Lab 7

7.1 Create a data set with 10 numerical variables *z1,...,z10* and 20 observations. Then transpose it without using `PROC TRANSPOSE`.

7.2 Transform the data set **z1** into **z2**.

7.3 Order alphabetically the variables in the data set **z**.

7.4 The data set **a** has the variables x and y.

(a) For every value of x find the most frequently appearing values of the variable y (do not care for *ex-aequo* situations, if they appear – just pick any of the most frequent values of y).

(b) For every value of x list all the most frequent values of the variable y.

(c) Find all such values of x for which there exists exactly one smallest y.

(d) Find all such values of x for which there are no repeating values of the variable y.

(e) Find all such values of x that have the largest number of distinct values of y.

(f) Find all values of the variable x for which the values of y form the set $\{1, \ldots, n\}$ for some $n \in \mathbb{N}$.

(g) Find all such x for which the distinct values of y form the set $\{1, \ldots, n\}$ for some $n \in \mathbb{N}$.

(h) Find such values of y that correspond to at least half of the values of x that appear in the data set **a**.

7.5 The data set **z3** has two variables *id*, *year* and *sales*.

- Find all values of *id* that did not appear before 1993.
- Find all values of *id* that appeared both in the first and in the last year.
- Find all values of *id* that were present in every year.

7.6 The data set **b** has the variables $a1, x1, a2, x2$.

(a) Treating $a1$ and $a2$ as grouping variables, for each group formed by $a1$ pick such values of $x1$ that are between the smallest and the largest value of $x2$ in the same group formed by $a2$.

(b) Identify the name of the group that appears **b** most frequently.

7.7 With a single SQL query, based on the data set **c**:

(a) find the month in which $r2$ has the largest number of missing values,

(b) find the month in which the values of $r1$ are most scattered.