

Przetwarzanie i analiza danych w systemie SAS®

Kolokwium nr 2 - 24.01.2008

- 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 uaktualniać.
- 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 plik `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. Przyjmijmy, że macierz kwadratowa znanego wymiaru jest dana w postaci zbioru sasowego. Za pomocą procedury `PROC IML` stworzyć zbiór sasowy zawierający wszystkie jej minory główne.
2. Dane są trzy zbiory sasowe, o strukturze takiej jak **klienci**, **auta** i **wypożyczenia** (w szczególności zakłada się, że zmienne o nazwach zaczynających się od *id* jednoznacznie identyfikują klientów, samochody i wypożyczenia). Używając języka SQL wypisać nazwiska klientów wraz z liczbami różnych marek przez nich wypożyczonych. (Uwaga: wśród samochodów dostępnych w wypożyczalni nie wszystkie są różnych marek.)
3. Przyjmijmy, że w pewnej bibliotece znajduje się pewna nieznana niezerowa liczba zbiorów sasowych „jednokolumnowych”, z jedyną zmienną będącą zmienną numeryczną. Napisać makro z parametrem będącym nazwą biblioteki, które poda medianę wszystkich obserwacji ze wszystkich zbiorów „jednokolumnowych” z tej biblioteki.
4. Zbiory o strukturze takiej jak **c41** i **c42** zawierają wyniki tej samej ankiety, którą badano pewną grupę osób na początku (zbiór **c41**) i pod koniec roku (zbiór **c42**). Za pomocą jednego `DATA STEP` wypisać do okienka Log liczbę osób, które w ciągu roku zmieniły odpowiedzi na wszystkie pytania z ankiety. (Można założyć znajomość liczby pytań i nazw zmiennych w zbiorach oraz posortowanie zbiorów względem identyfikatora osoby; może się zdarzyć, że nie wszystkie osoby brały udział w obydwu ankietach.)