LECTURE 08

21CSE295T

Statistical Data Analytics And Integration

Unit 4 SAS Programming and Automation

4.6	The MEANS Procedure
4.7	Specifying Descriptive Statistics
4.8	Limiting Decimal Places
4.9	PROC FREQ Procedure
4.1O	Two-Way and N-Way Tables
4.11	LIST Options
4.12	Output Delivery System (ODS)
4.13	HTML Outputs
4.14	PDF Outputs
4.15	RTF Outputs
4.16	EXCEL Outputs

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 08 TOPICS

Unit 4 SAS Programming and Automation

4.1	Reading Date and Time with Informats
4.2	Displaying Date and Time Values with Formats
4.3	Basics of SAS Functions
4.4	SAS Function Syntax
4.5	Converting Data with Functions
4.6	The MEANS Procedure
4.7	Specifying Descriptive Statistics
4.8	Limiting Decimal Places
4.9	PROC FREQ Procedure
4.1O	Two-Way and N-Way Tables
4.11	LIST Options
4.12	Output Delivery System (ODS)
4.13	HTML Outputs
4.14	PDF Outputs
4.15	RTF Outputs
4.16	EXCEL Outputs

4.6	The MEANS Procedure
4.7	Specifying Descriptive Statistics
4.8	Limiting Decimal Places
4.9	PROC FREQ Procedure
4.1O	Two-Way and N-Way Tables
4.11	LIST Options
4.12	Output Delivery System (ODS)
4.13	HTML Outputs
4.14	PDF Outputs
4.15	RTF Outputs
4.16	EXCEL Outputs

VirupakshanK@gmail.com

4.6 The MEANS Procedure

What is PROC MEANS?

`PROC MEANS` is a SAS procedure used to calculate descriptive statistics for numeric variables.

It can calculate:

- Mean
- Sum
- Minimum
- Maximum
- Standard deviation
- Variance
- Number of observations
- Median
- Range

```
proc means data=dataset;
    var variable-list;
run;
```

VirupakshanK@gmail.com

Example Dataset & Basic PROC MEANS Example

```
data student_marks;
    input Name $ Gender $ Department $ Marks;
    datalines;
Asha F Statistics 85
Ravi M Computer 78
Meena F Statistics 92
Arun M Computer 74
Kavya F Mathematics 88
;
run;
```

```
proc means data=student_marks;
    var Marks;
run;
```

```
data student_marks;
   input Name $ Gender $ Department $ Marks;
   datalines;
Asha F Statistics 85
Ravi M Computer 78
Meena F Statistics 92
Arun M Computer 74
Kavya F Mathematics 88
;
run;

proc means data=student_marks;
   var Marks;
run;
```

VirupakshanK@gmail.com

Example Dataset & Basic PROC MEANS Example

```
data student_marks;
    input Name $ Gender $ Department $ Marks;
    datalines;
Asha F Statistics 85
Ravi M Computer 78
Meena F Statistics 92
Arun M Computer 74
Kavya F Mathematics 88
;
run;
```

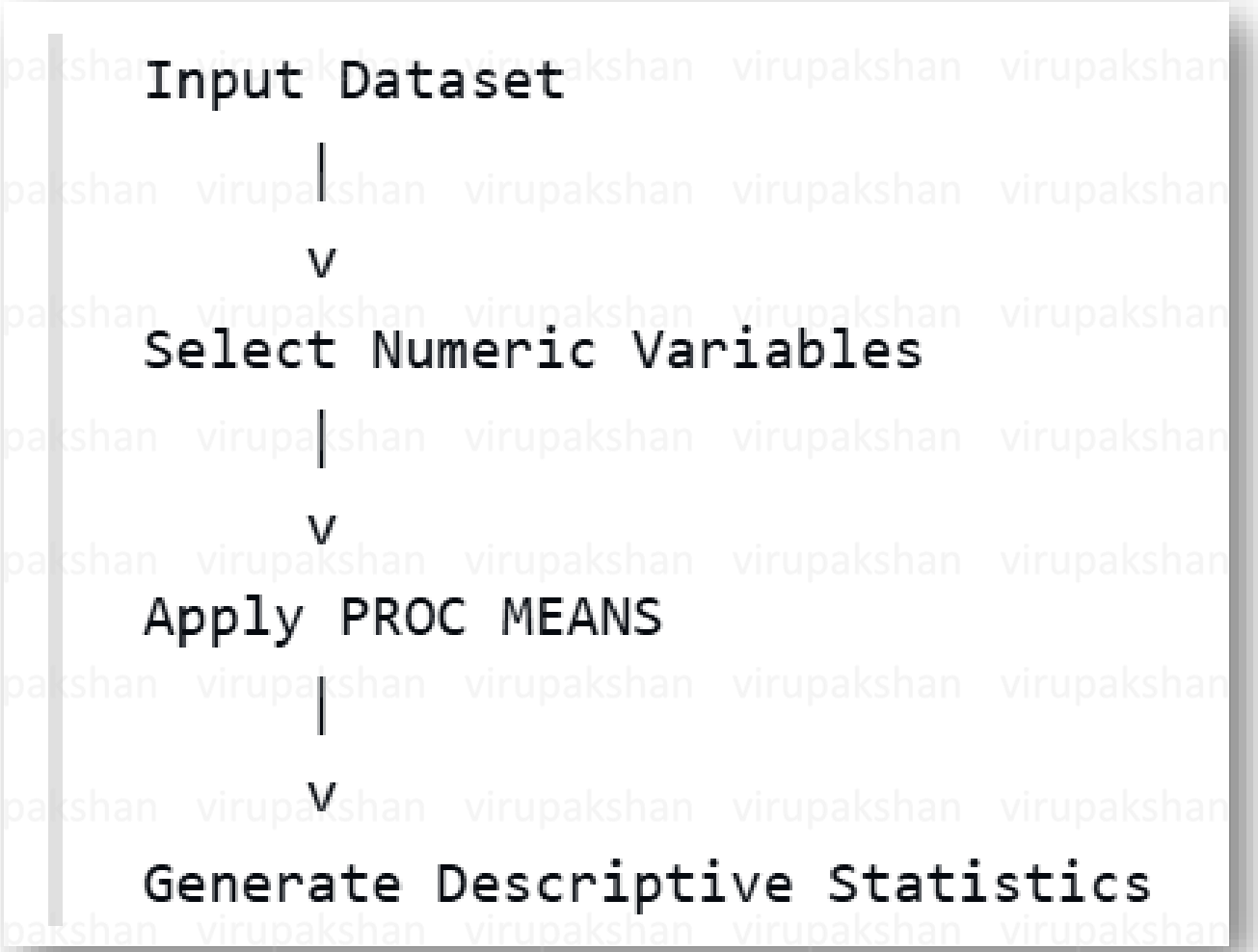
```
proc means data=student_marks;
    var Marks;
run;
```

Sample Output Concept

Statistic	Value
N	5
Mean	83.4
Std Dev	7.37
Minimum	74
Maximum	92

VirupakshanK@gmail.com

Diagram: PROC MEANS Workflow



VirupakshanK@gmail.com

4.7 Specifying Descriptive Statistics

Common Statistics in PROC MEANS

By default, `PROC MEANS` gives selected statistics. We can request specific statistics in the `PROC MEANS` statement.

Statistic Keyword	Meaning
N	Number of non-missing observations
NMISS	Number of missing observations
MEAN	Average
MEDIAN	Middle value
MIN	Minimum
MAX	Maximum
SUM	Total
STD	Standard deviation
VAR	Variance
RANGE	Maximum minus minimum
Q1	First quartile
Q3	Third quartile
CLM	Confidence limits for mean

VirupakshanK@gmail.com

Common Statistics in PROC MEANS

Formula: Mean

$$\bar{x} = \frac{\sum x_i}{n}$$

Where:

- $\bar{x}$ = mean
- x_i = each observation
- n = number of observations

Formula: Range

$$\text{Range} = \text{Maximum} - \text{Minimum}$$

Formula: Variance

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Formula: Standard Deviation

$$s = \sqrt{s^2}$$

VirupakshanK@gmail.com

Example: Specific Statistics

```
proc means data=student_marks n mean median min max std;
    var Marks;
run;
```

proc means data=student_marks n mean median min max std;
var Marks;
run;

Example: Statistics by Group

```
proc means data=student_marks mean min max;
    class Department;
    var Marks;
run;
```

proc means data=student_marks mean min max;
class Department;
var Marks;
run;

Explanation

- `class Department;` calculates statistics separately for each department.
- `var Marks;` specifies the numeric variable to analyze.

VirupakshanK@gmail.com

Example: Statistics by Group

```
proc means data=student_marks mean min max;
    class Department;
    var Marks;
run;
```

proc means data=student_marks mean min max;
class Department;
var Marks;
run;

Explanation

- `class Department;` calculates statistics separately for each department.
- `var Marks;` specifies the numeric variable to analyze.

Example Output Concept

Department	N	Mean	Minimum	Maximum
Computer	2	76.0	74	78
Mathematics	1	88.0	88	88
Statistics	2	88.5	85	92

VirupakshanK@gmail.com

4.8 Limiting Decimal Places

Sometimes SAS output contains many decimal places. We can limit decimals using options such as `MAXDEC`.

MAXDEC Option

The `MAXDEC=` option controls the maximum number of decimal places displayed in `PROC MEANS`.

Syntax

```
proc means data=dataset maxdec=n;
   var variable-list;
run;
```

Where `n` is the number of decimal places.

VirupakshanK@gmail.com

Example: Limit to Two Decimal Places

```
proc means data=student_marks mean std min max maxdec=2;
    var Marks;
run;
```

proc means data=student_marks mean std min max maxdec=2;
  var Marks;
run;

VirupakshanK@gmail.com

Example: Limit to Zero Decimal Places

proc means data=student_marks mean std min max maxdec=0;
   var Marks;
run;

```
proc means data=student_marks mean std min max maxdec=0;
    var Marks;
run;
```

VirupakshanK@gmail.com

Formatting Variables Before Analysis

You can also use formats to control how values appear.

```
data student_marks2;
    set student_marks;
    format Marks 8.2;
run;


proc print data=student_marks2;
run;
```

```
data student_marks2;
   set student_marks;
   format Marks 8.2;
run;

proc print data=student_marks2;
run;
```

VirupakshanK@gmail.com

Difference Between Rounding and Formatting

Method	Meaning
`round()` function	Changes or creates a rounded value
Format	Changes only the display
`maxdec=`	Controls decimal places in PROC MEANS output

VirupakshanK@gmail.com

4.9 PROC FREQ Procedure

VirupakshanK@gmail.com

What is PROC FREQ?

`PROC FREQ` is used to produce frequency tables.

It is mainly used for categorical variables such as:

- Gender
- Department
- Region
- Grade
- Yes/No responses

Basic Syntax

```sas
proc freq data=dataset;
    tables variable-list;
run;
```

VirupakshanK@gmail.com

Example Dataset

```
data students;
    input Name $ Gender $ Department $ Grade $;
    datalines;
Asha F Statistics A
Ravi M Computer B
Meena F Statistics A
Arun M Computer C
Kavya F Mathematics B
Nila F Computer A
;
run;
```

data students;
input Name $ Gender $ Department $ Grade $;
datalines;
Asha F Statistics A
Ravi M Computer B
Meena F Statistics A
Arun M Computer C
Kavya F Mathematics B
Nila F Computer A
;
run;

VirupakshanK@gmail.com

One-Way Frequency Tabled

```
proc freq data=students;
    tables Department;
run;
```

```
proc freq data=students;
  tables Department;
run;
```

Output Concept

Department	Frequency	Percent
Computer	3	50.00
Mathematics	1	16.67
Statistics	2	33.33

Percent Formula9

$$\text{Percent} = \frac{\text{Frequency}}{\text{Total Frequency}} \times 100$$

VirupakshanK@gmail.com

PROC FREQ Output Columns

Column	Meaning
Frequency	Count of observations
Percent	Percentage of total observations
Cumulative Frequency	Running total of frequencies
Cumulative Percent	Running total of percentages

VirupakshanK@gmail.com

4.10 Two-Way and N-Way Tables

VirupakshanK@gmail.com

Two-Way Tables

A **two-way table** shows the relationship between two categorical variables.

Syntax

```
proc freq data=dataset;
    tables variable1*variable2;
run;
```

proc freq data=dataset;
  tables variable1*variable2;
run;

Example: Gender by Department

```
proc freq data=students;
    tables Gender*Department;
run;
```

This shows how students are distributed by both gender and department.

Output Concept

Gender	Department	Frequency
F	Computer	1
F	Mathematics	1
F	Statistics	2
M	Computer	2

Cross-Tabulation Layout

	Department		
Gender	Computer	Mathematics	Statistics
F	1	1	2
M	2	0	0

VirupakshanK@gmail.com

N-Way Tables

An **n-way table** uses more than two categorical variables.

Syntax

```
proc freq data=dataset;
    tables variable1*variable2*variable3;
run;
```

Example: Gender by Department by Grade

```
proc freq data=students;
    tables Gender*Department*Grade;
run;
```

This produces a multi-dimensional frequency table.

VirupakshanK@gmail.com

When to Use N-Way Tables

Use **n-way tables** when you want to study frequency patterns across several categorical variables.

Example questions:

- How are grades distributed by gender and department?
- Which department has more male or female students with grade A?
- Is performance different across gender and department groups?

VirupakshanK@gmail.com

4.11 LIST Options

What is the LIST Option?

The `LIST` option in `PROC FREQ` displays two-way or n-way tables in a list format instead of a cross-tabular format.

This is useful when:

- There are many categories.
- The table is too wide.
- You want a compact summary.
- You are preparing output for reports.

```
proc freq data=dataset;
    tables variable1*variable2 / list;
run;
```

VirupakshanK@gmail.com

Example: Two-Way Table with LIST

```
proc freq data=students;
    tables Gender*Department / list;
run;
```

Output Concept

Gender	Department	Frequency	Percent
F	Computer	1	16.67
F	Mathematics	1	16.67
F	Statistics	2	33.33
M	Computer	2	33.33

Example: N-Way Table with LIST

```
proc freq data=students;
    tables Gender*Department*Grade / list;
run;
```

VirupakshanK@gmail.com

Other Useful PROC FREQ Options

Option	Purpose
`LIST`	Displays combinations in list format
`MISSING`	Includes missing values in frequency counts
`NOCUM`	Suppresses cumulative frequency and cumulative percent
`NOPERCENT`	Suppresses percentages
`NOROW`	Suppresses row percentages
`NOCOL`	Suppresses column percentages
`CHISQ`	Performs chi-square test

VirupakshanK@gmail.com

Example: Suppressing Percentages

proc freq data=students;
  tables Department / nocum nopercent;
run;

```
proc freq data=students;
    tables Department / nocum nopercent;
run;
```

NOCUM	Suppresses cumulative frequency and cumulative percent
NOPERCENT	Suppresses percentages

VirupakshanK@gmail.com

Example: Including Missing Values

```
proc freq data=students;
   tables Department / missing;
run;
```

```
proc freq data=students;
    tables Department / missing;
run;
```

`MISSING` Includes missing values in frequency counts

VirupakshanK@gmail.com

LECTURE 08 TOPICS

VirupakshanK@gmail.com

Unit 4 SAS Programming and Automation

4.1	Reading Date and Time with Informats
4.2	Displaying Date and Time Values with Formats
4.3	Basics of SAS Functions
4.4	SAS Function Syntax
4.5	Converting Data with Functions
4.6	The MEANS Procedure
4.7	Specifying Descriptive Statistics
4.8	Limiting Decimal Places
4.9	PROC FREQ Procedure
4.1O	Two-Way and N-Way Tables
4.11	LIST Options
4.12	Output Delivery System (ODS)
4.13	HTML Outputs
4.14	PDF Outputs
4.15	RTF Outputs
4.16	EXCEL Outputs

4.6	The MEANS Procedure
4.7	Specifying Descriptive Statistics
4.8	Limiting Decimal Places
4.9	PROC FREQ Procedure
4.1O	Two-Way and N-Way Tables
4.11	LIST Options
4.12	Output Delivery System (ODS)
4.13	HTML Outputs
4.14	PDF Outputs
4.15	RTF Outputs
4.16	EXCEL Outputs

VirupakshanK@gmail.com

LECTURE 08

21CSE295T

Statistical Data Analytics And Integration

Doubts?

THANK YOU!

Virupakshan K

+91-9840847872 | VIRU@KART.LA