1. This project will use data set cert.input08a and cert.input08b. Both data sets contain a common numeric variable named **ID**.

Write a program that will use a SAS DATA step to

Combine data sets **cert.input08a** and**cert.input08b** by matching values of **ID**

Write only observations that are in both data sets to new data set named **results.match08**

Write non-matching variables to **results.nomatch08**

And answer these 4 Questions:

How many obs in **results.match08**? 1200

How many var in **results.match08**? 117

How many obs in **results.nomatch08**? 2

How many var in **results.nomatch08**? 5

Proc sort data=cert.input08a out=work.input08a;

By ID;

Run;

Proc sort data=cert.input08b out=work.input08b;

By ID;

Run;

Data results.match08 results.nomatch08 (drop=ex: );

Merge work.input08a (in=a) work.input08b (in=b);

By ID;

If a and b then output results.match08;

Else output results.nomatch08;

Run;

Proc contents data=results.match08; run;

Proc contents data=results.nomatch08; run;

1. This project will use data set **cert.input36**. Write a SAS program that will clean the data in **cert.input36** as follows:

**Step 1:** Create temp dataset **cleandata36**

Convert all **group** values to upper case

Keep only observations with **group** equal to ‘A’ or ‘B’

**Step 2:** Determine the median value for the **kilograms** variable for each **group** (A,B) in the **cleandata36** dataset. Round median to nearest whole number.

**Step 3**: Create **results.output36** from **cleandata36**

Ensure all values for **kilograms** are between 40 and 200 inclusive

If the value is missing or out of range, replace with median for Group A or B

What is the MEAN **Kilograms** value for **Group** B in the **results.output36** dataset? 86.5

How many observations in **results.output36?** 4992

What is the MEAN **Kilograms** value for **Group** A in the **results.output36** dataset? 86.5

Data work.cleandata36;

Set cert.input36;

Group=upcase(group);

If upcase(group) in (‘A’,’B’);

Run;

Proc means data=work.cleandata36 median;

Class group;

Var kilograms;

Run;

Data results.output36;

Set cleandata36;

If kilograms < 40 or Kilograms > 200 then do;

If Group=’A’ then kilograms=79;

Else kilograms=89;

End;

Run;

Proc means data=results.output36 mean;

Class group;

Var kilograms;

Run;

1. This project will use dataset **cert.input04**. Write a SAS program that will create dataset **results.output04**.

-Round **Var1** and **Var2** to the nearest integer values.

-Multiply the newly rounded **Var1** by **Var2** to define **Var3**

-Add **Var12** through **Var19** (8 variables) together, treating missing values as 0. Assign the

Sum to **Var20**

For observation 16 what is the value of **Var3?** 80136

For observation 16 what is the value of **Var20?** 3175

Data results.output04;

Set cert.input04;

Var3=round(var1,1)\*round(var2,1);

Var20= sum(of var12-var19);

Run;

Proc print data=results.output04; run;

1. This project will use dataset **cert.input12** which contains a single observation with two variables, **salary** and **year**. Write a SAS program that will:

-Create an output dataset **results.output12**

-read **cert.input12** as input

-Increase the **salary** variable by 5.65% annually until it is greater than $500,000

-Increase the **year** variable by 1 with each annual increase.

-Create an output dataset **results.output12** that has one observation for each value

Of **year.** Each observation should have a **year** and **salary** variable.

What is the maximum **salary** that is less than 500k? 498737

What **year** does the above salary occur? 2027

Data results.output12;

Set cert.input12;

Do until (salary gt 500000);

Salary=salary\*1.0565;

Year+1;

Output;

End;

Run;

Proc print data=results.output12; run;

1. This project will use dataset **cert.input13**. This dataset contains 1001 observations and 2 variables, **Date1** an unformatted SAS date value and **Charnum** a character value representing a monetary amount. Write a SAS program that will:

-Save the new dataset as **results.output12**

-Create a new variable chdate that converts the **date1** variable to a character

Variable that is in the format ddmonyyyy

-Create a new variable **num1** that converts the **Charnum** variable to a numeric variable

What is the value of **Chdate** for observation 52? 30DEC1992

What is the average (mean) of the **num1** variable for the entire dataset? 51763

Data results.output13;

Set cert.input13;

Chdate=put (date1,date9.);

Num1=input(charnum, dollar7.);

Run;

Proc means data=results.output13;

Var num1;

Run;

1. This project will use dataset **cert.input27**. Write a SAS program that will:

-Create output dataset **results.output27a** as a subset of **cert.input27** where the

**Country** variables value is US or any variation i.e. us

**-**Sort **results.output27a** first by the variable **state** in ascending order, then by

**Postal\_Code** in descending order.

What is the value of **Employee\_ID** for observation 100 in **results.output27a?** 120781

Proc sort data=cert.input27 out=results.output27a (where=(upcase(country)=’US’));

By state descending Postal\_Code;

Run;

Proc print data=results.output27a; run;

**6b)** Continuing with the previous program add code that satisfies the following criteria:

-Create outpute dataset **results.output27b** from **cert.input27**

-sort the observations by **Postal\_Code** in descending order

-Remove duplicate values of **Postal\_Code** keeping only the first occurrence

What is the value of **Employee\_ID** for observation 98 in **results.output27b?** 120779

What is the value of **Employee\_ID** for observation 181 in **results.output27b**? 120272

Proc sort data=cert.input27 out=results.output27b nodupkey;

By descending postal\_code;

Run;

Proc print data=results.output27b; run;

1. Open the existing program, **program44.sas** from folder **C:\cert\errors**. This program is intended to:

-Create a new dataset using cert.input44 as input

-Drop variables **bp\_status, weight\_status,** and **smoking\_status**

-Create a new column, **chol\_status**, based on the following values

Less than 200 “Safe”

200-239 “High-Borderline”

240 and up “High”

-Should not calculate **chol\_status** for missing cholesterol values

How many variables are in the **work.out** dataset? 5

How many observations are in the “high” group? 1791

How many observations are in the “safe” group? 1405

Data out;

Set cert.input44;

**Length chol\_status $ 15;**

**Drop bp\_status** weight\_status smoking\_status;

If cholesterol **ne .** then do;

If cholesterol < 200 then chol\_status=’Safe’;

Else if cholesterol<= 239 then chol\_status=’High-Borderline’;

Else if cholesterol>239 then chol\_status=’High’;

**End;**

Run;

Proc contents data=work.out; run;

Proc freq **data=**work.out;

Table chol\_status;

Run;

1. Open the existing program, **program48.sas** from folder **C:\cert\errors**. This program is intended to:

-Create 3 groups for **Cvar** A-G is **Group=1** H-N for **Group=2** and O-Z for **Group=3**

-All variations should be in the same group i.e A vs a

-Calculate average **X** and  **Y** by **group**

What is the Average for **X** for **Group=2**? 47

What is the Average for **Y** for **Group=2**? 703

Data groups;

Set cert.input48;

If upcase(cvar) in (‘A’,’B’,’C’,’D’,’E’,’F’,’G’,) then group=1;

Else if upcase(cvar) in (‘H’,’I’,’J’,’K’,’L’,’M’,’N’) then group=2;

Else group=3

**new\_y=input (y,7.3);**

run;

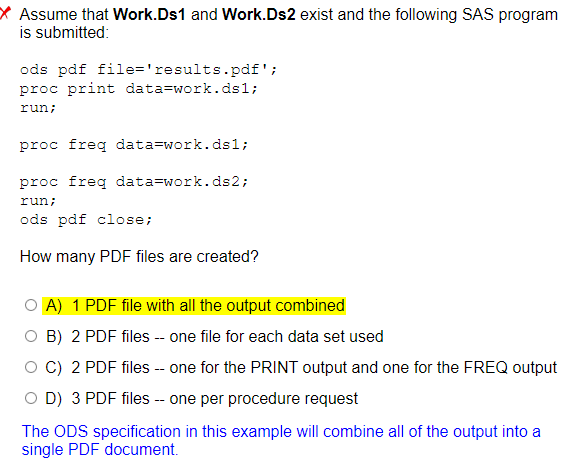
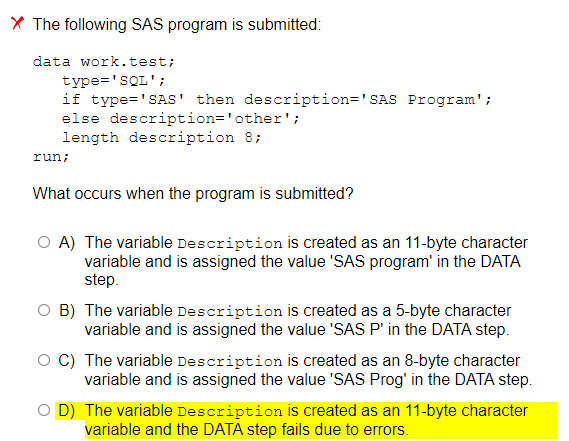
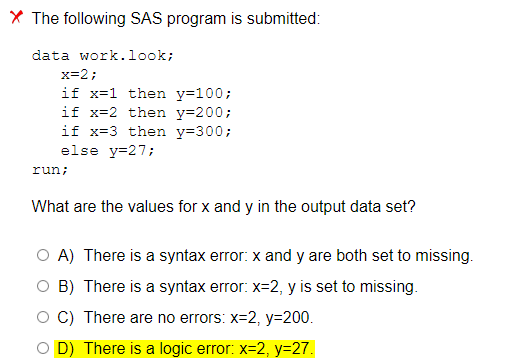
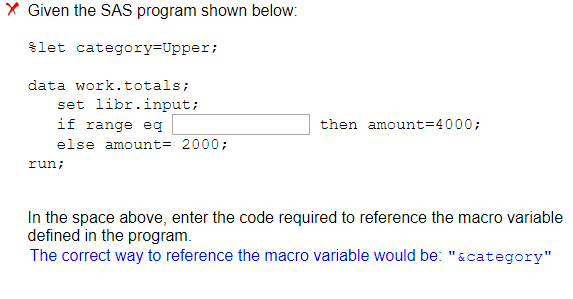
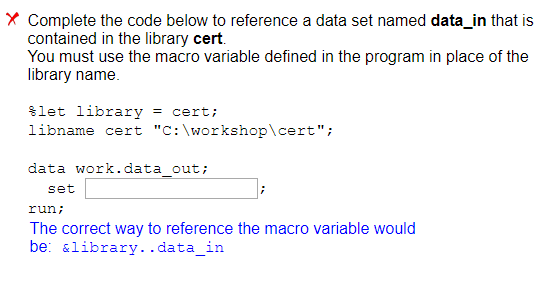
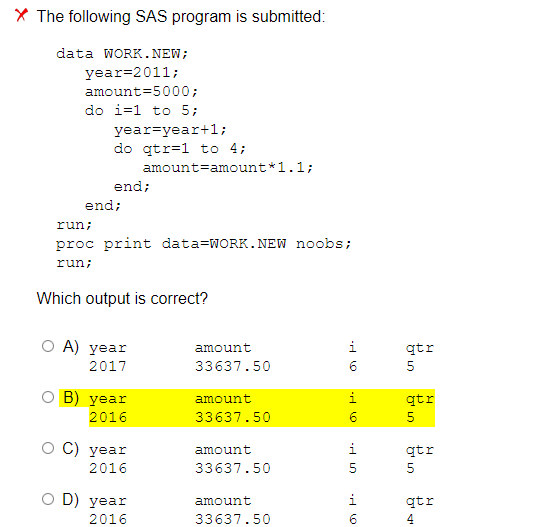
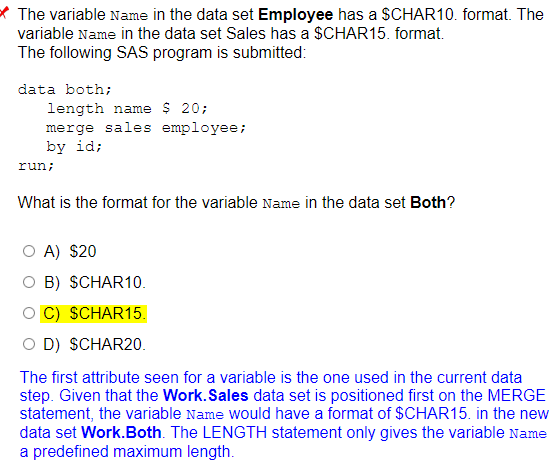
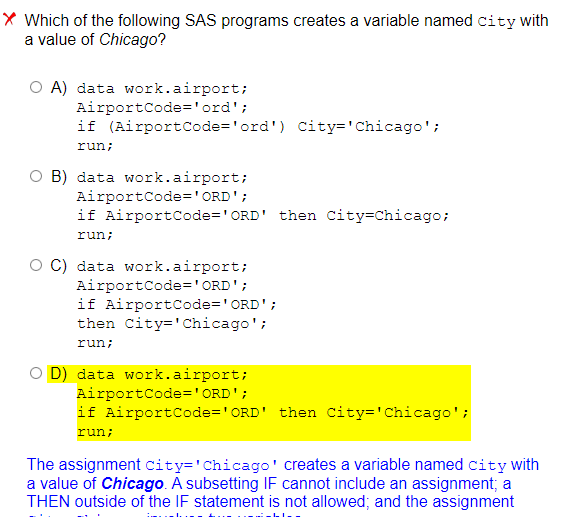
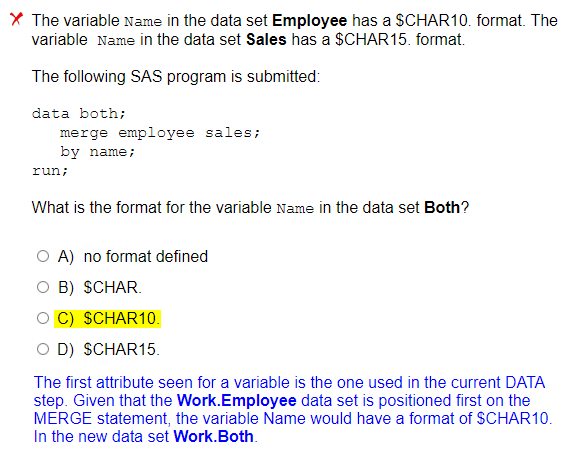
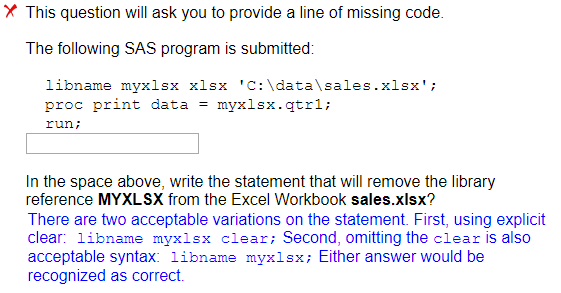
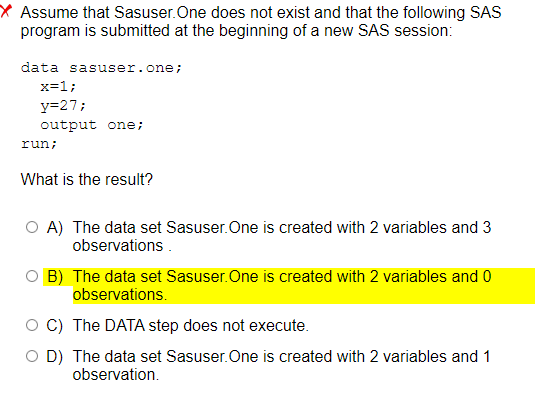
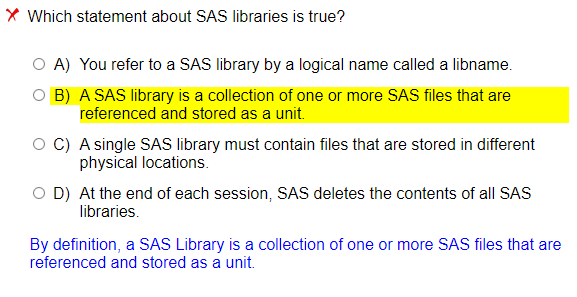
proc means data=groups **mean** maxdec=2;

where group = 2;

class group;

var x **new\_y**;

run;

/