

Prognozowanie w transporcie lotniczym (lab. 1)

PRZYGOTOWANIE DANYCH DO ANALIZY



Celem tych zajęć jest nauka umiejętności przygotowania arkusza danych w programie STATISTICA w sytuacji, gdy dane źródłowe są dostępne w arkuszu programu Excel.

Omówione zostaną sposoby przenoszenia danych z Excela do programu STATISTICA, opisywania arkusza danych, zmiany jednostek i innych czynności przygotowawczych do analizy.

Następnie przypomniane zostaną zasady graficznej prezentacji danych czasowych.

Przykład 1 (Plik Excela: *Ruch pasażerski USA 1990-2022 (M)*)

PRZYGOTOWANIE DANYCH W EXCELU

Proszę otworzyć w Excelu plik o nazwie *Ruch pasażerski USA 1990-2022 (M)*, dostępny na stronie domowej wykładowcy. Plik ten został pobrany ze strony www.transtats.bts.gov – amerykańskiego ministerstwa transportu (*United States Department of Transportation*). Plik ten, jak widać (fragment na rysunku obok), jest w formie nie pozwalającej na przeprowadzenie jakichkolwiek analiz statystycznych – w jednej kolumnie są w nim zawarte dane o miesiącu, roku i wielkości przewozów lotniczych w każdym okresie. Należy więc dokonać rozbicia danych z 1. kolumny do 3 oddzielnych kolumn. W tym celu należy wykonać następujące czynności:

- zaznaczamy całą kolumnę A;
- w zakładce *DANE* wybieramy polecenie *Tekst jako kolumny*, sprawdzamy czy zaznaczona jest opcja *Rozdzielany* i przechodzimy do kolejnego okna za pomocą przycisku *Dalej*;
- na liście „ograniczników” odznaczamy *Tabulator* wybierając *Przecinek*, po czym dokonujemy konwersji danych za pomocą przycisku *Zakończ*.

Uzyskany w ten sposób układ danych może być przedmiotem dalszych przekształceń, których dokonywać będziemy już w programie STATISTICA. Ale oczywiście także podjęcie jakichkolwiek analiz w Excelu wymagało takiego przeorganizowania danych.

SKOPIOWANIE DANYCH DO PROGRAMU STATISTICA

W pierwszej kolejności zajmiemy się poziomem przewozów w układzie miesięcznym. Takie dane pozwalają na przeprowadzenie analizy sezonowości, prognoz wybranymi metodami (np. trendów jednoimiennych okresów czy wyrównywania wykładniczego – metody te były omawiane na przedmiocie *Prognozowanie i symulacje*) i oczywiście przedstawienie ich w formie graficznej.

Aby skopiować dane z Excela do programu STATISTICA proszę zaznaczyć dane z kolumn A, B i C od wiersza 2 (a więc łącznie z nagłówkami kolumn) do końca danych miesięcznych.

Po skopiowaniu zaznaczonego w Excelu fragmentu arkusza otwieramy program STATISTICA, tworzymy nowy arkusz i wklejamy dane poleceniem *Edycja / Wklej z nagłówkami / Z nazwami zmiennych*.

PRZYGOTOWANIE ARKUSZA W PROGRAMIE STATISTICA

Teraz dokonamy następujących przekształceń arkusza danych:

- po pierwsze, zmieniamy nazwy zmiennych na: *Miesiąc*, *Rok*, *Pasażerskie przewozy lotnicze w USA*;
- po drugie, dane nie są ułożone chronologicznie, należy je więc posortować według dwóch zmiennych – jako pierwszą zmienną do sortowania wskazujemy *Rok*, jako drugą *Miesiąc*;
- z arkusza usuwamy zbędne, puste kolumny oraz ostatnie kilka pustych przypadków – ostatecznie dane powinny zawierać 3 kolumny i 392 przypadki;
- w kolejnym kroku dodajemy nową zmienną po zmiennej *Rok* i nazywamy ją *Data*. Następnie za pomocą *Operacji na danych* dokonujemy scalenia wartości zmiennej *Miesiąc* i *Rok*. Wartości zmiennej *Data* proszę wyświetlić w formacie *Mar-2010*, który następnie proszę zamienić na format użytkownika typu: *mmmm rrrr*;
- wygodniej będzie wyświetlić liczbę pasażerów w milionach – proszę dokonać odpowiedniego przekształcenia zmiennej 4, zmienić nazwę tej zmiennej dodając dopisek (w mln) – a wartości tej zmiennej wyświetlić z dokładnością do jednego miejsca po przecinku.

Na rysunku obok przedstawiono ostateczny wygląd arkusza danych, przygotowanego do analizy.

Proszę zwrócić uwagę na format zmiennej *Data* i ustawić analogiczny w swoim arkuszu danych.

| | 1 Miesiąc | 2 Rok | 3 Data | 4 Przewozy pasażerskie w USA (w mln) |
|---|--------------|----------|---------------|--|
| 1 | 1 | 1990 | Styczeń 1990 | 37,2 |
| 2 | 2 | 1990 | Luty 1990 | 36,1 |
| 3 | 3 | 1990 | Marzec 1990 | 44,2 |
| 4 | 4 | 1990 | Kwiecień 1990 | 41,4 |

Arkusz danych proszę **ZAPISAC** pod nazwą *Przewozy pasażerskie w USA (M)*.

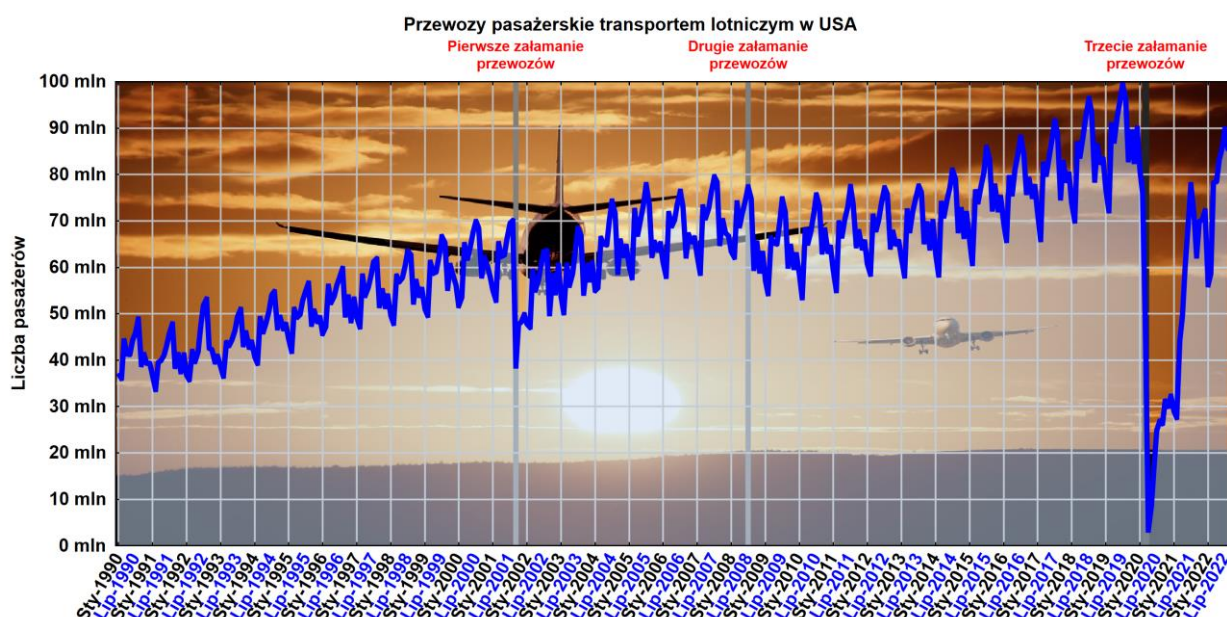
UWAGA! Proszę jeszcze uzupełnić do listopada 2024 r. z arkusza *Ruch pasażerski USA 2002-2024 (M)*



ANALIZA GRAFICZNA

Proszę wykonać wykres liniowy przewozów pasażerskich w USA. Gotowy wykres zamieszczono poniżej. A oto kilka wskazówek do formatowania wykresu:

- ponieważ wykres zawiera dość dużo obserwacji, warto zmienić jego proporcje – proszę po wywołaniu wszystkich opcji wykresu w grupie *Okno* ustawić szerokość i wysokość wykresu (w **centymetrach!**) na przykład na 20 i 12;
- proszę ustalić, któremu **numerowi przypadku** w arkuszu danych odpowiada **wrzesień 2001 i lipiec 2008** roku. Następnie w tych punktach na górnej osi proszę wstawić linie odniesienia, tak jak na wykresie i z takim opisem jak na wykresie;
- w analogiczny sposób proszę narysować linię odniesienia dla **marca 2020** (i odpowiednio opisać).
- pozostałe elementy formatowania wykresu były już tematem zajęć z *Prognozowania i symulacji*, a w razie problemów proszę poprosić o podpowiedź osobę prowadzącą zajęcia – **proszę starać się wykonać wykres w sposób zbliżony do poniższego wzorca, a o szczegóły proszę pytać!!!**



Wykres proszę zapisać, posługując się menu podręcznym i poleceniem *Zapisz wykres*.

Proszę ustalić, jakie wydarzenia miały miejsce we wrześniu 2001, lipcu 2008 roku, lutym/marcu 2020 r. w momentach wyróżnionych na wykresie i zamiast ogólnego stwierdzenia „pierwsze/drugie/trzecie załamanie przewozów” wpisać odniesienie do konkretnego wydarzenia z najnowszej historii.

Przykład 2

W pliku *Excelsa Ruch pasażerski USA 1990-2022 (M)* na końcu po wartościach miesięcznych znajdują się wartości roczne z lat 1990-2022 (opisane etykietą *All Rows*). Proszę wkleić je do programu *STATISTICA*, opisać odpowiednio arkusz z danymi rocznymi, a następnie **ZAPISAĆ** w swoim folderze pod nazwą **Przewozy pasażerskie w USA (R)**. Proszę sporządzić prezentację graficzną danych rocznych z lat 1990-20212 nanosząc nań dwie analogicznie opisane linie odniesienia (dla 2001, 2008 i 2020 r.) jak na wykresie dla danych miesięcznych.

UWAGA! Dane roczne z 2022 i 2023 r. proszę pobrać z drugiego arkusza.