Lab 8

8.1 You are given two data sets: **z1** and **z2**, both with a numerical variable x. Treating them as sequences (that is sets in which the order of elements matters), write a single DATA STEP that shows the number of differing elements in the sets, in the Log window.

8.2 **(Table look-up I)** What is the average of *sales* from **large** computed only for the values of *id* that occur in the set **small**?

8.3 What is the average of the variable *sales* from the data set **large** computed only for the observations with the numbers listed in the data set **numbers**?

8.4 Delete all the missing values from **dots**, that is transform **dots** into **without_dots**.

8.5 Create a data set **numbers** with 50 arbitrary numerical observations (with one numerical variable). Write a single DATA STEP that will create (based on the data set **numbers**) the data set **sums** with one numerical variable *sum* and 46 observations. The i-th observation in **sums** should be the sum of observations from **numbers** numbered $\{i, \ldots, i+4\}$.

8.6 Write a single DATA STEP that acts as the following SQL query:

```
proc sql noprint;
 create table aa as
  select *, count(y) as licznik
   from a
   group by x
   having count(y)>5
   ;
quit;
```

8.7 Based on the data sets **zb1**,..., **zb5** create a data set **result** that has the whole year 2007 in the variable *date*.

8.8 The data sets **jan**, **feb** and **mar** have two variables : *person* and *result*. Generate a data set with the most current results for every person (the names of the sets refer to the names of the months in which the measurements were taken).

8.9 The data sets **zx** and **zy** have single numerical variables (x and y respectively), and the set **zxy** has both x and y. Find the number of observations from **zxy** such that their values of x and y are equal to the values of x from **zx** and y from **zy** with the same observation number.

8.10 Use a single DATA STEP to update **first** with the data from **second**. (If a given year and month are missing in both sets, they should not be placed in the updated set. If, for a given year and month, there is no data in **first** (or in **second**), one should leave in the updated file the value of *sales* from **second** (**first**, respectively) - if it exists. If some year and month are present in both files, the value of *sales* from **first** should be increased by the relevant value from **second**.) One can assume that both files are sorted in ascending order with respect to *year* and that the names of sets, and the names and types of variables are known.