Exam questions for course:

Basic and advanced programming and statistics in SAS

– Master studies, SGH

Course ID: 223110-1234

Lecturer: Karol Przanowski PhD

The exam consists of a fast and skillful writing code in the SAS system on the computer in the presence of the examiner. You can make mistakes, but you have to know how to fix them. You are allowed to use SAS documentation and sample SAS codes made available during the lecture. Your goal is to convince the examiner that you have written similar codes on your own and can replicate them fast during the exam. Each examinee receives one question from each of three groups: data processing, reporting and statistics in SAS.

Every question begins with the sentence:

Using table SASHELP.CLASS write a code that ...

Data processing in SAS (number of points: 6):

1. creates a library engine v9 pointing to directory c:\ sas \
2. creates a library engine v8 pointing to directory c:\ sas \
3. creates a library engine odbc indicating the ODBC driver called CRM
4. reads it using a SET instruction in a single datastep pass
5. reads it and creates table RESULT with a new variable called TWO, which contains the first two letters of the name
6. selects the rows with age greater than 12 years
7. selects the rows with names starting with the letter A
8. creates a new column DATE_OF_BIRTH on the basis of age and today's date in the system
9. creates a new column and counts number of letters in the name
10. creates a new column: binary flag with 1 when the name contains letter a
11. creates a new column concatenating: name, age and gender, for example. Alfred-14-M
12. randomly multiplies each observation from 1 to 10 times
13. creates a RESULT table containing every second observation
14. sorts it by sex increasingly and age decreasingly
15. based on the processing in groups mechanism calculates the number of women and men
16. based on the processing in groups mechanism calculates minimum and maximum age for each gender
17. based on the processing in groups mechanism calculates average for each gender
18. using table SASHELP.AIR calculates the cumulative number of thousands of passengers, date by date - ascending, based on column AIR
19. using LAG function crates table RESULT containing only these rows for which the previous line was a woman

20. creates a new column as the sum of the ages of the previous two lines
21. creates table RESULT containing only observations with numbers from 5 to 7
22. creates table RESULT containing only the first observation
23. creates table RESULT containing only the last observation
24. creates new variables age1 to age20 and fills them the following way: eg. for age equal to 10 column age10 has a value of 1 and the other ones have 0
25. using SASHELP.ZHC select only those observations whose PCMS code is in table SASHELP.ZTC
26. adds column EUC to table SASHELP.ZHC from table SASHELP.ZTC using PCMS as joining key
27. tests joining tables SASHELP.ZHC and SASHELP.ZTC by PCMS, how many observations joined, how many are not joined in first and second tables
28. creates table RESULT containing columns from age11 to age16 and name. For example, if a student is 11 years old the variable age11 contains the student's gender and the remaining variables are empty
29. using table SASHELP.CARS creates table RESULT containing information about car models in multiple lines. For a given model, in rows, there are two additional columns named PROPERTY and VALUE. In the first column there is the name of the column from original table and its value in the second one.
30. counts the Levenstein distance between the name and the word Alfred
31. creates 19 tables from name1 to name19, each with one observation, the first with the first observation of the input table, and the last with the last
32. creates table RESULT containing observations repeated as many times as is the value of the variable age
33. creates a two-dimensional array in the IML taking the first two observations and columns age and weight and calculates its determinant

Reporting in SAS (number of points: 6):

1. creates a format for age group: up to 13 years included and above, using the MEANS procedure determine the number of students and the average values of weight
2. creates a two-dimensional table containing frequencies of age and sex
3. creates a table containing absolute and relative frequencies for age, includes cumulated values for both
4. creates an HTML report with a table with names on the left hand side and headings of sex; the table should be filled with total values of age
5. creates a report in PDF based on SASHELP.PRDSALE containing information about the current total sales with percentages by years and regions, adds summaries
6. creates a serial report for each region, i.e. creates as many HTML reports, as there are regions and every of these reports contains total sales divided by years and products; adds summaries
7. creates a list macrovariables name1 to name19 inserting names of the students in alphabetical order
8. creates a list macrovariables with names of students inserting each student's sex after his/her name
9. creates one long macrovariable with list of names of students in alphabetical order, separated with a #

10. creates a macro program with parameter sex which lists (PROC PRINT) observations for the selected gender
11. creates a bar graph of the average age for sex
12. creates a line graph of the time series containing number of passengers AIR from the set SASHELP.AIR
13. calculates percentiles for age: 5, 25, 50, 75 and 95
14. calculates standard deviation, average and the inter-quartile range
15. exports numbers 3,4,5 to EXCEL cells A1, A2 and A3
16. creates a data set containing file names from the directory c: \
17. creates text files with names as variable PRODUCT in table SASHELP.PRDSALE and inserts information about the current sales and year
18. creates a box plot for age by sex
19. creates a scatter plot for age and weight
20. calculates robust means for age: e.g. trimmed or winsorized mean
21. creates a listing for LaTeX
22. creates a table CONVERGENCE containing statistics of fit of age to normal distribution, using procedure UNIVARIATE

Statistics in SAS (number of points: 8):

1. creates a histogram of the age variable containing the gaussian curve
2. creates the chi-square test for gender and age, treating both as a nominal variables, counts p-value
3. calculates the V-Cramer statistic for age and sex
4. calculates the Gini coefficient for sex and age
5. calculates the Spearman correlation coefficient for age and weight
6. calculates the Pearson correlation coefficient for age and weight
7. calculates the Kolmogorov-Smirnov statistic to test the fit of empirical age distribution to the theoretical normal distribution
8. calculates the Shapiro-Wilk statistic to test the fit of empirical age distribution to the theoretical normal distribution
9. calculates the partial Pearson correlation coefficient for age and weight excluding the influence of height
10. calculates linear regression coefficients for age and weight
11. calculates determination coefficient of linear regression for age and weight
12. performs diagnostics of linear regression model for age and weight
13. examines homoscedasticity in linear regression model for age and weight
14. examines normality of residuals in linear regression model for age and weight
15. performs stepwise variable selection in multiple linear regression model of age dependent on weight and height
16. calculates collinearity measures in multiple linear regression model of age dependent on weight and height
17. calculates coefficients of determination for all models of multiple linear regression of age dependent on weight and height
18. determines outliers and/or influential observations in linear regression model of age and height

19. calculates the percent variance explained by principal components for the variables age, weight and height
20. calculates a forecast of age with weight equal to 80 in the linear regression model of age dependent on weight
21. determines clusters (concentration) variables (PROC VARCLUS) for all numeric variables in table SASHELP.CARS and the explained variance (ratio) of 80%