

Przetwarzanie i analiza danych w systemie SAS®

Kolokwium nr 2 - 22.01.2009

- 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. 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. (13pkt.) Załóżmy, że w pewnej bibliotece znajduje się niezerowa liczba zbiorów sasowych zawierających dokładnie jedną obserwację i dokładnie dwadzieścia zmiennych. (Może się zdarzyć, że w tej bibliotece są także zbiory o innych wymiarach.) Załóżmy też, że w zbiorach o kształcie podanym wyżej jedynymi zmiennymi są zmienne tekstowe o nieznanach z góry nazwach (jednakowych w każdym z tych zbiorów), a jedynymi wartościami tych zmiennych są `**` lub `---`. Napisać makro, z nazwą biblioteki będącą jedynym jego parametrem, które wypisze nazwy zmiennych przyjmujących jedynie wartości `**`. Przykładowe zbiory, które spełniają warunki zadania to **a**, **b** i **c** – w ich przypadku poprawnie działający program powinien wypisywać nazwy zmiennych `s2`, `k3` i `b2`.
2. (10pkt.) Dane są trzy zbiory sasowe, o strukturze takiej jak **klienci**, **auta** i **wypozyczenia** (w szczególności zakłada się, że zmienne o nazwach zaczynających się od *id* jednoznacznie identyfikują klientów, samochody i wypożyczenia). Używając języka SQL znaleźć nazwiska klientów, którzy wypożyczyli samochody wszystkich marek.
3. (10pkt.) Wiadomo, że pewien zbiór sasowy zawiera jedynie zmienne numeryczne. Za pomocą jak najmniejszej liczby kroków języka 4GL dla każdej obserwacji z tego zbioru znaleźć średnią z dwóch najmniejszych wartości, które nie są brakami danych.
4. (7pkt.) Macierz kwadratowa A wymiaru $n \times n$ (n jest znane) jest dana w postaci zbioru sasowego. Za pomocą procedury `PROC IML` znaleźć i zapisać do zbioru sasowego największy z minorów głównych macierzy A .