|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | **STAT S366F** | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **2022 Autumn Laboratory 4-5 (UG)**  **SAS Programming**   |  |  | | --- | --- | | **Date to release Assignment** | **Submission**  **Deadline** | | 26th September, 2022 | 11th October, 2022 | | |  | | --- | | Total Mark:  **/100** | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | LO TSZ KIN | | | | | | | |
| Student ID: |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Instructions:**   1. **Write your answer on the space provided**. 2. You may submit a single pdf file which consists of all solutions in numeral order. 3. You are advised to keep a copy of your laboratory answer booklet in case of loss. | |  |  |  | | --- | --- | --- | | For Instructor Uses only | **Q. No.** | **Mark** | | **1** | **/50** | | **2** | **/50** | |  |  | |  |  | |  |  | |  |  | |  |  | | **Total** | |  | |

|  |
| --- |
| **Laboratory 4-5: Working on your data** |

**Learning Outcome:**

Upon completion of this laboratory, learners should be able to

* Write programs to load data from different file formats
* Create and modify variables
* Select appropriate data format for data storage
* Demonstrate the ability to perform data summarization tasks

|  |
| --- |
| **Question 1. If-then-do statements (50%)** |

1. Import the dataset ‘lab4.dat’ into SAS
2. Read variables **Model**, **Year**, **Brand**, and **Price**.
3. Create a new variable called **cost** and group the observations according to the following criteria: IF the price is smaller than $10,000, assign LOW. If the price is higher than $30,000, assign HIGH, otherwise assign MEDIUM. Missing price should be displayed as “MISSING”.
4. If the model is “Corvette”, assign **brand** as “Chevy”. If the model is “Miata”, assign the brand as “Mazda”. If the model is “Supra”, assign the brand as “Toyota”.
5. Create a new variable called **country**. Assign the observations with brand “Mazda”, “Honda” and “Toyota” as “Japan”. For the remaining ones, assign “US”.
6. Print the subset of observations with **country** “Japan”.

(25%)

|  |
| --- |
| **Write down your codes and paste the output obtained from SAS here**:  Data Lab4;  Infile '/home/u62236541/SAS totural/Lab 4/lab4\_2022.dat';  INPUT Model $1-9  Year 10-13  Brand $15-20  price :COMMA9.;  Length cost $10;  Length country $5;  If price = '.' then cost = '.';  else IF price < 10000 then cost = "LOW";  else if price > 30000 then cost = "HIGH";  else cost = "MISSING";  IF Model = "Corvette" Then do;  Brand = "Chevy";  country = "USA";  END;  Else IF Model = "Miata" then do;  Brand = "Mazda";  country = "Japan";  END;  Else if Model = "Supra" then do;  Brand = "Toyota";  country = "Japan";  END;  IF Brand in ("Mazda", "Toyota" ,"Honda") then country = "Japan";  Else country = "USA";  IF country = "Japan";  RUN;  PROC Print Data = Lab4;  Format Price Dollar10.; |

1. Briefly explain how your code works. You should explain them line by line. (25%)

|  |
| --- |
| Line 1: Is set the table name as “Lab 4”  Line 2 : Set the raw data read from Location  Line 3- Line 8: Mean Turn the Raw data into read data with the name and set some of that to string ,  Line 10-13 : Set the Cost as high or low from reading Price which is lower then 10000 or higher then 30000 .Or if data is “.” Mean missing.  Line 15-26 : Set the Model , brand and country as If the model is “Corvette”, assign **brand** as “Chevy”. If the model is “Miata”, assign the brand as “Mazda”. If the model is “Supra”, assign the brand as “Toyota”.  Line 27-29 : Set a new field as country which brand “Mazda”, “Honda” and “Toyota” as “Japan”. For the remaining ones, assign “US”.  Line 33-34 is print the data out |

Use comments to separate the codes for each step clearly. Styles and commenting do count on the overall marks.

|  |
| --- |
| **Question 2. Using iterative DO, Arrays and retain function (50%)** |

1. Open your previously submitted Lab3.
2. Create an array to read variables OldMaths, OldEnglish, OldChinese, OldPhysics and OldICT.
3. Use iterative DO and arrays to perform the following operation: If ANY of the scores OldMaths, OldEnglish, OldChinese, OldPhysics or OldICT is smaller than 0, set them to zero and store them in new variables Maths, English, Chinese, Physics and ICT. Othewise, store the original value in the new variables.
4. Create a new variable MAXSCORE, which finds the maximum of Maths, English, Chinese, Physics and ICT for each record.
5. For each variable Maths, English, Chinese, Physics and ICT, calculate the **cumulative score** over each row and save them as variables MathsTotal, EnglishTotal, ChineseTotal, PhysicsTotal and ICTTotal, respectively.
6. Print a table for all variables **excluding** OldMaths, OldEnglish, OldChinese, OldPhysics and OldICT with a title ‘Lab45’.

(25%)

|  |
| --- |
| **Write down your codes and paste the output obtained from SAS here**:  LIBNAME LAB3 '/home/u62236541';  data lab3.partA;  infile "/home/u62236541/SAS totural/Lab 3/lab3\_2022.dat" DLM = ' ';  Input Name $ OldMaths OldEnglish OldChinese OldPhysics OldICT;  Array array1(5) OldMaths OldEnglish OldChinese OldPhysics OldICT;  Array newarray (5) Maths English Chinese Physics ICT;  Array totalarr(5) Mathtotal ENGtotal Chitotal PHYtotal ICTtotal;  Do i = 1 to 5;  if array1(i) ='.' then newarray(i) = '.';  else If array1(i) < 0 then newarray(i) = 0;  else newarray(i) = array1(i) ;  totalarr(i)=newarray(i);  END;    MAXSCORE = max(newarray(1),newarray(2),newarray(3),newarray(4),newarray(5));  DROP OldMaths OldEnglish OldChinese OldPhysics OldICT i;  RUN;  proc print data = Lab3;  var Lab3;  Title 'Summary performance of students';    RUN; |

1. Briefly explain how your code works. You should explain them line by line. (25%)

|  |
| --- |
| Line 1-3 : Mean Read data from Library where data name = “lab3.partA” and from the location .  Line 6-8 :Save the data as array and Name it .  Line 9-14 mean : If ANY of the scores OldMaths, OldEnglish, OldChinese, OldPhysics or OldICT is smaller than 0, set them to zero and store them in new variables Maths, English, Chinese, Physics and ICT. Othewise, store the original value in the new variables.  Line 16 mean calculate the Max value from the array  Line 17: Delete the data from the table .  Line 21-25 : Print the data out and Name the title as ‘Summary performance of students'. |

Use comments to separate the codes for each step clearly. Styles and commenting do count on the overall marks.

-END-