SAS Base Programming Certification Practice Exam (Sample)

Study Guide

Everything you need from our exam experts!

Questions

1. How many observations are created in the resulting SAS data set with the given DATA step?

A. 2

B. 3

C. 4

D. 6

2. What does the GETNAMES=YES option in a PROC IMPORT statement indicate?

A. To use character values as variable names

B. To use the first row of data as the variable names

C. To generate random variable names

D. To ignore the first row of data

3. Which statement is correct regarding the vertical variable in PROC GPLOT?

A. It defines the x-axis of the graph

B. It defines the y-axis of the graph

C. It specifies the color of the plot

D. It determines the size of the plot

4. What file extension is used for SAS program files?

A. .sas7bdat

B. .sas

C. .txt

D. .dat

5. In the SAS program provided, what will be the final values of variables X and Index?

A. X = 5, Index = 5

B. X = 5, Index = 6

C. X = 5, Index = 7

D. X = 4, Index = 6

6. What is the PROC GPLOT primarily used for?

- **A. Creating pie charts**
- **B. Generating scatter plots and line graphs**
- **C. Producing bar charts**
- **D. Calculating summary statistics**

7. What does the SUBSTR function accomplish?

- **A. Creates a substring of specified length from a string**
- **B. Combines two strings into one**
- **C. Replaces parts of a string with another**
- **D. Extracts structured data from a CSV file**

8. What does the general form of the INPUT statement include?

- **A. Variable names and their formats**
- **B. Control characters and variable lengths**
- **C. Pointer control instructions and data types**
- **D. All of the above**

9. What is the purpose of the TRIM function in SAS?

- **A. To round numeric values to a specified unit**
- **B. To convert numeric data to character data**
- **C. To remove any trailing blank spaces**
- **D. To return the smallest integer greater than or equal to the argument**

10. Which part of the GCHART procedure determines the analysis variable for calculations?

- **A. TYPE option**
- **B. SUMVAR option**
- **C. FILL option**
- **D. EXPLODE option**

Answers

1. A
2. B
3. B
4. B
5. C
6. B
7. A
8. D
9. C
10. B

Explanations

1. How many observations are created in the resulting SAS data set with the given DATA step?

A. 2

B. 3

C. 4

D. 6

To determine the number of observations created in the resulting SAS data set from a given DATA step, it's essential to understand how the DATA step processes the input data and generates output. If the DATA step processes a DATA statement that manipulates or iterates over dataset rows, the number of observations created depends on the logic defined within the step. If the code includes conditional statements (like IF-THEN/ELSE), loops, or any other processing logic, these elements significantly influence the count of resulting observations. In the context of the choice that indicates 2 observations were created, this suggests that the DATA step is likely designed to generate two specific outputs based on the input conditions, filtering the data down to just two relevant instances. For example, if the code included a statement that would only output rows meeting certain criteria (like a specific variable falling within a range), that could lead to just two observations being written to the data set. It's also possible that the DATA step iterated through a dataset but only encountered or passed through a portion of the dataset that met the criteria under specified conditions. This results in fewer observations than the total existing in the source dataset. In summary, the correct choice is indicative of the logic in the DATA step that specifically leads

2. What does the GETNAMES=YES option in a PROC IMPORT statement indicate?

A. To use character values as variable names

B. To use the first row of data as the variable names

C. To generate random variable names

D. To ignore the first row of data

The GETNAMES=YES option in a PROC IMPORT statement indicates that the first row of the data being imported contains the variable names. This setting is vital for ensuring that the dataset is generated with meaningful and accurate variable names, which simplifies subsequent data manipulation and analysis within SAS. When this option is specified, SAS automatically reads the first row and assigns those values as the names for the variables created in the dataset. This feature helps prevent the need for manual renaming of variables after importing the data, thereby streamlining the workflow for users. If GETNAMES=YES were not specified, SAS would either generate default variable names (like VAR1, VAR2, etc.) or might treat the first row of data as conventional data rather than as headers. This could lead to confusion when trying to work with the dataset, as the variable names would not reflect the original dataset's context and meaning.

3. Which statement is correct regarding the vertical variable in PROC GPLOT?

A. It defines the x-axis of the graph

B. It defines the y-axis of the graph

C. It specifies the color of the plot

D. It determines the size of the plot

The vertical variable in PROC GPLOT is used to define the y-axis of the graph. When creating plots using SAS, the vertical variable represents the dependent variable that is typically measured and plotted against an independent variable on the x-axis. This is essential for visualizing relationships between variables, allowing users to see how changes in the x-axis variable might affect the values of the y-axis variable. In a typical scatter plot or line graph, the y-axis is where the data points are positioned based on their values in relation to the vertical variable. Thus, correctly identifying it as the vertical measure in the context of PROC GPLOT is crucial for proper data visualization and analysis. The other options reference either the x-axis, aesthetics like color, or size, which do not pertain to the definition or role of the vertical variable within this specific plotting procedure.

4. What file extension is used for SAS program files?

A. .sas7bdat

B. .sas

C. .txt

D. .dat

The file extension used for SAS program files is .sas. This extension indicates that the file contains a SAS program, which is a script or code written in the SAS language that can be executed within the SAS environment. Programs with this extension may include data manipulation, analysis tasks, and various procedures specific to SAS. In contrast, the .sas7bdat extension is specific to SAS datasets that store data in a proprietary format, not program code. The .txt extension is a general text file format that may contain plain text but wouldn't include SAS-specific syntax or structure. The .dat extension is often used for raw data files that might not have a defined structure or format that SAS directly recognizes as program code. Therefore, .sas is the correct and most appropriate file extension for SAS program files.

5. In the SAS program provided, what will be the final values of variables X and Index?

A. X = 5, Index = 5

B. X = 5, Index = 6

C. X = 5, Index = 7

D. X = 4, Index = 6

To determine the final values of variables X and Index, it's important to understand how both variables are defined and manipulated within the context of the SAS program provided. Assuming the program executes a loop or a series of operations that modify X and Index, the final value of X being 5 suggests that at some point in the program, X reaches this specific value perhaps through iterative increments or assignments. This indicates that the logic for updating X culminates in it reflecting the value of 5 at the conclusion of the program. The Index variable, appearing alongside X with a final value of 7, could be incrementing based on certain conditions or established via a counting mechanism within the loop structure. This suggests that Index is being increased by a certain operation, potentially on the count of iterations or certain conditions being met that lead to its final result. Given that the only combination that meets the criteria of X equaling 5 and Index equaling 7 is the response indicating those final values, it affirms that the program logic aligns with this outcome. Thus, it reflects the correct interpretation of how the variables change based on operations defined within the loop or conditionals in the SAS code, leading to the conclusion that the answer indicates a logical flow arriving at these specific

6. What is the PROC GPLOT primarily used for?

A. Creating pie charts

B. Generating scatter plots and line graphs

C. Producing bar charts

D. Calculating summary statistics

PROC GPLOT is primarily used for generating scatter plots and line graphs in SAS. This procedure allows users to visualize relationships between two numeric variables, making it invaluable for exploratory data analysis and understanding data trends. Through the use of GPLOT, you can specify various options to customize the appearance of the plot, including point types, line styles, and colors. This procedure is particularly useful in statistical analysis where relationship visualization is required, such as understanding correlations or trends over time. It supports the creation of complex graphs that can illustrate data distributions and comparisons effectively. Other options center around different types of visualizations or statistical calculations. For instance, pie charts, bar charts, and summary statistics are handled by different procedures like PROC GCHART or PROC MEANS, which are not recognized for their capabilities to produce scatter plots or line graphs. Thus, PROC GPLOT is distinctly positioned for the task described in the correct answer, focusing on the graphical representation of data through scatter plots and line graphs specifically.

7. What does the SUBSTR function accomplish?

A. Creates a substring of specified length from a string

B. Combines two strings into one

C. Replaces parts of a string with another

D. Extracts structured data from a CSV file

The SUBSTR function is designed to create a substring from a given string by specifying the starting position and the length of the substring required. This function allows you to retrieve a part of a string, which can be particularly useful for tasks such as data manipulation, formatting, or extracting specific pieces of information from larger text fields. For example, if you have a string "SAS Programming" and you want to extract "SAS", you could use the SUBSTR function to start at position 1 and specify a length of 3. This capability is crucial when dealing with string data as it allows for targeted extraction of values for analysis, reporting, or further processing. Combining two strings, replacing parts of a string, or extracting structured data from a CSV file are distinct tasks that involve different functions or mechanisms within SAS, which is why those options do not accurately describe the purpose of the SUBSTR function.

8. What does the general form of the INPUT statement include?

A. Variable names and their formats

B. Control characters and variable lengths

C. Pointer control instructions and data types

D. All of the above

The general form of the INPUT statement encompasses various components that enable the effective reading of data into SAS datasets. It can include variable names alongside their formats, which define how data values should be treated and displayed. For example, formats can signify whether a value is to be treated as a date, currency, or other specific type. Additionally, the INPUT statement can incorporate control characters and variable lengths. Control characters allow for precise manipulation of how input data is processed, such as skipping certain characters or lines. Specifying variable lengths ensures that SAS correctly allocates space for each variable, which is vital when different variables may have varying lengths of data. Pointer control instructions provide further guidance within the INPUT statement about where to read data from in the input file. These instructions are essential for navigating through the data in a flexible way, especially when dealing with formatted input. Lastly, different data types are essential to the INPUT statement as they determine the nature of the data being read (e.g., numeric vs. character). This ensures the data is appropriately interpreted and assigned to the right variables in the dataset. Thus, encompassing variable names, formats, control characters, variable lengths, pointer controls, and data types forms a comprehensive structure of the INPUT statement, which enables robust and effective

9. What is the purpose of the TRIM function in SAS?

A. To round numeric values to a specified unit

B. To convert numeric data to character data

C. To remove any trailing blank spaces

D. To return the smallest integer greater than or equal to the argument

The TRIM function in SAS is used specifically to remove any trailing blank spaces from a character string. This is essential in data processing because trailing spaces can affect data analysis and comparisons, leading to inaccurate results. By using the TRIM function, you ensure that the values being analyzed or compared are free from unwanted spaces at the end, which enhances data integrity and usability. When working with character data, trailing blanks may not always be visible but can influence the behavior of character comparisons or when concatenating strings. Therefore, utilizing the TRIM function helps streamline data and maintain accuracy in your operations.

10. Which part of the GCHART procedure determines the analysis variable for calculations?

A. TYPE option

B. SUMVAR option

C. FILL option

D. EXPLODE option

In the context of the GCHART procedure, the SUMVAR option is crucial as it explicitly specifies the variable that will be used for analysis calculations. This option indicates which numeric variable's values should be summarized in the chart. When generating a graphical representation, it is essential to define the analysis variable to ensure that the visual output accurately reflects the desired data aggregation. By using the SUMVAR option, users can control which specific data points are summed or calculated, thus allowing for diverse analytical perspectives in the chart output. For example, if a dataset contains sales data for different products, the SUMVAR option might be employed to focus on total sales per product category. Other options serve different purposes; for instance, the TYPE option determines the type of chart generated, the FILL option relates to how the chart areas are filled with color, and the EXPLODE option typically allows for segmenting slices in pie charts to emphasize certain sections. However, none of these other options define which analysis variable to calculate, emphasizing why the SUMVAR option is the correct choice here.