

Przetwarzanie i analiza danych w systemie SAS®

Kolokwium nr 1 - 25.11.2013

- Zapisać plik **szablon.sas** pod nazwą **nazwisko.sas** (gdzie **nazwisko** to nazwisko piszącego kolokwium). Na początku pliku należy wpisać w komentarzu własne imię i nazwisko.
- Rozwiązania zadań należy wpisywać do pliku **nazwisko.sas**. Plik należy we własnym interesie często zapisywać.
- Rozpakowane pliki z danymi wejściowym mają być umieszczone w bibliotece o nazwie **KOLO**.
- Wszelkie tworzone przez siebie zbiory sasowe należy umieszczać w bibliotece **WORK**.
- Tworzone programy powinny działać poprawnie bez żadnych zmian dla dowolnych zbiorów o takiej samej strukturze (tj. o takich samych zmiennych i ich atrybutach), jak wymienione w treści zadań. W szczególności, rozwiązania będą testowane na zbiorach różnych od podanych.
- Rozwiązania (tylko zawartość pliku **nazwisko.sas**, bez tworzonych zbiorów sasowych) należy przesłać na adres prowadzącego laboratorium, tzn.:

A.Sakowicz@mini.pw.edu.pl lub K.Szpojankowski@mini.pw.edu.pl lub matysiak@mini.pw.edu.pl

List ma nie zawierać żadnych załączników – innymi słowy kod program musi być bezpośrednio wklejony do listu. We własnym interesie należy poczekać na potwierdzenie odbioru.

- Powodzenia!

1. (4pkt) Zbiór sasowy **a** ma znaną liczbę (50) zmiennych o nazwach x_1, \dots, x_{50} . W obserwacjach ze zbioru **a** są braki danych. Pisząc jeden DATA STEP przekształcić zbiór **a** do zbioru **aa**, tzn. przesunąć w wierszu wszystkie braki danych na prawo, a niebrakujące dane na lewo, z zachowaniem kolejności niebrakujących danych w wierszu.
2. (4pkt) Zakładamy, że tabela **b** zawiera rezultaty (zmienna *w*) pewnego testu, któremu poddawane były różne osoby (zmienna *o*) w różnych kwartałach (zmienna *k*). Używając tylko PROC SQL znaleźć liczbę kwartałów, w których co najmniej jedna osoba osiągnęła wynik lepszy od swojej średniej.
3. (6pkt) Wczytać plik tekstowy **c.txt** do zbioru sasowego **c**. Wiadomo, że plik tekstowy **c.txt** zawiera nieznaną liczbę trzYWierszowych bloków. W pierwszym wierszu każdego kolejnego bloku znajduje się data (patrz zmienna *data* w zbiorze **c**), a potem cztery liczby odpowiadające czterem kwartałom roku z tejże daty. Zmienna *liczba* ma w każdym wierszu zawierać liczbę związaną z tym kwartałem, do którego należy data z początku pierwszego wiersza danego bloku. Liczby znajdujące się w drugim wierszu każdej kolejnej trójki wierszy dają numery kolumn: znaki zawarte w trzecim wierszu trójki pomiędzy tymi kolumnami tworzą zawartość zmiennej *tekst*. (Można założyć, że maksymalna różnica liczb z drugiego wiersza jest osiągnięta w pierwszym bloku.)
4. (6pkt) Zbiór sasowy **d** zawiera daty początkowe (gwiazdka w kolumnie *in*) i końcowe (gwiazdka w kolumnie *out* - oznacza ona, że tego dnia dana osoba nie była już pacjentem) pobytów w szpitalu pacjentów o numerach identyfikacyjnych *id* (zbiór jest posortowany rosnąco po *dacie*). Za pomocą jak najmniejszej liczby kroków (zarówno PROC jak i DATA STEPów) wypisać do okienka Log średnie dzienne obłożenie szpitala (tj. średnią z liczby pacjentów przebywających w szpitalu) w całym okresie od pierwszej do ostatniej daty występującej w zbiorze **d**.