

Prognozowanie i symulacje (lab. 2)

Program STATISTICA / arkusz danych czasowych

Przykład 1: Od danych do informacji – przygotowanie danych i prosta analiza

Na stronie <https://stat.gov.pl/obszary-tematyczne/inne-opracowania> co miesiąc pojawia się *Biuletyn Statystyczny GUS*. Z najnowszego biuletynu dostępnego na stronie proszę pobrać dane meteorologiczne (tabela 6) i otworzyć je w *Excelu*. Celem prostej analizy będzie porównanie temperatury w sierpniu 2024 na tle tych samych miesięcy z poprzednich lat.

Aby przeprowadzić analizę wykonamy następujące czynności:

- usuwamy kolumny nie zawierające żadnych danych (co drugą kolumnę od C, E, itd.).
- blokujemy okienko tak, by zawsze widoczne były nazwy miejscowości.
- za pomocą operacji filtrowania proszę wyselekcjonować tylko dane ze sierpniów (ze wszystkich lat) i przekopiować je do nowego arkusza wraz z datami i nazwami miast.
- stosując formatowanie warunkowe proszę pokolorować temperatury sierpnia (dla każdej miejscowości oddzielnie: niskie temperatury na niebiesko, wysokie – na czerwono), tak by łatwo dokonać wzrokowej oceny temperatury w sierpniu 2024 r. – czy tegoroczny sierpień był chłodny, typowy czy ciepły na tle poprzednich lat?
- w analogiczny sposób proszę przeprowadzić analizę poziomu opadów w sierpniu 2024 na tle sierpniów z poprzednich lat.

Przykład 2: Formatowanie danych liczbowych, wyników i dat – formaty niestandardowe

Proszę otworzyć plik danych *Wypadki w Polsce od 1998 roku (M)* i kliknąć dwukrotnie nagłówek zmiennej *DATA*. W oknie informacji o każdej zmiennej można zmieniać format wyświetlania jej wartości w arkuszu.

- Proszę zmienić format wyświetlania na *DATA / Format Daty Windows*, zamknąć okno i zobaczyć jak teraz wyglądają wartości w kolumnie *DATA*. Następnie proszę zmienić format na *Mar-2010* i zobaczyć wygląd dat.
- Czy można rozszerzyć zapis daty, do pełnych nazw miesięcy? Proszę wybrać format *Użytkownika*, w polu obok pojawia się kod zapisu daty w postaci *mmm-yyy*. Zmieniamy go na następujący: *mmm yy*. Jak będą wyglądały teraz daty w arkuszu?
- Zmieniamy jeszcze raz format zmiennej *DATA* – tym razem na *Ogólny*. Czy wiedzą Państwo, jakie jest znaczenie „dziwnych” wartości liczbowych, które pojawiły się w kolumnie *DATA* (programy komputerowe „pamiętają” daty w zupełnie inny sposób niż ludzie ☺).
- Wracamy do „najładniejszego” formatu *mmm yy* i zapisujemy plik danych przed zamknięciem.

Przykład 3: Operacje na arkuszu danych i formuły arkusza danych (przypomnienie)

Globalne operacje na arkuszu danych wykonuje się za pomocą poleceń, które znaleźć można pod przyciskami

Zmienne ▾ Przypadki ▾

znajdującymi się na pasku nad arkuszem danych.

Proszę otworzyć plik danych *Wypadki w Polsce od 1998 roku (M)*, na końcu arkusza danych dodać nową zmienną i nazwać ją *Wskaźnik ciężkości wypadków*.

Formuły w programie *STATISTICA* wpisujemy w tym samym oknie, które otwiera się po dwukrotnym kliknięciu nagłówka zmiennej. Proszę zaproponować odpowiednią formułę i za jej pomocą wyliczyć *Wskaźnik ciężkości wypadków*, który określa się jako liczbę osób zabitych w przeliczeniu na 100 osób poszkodowanych (łącznie zabitych i rannych).

Wyświetlanie wartości tej zmiennej ustawiamy z dokładnością do jednego miejsca po przecinku.

Proszę podać, ile osób spośród 100 poszkodowanych w wypadkach drogowych poniosło śmierć w:

- listopadzie 2000 r., listopadzie 2010 r., listopadzie 2020 r.

Czym można wytłumaczyć nagły skok tego wskaźnika w listopadzie 2020 r.?

Przykład 4: Wklejanie danych z Excela do programu STATISTICA i operacje na danych

W tym punkcie poznamy sposób wklejania danych z arkusza *Excela* do arkusza programu *STATISTICA*. Następnie stworzymy zmienną *DATA*, opisującą poszczególne okresy czasowe.

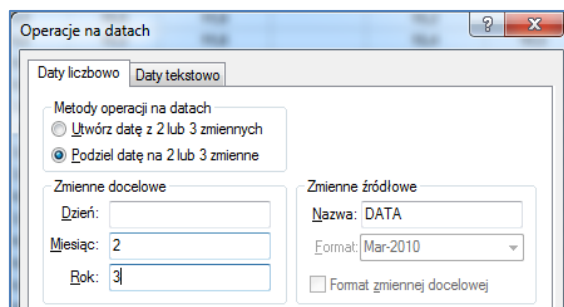
- Proszę otworzyć plik *Excela* o nazwie *Stopa bezrobocia*.
- Zanim wkleimy dane do programu *STATISTICA* musimy sprawdzić, czy wśród danych występują tylko wartości liczbowe – jak widać, brak danych w utworzonym arkuszu jest oznaczany za pomocą symbolu „,” – proszę zaznaczyć cały obszar danych i za pomocą polecenia *ZAMIEN* (można wykorzystać odpowiedni skrót z klawiatury) usunąć symbole „,”.
- Kopiujemy cały obszar danych (wraz z nazwami krajów i datami, ale bez informacji w wierszach 1-7);
- W programie *STATISTICA* tworzymy nowy arkusz danych, umieszczamy kursor w pierwszej komórce i wklejamy dane za pomocą polecenia: *EDYCJA / WKLEJ Z NAGŁÓWKAMI / Z NAZWĄ ZMIENNYCH*.
- Program *STATISTICA* nie rozpoznał „niekonwencjonalnego” sposobu zapisu daty stosowanego przez *EUROSTAT*, dlatego też stworzymy nową zmienną, zawierającą „aktywny”, „edytowalny” opis daty – w tym celu należy wstawić nową zmienną po zmiennej nr 1.
- Proszę nadać tej zmiennej nazwę *DATA* i oraz format wyświetlania *Data / Mar-2010* (czyli trzyliterowy skrót miesiąca – rok).
- W pierwszej komórce arkusza w kolumnie *DATA* wpisujemy *Sty-2000*, w drugiej *Lut-2000*, a następnie zaznaczamy te dwie komórki i przeciągamy je jak formułę w *Excelu* do końca arkusza, wypełniając w ten sposób datami całą kolumnę – na koniec usuwamy zbędną już kolumnę nr 1.

Prognozowanie i symulacje (lab. 2)

Program STATISTICA / arkusz danych czasowych

Przykład 5: Agregacja danych i operacje na datach – część 1

Zdarza się tak, że dane otrzymane w pewnym układzie czasowym (np. miesięcznym) trzeba przekształcić do innej postaci (np. rocznej). W tym celu można wykorzystać narzędzie analizy danych znane z zajęć ze statystyki opisowej, wyliczając średnią (lub inne statystyki) w grupach określonych przez zmienną *Rok*.



- Proszę otworzyć plik: *Stopa bezrobocia w krajach UE 2000-2024 (M)*.
- Musimy wyodrębnić ze zmiennej *DATA* informację o roku – proszę w tym celu po zmiennej nr 1 dodać dwie nowe zmienne, i nazwać je *MIESIĄC* i *ROK*.
- Proszę zaznaczyć zmienną *DATA* i odnaleźć za pomocą przycisku *Zmienne polecenie Operacje na datach*, wybrać takie ustawienia jak na rysunku powyżej i zatwierdzić operację przyciskiem *OK*. – program automatycznie wypełni kolumny *MIESIĄC* i *ROK* odpowiednimi wartościami.
- Teraz wybieramy polecenie *STATYSTYKA / STATYSTYKI PODSTAWOWE I TABELĘ* a następnie **PRZEKROJE, PROSTA ANOVA**. Na liście zmiennych zależnych wskazujemy wszystkie zmienne zawierające dane o bezrobociu (a więc od 4 do końca), a na liście zmiennych grupujących zmienną *ROK*.
- Zatwierdzamy wybór i za pomocą przycisku *OK* przechodzimy do okna wynikowego, gdzie w zakładce *Statystyki opisowe* odznaczamy wszystkie miary – średnia jest tu wyliczana obligatoryjnie. Wywołujemy arkusz wynikowy ze średnią roczną stopą bezrobocia w poszczególnych latach 2000-2022.
- Usuwamy ostatni przypadek, „wyciągamy” wyniki ze skróty do osobnego okna (w tym celu klikamy prawym przyciskiem myszki arkusz wyników w skróty i wybieramy *Pobierz do oddzielnego okna / Przenieś*), następnie wybieramy polecenie *DANE → ARKUSZ WEJŚCIOWY* i zapisujemy wyodrębniony arkusz pod nazwą *Bezrobocie w UE dane roczne* – plik ten będziemy jeszcze wykorzystywać na zajęciach.

Przykład 6: Agregacja danych – część 2 (Dane: Wypadki w Polsce od 1998 roku (M))

Stosując ponownie analizę służącą do wyznaczania **statystyk opisowych w grupach** dokonamy przekształcenia danych miesięcznych dotyczących wypadków w Polsce na dane roczne. Tym razem jednak nie będziemy wyliczać średnich – dla danych o liczbie wypadków logiczniej dokonać sumowania.

Tym razem wykorzystamy analizę **PRZEKROJE UPROSZCZONE** – w zakładce *PODSTAWOWE* na liście zmiennych zależnych wskazujemy dane liczbowe dotyczące *WYPADKÓW*, *RANNYCH* i *ZABITYCH* zaś na liście zmiennych grupujących wybieramy *ROK*. Przechodzimy do kolejnego okna, gdzie w zakładce *STATYSTYKI OPISOWE* odznaczamy wszystkie statystyki, wybierając tylko opcję *SUMY*. Następnie wywołujemy arkusz z wynikami. Zawiera on sumy poszczególnych zdarzeń dla każdego roku oraz wartość sumaryczną z całego badanego okresu.

- Proszę podać łączną liczbę osób, które poniosły śmierć w wypadkach drogowych w całym badanym okresie, to jest w latach 1998-2024 –, proszę podać liczbę osób rannych w wypadkach drogowych w tym okresie
- Przeciągamy arkusz ze skróty wyników na pulpit programu *STATISTICA* i zapisujemy jako arkusz danych (albo „wyciągamy” obiekt ze skróty w sposób opisany w przykładzie 6). Usuwamy wiersz zawierający sumę wszystkich wartości. Aby program traktował ten nowy arkusz jako przedmiot analizy wybieramy polecenie *DANE → ARKUSZ WEJŚCIOWY*. Zapisujemy plik pod nazwą *Wypadki w Polsce – dane roczne*.

Przykład 7: Transponowanie danych (Dane: Energia odnawialna w Europie 2004-2023 (R) – plik STATISTICA)

1. W programie *STATISTICA*: Układ tego arkusza pozwala prowadzić analizę porównawczą państw w poszczególnych latach, ale nie można analizować danych jako szeregu czasowego. Aby zamienić położenie wierszy i kolumn proszę wykorzystać polecenia: *DANE → TRANSPONUJ → PLIK*. Przy tej operacji program „gubi” niestety formatowanie arkusza – proszę więc ponownie, w całym arkuszu, ustalić dokładność wyświetlania liczb na jedno miejsce po przecinku.
2. W programie *Excel*: Proszę otworzyć plik *Excelsa* o tej samej nazwie. Dane z pierwszego arkusza (*Data*) proszę poddać analogicznemu przekształceniu jak w p. 1, wykorzystując tym razem narzędzia dostępne w programie *Excel*. **Wskazówka:** W *Excelu* można „obrócić” dane kopiując je i wklejając w specjalny sposób. Proszę poszukać takiej opcji „specjalnego” wklejania i ją zastosować.