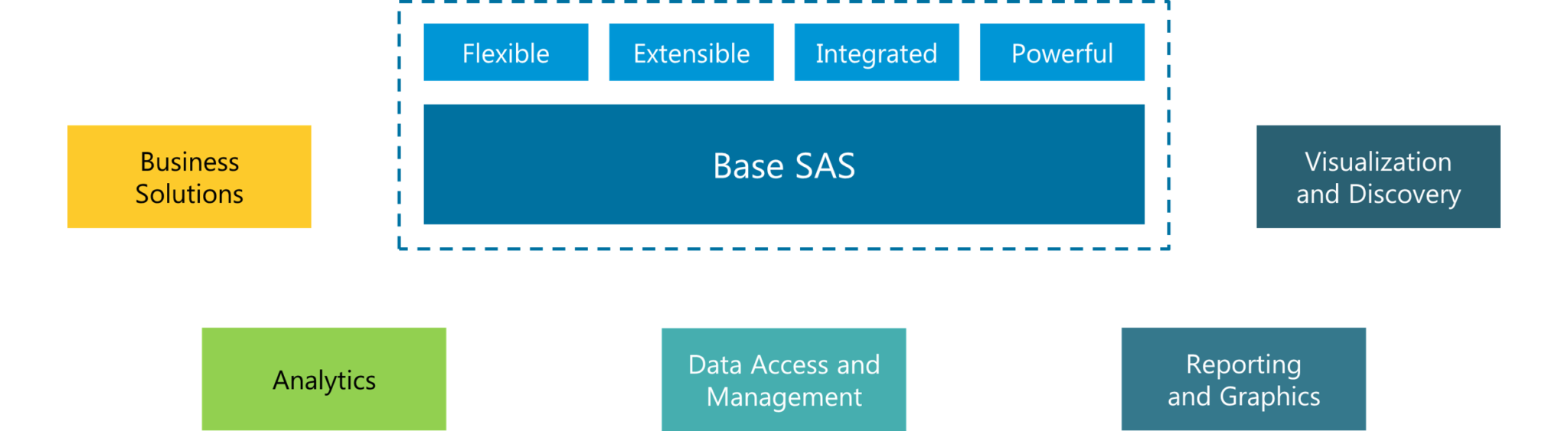
**1. What is SAS?**

**Answer:** SAS (Statistical Analytics System) is a software suite for advanced analytics, multivariate analyses, business intelligence, data management and predictive analytics.

It is developed by SAS Institute. SAS provides a graphical point-and-click user interface for non-technical users and more advanced options through the SAS language.

**2. What are the features of SAS?**

**Answer:** The following are the features of SAS:

**Figure:** *SAS Interview Questions – Features of SAS*

1. **Business Solutions:**SAS provides business analysis that can be used as business products for various companies to use.
2. **Analytics:**SAS is the market leader in the analytics of various business products and services.
3. **Data Access & Management:**SAS can also be use as a DBMS software.
4. **Reporting & Graphics:**Hello SAS helps to visualize the analysis in the form of summary, lists and graphic reports.
5. **Visualization:**We can visualize the reports in the form of graphs ranging from simple scatter plots and bar charts to complex multi-page classification panels.

**3. Compare SAS with other data analytics tools.**

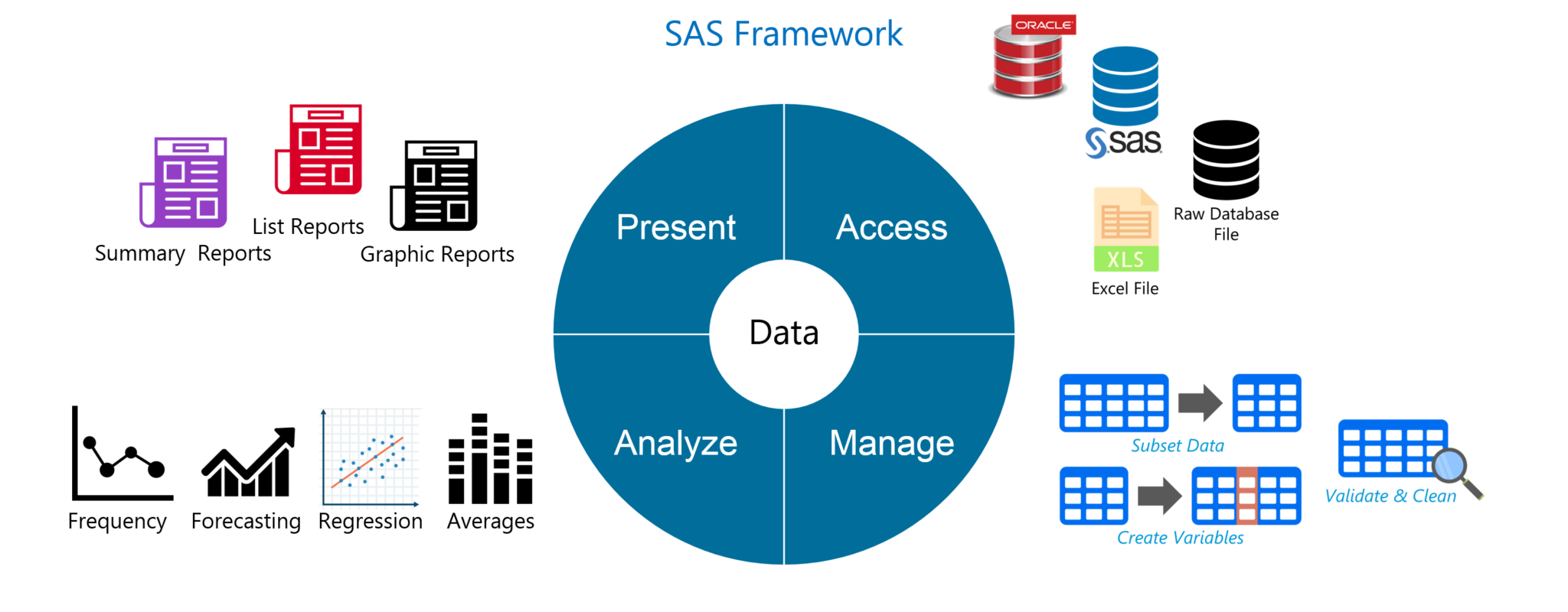
**Answer:** We will compare SAS with the popular alternatives in the market based on the following aspects:

1. **Ease of Learning**   
   SAS is easy to learn and provides easy option (PROC SQL) for people who already know SQL. R on the other hand has a very steep learning curve as it is a low level programming language.
2. **Data Handling Capabilities**  
   SAS is on par with all leading tools including R & Python when it comes to handling huge amount of data and options for parallel computations.
3. **Graphical Capabilities**  
   SAS provides functional graphical capabilities and with a little bit of learning, it is possible to customize on these plots.
4. **Advancements in Tool**  
   SAS releases updates in controlled environment, hence they are well tested. R & Python on the other hand, have open contribution and there are chances of errors in latest developments.
5. **Job Scenario**  
   Globally, SAS is the market leader in available corporate jobs. In India, SAS controls about 70% of the data analytics market share compared to 15% for R.

**GET STARTED WITH SAS**

**4. Mention few capabilities of SAS Framework.**

**Answer:**The following are the four capabilities in SAS Framework:

**Figure:***SAS Interview Questions – SAS Framework*

1. **Access:**As we can learn from the figure, SAS allows us to access data from multiple sources like an Excel file, raw database, Oracle database and SAS Datasets.
2. **Manage:** We can then manage this data to subset data, create variables, validate and clean data.
3. **Analyze:** Further, analysis happens on this data. We can perform simple analyses like frequency and averages and complex analyses including regression and forecasting. SAS is the gold standard for statistical analyses.
4. **Present:** Finally we can present our analysis in the form of list, summary and graphic reports. We can either print these reports, write them to data file or publish them online.

**5. What is the function of *output* statement in a SAS Program?**

**Answer:**You can use the OUTPUT statement to save summary statistics in a SAS data set. This information can then be used to create customized reports or to save historical information about a process.

You can use options in the OUTPUT statement to

1. Specify the statistics to save in the output data set,
2. Specify the name of the output data set, and
3. Compute and save percentiles not automatically computed by the CAPABILITY procedure.

**6. What is the function of *Stop* statement in a SAS Program?**

**Answer:***Stop statement*causes SAS to stop processing the current data step immediately and resume processing statement after the end of current data step.

**7. What is the difference between using drop = data set option in data statement and set statement?**

**Answer:**If you don’t want to process certain variables and you do not want them to appear in the new data set, then specify drop = data set option in the *set* statement.

Whereas If want to process certain variables and do not want them to appear in the new data set, then specify drop = data set option in the *data* statement.

**8. Given an unsorted data set, how to read the last observation to a new data set?**

**Answer:**We can read the last observation to a new data set using end= data set option.

For example:

data work.calculus;  
set work.comp end=last;  
If last;  
run;

Where *calculus* is a new data set to be created and *comp* is the existing data set. *last* is the temporary variable (initialized to 0) which is set to 1 when the set statement reads the last observation.

**9. What is the difference between reading data from an external file and reading data from an existing data set?**

**Answer:**The main difference is that while reading an existing data set with the SET statement, SAS retains the values of the variables from one observation to the next. Whereas when reading the data from an external file, only the observations are read. The variables will have to re-declared if they need to be used.

**10. How many data types are there in SAS?**

**Answer:**There are two data types in SAS. Character and Numeric. Apart from this, dates are also considered as characters although there are implicit functions to work upon dates.

**11. What is the difference between SAS functions and procedures?**

**Answer:**Functions expect argument values to be supplied across an observation in a SAS data set whereas a procedure expects one variable value per observation.

For example:

data average ;  
set temp ;  
avgtemp = mean( of T1 – T24 ) ;  
run ;

Here arguments of mean function are taken across an observation. The mean function calculates the average of the different values in a single observation.

proc sort ;  
by month ;  
run ;  
proc means ;  
by month ;  
var avgtemp ;  
run ;

Proc means is used to calculate average temperature by month (taking one variable value across an observation). Here, the procedure *means*on the variable month.

**12. What are the differences between sum function and using “+” operator?**

**Answer:**SUM function returns the sum of non-missing arguments whereas “+” operator returns a missing value if any of the arguments are missing.

Example:  
data mydata;  
input x y z;  
cards;  
33 3 3  
24 3 4  
24 3 4  
. 3 2  
23 . 3  
54 4 .  
35 4 2  
;  
run;

data mydata2;  
set mydata;  
a=sum(x,y,z);  
p=x+y+z;  
run;

In the output, value of p is missing for 4th, 5th and 6th observation as:

a p

39 39

31 31

31 31

5 .

26 .

58 .

41 41

**13. What are the differences between PROC MEANS and PROC SUMMARY?**

**Answer:**PROC MEANS produces subgroup statistics only when a BY statement is used and the input data has been previously sorted (using PROC SORT) by the BY variables.

PROC SUMMARY automatically produces statistics for all subgroups, giving you all the information in one run that you would get by repeatedly sorting a data set by the variables that define each subgroup and running PROC MEANS. PROC SUMMARY does not produce any information in your output. So you will need to use the OUTPUT statement to create a new DATA SET and use PROC PRINT to see the computed statistics.

**14. Give an example where SAS fails to convert character value to numeric value automatically?**

**Answer:**Suppose value of a variable PayRate begins with a dollar sign ($). When SAS tries to automatically convert the values of PayRate to numeric values, the dollar sign blocks the process. The values cannot be converted to numeric values.

Therefore, it is always best to include *INPUT* and *PUT* functions in your programs when conversions occur.

**LEARN SAS FROM EXPERTS**

**15. How do you delete duplicate observations in SAS?**

**Answer:**There are three ways to delete duplicate observations in a dataset:

1. By using nodups in the procedureProc sort data=SAS-Dataset nodups;  
   by var;  
   run;

2. By using an SQL query inside a procedure

Proc sql;  
Create SAS-Dataset as select \* from Old-SAS-Dataset where var=distinct(var);  
quit;

3. By cleaning the data

Set temp;  
By group;  
If first.group and last.group then  
Run;

**16. How does PROC SQL work?**

**Answer:**PROC SQL is a simultaneous process for all the observations. The following steps happen when PROC SQL is executed:

1. SAS scans each statement in the SQL procedure and check syntax errors, such as missing semicolons and invalid statements.
2. SQL optimizer scans the query inside the statement. The *SQL Optimizer* decides how the SQL query should be executed in order to minimize run time.
3. Any tables in the *FROM*statement are loaded into the data engine where they can then be accessed in memory.
4. Code and Calculations are executed.
5. Final Table is created in memory.
6. Final Table is sent to the output table described in the SQL statement.

**17. Briefly explain Input and Put function?**

**Answer:**Input function – Character to numeric conversion- Input(source,informat)

put function –  Numeric to character conversion- put(source,format)

**18. What would be the result of the following SAS function (given that 31 Dec, 2000 is Sunday)?  
    Weeks  = intck (‘week’,’31 dec 2000’d,’01jan2001’d);  
Years    = intck (‘year’,’31 dec 2000’d,’01jan2001’d);  
    Months = intck (‘month’,’31 dec 2000’d,’01jan2001’d);**

**Answer:**Here, we will calculate the weeks between 31st December, 2000 and 1st January, 2001. 31st December 2000 was a Sunday. So 1st January 2001 will be a Monday in the same week. Hence, Weeks = 0

Years = 1, since both the days are in different calendar years.

Months = 1 ,since both the days are in different months of the calendar.

**19. Suppose the variable address stores the following expression:**

**209 RADCLIFFE ROAD, CENTER CITY, NY, 92716**

**What would be the result returned by the scan function in the following cases?  
 a=scan(address,3);**  
**b=scan(address,3,’,’);**

**Answer:**a=Road; b=NY

**20. What is the length assigned to the target variable by the scan function?**

**Answer:**200

**21. Name few SAS functions?**

**Answer:**Scan, Substr, trim, Catx, Index, tranwrd, find, Sum.

**22. What is the work of tranwrd function?**

**Answer:***TRANWRD* function replaces or removes all occurrences of a pattern of characters within a character string.

**23. Consider the following SAS Program**

data finance.earnings;  
Amount=1000;  
Rate=.075/12;  
do month=1 to 12;  
Earned+(amount+earned)\*(rate);  
end;  
run;

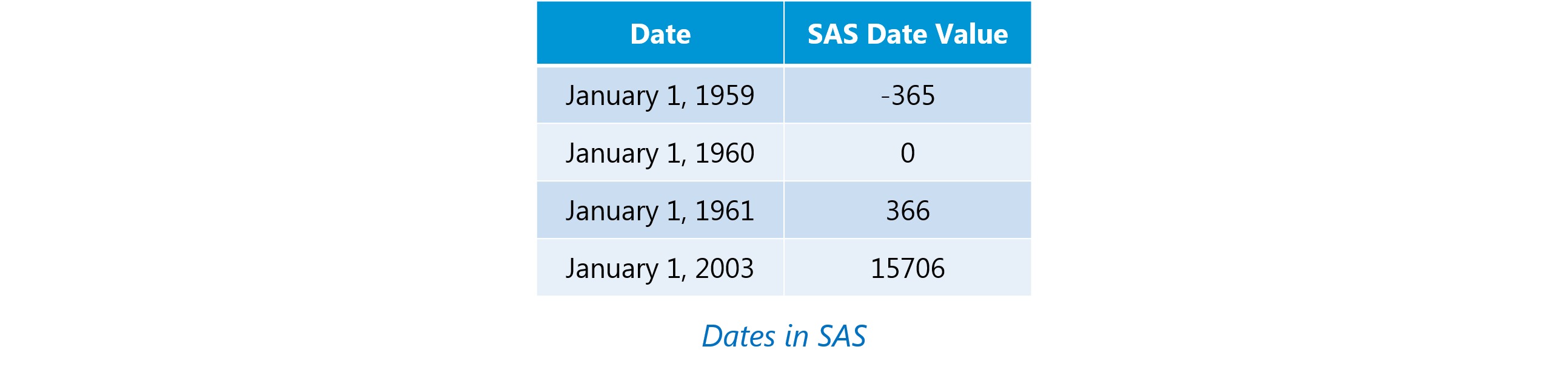
**What would be the value of month at the end of data step execution and how many observations would be there?**

**Answer:**Value of month would be 13

No. of observations would be 1

**24. How do dates work in SAS data?**

Data is central to every data set. In SAS, data is available in tabular form where variables occupy the column space and observations occupy the row space.

* SAS treats numbers as numeric data and everything else falls under character data. Hence SAS has two data types numeric and character.
* Apart from these, dates in SAS are represented in a special way compared to other languages.  
  

**Figure:** *SAS Interview Questions – SAS Dates*

* A SAS date is a numeric value equal to the number of days since January 1, 1960.
* Apart from Date Values, there are many tools to work on dates such as informats for reading dates, functions for manipulating dates and formats for printing dates.

**25. Consider the following SAS Program**

data finance;  
Amount=1000;  
Rate=.075/12;  
do month=1 to 12;  
Earned+(amount+earned)\*(rate);  
output;  
end;  
run;

**How many observations would be there at the end of data step execution?**

**Answer:**12

**26. How do you use the do loop if you don’t know how many times you should execute the do loop?**

**Answer:**We can use ‘*do until’* or *‘do while’*to specify the condition.

**27. What is the difference between *do while* and *do until*?**

**Answer:**An important difference between the *DO UNTIL* and *DO WHILE* statements is that the DO WHILE expression is evaluated at the top of the DO loop. If the expression is false the first time it is evaluated, then the DO loop never executes. Whereas DO UNTIL  executes at least once.

**28. How do you specify the number of iterations and specific condition within a single do loop?**

**Answer:**

data work;  
do i=1 to 20 until(Sum>=20000);  
Year+1;  
Sum+2000;  
Sum+Sum\*.10;  
end;  
run;

This iterative DO statement enables you to execute the DO loop until Sum is greater than or equal to *20000*or until the DO loop executes 10 times, whichever occurs first.

**29. What are the parameters of *Scan* function?**

**Answer:**This is how the scan function is used.

scan(argument,n,delimiters)

Here, *argument* specifies the character variable or expression to scan,

*n*specifies which word to read, and

*delimiters* are special characters that must be enclosed in single quotation marks.

**30. If a variable contains only numbers, can it be a character data type?**

**Answer:**Yes, it depends on how you use the variable. There are some numbers we will want to use as a categorical value rather than a quantity. An example of this can be a variable called “Foreigner” where the observations have the value “0” or “1” representing not a foreigner and foreigner respectively. Similarly, the ID of a particular table can be in number but does not specifically represent any quantity. Phone numbers is another popular example.

**31. If a variable contains letters or special characters, can it be numeric data type?**

**Answer:**No, it must be character data type.

**32. What can be the size of largest dataset in SAS?**

**Answer:**The number of observations is limited only by computer’s capacity to handle and store them.

Prior to SAS 9.1, SAS data sets could contain up to 32,767 variables. In SAS 9.1, the maximum number of variables in a SAS data set is limited by the resources available on your computer.

**33. Give some examples where PROC REPORT’s defaults are different than PROC PRINT’s defaults?**

**Answer:**

* No Record Numbers in Proc Report
* Labels (not var names) used as headers in Proc Report
* REPORT needs NOWINDOWS option

**34. Give some examples where PROC REPORT’s defaults are same as PROC PRINT’s defaults?**

**Answer:**

* Variables/Columns in position order.
* Rows ordered as they appear in data set.

**GET SAS CERTIFIED TODAY**

**35. What is the purpose of trailing @ and @@? How do you use them?**

**Answer:**The trailing @ is also known as a column pointer. By using the trailing @, in the *Input* statement gives you the ability to read a part of your raw data line, test it and then decide how to read additional data from the same record.

* The single trailing @ tells the SAS system to “hold the line”.
* The double trailing @@ tells the SAS system to “hold the line more strongly”.
* An *Input* statement ending with @@ instructs the program to release the current raw data line only when there are no data values left to be read from that line. The @@, therefore, holds the input record even across multiple iteration of the data step.

**36. What is the difference between Order and Group variable in proc report?**

**Answer:**

* If the variable is used as group variable, rows that have the same values are collapsed.
* Group variables produce list report whereas order variable produces summary report.

**37. Give some ways by which you can define the variables to produce the summary report (using proc report)?**

**Answer:**All of the variables in a summary report must be defined as group, analysis, across or computed variables.

**38. What are the default statistics for means procedure?**

**Answer:***n*-count, mean, standard deviation, minimum, and maximum

**39. How to limit decimal places for variable using PROC MEANS?**

**Answer:**By using MAXDEC**=**option

**40. What is the difference between CLASS statement and BY statement in proc means?**

**Answer:**

* Unlike CLASS processing, BY processing requires that your data already be sorted or

indexed in the order of the BY variables.

* BY group results have a layout that is different from the layout of CLASS group results.

**41. What is the difference between PROC MEANS and PROC Summary?**

**Answer:**The difference between the two procedures is that PROC MEANS produces a report by default. By contrast, to produce a report in PROC SUMMARY, you must include a PRINT option in the PROC SUMMARY statement.

**42. How to specify variables to be processed by the FREQ procedure?**

**Answer:**By using TABLES Statement.

**43. Describe CROSSLIST option in TABLES statement?**

**Answer:**Adding the CROSSLIST option to TABLES statement displays crosstabulation tables in ODS column format.

**44. How to create list output for crosstabulations in proc freq?**

**Answer:**To generate list output for crosstabulations, add a slash (/) and the LIST option to the TABLES statement in your PROC FREQ step.

TABLES *variable-1\*variable-2 <\* … variable-n>*/ LIST;

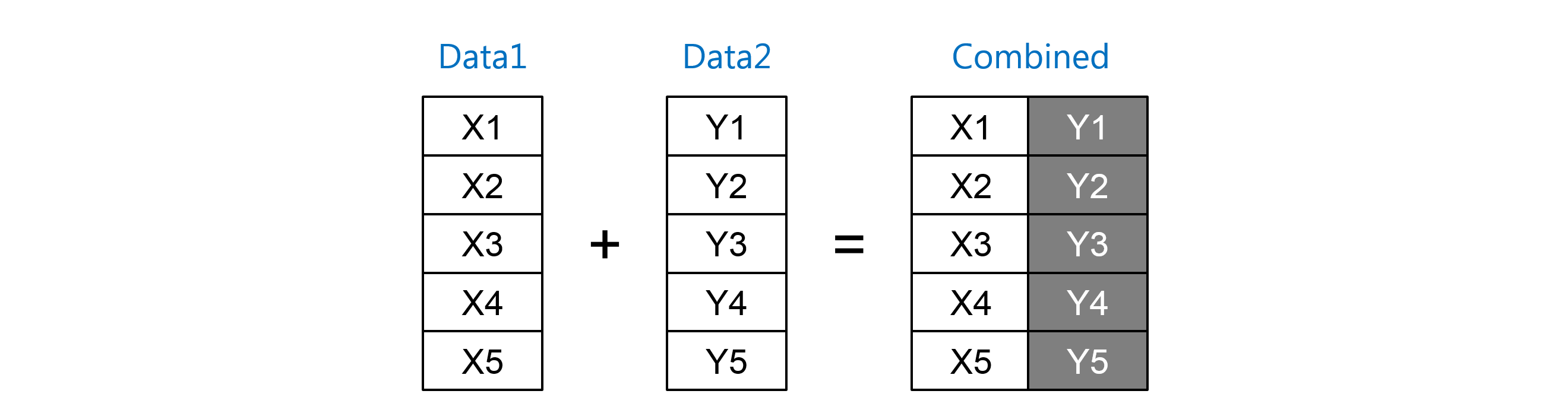
**45. Where do you use PROC MEANS over PROC FREQ?**

**Answer:**We will use PROC MEANS for numeric variables whereas we use PROC FREQ for categorical variables.

**46. Explain how merging helps to combine data sets.**

**Answer:**Merging combines observations from two or more SAS data sets into a single observation in a new data set.

A one-to-one merge, shown in the following figure, combines observations based on their position in the data sets. You use the MERGE statement for one-to-one merging.



We can merge the datasets on one-to-one fashion with the below code.

|  |  |
| --- | --- |
| 1  2  3 | data combined;      merge data1 data2;  run; |

**47. Consider the following SAS Program:**

data concat;  
set a b;  
run;

**where format of variable Revenue in dataset a is dollar10.2 and format  of variable Revenue in dataset b is dollar12.2 . What would be the format of Revenue in resulting dataset (concat)?**

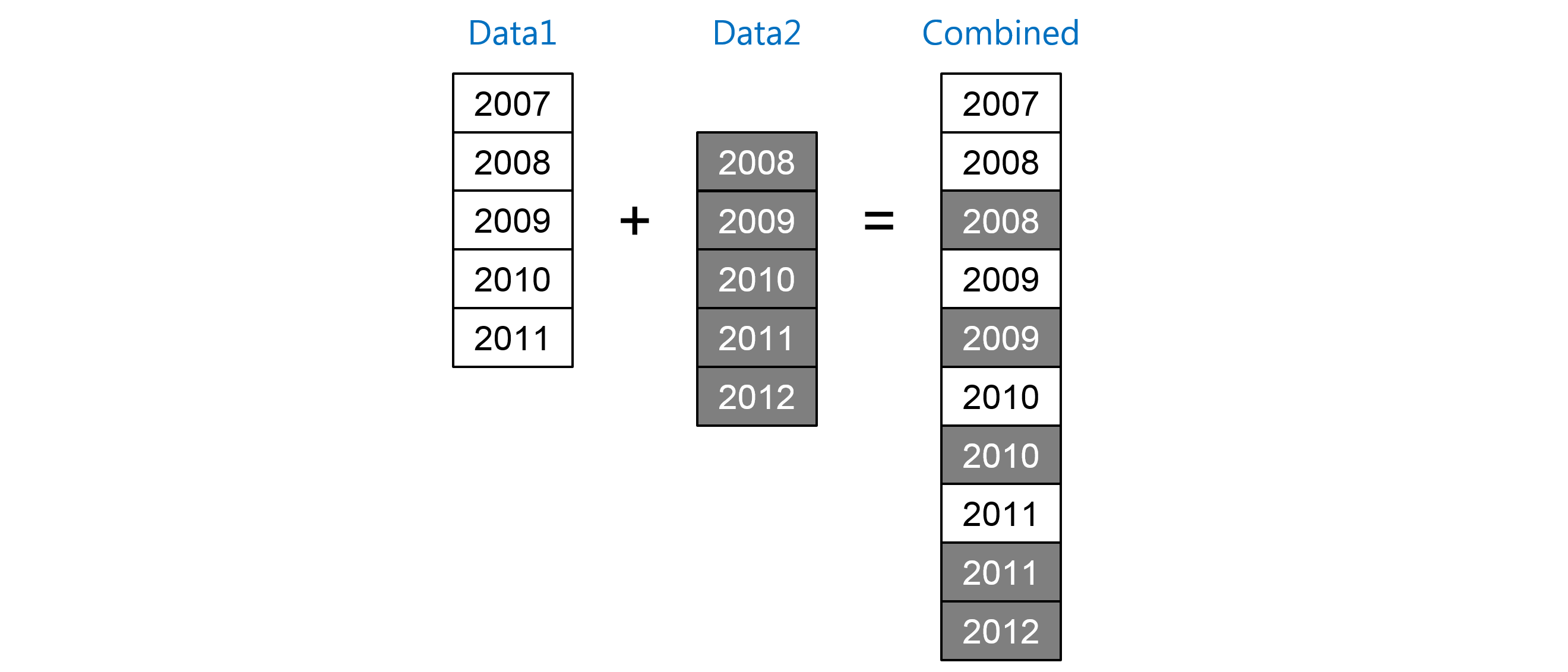
**Answer:**dollar10.2

The format will be the variable name ‘dollar’ which will be of length 10 numbers followed by two numbers after the decimal point.

**48. What is interleaving in SAS?**

**Answer:**Interleaving combines individual, sorted SAS data sets into one sorted SAS data set. For each observation, the following figure shows the value of the variable by which the data sets are sorted. You interleave data sets using a SET statement along with a BY statement.

In the following example, the data sets are sorted by the variable Year.

We can sort and then join the datasets on Year with the below code.

|  |  |
| --- | --- |
| 1  2  3  4 | data combined;      set data1 data2;      by Year;  run; |

**49. I have a dataset concat having variable a b & c. How to rename a b to e & f?**

**Answer:**We will use the following code to rename a b to e f

data concat(rename=(a=e b=f));  
set concat;  
run;

**50. What is the difference between One to One Merge and Match Merge? Give an example.**

**Answer:**If both data sets in the merge statement are sorted by id(as shown below) and each observation in one data set has a corresponding observation in the other data set, a one-to-one merge is suitable.

data mydata1;  
input id class $;  
cards;  
1 Sa  
2 Sd  
3 Rd  
4 Uj  
;

data mydata2;  
input id class1 $;  
cards;  
1 Sac  
2 Sdf  
3 Rdd  
4 Lks  
;

data mymerge;  
merge mydata1 mydata2;  
run;

If the observations do not match, then match merging is suitable

data mydata1;  
input id class $;  
cards;  
1 Sa  
2 Sd  
2 Sp  
3 Rd  
4 Uj  
;

data mydata2;  
input id class1 $;  
cards;  
1 Sac  
2 Sdf  
3 Rdd  
3 Lks  
5 Ujf  
;

data mymerge;  
merge mydata1 mydata2;  
by id  
run;