[SAS Basic LEX](https://lex.infosysapps.com/web/en/viewer/web-module/lex_21013665361891790000?collectionId=lex_14701070819211012000&collectionType=Course&pathId=lex_24461790627328750000)

Merge l-130-131

SAS Basic Merge

DATA Employees;

INPUT Eid name $ Sal Did;

DATALINES;

101 Vinod 500 10

102 Avi 600 20

103 Navin 800 10

104 Ajinkya 750 20

RUN;

DATA Departments;

INPUT Did Dname $;

DATALINES;

10 HR

20 MFG

RUN;

PROC SORT DATA=Employees;

BY Did;

RUN;

PROC SORT DATA=Departments;

BY Did;

RUN;

DATA Emp\_Dept;

MERGE Employees Departments;

BY Did;

Report

* statistical analysis in SAS
  + descriptive statistics
  + frequency analysis
  + regression analysis
* reporting in SAS

The analytics team has been asked to study the sales transaction data and share their understanding with the management. The analytics team decides to generate the following information from the sales transaction data:

* number of sales
* average sales amount
* minimum price for any product
* maximum price for any product, etc.

Let us solve this by learning about **descriptive statistics** and use of **MEANS**procedure in SAS.

Company has sample sales data in dataset Year\_sales. Descriptive statistics like number of observation in dataset, total sales amount, maximum or minimum sale amount and average units sold by each sales person can be used to summarize sales data and present it in more meaningful way.

Descriptive statistics are important because if you simply present the raw data it would be hard to visualize what the data is showing, especially if there are more values. For example average sales of each sales representative will simply represent performance of sales representatives.

To learn more about descriptive analysis click [here](https://en.wikipedia.org/wiki/Descriptive_statistics).

Proc mean

PROC MEANS is used in a variety of analytical, business intelligence, reporting and data management situations to produce descriptive statistics.

**Basic and Default statistics:**

By default, PROC MEANS analyzes all numeric variables in your dataset and returns 5 default statistical measures as given below:

* N: Number of observations processed
* MEAN: Average of the analysis variable
* STD: Standard Deviation
* MAX: Largest Value
* MIN: Smallest Value

PROC MEANS DATA=datasetname <options>;

<statements>;

RUN;

**MEANS statement**

VAR: It is used to specify a variable for which descriptive statistics need to be generated. If not specified, descriptive statistics gets generated for all numerical variables available in dataset.

**MEANS options**

It is used to specify statistics you want to generate. MEANS procedure can generate N, MEAN, MODE, MEDIAN, SUM, STD, MIN, MAX etc.

FREQ

From the given sample sales data, the analytics team want to understand the frequency (no. of times a product was purchased) or the percentage of sales done by each sales representative, frequency analysis is the best practice.

To learn more about frequency analysis click [here](https://en.wikipedia.org/wiki/Frequency_analysis).

FREQ returns frequency which is the number of times an event occurs. It also returns:

* cumulative frequency (accumulation or the addition of successive frequencies)
* percentage of frequency
* percentage of cumulative frequency

The result is produced in matrix forma

PROC FREQ DATA=SASdataset;

TABLES variable(s) / options;

OUTPUT OUT=SASdataset;

RUN;

# Frequency procedure statement

* TABLES: To specify variable for which frequency analysis needs to be done
* OUT: To store result data in dataset format

PROC FREQ DATA=Year\_sales ;

TABLES Type;

REG

The company also maintains data of customers. The management wishes to know if there is any relation among the attributes, weight and age of customers. They would like an analysis that allows them to predict the weight when provided with the age of the customer.

To solve this, you need to understand the importance of **regression analysis** and the use of **REG**procedure in SAS

For more information on regression analysis visit [here](https://en.wikipedia.org/wiki/Regression_analysis).

In SAS, regression analysis can be done using REG procedure. Before using REG procedure you must have dependent variable and regressor variable.

* REG procedure helps analyst in finding the relation between two attributes and analyzing how one (independent) attribute influences other (dependent)
* It generates the regression equation which can be used to predict the target variable
* It also generates Fit plot which shows the positive slope of the fitted line and residual i.e. difference between the observed value of the dependent variable and the predicted value

PROC REG DATA=SASdataset options;

MODEL dependent(s) = regressors / options;

OUTPUT OUT=SASdataset;

RUN;

# REG procedure statement

* MODEL: to specify dependent and independent variables
* OUT: to specify dataset if result needs to be recorded in SAS

DATA Regression\_data;

INPUT AGE WEIGHT\_KG HEIGHT\_CM;

DATALINES;

1 10.2 76.1

2 12.3 85.6

3 14.6 94.9

4 16.7 102.9

5 18.7 109.9

6 20.7 116.1

7 22.9 121.7

8 25.3 127.0

9 28.1 132.2

10 31.4 137.5

11 32.2 140.0

12 37.0 147.0

13 40.9 153.0

14 47.0 160.0

15 52.6 166.0

16 58.0 171.0

17 62.7 175.0

18 65.0 177.0

RUN;

PROC REG DATA=Regression\_data ;

MODEL WEIGHT\_KG =AGE;

RUN;

Creating Reports using SAS

ODS is designed to overcome the limitations of traditional SAS output “proc print” . It provides a method of delivering output in a variety of formats and makes the formatted output easy to access. With ODS, you can create various file types including HTML, Rich Text Format, Portable Document Format, and SAS datasets. You can also take advantage of ODS features to easily convert existing SAS code so that it can create these file types. In addition, the variety of styles available within ODS enables you to enhance presentation output by controlling the report’s overall color scheme, font and size.

ODS output\_filetype FILE = 'filename and path';

<procedure step>;

ODS output\_filetype CLOSE;

# Output file types in SAS

* RTF: Rich Text Format
* PDF: Portable Document Format
* HTML: Hypertext Markup Language

ODS HTML FILE=Frequency\_output.html;

PROC FREQ DATA=Year\_sales ;

TABLES Type;

RUN;

ODS HTML CLOSE;