

LECTURE 02

21CSE295T

Statistical Data Analytics And Integration

Unit 1 Introduction to Statistical Analysis

1.12	SAS Statements
1.13	Structure
1.14	Libraries
1.15	Rules
1.16	Datasets
1.17	Representation of Missing Values
1.18	Assigning Libraries
1.19	PROC CONTENTS
1.20	Visualizing Descriptors and Dataset Attributes

Virupakshan K

+91-9840847872 | VIRU@KART.LA

SYLLABUS

21CSE295T Statistical Data Analytics And Integration

VirupakshanK@gmail.com

Unit 1 Introduction to Statistical Analysis

9 hours

Computational Statistics- Statistical Pattern Recognition, Nonparametric Regression, Exploratory Data Analysis, Monte Carlo Methods for Inferential Statistics-Basic concepts of statistics, Types of data (categorical, numerical), Populations and samples, Measures of central tendency (mean, median, mode), Measures of variability (range, variance, standard deviation), Descriptive vs. inferential statistics.

Tutorial:

- SAS Statements, Structure, Libraries, Rules, Datasets, Representation of missing values, Assigning Libraries, PROC CONTENTS, visualize the descriptor and datasets attributes

Unit 2 Data Analysis Using SAS

9 hours

Basic Reports – Selecting Variables – Identifying observations – Subsetting the Data – Limiting the observations-SORT Procedure – Generating Column totals – Specifying titles and footnotes in procedure output – Assigning Descriptive labels and permanent label -How SAS Processes Programs – Compilation Phase – PDV – Descriptor Portion-Execution Phase – Data Portion – Debugging a Data Step – Testing Programs-BY-Group Processing Definitions – Preprocessing Data – Determine Whether the Data Requires Preprocessing – Sorting Observations for BY-Group Processing-FIRST, and LAST, DATA Step Variables – How the DATA Step Identifies BY Groups – Assignment Statements – Operators in SAS Expressions

Tutorial:

- Introduction of report and PRINT Procedure-Hands on practice of VAR, ID, WHERE statement, where expression[eq, ne, gt, lt, ge, le, AND, OR, CONTAINS]

Unit 3 Advanced Statistical Techniques in SAS

9 hours

Determining the Structure and Contents of Data Sets – Testing Program – Methods of Combining SAS Data Sets-One-to-One Reading – Concatenating – Match-Merging – Renaming Variables – Excluding Unmatched Observations -Selecting Matching Observations -Basics of DO Loops – Executing DO Loops -OUTPUT Statements – Nesting DO Loops-Iteratively Processing Observations from a Data Set -DO UNTIL Statement – DO WHILE Statement Defining and Processing Arrays –Arrays -Applying SAS Formats and Informats – Specifying SAS Formats

Tutorial:

- One-to-One Reading to Combine Data Sets-Concatenating to Combine Data Sets-Match-Merging to Combine Data Sets-Renaming Variables-Processing Iterative DO Loops-Indenting and Nesting DO Groups

Unit 4 SAS Programming and Automation

9 hours

Definitions – Reading Date and Time with Informats – Displaying Date and Time Values with Formats-Basics of SAS Functions - SAS Functions Syntax – Converting Data with Functions- The MEANS Procedure – Specifying Descriptive Statistics – Limiting Decimal Places – Specifying Variable Using VAR Statement – Summarized Data Set Using OUTPUT Statement – The FREQ Procedure-Specifying Variable Using TABLES Statements- Two-Way and N-Way Tables – LIST Options-The Output Delivery System (ODS) – HTML Outputs – PDF Outputs-RTF Outputs – EXCEL Outputs – The EXPORT Procedure

Tutorial:

- Reading Date and Time with Formats and Informats- Execution of Data Values Using Numeric Functions-Creating a Descriptive Statistics by Using PROC MEANS-Creating a Descriptive Statistics by Using PROC FREQ-Creating Outputs with HTML, PDF, RTF, EXCEL Formats- Exporting a External File

Unit 5 Advanced SAS Topics

9 hours

Creating a Default List Report – Windowing/Non Windowing Mode - Define Statement-Defining Variables – Column Statement – Selecting Observations – Defining the Variables-Group Processing the Variables by Using Different Options – Computing the Variable-Bar Chart – Box Plot – Mean Measurement Over the Time using Line Plot – Spaghetti Plot – Survival Plot – Waterfall Plot – Forest Plot-Introduction to Macro Facility – SAS Programs and Macro Processing – Macro Variables – Macro Processing – Scopes of Macro Variables – Macro Expressions-Scopes of Macro Variables – Macro Expressions-Macro Quoting – Interfaces with Macro Facility -Storing and Reusing Macros.

Tutorial:

- Creating Graphs by Using Bar Chart -Creating Graphs Using Various Types of Plots-Generating SAS Code Using Macros-Defining User-Defined Macro Variables-Creating a Tables Using PROC SQL with different Clauses- Generating a Cartesian Product -Combining Tables Using Multiple Joins

LECTURE 02 TOPICS

VirupakshanK@gmail.com

Unit 1 Introduction to Statistical Analysis

9 hours

Computational Statistics- Statistical Pattern Recognition, Nonparametric Regression, Exploratory Data Analysis, Monte Carlo Methods for Inferential Statistics-Basic concepts of statistics, Types of data (categorical, numerical), Populations and samples, Measures of central tendency (mean, median, mode), Measures of variability (range, variance, standard deviation), Descriptive vs. inferential statistics.

Tutorial:

- SAS Statements, Structure, Libraries, Rules, Datasets, Representation of missing values, Assigning Libraries, PROC CONTENTS, visualize the descriptor and datasets attributes

Unit 2 Data Analysis Using SAS

9 hours

Basic Reports – Selecting Variables – Identifying observations – Subsetting the Data – Limiting the observations-SORT Procedure – Generating Column totals – Specifying titles and footnotes in procedure output – Assigning Descriptive labels and permanent label -How SAS Processes Programs – Compilation Phase – PDV – Descriptor Portion-Execution Phase – Data Portion – Debugging a Data Step – Testing Programs-BY-Group Processing Definitions – Preprocessing Data – Determine Whether the Data Requires Preprocessing – Sorting Observations for BY-Group Processing-FIRST. and LAST. DATA Step Variables – How the DATA Step Identifies BY Groups – Assignment Statements – Operators in SAS Expressions

Tutorial:

- Introduction of report and PRINT Procedure-Hands on practice of VAR, ID, WHERE statement, where expression[eq, ne, gt, lt, ge, le, AND, OR, CONTAINS]

Unit 3 Advanced Statistical Techniques in SAS

9 hours

Determining the Structure and Contents of Data Sets – Testing Program – Methods of Combining SAS Data Sets-One-to-One Reading – Concatenating – Match-Merging – Renaming Variables – Excluding Unmatched Observations -Selecting Matching Observations -Basics of DO Loops – Executing DO Loops -OUTPUT Statements – Nesting DO Loops-Iteratively Processing Observations from a Data Set -DO UNTIL Statement – DO WHILE Statement Defining and Processing Arrays –Arrays -Applying SAS Formats and Informats – Specifying SAS Formats

Tutorial:

- One-to-One Reading to Combine Data Sets-Concatenating to Combine Data Sets-Match-Merging to Combine Data Sets-Renaming Variables-Processing Iterative DO Loops-Indenting and Nesting DO Groups

Unit 4 SAS Programming and Automation

9 hours

Definitions – Reading Date and Time with Informats – Displaying Date and Time Values with Formats-Basics of SAS Functions - SAS Functions Syntax – Converting Data with Functions- The MEANS Procedure – Specifying Descriptive Statistics – Limiting Decimal Places – Specifying Variable Using VAR Statement – Summarized Data Set Using OUTPUT Statement – The FREQ Procedure-Specifying Variable Using TABLES Statements - Two-Way and N-Way Tables – LIST Options-The Output Delivery System (ODS) – HTML Outputs – PDF Outputs-RTF Outputs – EXCEL Outputs – The EXPORT Procedure

Tutorial:

- Reading Date and Time with Formats and Informats- Execution of Data Values Using Numeric Functions-Creating a Descriptive Statistics by Using PROC MEANS-Creating a Descriptive Statistics by Using PROC FREQ-Creating Outputs with HTML, PDF, RTF, EXCEL Formats-Exporting a External File

Unit 5 Advanced SAS Topics

9 hours

Creating a Default List Report – Windowing/Non Windowing Mode - Define Statement-Defining Variables – Column Statement – Selecting Observations – Defining the Variables-Group Processing the Variables by Using Different Options – Computing the Variable-Bar Chart – Box Plot – Mean Measurement Over the Time using Line Plot – Spaghetti Plot – Survival Plot – Waterfall Plot – Forest Plot-Introduction to Macro Facility – SAS Programs and Macro Processing – Macro Variables – Macro Processing – Scopes of Macro Variables – Macro Expressions-Scopes of Macro Variables – Macro Expressions-Macro Quoting – Interfaces with Macro Facility -Storing and Reusing Macros.

Tutorial:

- Creating Graphs by Using Bar Chart -Creating Graphs Using Various Types of Plots-Generating SAS Code Using Macros-Defining User-Defined Macro Variables-Creating a Tables Using PROC SQL with different Clauses- Generating a Cartesian Product -Combining Tables Using Multiple Joins

1.12.1 SAS Statements

SAS, Statistical Analysis System, is a software suite used for data management, statistical analysis, reporting, and predictive modeling.

SAS programs are written using statements.

What is a SAS Statement?

A SAS statement is an instruction given to SAS (Statistical Analysis System).

Basic Structure of a SAS Program

```
DATA dataset_name;  
    INPUT variable1 variable2 variable3;  
    DATALINES;  
    values;  
RUN;  
  
PROC procedure_name DATA=dataset_name;  
RUN;
```

Important rule:

Every SAS statement ends with a semicolon ;

1.12.2 SAS Statements: Example

```
DATA students;  
    INPUT Name $ Age Marks;  
    DATALINES;  
Ravi 20 85  
Anu 19 90  
Kiran 21 75  
;  
RUN;  
  
PROC PRINT DATA=students;  
RUN;
```

SAS Statement	Meaning
DATA students;	Creates a <u>dataset</u> named students
INPUT Name \$ Age Marks;	Defines <u>variables</u>
DATALINES;	Indicates <u>start</u> of raw data
RUN;	<u>Executes</u> the step
PROC PRINT DATA=students;	<u>Prints</u> the dataset

1.13 Structure

VirupakshanK@gmail.com

SAS Program Flow

SAS Program

```
|  
|-- DATA Step  
|  
|   |-- Create or modify data  
|  
|-- PROC Step  
|  
|   |-- Analyze or report data
```

VirupakshanK@gmail.com

1.14 Libraries

A **SAS library** is a collection of SAS files stored in a particular location.

There are two main types: **Temporary & Permanent**

1. Temporary Library

The default temporary library is WORK.

Datasets stored in WORK are deleted when the SAS session ends.

Example:

```
DATA work.students;
    INPUT Name $ Marks;
    DATALINES;
Ravi 85
Anu 90
;
RUN;
```

2. Permanent Library

A permanent library stores datasets even after the SAS session ends.

Example:

```
LIBNAME mylib "C:\SASData";
```

Then create a dataset:

sas

```
DATA mylib.students;
    INPUT Name $ Marks;
    DATALINES;
Ravi 85
Anu 90
;
RUN;
```

1.15 Rules : SAS Naming Rules

Dataset and Variable Names

Common SAS naming rules:

1. Must begin with a letter or underscore
2. Can contain letters, numbers, and underscores
3. Cannot contain spaces
4. Usually up to 32 characters
5. Should not use special symbols
6. Should avoid reserved words

Valid names:

```
student_marks  
Age  
_total  
marks2025
```

Invalid names:

```
student marks  
2marks  
total%  
class-age
```

1.16 Datasets

A **SAS dataset** is a table made up of rows and columns.

- **Rows** are called observations
- **Columns** are called variables

Dataset Structure

SAS Dataset

```
|  
|-- Descriptor Portion  
|  
|   |-- Dataset name  
|   |-- Variable names  
|   |-- Variable types  
|   |-- Variable lengths  
|  
|-- Data Portion  
|  
|   |-- Actual observations
```

Example Dataset

Name	Age	Marks
Ravi	20	85
Anu	19	90
Kiran	21	75

Variables:

- Name
- Age
- Marks

Observations:

- Ravi's record
- Anu's record
- Kiran's record

1.17.1 Representation of Missing Values

Missing values occur when data is unavailable, unknown, not recorded, or not applicable.

Example:

Name	Age	Marks
Ravi	20	85
Anu	.	90
Kiran	21	.

Here, Anu's age and Kiran's marks are missing.

Missing Values in SAS

Numeric Missing Values

In SAS, numeric missing values are represented by a period:

```
DATA students;
    INPUT Name $ Age Marks;
    DATALINES;
Ravi 20 85
Anu . 90
Kiran 21 .
;
RUN;
```

Character Missing Values

Character missing values are represented by a blank space.

```
DATA students;
    INPUT Name $ Gender $ Marks;
    DATALINES;
Ravi M 85
Anu . 90
Kiran M 75
;
RUN;
```

1.17.2 Representation of Missing Values

Why Missing Values Occur

- Data entry errors
- Non-response in surveys
- Sensor failure
- Not applicable values
- Data corruption

Handling Missing Values

Common methods:

- 1.Delete rows with missing values
- 2.Replace with mean, median, or mode
- 3.Use model-based imputation
- 4.Use special missing value categories
- 5.Investigate the cause before deciding

EXAMPLE:

Dataset:

Student	Marks
A	85
B	.
C	90
D	70

Mean excluding missing value:

$$\frac{80 + 90 + 70}{3} = 80$$

The missing value may be replaced with 80 if mean imputation is appropriate.



Student	Marks
A	85
B	80
C	90
D	70

1.18.1 Assigning Libraries

Assigning libraries means **creating a reference name**, called a **libref**, for a folder location where SAS datasets are stored.

LIBNAME Statement

Syntax:

```
LIBNAME libref "folder-path";
```

Example:

```
LIBNAME college "C:\SASData\College";
```

Here:

- **college** is the **libref**
- **C:\SASData\College** is the physical **folder path**

Important Notes

- Libref usually has up to 8 characters.
- The path must exist.
- The folder path should be enclosed in quotes.
- Temporary datasets are stored in WORK.
- Permanent datasets use a user-defined library.

1.18.2 Assigning Libraries

Creating a Permanent Dataset

```
LIBNAME college "C:\SASData\College";

DATA college.students;
  INPUT Name $ Age Marks;
  DATALINES;
Ravi 20 85
Anu 19 90
Kiran 21 75
;
RUN;
```

This dataset is stored permanently in the **college** library.

Accessing the Dataset

```
PROC PRINT DATA=college.students;
RUN;
```

Clearing a Library

```
LIBNAME college CLEAR;
```

This removes the library assignment but does not delete the physical files.

```
LIBNAME college "C:\SASData\College";
|
v
college.students
|
v
Stored in C:\SASData\College
```

1.19.1 PROC CONTENTS

PROC CONTENTS displays the structure and metadata of a SAS dataset.

It tells us:

- Dataset name
- Number of observations
- Number of variables
- Variable names
- Variable types
- Variable lengths
- Variable labels
- Formats
- Informats
- Creation date

Syntax

```
PROC CONTENTS DATA=dataset_name;  
RUN;
```

Example

```
PROC CONTENTS DATA=college.students;  
RUN;
```

Sample Output Concept

```
Dataset Name: COLLEGE.STUDENTS  
Number of Observations: 3  
Number of Variables: 3
```

Variable	Type	Length
Name	Character	8
Age	Numeric	8
Marks	Numeric	8

Example Dataset

```
DATA students;  
    INPUT Name $ Age Marks;  
    DATALINES;  
    Ravi 20 85  
    Anu 19 90  
    Kiran 21 75  
    ;  
RUN;  
  
PROC CONTENTS DATA=students;  
RUN;
```

1.19.2a PROC CONTENTS : RESULTS for PROC CONTENTS

```
1 DATA students;
2     INPUT Name $ Age Marks;
3     DATALINES;
4 Ravi 20 85
5 Anu 19 90
6 Kiran 21 75
7 ;
8 RUN;
9
10 PROC CONTENTS DATA=students;
11 RUN;
```

viru_print_contents_04.sas

CODE LOG RESULTS OUTPUT DATA

Table of Contents

The CONTENTS Procedure

Data Set Name	WORK.STUDENTS	Observations	3
Member Type	DATA	Variables	3
Engine	V9	Indexes	0
Created	05/31/2026 17:11:45	Observation Length	24
Last Modified	05/31/2026 17:11:45	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
2	Age	Num	8
3	Marks	Num	8
1	Name	Char	8

1.19.2b PROC CONTENTS : RESULTS for PROC PRINT

```
1 DATA students;  
2     INPUT Name $ Age Marks;  
3     DATALINES;  
4 Ravi 20 85  
5 Anu 19 90  
6 Kiran 21 75  
7 ;  
8 RUN;  
9  
10 PROC PRINT DATA=students;  
11 RUN;
```

viru_print_contents_04.sas

CODE LOG RESULTS OUTPUT DATA

Table of Contents

Obs	Name	Age	Marks
1	Ravi	20	85
2	Anu	19	90
3	Kiran	21	75

1.19.3 PROC CONTENTS

Why Use PROC CONTENTS?

PROC CONTENTS is useful for:

- Checking dataset structure
- Identifying variable types
- Finding variable lengths
- Debugging errors
- Understanding unknown datasets
- Verifying imported data

Difference Between PROC PRINT and PROC CONTENTS

Feature	PROC PRINT	PROC CONTENTS
Purpose	Displays actual data	Displays metadata
Shows rows?	Yes	No
Shows variable types?	Limited	Yes
Used for	Viewing data	Understanding structure

1.20.1 Visualizing Descriptors and Dataset Attributes

Visualizing dataset attributes means understanding the structure, variables, types, missing values, distributions, and relationships in a dataset using tables, reports, and charts.

Attribute View

Dataset: STUDENTS

Variable	Type	Length	Description
Name	Character	8	Student name
Age	Numeric	8	Student age
Marks	Numeric	8	Exam marks

Dataset Attributes

Common attributes include:

- Dataset name
- Number of rows
- Number of columns
- Variable names
- Data types
- Variable lengths
- Labels
- Formats
- Missing values
- Unique values
- Distribution of values

1.20.2 Visualizing Descriptors and Dataset Attributes

Conceptual Diagram

Dataset

```
|  
|-- Metadata  
|    |-- PROC CONTENTS  
|  
|-- Actual Data  
|    |-- PROC PRINT  
|  
|-- Categorical Summary  
|    |-- PROC FREQ  
|  
|-- Numeric Summary  
|    |-- PROC MEANS / PROC UNIVARIATE
```

1.20.3 VDDA: SAS Procedures for Viewing Attributes

1. PROC CONTENTS

Shows metadata.

```
sas  
  
PROC CONTENTS DATA=students;  
RUN;
```

2. PROC PRINT

Displays actual observations.

```
sas  
  
PROC PRINT DATA=students;  
RUN;
```

3. PROC FREQ

Displays frequency counts for categorical variables.

```
sas  
  
PROC FREQ DATA=students;  
TABLES Gender;  
RUN;
```

4. PROC MEANS

Displays summary statistics for numeric variables.

```
sas  
  
PROC MEANS DATA=students;  
VAR Marks;  
RUN;
```

5. PROC UNIVARIATE

Provides detailed distributional analysis.

```
sas  
  
PROC UNIVARIATE DATA=students;  
VAR Marks;  
RUN;
```

1.20.4 Visualizing Descriptors and Dataset Attributes: Example

Example Dataset

sas

```
DATA students;  
  INPUT Name $ Gender $ Age Marks;  
  DATALINES;  
Ravi M 20 85  
Anu F 19 90  
Kiran M 21 75  
Meena F 20 88  
;  
RUN;
```

Print Data

sas

```
PROC PRINT DATA=students;  
RUN;
```

Check Structure

sas

```
PROC CONTENTS DATA=students;  
RUN;
```

Visualizing Attributes

Frequency of Gender

sas

```
PROC FREQ DATA=students;  
  TABLES Gender;  
RUN;
```

Summary of Marks

sas

```
PROC MEANS DATA=students;  
  VAR Marks;  
RUN;
```

1.20.5 Visualizing Descriptors and Dataset Attributes (CODE 1)

CODE LOG RESULTS OUTPUT DATA

```
1 DATA students;
2   INPUT Name $ Gender $ Age Marks;
3   DATALINES;
4 Ravi M 20 75
5 Anu F 19 90
6 Kiran M 21 75
7 Meena F 20 88
8 ;
9 RUN;
10
11 * PROC PRINT DATA=students;
12 * PROC CONTENTS DATA=students;
13
14 * PROC FREQ DATA=students;
15   * TABLES Gender;
16
17 * PROC MEANS DATA=students;
18 PROC UNIVARIATE DATA=students;
19   VAR Marks;
20
21 RUN;
```

viru_08_SAS_Procs.sas

CODE LOG RESULTS OUTPUT DATA



Table of Contents

The UNIVARIATE Procedure
Variable: Marks

Moments			
N	4	Sum Weights	4
Mean	82	Sum Observations	328
Std Deviation	8.1240384	Variance	66
Skewness	0.05222063	Kurtosis	-5.6992654
Uncorrected SS	27094	Corrected SS	198
Coeff Variation	9.90736391	Std Error Mean	4.0620192

Basic Statistical Measures			
Location		Variability	
Mean	82.00000	Std Deviation	8.12404
Median	81.50000	Variance	66.00000
Mode	75.00000	Range	15.00000
		Interquartile Range	14.00000

Tests for Location: Mu0=0		
Test	Statistic	p Value
Student's t	t 20.187	Pr > t 0.0003
Sign	M 2	Pr >= M 0.1250
Signed Rank	S 5	Pr >= S 0.1250

Tests for Location: Mu0=0

Test	Statistic	p Value
Student's t	t 20.187	Pr > t 0.0003
Sign	M 2	Pr >= M 0.1250
Signed Rank	S 5	Pr >= S 0.1250

Quantiles (Definition 5)

Level	Quantile
100% Max	90.0
99%	90.0
95%	90.0
90%	90.0
75% Q3	89.0
50% Median	81.5
25% Q1	75.0
10%	75.0
5%	75.0
1%	75.0
0% Min	75.0

Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
75	3	75	1
75	1	75	3
88	4	88	4
90	2	90	2

1.20.6 Visualizing Descriptors and Dataset Attributes: (CODE 2)

```
1 TITLE "Tutorial 1: PROC PRINT and PROC CONTENTS";
2
3 DATA students;
4     INPUT Name $ Age Gender $ Marks;
5     DATALINES;
6 Ravi 20 M 85
7 Anu 19 F 90
8 Kiran 21 M 75
9 Meena 20 F 88
10 ;
11 RUN;
12
13 PROC PRINT DATA=students;
14     TITLE "Student Dataset";
15 RUN;
16
17 PROC CONTENTS DATA=students VARNUM;
18     TITLE "Structure of Student Dataset";
19 RUN;
20
21 TITLE;
```

CODE LOG RESULTS OUTPUT DATA

Table of Contents

Student Dataset

Obs	Name	Age	Gender	Marks
1	Ravi	20	M	85
2	Anu	19	F	90
3	Kiran	21	M	75
4	Meena	20	F	88

Structure of Student Dataset

The CONTENTS Procedure

Data Set Name	WORK.STUDENTS	Observations	4
Member Type	DATA	Variables	4
Engine	V9	Indexes	0
Created	05/31/2026 18:52:08	Observation Length	32
Last Modified	05/31/2026 18:52:08	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information

Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1

Lecture 02 Topics

Unit 1 Introduction to Statistical Analysis

1.12	SAS Statements
1.13	Structure
1.14	Libraries
1.15	Rules
1.16	Datasets
1.17	Representation of Missing Values
1.18	Assigning Libraries
1.19	PROC CONTENTS
1.20	Visualizing Descriptors and Dataset Attributes

LECTURE 02 TOPICS

VirupakshanK@gmail.com

Unit 1 Introduction to Statistical Analysis

9 hours

Computational Statistics- Statistical Pattern Recognition, Nonparametric Regression, Exploratory Data Analysis, Monte Carlo Methods for Inferential Statistics-Basic concepts of statistics, Types of data (categorical, numerical), Populations and samples, Measures of central tendency (mean, median, mode), Measures of variability (range, variance, standard deviation), Descriptive vs. inferential statistics.

Tutorial:

- SAS Statements, Structure, Libraries, Rules, Datasets, Representation of missing values, Assigning Libraries, PROC CONTENTS, visualize the descriptor and datasets attributes

Unit 2 Data Analysis Using SAS

9 hours

Basic Reports – Selecting Variables – Identifying observations – Subsetting the Data – Limiting the observations-SORT Procedure – Generating Column totals – Specifying titles and footnotes in procedure output – Assigning Descriptive labels and permanent label -How SAS Processes Programs – Compilation Phase – PDV – Descriptor Portion-Execution Phase – Data Portion – Debugging a Data Step – Testing Programs-BY-Group Processing Definitions – Preprocessing Data – Determine Whether the Data Requires Preprocessing – Sorting Observations for BY-Group Processing-FIRST. and LAST. DATA Step Variables – How the DATA Step Identifies BY Groups – Assignment Statements – Operators in SAS Expressions

Tutorial:

- Introduction of report and PRINT Procedure-Hands on practice of VAR, ID, WHERE statement, where expression[eq, ne, gt, lt, ge, le, AND, OR, CONTAINS]

Unit 3 Advanced Statistical Techniques in SAS

9 hours

Determining the Structure and Contents of Data Sets – Testing Program – Methods of Combining SAS Data Sets-One-to-One Reading – Concatenating – Match-Merging – Renaming Variables – Excluding Unmatched Observations -Selecting Matching Observations -Basics of DO Loops – Executing DO Loops -OUTPUT Statements – Nesting DO Loops-Iteratively Processing Observations from a Data Set -DO UNTIL Statement – DO WHILE Statement Defining and Processing Arrays –Arrays -Applying SAS Formats and Informats – Specifying SAS Formats

Tutorial:

- One-to-One Reading to Combine Data Sets-Concatenating to Combine Data Sets-Match-Merging to Combine Data Sets-Renaming Variables-Processing Iterative DO Loops-Indenting and Nesting DO Groups

Unit 4 SAS Programming and Automation

9 hours

Definitions – Reading Date and Time with Informats – Displaying Date and Time Values with Formats-Basics of SAS Functions - SAS Functions Syntax – Converting Data with Functions- The MEANS Procedure – Specifying Descriptive Statistics – Limiting Decimal Places – Specifying Variable Using VAR Statement – Summarized Data Set Using OUTPUT Statement – The FREQ Procedure-Specifying Variable Using TABLES Statements - Two-Way and N-Way Tables – LIST Options-The Output Delivery System (ODS) – HTML Outputs – PDF Outputs-RTF Outputs – EXCEL Outputs – The EXPORT Procedure

Tutorial:

- Reading Date and Time with Formats and Informats- Execution of Data Values Using Numeric Functions-Creating a Descriptive Statistics by Using PROC MEANS-Creating a Descriptive Statistics by Using PROC FREQ-Creating Outputs with HTML, PDF, RTF, EXCEL Formats-Exporting a External File

Unit 5 Advanced SAS Topics

9 hours

Creating a Default List Report – Windowing/Non Windowing Mode - Define Statement-Defining Variables – Column Statement – Selecting Observations – Defining the Variables-Group Processing the Variables by Using Different Options – Computing the Variable-Bar Chart – Box Plot – Mean Measurement Over the Time using Line Plot – Spaghetti Plot – Survival Plot – Waterfall Plot – Forest Plot-Introduction to Macro Facility – SAS Programs and Macro Processing – Macro Variables – Macro Processing – Scopes of Macro Variables – Macro Expressions-Scopes of Macro Variables – Macro Expressions-Macro Quoting – Interfaces with Macro Facility -Storing and Reusing Macros.

Tutorial:

- Creating Graphs by Using Bar Chart -Creating Graphs Using Various Types of Plots-Generating SAS Code Using Macros-Defining User-Defined Macro Variables-Creating a Tables Using PROC SQL with different Clauses- Generating a Cartesian Product -Combining Tables Using Multiple Joins

LECTURE 02

21CSE295T

Statistical Data Analytics And Integration

Unit 1 Introduction to Statistical Analysis

Doubts?

THANK YOU!

Virupakshan K

+91-9840847872 | VIRU@KART.LA