

Przetwarzanie i analiza danych w systemie SAS®

Kolokwium nr 1 - 19.11.2012

- 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 kolokwium, tzn.:

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. Dla zbioru sasowego **a** napisać program, składający się z jak najmniejszej liczby kroków 4GL, który przekształci ten zbiór do zbioru **aa**, który zawiera informację o liczbie „słów” j -literowych w obrębie grup wyznaczonych przez g . Założenia:
 - zbiór wejściowy jest posortowany rosnąco względem g ,
 - minimalna liczba znaków w słowie ze zmiennej s to 1 (nie ma pustych słów),
 - maksymalna liczba znaków w słowie ze zmiennej s to 30.
2. Zbiór sasowy **b** ma jedną zmienną numeryczną x , która zawiera braki danych. Luką nazywamy ciąg kolejnych braków danych. Napisać jeden DATA STEP, który wczyta zbiór **b** i zastąpi każdą lukę sumą (niebrakujących) wartości bezpośrednio sąsiadujących z luką (wynik działania takiego DATA STEPu dla zbioru **b** to zbiór **bb**). Założenia:
 - jeden izolowany brak danych też jest luką,
 - pierwsza i ostatnia wartość zmiennej nie jest brakiem danych.
3. Zbiór sasowy **c** zawiera rezultaty pewnego testu, któremu poddawane były różne osoby (zmienna o). Termin przeprowadzania testu wyznaczony jest przez rok (zmienna r) i kwartał (zmienna k); wyniki zawiera zmienna w . Za pomocą pojedynczego zapytania SQL, dla każdej osoby biorącej udział w teście (łącznie z osobą A), podać liczbę wyników uzyskanych przez tę osobę, które były niegorsze niż najgorszy wynik osoby A ?
4. Plik tekstowy **plikD.txt** zawiera identyfikatory zawodników, wyniki przez nich uzyskiwane wraz z datami ich uzyskania. Dane występują w dziewięciowierszowych blokach: najpierw trzy identyfikatory, następnie trzy daty (odpowiadające kolejnym identyfikatorom) i trzy wyniki (tak jak daty, odpowiadające kolejnym identyfikatorom). Za pomocą jednego DATA STEPu wczytać **plikD.txt** do zbioru sasowego **D**.