

MS3251 Analytics Using SAS Assignment 1

程序代写代做 CS编程辅导

- You must complete the assignment by yourself. Exchanging ideas with classmates is encouraged, but the line between discussion and collaboration. Showing your work to others is a non-acceptable collaboration. All identified collaboration will be marked.
- Complete all tasks and save your SAS code into one PDF file. You must mark each question's answer with a comment statement in the code, such as `/*Question 1*/`. Save the file as `nnnnnnnn.pdf` where `nnnnnnnn` is your full name.
- You should set all irrelevant statements in your code as SAS comment statements. Be aware that your SAS code created under a non-English operating system may contain extraneous characters when viewed in an English operating system. You are responsible for ensuring that your submitted code in the PDF file is free of these characters. **All extraneous characters in your submitted code will be considered errors.**
- When running in SAS OnDemand Studio, your submitted SAS code should be error-free.
- You must submit your code file via the link Assignment 1 under the Assignments section of the course on Canvas. If you submit your file more than once, only the latest submitted file will be marked. The assignment is due at 06:00 a.m. on 1 November 2023. Two marks shall be deducted for every 1 minute or less late. The submission link will be closed at 06:40 a.m. on 1 November 2023. Submission by other methods will not be accepted.

QQ: 749389476

Question 1

Suppose the raw data file StockPrice.csv (not available in Canvas) has already been referred to by the FILENAME reference 'StockPrice' in SAS Studio. StockPrice.csv contains the closing indexes of 1,000 stocks. For each stock listed in the file, the respective closing indexes are grouped into multiple records (rows) in the file. The exact number of records varies from stock to stock. The first field of each stock record group is the stock identity (4 characters exactly, no embedded blank). It is then followed by its 100 closing indexes (standard numeric). The last field of each record is always ended with a number or a comma. The fields in each record are separated by commas. There is no missing value in the records. The table below shows a subset of records in StockPrice.csv:

ABCD,1.41,1.78,1.28,1.73,1.54,0.98,1.93,1.99,2.82,3.10,3.59,3.68,1.54,1.06,1.21,1.25,1.39,1.07,1.30,1.18,1.15,1.38,1.01,0.56,1.12,1.06,1.02,1.02,0.39,0.33,1.77,2.13,2.86,2.44,2.33,2.22,1.17,1.12,1.05,1.24,1.35,2.21,1.26,1.30,1.39,1.72,1.23,0.97,1.89,2.17,3.06,3.34,3.99,3.74,1.88,1.58,2.30,2.73,2.75,3.04,1.18,1.13,1.95,2.25,2.82,2.22,1.47,1.68,1.38,1.15,1.81,1.21,1.45,1.12,1.28,1.73,2.46,1.97,1.61,0.83,0.42,0.24,0.42,0.25,1.08,1.09,0.75,0.80,0.95,1.31,1.63,1.99,1.89,2.03,2.60,2.38,2.00,1.73,2.04,2.98
EFGH,10.94,10.49,10.33,10.27,10.26,10.68,10.44,10.58,10.98,10.12,10.17,10.99,10.90,10.32,10.84,10.13,10.37,10.35,10.65,10.77,10.91,10.20,10.19,10.23,10.69,10.91,10.18,10.63,10.24,10.86,10.86,10.17,10.83,10.13,10.86,10.47,10.12,10.41,10.67,10.34,10.02,10.31,10.95,10.39,10.31,10.65,10.71,10.87,10.40,10.89,10.41,10.80,10.96,10.74,10.33,10.26,10.48,10.10,10.17,10.81,10.05,10.58,10.91,10.32,10.56,10.69,10.54,10.34,10.63,10.45,10.98,10.80,10.93,10.15,10.89,10.25,10.18,10.48,11.00,10.68,10.39,10.00,10.23,10.38,10.48,10.07,10.68,10.18,10.54,10.44,10.50,10.08,10.84,10.47,10.52,10.40,10.30,10.03,10.24,10.86

程序代写代做 CS编程辅导

Write a SAS DATA step that will perform the following activities:

- Create four SAS data sets. Name each data set as Allstock, Low_volatile, Mid_volatile, and High_volatile, respectively. These four data sets must be kept in the Work library of SAS Studio. You must create all four SAS data sets in this DATA step.
- Read the record from the data file to SAS to create one observation for each stock regardless of the number of records in the data file.
- Compute the minimum, maximum, and standard deviation of the 100 closing indexes for each stock.
- Compute the standard deviation of the 100 closing indexes for each stock. Ensure that only 4 decimal places of standard deviation are kept in the created data set.
- Send one observation for each stock to the 4 SAS data sets created in (1) according to the following specifications:
 - Data set Allstock contains the observations of all stocks.
 - Data set Low_volatile contains only observations for the stocks with a standard deviation of the respective indexes less or equal to 5.
 - Data set Mid_volatile contains only observations for stocks with a standard deviation of the respective indexes higher than 5 but lower than 10.
 - Data set High_volatile contains only observations for the stocks with a standard deviation of the respective indexes equal to 10 or higher.
- Each of the four SAS data sets must contain only these four variables: stock identity, maximum, minimum, and the standard deviation of its 100 closing indexes.

You must accomplish the above activities with only one DATA step. Using any other procedures is not allowed. The above activities may be carried out in any order within the DATA step. You may name the variables in any way you want as long as they are valid SAS variable names and meaningful. The variables in the SAS data sets can be arranged in any order.

Question 2

Suppose the raw data file Survey.txt (not available in Canvas) contains a recent household survey result. The Filename reference 'Surref' has already been referred to in SAS Studio. The data file is hierarchical in structure. It consists of a header record for a household and is immediately followed by one record (row) for each household member, if applicable. For example, a household of three members will have three records exactly after its header record. A household may have no identified member at the time of the survey. In that case, the household will have only a header record in the file. There is no missing value in the data file. The fields in each record are separated by commas. A subset of records in Survey.txt is displayed below:

A1234567BC012,A
 15/FEB/1980,Y,Male,Married,B,FT,55000
 3/JUN/1982,N,Female,Married,3,UE,0
 24/JAN/2005,N,Male,Unknown,2,NA,0
 D135EG023456789,
 19/OCT/1950,Y,Fe,5000
 X123A567F9,A
 B2345234CC,A
 21/MAY/1975,N,Ma,000
 30/JUN/1978,Y,Fe,10000



The fields are arranged in the following order in each record:

Header record of a household:

Field	Format
Household identity	10-15 characters, always begin with an alphabetical letter in uppercase, such as A, B, etc.
Type of housing	Character 'A' for private and 'E' for public

Household member's record (if present):

Field	Format
Date of birth	10-11 characters in the form dd/mm/yyyy where dd is the day value which may be in the form of 1 or 2 digits.
Householder indicator	Character 'Y' for yes, and 'N' for no.
Gender	Character 'F' for female, or 'M' for male.
Marital status	Characters of 'Married', 'Single', 'Divorced', or 'Unknown'.
Achieved education level	Standard numeric with 0 for none, 1 for primary, 2 for secondary, 3 for tertiary.
Employment status	Characters of 'FT', 'PT', 'UE', or 'NA'.
Monthly income	Standard numeric.

Write a DATA step that will perform the following activities:

- Create a SAS data set named Members. Members must be stored in the Work library of SAS Studio.
- Read the records from Survey.txt into Members.
- Create only one observation for each household in Survey.txt.
- Each created observation in Members must contain only these variables but not necessarily in the given order:
 - The identity of the respective household
 - The type of housing of the respective household
 - The number of members in the household (0 for a household without members)
 - The number of male members (0 for a household without members)
 - The maximum achieved education level among all members in the household (0 for a household without members)
 - The total monthly income of the household (0 for a household without members)
 - The number of members in the household is at least 18 years old as of 1 January 2023 (0 for a household without members) {Hint: Use function INTCK to compute

the number of years between two dates. For example, `INTCK('YEAR', '31Jan2010'd, '1Jan2012'd, 'c')` returns a value of 1, because there is 1 full year between the two dates.}

The first few observations are shown below for reference:

Household Identity	Type of Household	Number of Male Members in the Household	Maximum Achieved Education Level	Total Monthly Income	Number of Members >= 18 years
A1234567BC012	A	2	3	55000	2
D135EG023456789	B	0	0	5000	1
X123A567F9	A	0	0	0	0
B2345234CC	A	1	2	40000	2

The labels displayed in the above picture are for illustration purposes only. You do not need to include them in your DATA step.

You must accomplish the above activities with only one DATA step. Using any other procedures is not allowed. The activities may be carried out in any order within the DATA step. You may name the variables in any way you want as long as they are valid SAS variable names and meaningful. The variables in the SAS data sets can be arranged in any order.

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>