

Przetwarzanie i analiza danych w systemie SAS®

Kolokwium nr 2 - 24.01.2011

- 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. (18p.) Napisać makro, które dla zadanego zbioru będzie tworzyło jego kopię ze zoptymalizowanymi długościami wszystkich zmiennych tekstowych z tego zbioru. Przez optymalizację rozumie się ustalenie najmniejszej długości zmiennej tekstowej, przy której żadna z jej wartości nie ulega ucięciu (ucięciu mogą ulec jedynie ewentualnie końcowe spacje). Nazwa zbioru ma być zadawana w postaci *bib.zb*, gdzie *bib* oznacza nazwę biblioteki, a *zb* właściwą nazwę zbioru.
2. (6p.) Przypuśćmy, że mamy zbiór o strukturze podobnej do struktury zbioru **z**. Zbiór zawiera informacje o aktywności klientów (jednoznacznie identyfikowanych przez zmienną *id*) w ciągu kilku lat (zmienna *year*). Brak rekordu z danym rokiem i danym identyfikatorem klienta oznacza brak aktywności danego klienta w danym roku. Napisać pojedyncze zapytanie SQL, które znajdzie tych klientów, którzy byli aktywni w co najmniej trzydziestu procentach lat, które występują w zbiorze **z**.
3. (6p.) 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.
4. (10p.) 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 STEPu 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.)
5. (Ekstra problem za dodatkowe 2 pkt) Do zadania 2: napisać pojedyncze zapytanie SQL, które znajdzie wszystkich klientów, którzy byli aktywni w pierwszym i ostatnim roku występującym w zbiorze **z**.