

Statystyczne Metody Wspomagania Decyzji

laboratorium 1

*Cel: ogólne zapoznanie z SAS Enterprise Guide;
różne metody wprowadzania danych; podstawy eksploracji danych*

Zadanie 1

Utworzyć własną bibliotekę, w której będą przechowywane zbiory danych.

Zadanie 2

Badania demograficzne przeprowadzone w 1988 roku w USA wykazały, że wśród kobiet (mających 18 i więcej lat) było: 17364 tys. panien, 56128 tys. mężatek, 11239 tys. wdów i 8170 tys. rozwódek.

- Wprowadzić powyższe dane do systemu SAS (wpisując je bezpośrednio do tabeli danych oraz korzystając z **Program Editor**).
- Utworzyć wykres blokowy dla stanu cywilnego danej grupy kobiet. Porównać różne formy opisu wykresu.
- Utworzyć wykres kołowy dla stanu cywilnego danej grupy kobiet. Porównać różne formy opisu wykresu.

Zadanie 3

Koncern paliwowy planuje otworzyć nową stację benzynową w pewnym mieście. Rozważane są cztery możliwe lokalizacje stacji - w południowej, północnej, zachodniej i wschodniej dzielnicy miasta (oznaczenia dzielnic: S - południowa, N - północna, W - zachodnia i E - wschodnia). W ramach badania opinii społecznej odnośnie preferowanej lokalizacji stacji zapytano o to tysiąc kierowców. Ich odpowiedzi znajdują się w pliku **petrol.xls**.

- Pobrać plik (**Import Data**).
- Utworzyć wykres blokowy dla badanych preferencji.
- Utworzyć wykres kołowy dla badanych preferencji.

Zadanie 4

Wytrzymałość na ciśnienie wewnętrzne szkła butelek jest ich ważną charakterystyką jakościową. W celu zbadania wytrzymałości butelek umieszcza się je w maszynie hydrostatycznej, po czym zwiększa się ciśnienie aż do zniszczenia butelki. Plik **bottles.xls** zawiera dane opisujące graniczną wytrzymałość na ciśnienie wewnętrzne (mierzone w psi) szkła badanej partii butelek.

- Utworzyć zmienną o nazwie *cisnienie*, opisującą wytrzymałość na ciśnienie wewnętrzne szkła butelek mierzone w Mpa (pamiętając, że 1psi = 0.0068947 Mpa).
- Utworzyć wykres skrzynkowy dla wytrzymałości butelek.
- Utworzyć histogram dla danych opisujących wytrzymałość butelek.
- Wyznaczyć i zinterpretować podstawowe statystyki próbkowe (średnia, mediana, wariancja, odchylenie standardowe, rozstęp, kwartyły, rozstęp międzykwartyłowy, wartości ekstremalne, współczynnik asymetrii, kurtoza, współczynnik zmienności) dla danych opisujących wytrzymałość butelek.
- Obliczyć 5, 10, 25, 50, 75, 90 i 95 percentyl dla rozważanych danych.
- Wyznaczyć średnią uciętą i średnią winsorowską dla danych opisujących wytrzymałość butelek. Porównać średnią uciętą i winsorowską ze średnią arytmetyczną i medianą. Prześledzić, jak zmienia się wartość średniej uciętej i średniej winsorowskiej wraz ze zmianą stopnia ucięcia próbek.
- Zbadać (różnymi metodami), czy wytrzymałość butelek ma rozkład normalny.
- Wyznaczyć 95% przedział ufności dla średniej wytrzymałości butelek oraz dla odchylenia standardowego wytrzymałości butelek.
- Na poziomie istotności 0.05 stwierdzić, czy wytrzymałość butelek różni się istotnie od 1.8 MPa.

Zadanie 5

Zmienne, znajdujące się w pliku **cardata.xls**, zawierają dane dotyczące parametrów kilku wybranych marek samochodów.

- Zmienna *mpg* zawiera dane odpowiadające liczbie mil, przejechanych przez dany samochód na galonie paliwa. Pamiętając, że 1 mila = 1609 m oraz że 1 galon (amerykański) = 3.785 l, utworzyć zmienną *zp* opisującą zużycie paliwa mierzone w litrach na 100 kilometrów.
- Utworzyć wykres skrzynkowy zużycia paliwa.
- Utworzyć histogram dla danych opisujących zużycie paliwa.
- Wyznaczyć i zinterpretować podstawowe statystyki próbkowe (średnia, mediana, wariancja, odchylenie standardowe, rozstęp, kwartyle, rozstęp międzykwartyłowy, wartości ekstremalne, współczynnik asymetrii, kurtoza, współczynnik zmienności) dla danych opisujących zużycie paliwa.
- Obliczyć 5, 10, 25, 50, 75, 90 i 95 percentyl dla rozważanych danych.
- Wyznaczyć średnią uciętą i średnią winsorowską zużycia paliwa.
- Zbadać (różnymi metodami), czy zużycie paliwa ma rozkład normalny.
- Zbudować wykres dystrybuanty empirycznej zużycia paliwa.
- Na poziomie istotności 0.05 stwierdzić, czy zużycie paliwa różni się istotnie od 8 litrów na 100 kilometrów.
- Utworzyć nową zmienną charakteryzującą zużycie paliwa przez poszczególne samochody w sposób opisowy, przyjmując następujące kody:

Zużycie paliwa [litry na 100 km]	Kod opisujący kategorię zużycia paliwa
nie więcej niż 7	mało
więcej niż 7 ale nie więcej niż 10	przeciętnie
więcej niż 10	dużo

- Utworzyć wykres kołowy dla nowo utworzonej zmiennej i wskazać, jaki procent badanych samochodów należy do każdej kategorii.

Zadanie 6

Przeprowadzić wstępną analizę statystyczną (jak w zadaniu 6) danych dotyczących zużycia paliwa oddzielnie dla samochodów produkowanych w Europie, Ameryce i Japonii (wykorzystać zmienne *origin* i *carmakers*).

Zadanie 7

Przeprowadzić wstępną analizę statystyczną (jak w zadaniu 6) danych dotyczących zużycia paliwa wyłącznie dla samochodów mających 6 lub 8 cylindrów.

Zadanie 8

Przeprowadzić wstępną analizę statystyczną (jak w zadaniu 6) danych dotyczących zużycia paliwa wyłącznie dla samochodów ważących mniej niż 2500 funtów.

Zadanie 9

W pliku **swiat95.xls** znajdują się dane (głównie demograficzne) dotyczące wybranych 109 państw.

- Zbadać normalność rozkładu wielkości populacji (zmienna *populatin*). Powtórzyć badanie dla zlogarytmowanej wielkości populacji. Wyciągnąć wnioski.
- Utworzyć wykres słupkowy opisujący rozkład dominującej religii (zmienna *religion*).
- Porównać rozkłady średniej długości życia mężczyzn (zmienna *lifeexpm*) oraz kobiet (zmienna *lifeexpf*) w zależności od dominującej religii.
- Porównać rozkłady średniej długości życia mężczyzn (zmienna *lifeexpm*) oraz kobiet (zmienna *lifeexpf*) w zależności od regionu (zmienna *region*).
- Porównać rozkład odsetka mężczyzn umiejących czytać (zmienna *lit_male*) oraz odsetka kobiet umiejących czytać (zmienna *lit_fema*) w zależności od dominującej religii.
- Porównać rozkład odsetka mężczyzn umiejących czytać (zmienna *lit_male*) oraz odsetka kobiet umiejących czytać (zmienna *lit_fema*) w zależności od regionu.
- Na podstawie wykresu rozrzutu stwierdzić, czy istnieje zależność między średnią długością życia mężczyzn i kobiet.
- Na podstawie wykresu rozrzutu stwierdzić, czy istnieje zależność między wielkością odsetka mężczyzn i kobiet umiejących czytać.
- Na podstawie wykresu rozrzutu stwierdzić, czy istnieje zależność między wielkością populacji miejskiej (zmienna *urban*) i gęstością zaludnienia (zmienna *density*).