Introduction to the SAS System

Lab 5

5.1 We are given a data set **a** with a grouping variable x and an additional numeric variable u. Find the largest (absolute) difference of consecutive values of y in every group formed by x.

5.2 We are given a data set **b** with a grouping variable x and numeric variables $y1, \ldots, y10$. For every value of x find ten largest values of the numbers stored in variables $y1, \ldots, y10$. That is, if there is a value of x with, say, 6 observations, one needs to determine ten largest values of the 60 numbers appearing in $y1, \ldots, y10$.

5.3 We are given a data set **c** with a grouping variable x and numeric variables $y1, \ldots, y10$. Each group generated by x has 5 rows. Replace the missing values in every group by the arithmetic mean of the non-missing $y1, \ldots, y10$ values from the group.

5.4 Consider a data set **values** with a single character variable, with the values of the form:

"value_1, value_2, value_3, ..., value_n",

where *value* is a number, and n differs from row to row. Create a data set with a single numeric variable, storing split values from **values**. Each of the split values should make a separate row in the new data set.

5.5 Consider a system generating alerts for clients at some moments of time. There is a table **ALERT_CLIENT** with three columns: *alert_id*, *client_id* and *alert_time* (alert generation time).

- Generate a report (create a data set) that lists all the alerts for which the associated client had a previous alert more recently than two years ago.
- Do the same but without sorting the table (it may be very big). You can now assume that:
  - the values in *alert_id* are not duplicated,
  - they are strictly decreasing,
  - the values in *alert_date* never increase.

For the attached data set:

- *alert_id* = 5028 should not be reported (as the previous *alert_date* for *client_id* = 1001 is 02/15/2018, which is earlier than two years from 03/02/2020),
- *alert_id* = 4982 should be reported (as the previous *alert_date* for *client_id* = 1014 is 05/30/2017, which is more recent than two years from 04/26/2019).

5.6 Data set **d** contains variable **year** and variables **y1**,..., **y12** with results for each month. Create a dataset, which contains variable **date** being the first day in a given month (use `date9.` format) and variable **y** being the value for a respective month (see data set **dtrans**).

5.7 Assume you are given the data set **dtrans**. Transform it into a data set of the same structure as **d**.