SAS Programming: Combining and Modifying Data Sets

Introduction

Data analysis and manipulation are essential skills for anyone working with data, regardless of their field. SAS is a powerful programming language that can be used to perform a wide range of data-related tasks, from simple data cleaning and transformation to complex statistical analysis and machine learning.

This book is a comprehensive guide to SAS programming, covering everything from the basics of the language to advanced topics such as data mining and machine learning. It is written for beginners and experienced programmers alike, and it provides a wealth of examples and exercises to help readers learn the material.

In the first part of the book, you will learn the basics of SAS programming, including how to create and manipulate data sets, perform data transformations, and create charts and graphs. You will also learn about the different SAS procedures that can be used to perform statistical analysis and machine learning.

In the second part of the book, you will learn more advanced topics, such as data mining and machine learning. You will learn how to use SAS to build predictive models, identify patterns in data, and make predictions. You will also learn about the different SAS tools that can be used to manage and analyze data, such as SAS Enterprise Guide and SAS Cloud Analytics Services.

By the end of this book, you will be able to use SAS to perform a wide range of data-related tasks, from simple data cleaning and transformation to complex statistical analysis and machine learning. You will also

be familiar with the different SAS tools that can be used to manage and analyze data.

Whether you are a beginner or an experienced programmer, this book will help you to learn SAS and use it to its full potential.

Book Description

SAS Programming: Combining and Modifying Data Sets is a comprehensive guide to SAS programming, covering everything from the basics of the language to advanced topics such as data mining and machine learning. It is written for beginners and experienced programmers alike, and it provides a wealth of examples and exercises to help readers learn the material.

In the first part of the book, you will learn the basics of SAS programming, including how to create and manipulate data sets, perform data transformations, and create charts and graphs. You will also learn about the different SAS procedures that can be used to perform statistical analysis and machine learning.

In the second part of the book, you will learn more advanced topics, such as data mining and machine learning. You will learn how to use SAS to build

predictive models, identify patterns in data, and make predictions. You will also learn about the different SAS tools that can be used to manage and analyze data, such as SAS Enterprise Guide and SAS Cloud Analytics Services.

By the end of this book, you will be able to use SAS to perform a wide range of data-related tasks, from simple data cleaning and transformation to complex statistical analysis and machine learning. You will also be familiar with the different SAS tools that can be used to manage and analyze data.

Whether you are a beginner or an experienced programmer, **SAS Programming: Combining and Modifying Data Sets** will help you to learn SAS and use it to its full potential.

What you'll learn:

- The basics of SAS programming
- How to create and manipulate data sets

- How to perform data transformations
- How to create charts and graphs
- The different SAS procedures that can be used to perform statistical analysis and machine learning
- How to use SAS to build predictive models
- How to identify patterns in data
- How to make predictions
- The different SAS tools that can be used to manage and analyze data

Who this book is for:

- Beginners who want to learn SAS programming
- Experienced programmers who want to learn more about SAS
- Data analysts who want to use SAS to analyze data
- Data scientists who want to use SAS to build predictive models

Chapter 1: Data Manipulation Essentials

1. Understanding Data Structures

Data structures are the foundation of data manipulation in SAS. They determine how data is stored and organized, and they play a crucial role in the efficiency and performance of data manipulation operations.

SAS supports a variety of data structures, including:

- **Tables:** Tables are the most common data structure in SAS. They are similar to spreadsheets, with rows and columns. Each row represents a record, and each column represents a variable.
- **Arrays:** Arrays are one-dimensional data structures that store a collection of values of the same data type. Arrays are indexed by integers,

and they can be used to store data that is related in some way.

- **Sets:** Sets are unordered collections of unique values. They can be used to store data that needs to be kept unique, such as customer IDs or product codes.
- **Hash tables:** Hash tables are data structures that store key-value pairs. They are used to store data that needs to be accessed quickly and efficiently, such as customer information or product data.

The choice of which data structure to use depends on the specific data manipulation task that needs to be performed. Tables are the most versatile data structure, and they can be used for a wide variety of tasks. Arrays are useful for storing data that is related in some way, such as the sales data for a particular product. Sets are useful for storing data that needs to be kept unique, such as customer IDs or product codes. Hash tables are useful for storing data that needs to be

accessed quickly and efficiently, such as customer information or product data.

Understanding the different data structures available in SAS is essential for efficient data manipulation. By choosing the right data structure for the task at hand, you can improve the performance and efficiency of your SAS programs.

Chapter 1: Data Manipulation Essentials

2. Basic Data Transformation Techniques

Data transformation is the process of converting data from one format or structure to another. This can be done for a variety of reasons, such as to clean the data, prepare it for analysis, or to make it more compatible with a particular software program.

SAS provides a wide range of data transformation functions that can be used to perform a variety of tasks, such as:

- **Renaming variables:** The RENAME statement can be used to rename one or more variables in a data set.
- **Creating new variables:** The DATA step can be used to create new variables based on existing variables.

- **Modifying variable values:** The MODIFY statement can be used to modify the values of one or more variables in a data set.
- **Recoding variables:** The RECODE statement can be used to recode the values of a variable into a new set of values.
- **Sorting data:** The SORT procedure can be used to sort data in ascending or descending order by one or more variables.

Data transformation is an essential skill for data analysts and programmers. By understanding how to transform data, you can improve the quality of your data and make it more useful for analysis.

Here are some examples of how data transformation can be used in practice:

- **Cleaning data:** Data cleaning is the process of removing errors and inconsistencies from data. This can involve tasks such as removing

duplicate records, correcting data entry errors, and dealing with missing values.

- **Preparing data for analysis:** Data preparation is the process of getting data into a format that is suitable for analysis. This can involve tasks such as transforming data into a consistent format, creating new variables, and recoding variables.
- **Making data more compatible with a particular software program:** Different software programs have different requirements for the format of data. Data transformation can be used to convert data into a format that is compatible with a particular software program.

By understanding how to transform data, you can make your data more useful and valuable.

Chapter 1: Data Manipulation Essentials

3. Merging and Appending Data Sets

Data merging and appending are two important techniques for combining data from multiple sources. Merging combines data from two or more data sets that have common variables, while appending combines data from two or more data sets that have the same variables.

There are two main types of merges: inner merges and outer merges. An inner merge only includes rows that have matching values in both data sets. An outer merge includes all rows from both data sets, even if they do not have matching values.

There are also two main types of appends: horizontal appends and vertical appends. A horizontal append adds the rows from one data set to the bottom of

another data set. A vertical append adds the columns from one data set to the right of another data set.

Data merging and appending can be used for a variety of purposes, such as:

- Combining data from multiple sources into a single data set
- Adding new variables to an existing data set
- Removing duplicate rows from a data set
- Reshaping a data set

SAS provides a variety of procedures that can be used to merge and append data sets. The most commonly used procedures are the PROC SORT, PROC MERGE, and PROC APPEND procedures.

The PROC SORT procedure is used to sort data sets by one or more variables. Sorting data sets can be useful for merging and appending data sets, as it ensures that the data is in the same order.

The PROC MERGE procedure is used to merge two or more data sets. The PROC MERGE procedure can be used to perform inner merges, outer merges, and self-merges.

The PROC APPEND procedure is used to append two or more data sets. The PROC APPEND procedure can be used to perform horizontal appends and vertical appends.

Data merging and appending are powerful techniques that can be used to combine data from multiple sources and create new data sets. By understanding how to use these techniques, you can improve your data analysis and reporting capabilities.

This extract presents the opening three sections of the first chapter.

Discover the complete 10 chapters and 50 sections by purchasing the book, now available in various formats.

Table of Contents

This extract presents the opening three sections of the first chapter.

Discover the complete 10 chapters and 50 sections by purchasing the book, now available in various formats.