,2
6	2004	5 530	5 842	5 712	3 368	68,4	77,9	90,8	94,1
7	2005	5 318	5 361	5 444	3 336	65,8	71,5	86,5	93,2
8	2006	4 709	5 091	5 243	3 298	58,3	67,9	83,3	92,1
9	2007	4 620	4 949	5 583	3 059	57,2	66,0	88,7	85,4
10	2008	4 275	4 477	5 437	2 645	52,9	59,7	86,4	73,9
11	2009	4 273	4 152	4 572	2 337	52,9	55,3	72,6	65,3
12	2010	3 992	3 648	3 908	1 905	49,4	48,6	62,1	53,2
13	2011	4 111	4 009	4 189	1 960	50,9	53,4	66,6	54,7

$=B2/B\$2*100$

Zastosowanie narzędzie *Formatowania warunkowego*, dostępnych w programie *Excel*, pozwala w ciekawy sposób przedstawić zmiany skutków wypadków drogowych w postaci „mapy kolorów”

## Przykładowe wnioski:

W latach 2000-2011 nastąpił spadek liczby ofiar wypadków drogowych we wszystkich krajach. Największy we Francji – niemal dwukrotny (dokładnie o 49,1%), podobnie w Niemczech i Wielkiej Brytanii. W Polsce liczba ofiar wypadków drogowych spadła w 2011 roku w stosunku do roku 2000 o jedną trzecią (33,4%).

# Indeksy dynamiki o zmiennej podstawie

Wyznaczanie indeksu o zmiennej podstawie polega na odniesieniu wartości z danego roku do wartości z roku poprzedniego (a następnie pomnożeniu przez 100). Wartości indeksów względem w ujęciu rok do roku przedstawiono poniżej.

	A	B	C	D	E	F	G	H	I
1		Francja	Niemcy	Polska	Wielka Brytania	Francja	Niemcy	Polska	Wielka Brytania
2	2000	8 079	7 503	6 294	3 580				
3	2001	8 162	6 977	5 534	3 598	101,0	93,0	87,9	100,5
4	2002	7 655	6 842	5 826	3 581	93,8	98,1	105,3	99,5
5	2003	6 058	6 613	5 642	3 658	79,1	96,7	96,8	102,2
6	2004	5 530	5 842	5 712	3 368	91,3	88,3	101,2	92,1
7	2005	5 318	5 361	5 444	3 336	96,2	91,8	95,3	99,0
8	2006	4 709	5 091	5 243	3 298	88,5	95,0	96,3	98,9
9	2007	4 620	4 949	5 583	3 059	98,1	97,2	106,5	92,8
10	2008	4 275	4 477	5 437	2 645	92,5	90,5	97,4	86,5
11	2009	4 273	4 152	4 572	2 337	100,0	92,7	84,1	88,4
12	2010	3 992	3 648	3 908	1 905	93,4	87,9	85,5	81,5
13	2011	4 111	4 009	4 189	1 960	103,0	109,9	107,2	102,9

$$=B3/B2*100$$

## Przykładowe wnioski:

Odejmując od indeksów dynamiki wartość 100 otrzymujemy zmiany procentowe.

Zmiany liczby ofiar wypadków drogowych w badanym okresie w Niemczech były bardzo systematyczne, dopiero w roku 2011 zanotowano wzrost w stosunku do roku 2010. W Polsce, okresy spadkowe, były przeplatane latami, w których liczba wypadków wzrastała.

# Prezentacja graficzna indeksów dynamiki

## Krok 1 – kopiowanie danych z *Excela* do programu *STATISTICA*

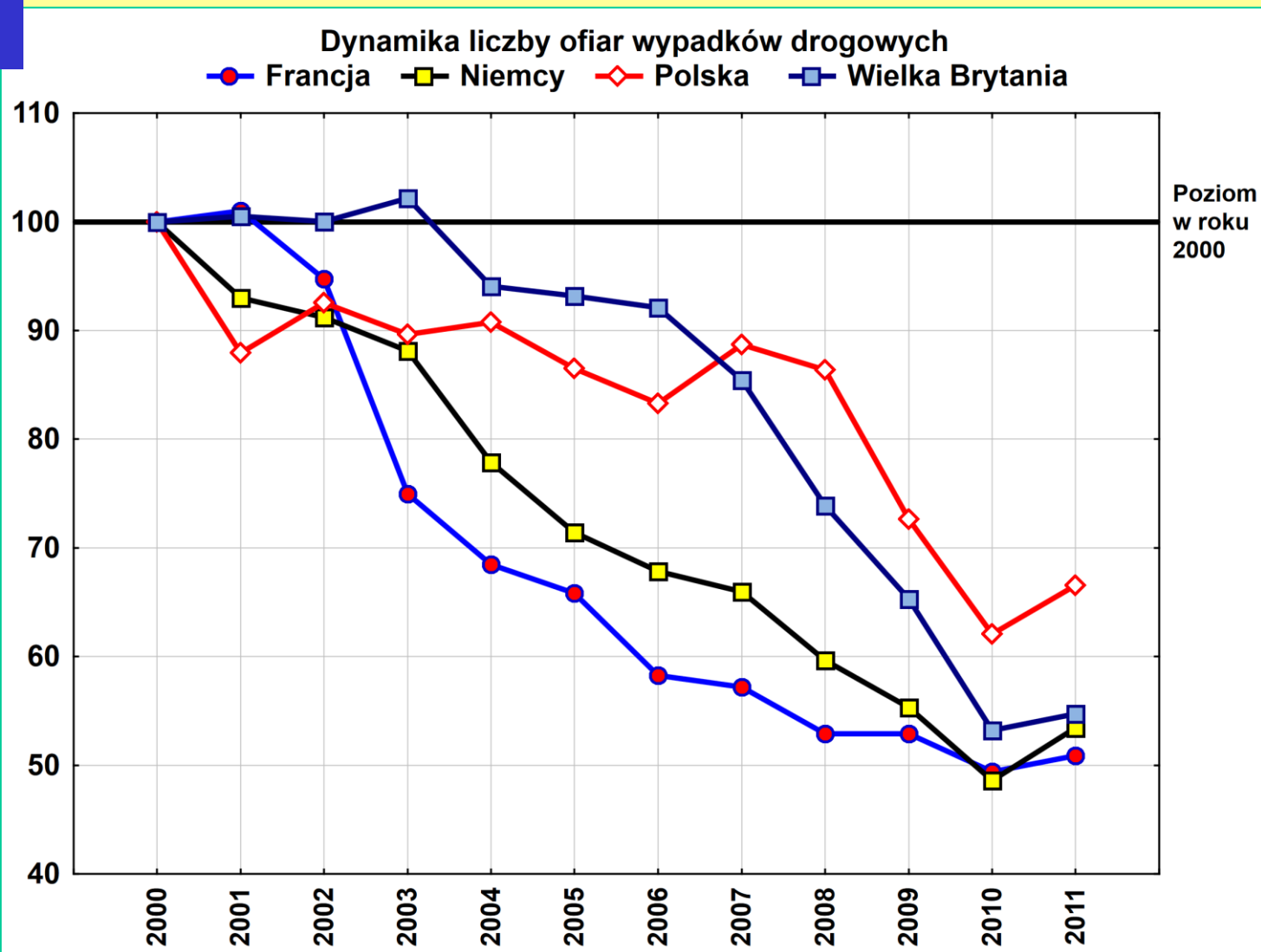
Po zaznaczeniu odpowiedniego obszaru danych w arkuszu *Excela* wklejamy je do programu *STATISTICA* wykorzystując polecenie *Edycja / Wklej z nagłówkami / ...* i w zależności od tego, czy dane posiadają nazwy przypadków i/lub zmiennych wskazujemy odpowiednią opcję.

## Krok 2 – wykorzystanie wykresów liniowych lub wykresów słupkowych (z poziomem odchylenia)

Wyniki można przedstawić w formie wielokrotnego wykresu liniowego (lepszą opcją dla indeksów o stałej podstawie) lub wykresu słupkowego (lepszą opcją dla indeksów o zmiennej podstawie)

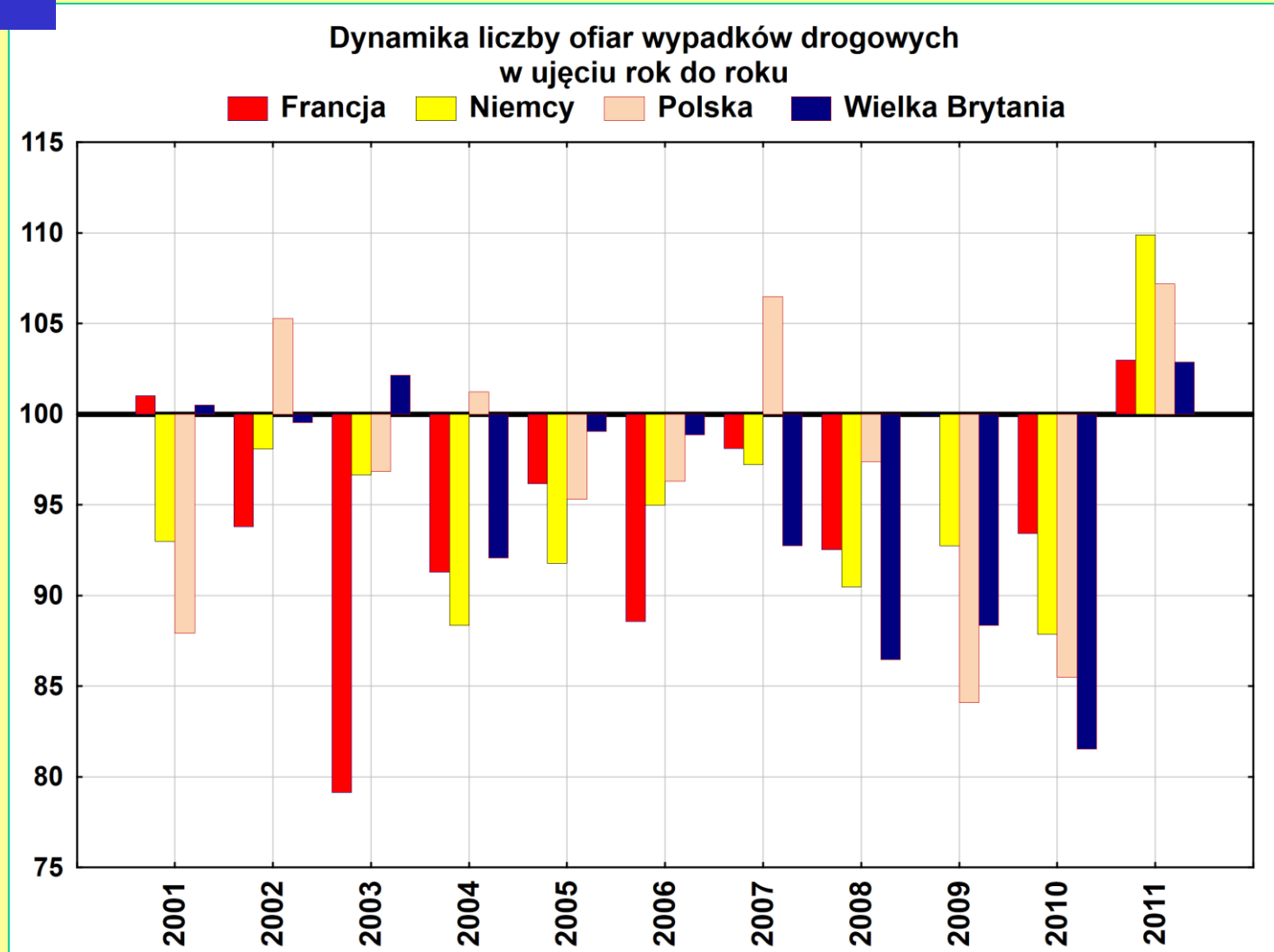
# Wykresy indeksów dynamiki (1)

Indeksy o stałej  
podstawie –  
wykres liniowy



# Wykresy indeksów dynamiki (2)

Indeksy o zmiennej podstawie – wykres słupkowy



# Czas **linearny** i czas **cykliczny**

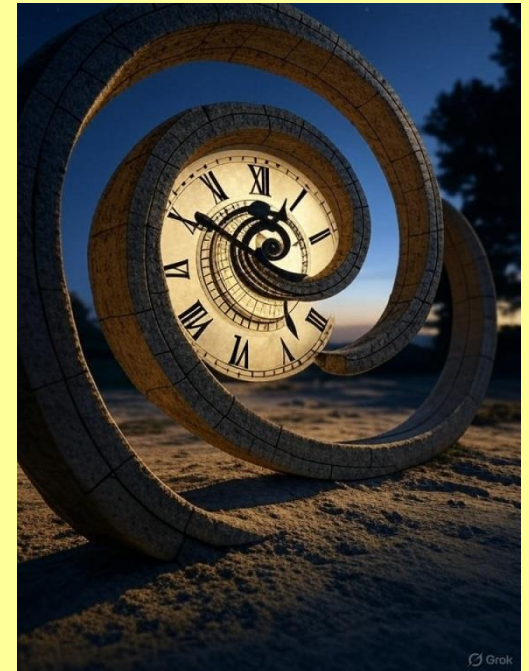
**ANALIZA DYNAMIKI** dotyczy tego aspektu naszego życia, który jest **niepowtarzalny, liniowy** – a jak mawiał filozof „nie wchodzi się dwa razy do tej samej rzeki”.

Czas **linearny** to czas zmian, rozwoju, a z punktu widzenia psychiki także **lęku przed przyszłością**.

Ale zdarzenia charakteryzują się także **cyklicznością**: pory dnia, pory roku to powtarzające się zjawiska naturalne, które dają nam poczucie **powtarzalności i bezpieczeństwa**.

Cywilizacja, kultura, a zwłaszcza religia akcentuje **zdarzenia cykliczne** (np. święta, wakacje, dni robocze i wolne, sezony piłkarskie...),  
co **łagodzi nasz lęk przed przyszłością**.

Do analizy cykliczności na podstawie danych,  
służy **ANALIZA SEZONOWOŚCI**.



# Znaczenie analizy sezonowości

Zrozumienie regularnej cykliczności, nazywanej sezonowością i umiejętność jej opisu jest bardzo ważna z punktu widzenia sprawnego ZARZĄDZANIA.

Nieuwzględnianie sezonowości w opisie zjawisk społecznych i gospodarczych może prowadzić do błędnych decyzji, może być też wykorzystywane do „kłamstwa statystycznego”.

Przykład dotyczy zarządzania BEZPIECZEŃSTWEM RUCHU DROGOWEGO (BRD).

Celem analizy będzie ocena zagrożenia wypadkami dla różnych uczestników ruchu drogowego (pieszych oraz kierowców i pasażerów).

# Wskaźniki sezonowości

Wskaźniki sezonowości mogą mieć charakter **addytywny** lub **multiplikatywny**.

Wskaźniki **addytywne** pozwalają na ocenienie o ile dla danego typu okresu wartości odchylają się od modelu bez efektu sezonowości (są to więc odchylenia bezwzględne, czyli +/-).

Wskaźniki **multiplikatywne** pokazują stosunek poziomu zjawiska w danym typie okresu do poziomu w modelu bez sezonowości, tak więc pozwalają ocenić „ile razy” w danym okresie badane zjawisko wzrasta (spada).

Wyliczanie wskaźników sezonowości jest skomplikowaną rachunkowo procedurą, dlatego warto korzystać ze specjalistycznego oprogramowania.

W programie *STATISTICA*, analizę sezonowości wykonuje się w następujący sposób:

- polecenie *STATYSTYKA / ZAAWANSOWANE MODELE LINIOWE I NIELINIOWE / SZEREGI CZASOWE I PROGNOZOWANIE / DEKOMPOZYCJA SEZONOWA*.
- wybieramy analizowaną zmienną, ustawiamy liczbę okresów w jednym cyklu sezonowości i wskazujemy rodzaj wskaźników.

# Wskaźniki sezonowości

Na rysunku przedstawiono **multiplikatywne** wskaźniki sezonowości dla liczby ofiar wypadków:

- wśród kierowców i pasażerów;
- wśród pieszych.

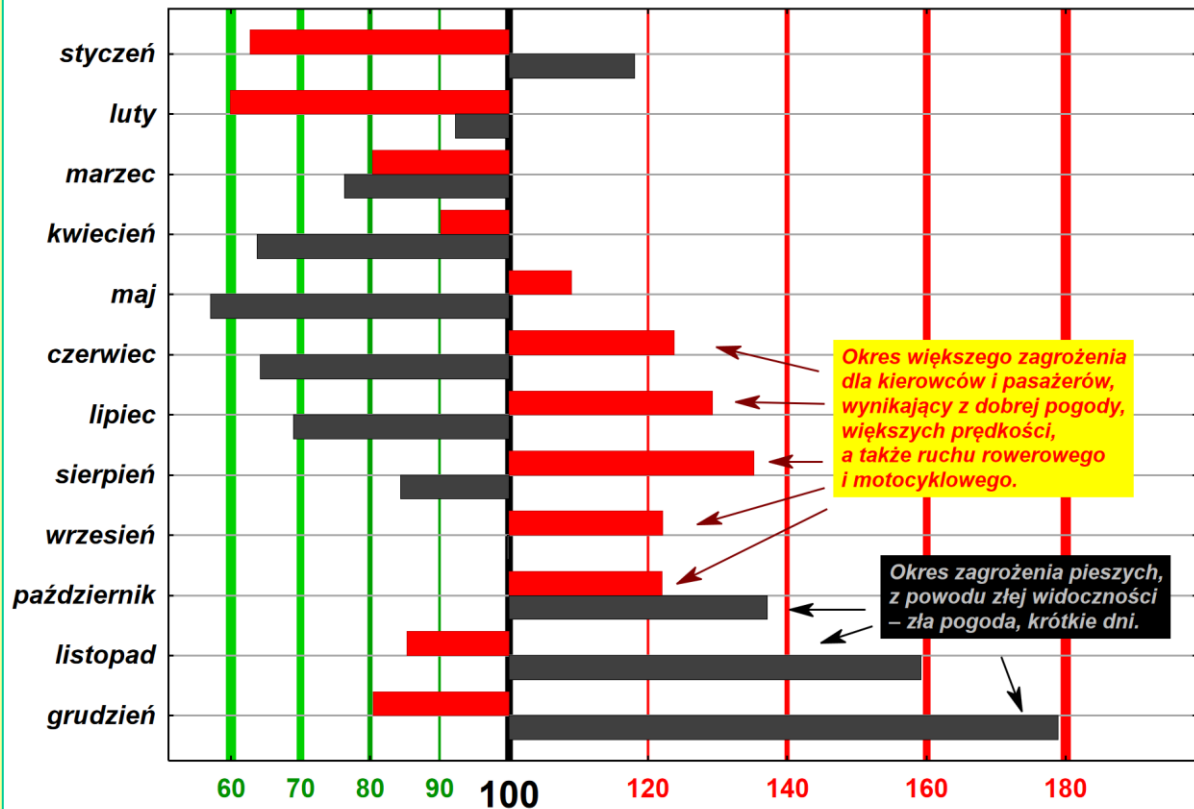
## Sezonowość liczby ofiar wypadków drogowych

■ KIERUJĄCY I PASAŻEROWIE ■ PIESI

Zagrożenie dla pieszych oraz pozostałych uczestników ruchu drogowego ma zupełnie inny typ sezonowości. **PIESI** są najbardziej narażeni w okresie jesienno-zimowym z kulminacją w grudniu – w miesiącu tym ginie ok. 3 razy więcej pieszych niż w maju czy czerwcu.

**KIEROWCY I PASAŻEROWIE** są najbardziej narażeni w okresie od maja do października, z kulminacją w sierpniu, kiedy to ginie ok. 2 razy więcej osób niż w styczniu czy lutym.

Całościowo, najbardziej niebezpiecznym miesiącem na drogach jest **październik**.



Wskaźniki sezonowości (poziom średnioroczny = 100)

Źródło: Obliczenia własne na podstawie danych z lat 2010-2023 udostępnianych na stronie [www.policja.pl](http://www.policja.pl)

# Interpretacja wyników analizy sezonowości liczby wypadków

Wskaźnik sezonow.
-832
-1090
-725
-384
225
436
552
602
454
630
53
78

Na podstawie **addytywnych** wskaźników sezonowości stwierdzić możemy, iż najczęściej wypadków – przeciętnie o 630 więcej niż poziom średnioroczny zdarza się w październiku, duża liczba wypadków występuje też w lipcu i sierpniu. Do najbezpieczniejszych miesięcy należą luty (średnio o ok. 1090 wypadków mniej niż przeciętna miesięczna) oraz styczeń i marzec.

W październiku jest, średnio rzecz biorąc, o ok. 1700 wypadków więcej niż w lutym.

Niemalże identyczne wnioski o charakterze sezonowości można wyciągnąć na podstawie analizy **multiplikatywnej**. Tyle, że otrzymujemy informację o wahaniach względnych. Dla przykładu, w październiku jest najczęściej wypadków, o ok. 16,6% wypadków więcej niż przeciętna miesięczna, a w lutym o ok. 30,2% mniej.

Jeżeli chcemy porównać dwa okresy – na przykład październik do lutego – wskaźniki należy przez siebie podzielić. W październiku jest zwykle  $116,6/69,8 = 1,67$  razy więcej wypadków niż w lutym (czy inaczej o 67% więcej) .

Wskaźnik sezonow.
76,6
69,8
79,9
89,6
106,1
112,1
115,1
116,3
112,7
116,6
101,4
103,8