

SAS: A COMPLETE, COMPREHENSIVE AND INTEGRATED PLATFORM FOR STATISTICAL ANALYSIS

Latika Sharma* and Nitu Mehta Ranka**

ABSTRACT

This paper introduces software called SAS (Statistical Analysis System) which has many modules having utility in statistical analysis as well as data processing. SAS is a comprehensive software which deals with many problems related to Statistical Analysis, Spreadsheet, Data Creation, etc. It is a layered, multivendor architecture. The SAS System provides extensive statistical capabilities, including tools for both specialized and enterprise-wide analytical needs. Research institutes, laboratories, market research firms, universities, pharmaceutical companies, government agencies, banks, and insurance companies can all take advantage of these statistical capabilities. This paper focuses on traditional analysis of variance to exact methods to statistical visualization techniques, the SAS System provides the tools required to analyze data and help to make the right decisions. Software includes a wide range of statistical analyses, including analysis of variance, regression analysis, categorical data analysis, multivariate analysis, survival analysis, psychometric analysis, cluster analysis, and nonparametric analysis. SAS has Extensive utility in research studies pertaining to Economics. It can handle voluminous data of macro level studies with ease and perfection. Forecasting of economic variables with precision through time series data analysis is made possible by SAS.

INTRODUCTION

Excellent introduction to some of the ideas of empirical financial economics and Statistics of their technical exploration and investigation with SAS is provided by Boehmer et al. in "Using SAS in Financial Research". It provides insightful discussion of some of the most popular research topics in finance and statistics such as random walk theory, overreaction hypothesis, the capital asset pricing model, price-earnings relationship, forecasting market performance, event studies, including analysis of variance, regression analysis, categorical data analysis, multivariate analysis, survival analysis, psychometric analysis, cluster analysis, and nonparametric analysis and others. Ease and fluency of SAS in implementing the methodology of different types of empirical research in financial Economics and Statistics comes out naturally throughout in this paper. This implies some intermediate steps and media employed in order to access the original databases and handle initial data preparation. Being not necessarily straightforward and done outside of SAS, this makes

the case about having SAS working directly with the aforementioned databases for analysis as well as data access and preparation in uninterrupted flow.

RESEARCH METHODOLOGY

What is the SAS System?

The SAS System is a modular, integrated, and hardware-independent application. It is used as an information delivery system by business organizations, governments, and universities worldwide. SAS is used for virtually every aspect of information management in organizations, including decision support, project management, financial analysis, quality improvement, data warehousing, report writing, and presentations. However, this guide will focus on just one aspect of SAS: its ability to perform the types of statistical analyses that are appropriate for research in the Economics and education.

RESULTS AND DISCUSSION

Basic Structure of SAS

* Assistant Professor, Department of Agriculture Economics and Management, RCA, MPUAT, Udaipur.

** Senior Research Fellow, Department of Agriculture Economics and Management, RCA, MPUAT, Udaipur.

There are two main components to most SAS programs - the data step(s) and the procedure step(s). The data step reads data from external sources, manipulates and combines it with other data set and prints reports. The data step is used to prepare our data for use by one of the procedures (often called "procs"). SAS is very lenient about the format of its input - statements can be broken up across lines, multiple statements can appear on a single line, and blank spaces and lines can be added to make the program more readable. The procedure steps perform analysis on the data, and produce (often huge amounts of) output. The most effective strategy for learning SAS is to concentrate on the details of the data step, and learn the details of each procedure as we have a need for them.

Using the SAS System for Statistical Analyses

SAS is a particularly powerful tool for statistician, Economist and educators because it allows them to easily perform virtually any type of statistical analysis that may be required in their

research. SAS is comprehensive enough to perform the most sophisticated multivariate analyses, but is so easy to use that undergraduates can perform simple analyses after only a short period of instruction. In a sense, the SAS System may be viewed as a library of prewritten statistical algorithms. By submitting a brief SAS program, we can access a procedure from the library and use it to analyze a set of data.

Example

An experiment was conducted to study the hybrid seed production of bottle gourd (*Lagenaria siceraria* (Mol) Standl) Cv. Pusa hybrid-3 under open field conditions during Kharif-2005 at Indian Agricultural Research Institute, New Delhi. The main aim of the investigation was to compare natural pollination and hand pollination. The data were collected on 10 randomly selected plants from each of natural pollination and hand pollination on number of fruit set for the period of 45 days, fruit weight (kg), seed yield per plant (g) and seedling length (cm). The data obtained is as given below-

Group	No. of fruit Set (45days)	Fruit weight (kg)	Seed yield/plant (g)	Seedling length (cm)
1	7.0	1.85	147.70	16.86
1	7.0	1.86	136.86	16.77
1	6.0	1.83	149.97	16.35
1	7.0	1.89	172.33	18.26
1	7.0	1.80	144.46	17.90
1	6.0	1.88	138.30	16.95
1	7.0	1.89	150.58	18.15
1	7.0	1.79	140.99	18.86
1	6.0	1.85	140.57	18.39
1	7.0	1.84	138.33	18.58
2	6.3	2.58	224.26	18.18
2	6.7	2.74	197.50	18.07
2	7.3	2.58	230.34	19.07
2	8.0	2.62	217.05	19.00
2	8.0	2.68	233.84	18.00
2	8.0	2.56	216.52	18.49
2	7.7	2.34	211.93	17.45
2	7.7	2.67	210.37	18.97
2	7.0	2.45	199.87	19.31
2	7.3	2.44	214.30	19.36

{Here 1 denotes natural pollination and 2 denotes the hand pollination}

To Obtain mean, standard deviation, minimum

and maximum values of all the characters, for each group as well as a whole. To Obtain mean, median, coefficient of skewness, coefficient of kurtosis of all the characters and each group separately. Also

test whether the data follows a normal distribution or not for all the characters separately for each of the two groups. Prepare a discrete frequency table and 2 way frequency tables for all the characters for the above data on group. To create a stem and leaf plot and box plot for all the characters. Prepare a histogram for the grouped frequency distribution on seed yield per plant.

Solution:

For the analysis of the data SAS commands are given in the sequel.

{Here Number of fruit (45 days) is termed as fs45, Fruit weight (kg) is termed as fw, Seed yield/plant (g) is termed as syp and Seedling length (cm) is termed as sl. It may, however, be noted that one can retain the same name or can code in any other fashion}.

```
data descriptive_stats;
input group fs45 fw syp sl;
label fs45 = "No. of fruit Set (45days)";
label fw = "Fruit weight (kg)";
label syp = "Seed yield/plant (g)";
label sl = "Seedling length (cm)";
```

1. To obtain mean, standard deviation, minimum and maximum values of all the characters.;

```
proc means data=descriptive_stats;
title 'mean, standard deviation, minimum and
maximum values of all characters';
var fs45 fw syp sl;
run;
```

2. To obtain mean, standard deviation, minimum and maximum values of all the characters for each group separately.;

```
proc means data=descriptive_stats;
title 'mean, standard deviation, minimum and
maximum values of all characters for each group
separately';
```

```
var fs45 fw syp sl;
by group;
run;
```

3. To obtain mean, median, coefficient of skewness, coefficient of kurtosis of all the

characters.;

```
proc summary print mean median std skewness
kurtosis data=descriptive_stats;
title 'mean, median, coefficient of skewness,
coefficient of kurtosis of all the characters';
var fs45 fw syp sl;
run;
```

4. To obtain mean, median, coefficient of skewness, coefficient of kurtosis of all the characters for each of the group separately.;

```
proc summary print mean median std skewness
kurtosis data=descriptive_stats;
title 'mean, median, coefficient of skewness,
coefficient of kurtosis of all characters for each of
the group separately';
class group;
var fs45 fw syp sl;
run;
```

5. To test whether the data follows a normal distribution or not for all the characters and for each of the two groups separately.;

```
proc univariate normal data=descriptive_stats;
title 'test whether the data follows a normal
distribution or not for all the characters and for each
of the two groups separately';
class group;
var fs45 fw syp sl;
run;
```

6. To prepare a discrete frequency table for all the characters for the above data on group.;

```
proc freq data=descriptive_stats;
title 'discrete frequency table for all the
characters for the above data on group';
by group;
run;
```

7. To prepare 2-way frequency table between group and fruit set after 45 days.;

```
proc freq data=descriptive_stats;
title '2-way frequency table between group and
fruit set after 45 days';
table group*fs45;
```

```
run;
```

8. To create a stem and leaf plot and box plot for all the characters.;

```
proc univariate data=descriptive_stats plot;
  title 'a stem and leaf plot and box plot for all
the characters';
```

```
var fs45 fw syp sl;
```

```
run;
```

9. To create a stem and leaf plot and box plot for all the characters for each group separately.;

```
proc univariate data=descriptive_stats plot;
  title 'a stem and leaf plot and box plot for all
the characters for each group separately';
```

```
class group;
```

```
var fs45 fw syp sl;
```

```
run;
```

10. To prepare a histogram for the grouped frequency distribution for seed yield/plant (g) in question number 10.;

```
proc univariate data=seedyield noprint;
```

```
label syp = "seed yield/plant (g)";
```

```
histogram syp/ midpoints = 140 to 240 by 10;
```

```
run;
```

After submitting each of these SAS statements the log of the program can be seen in the log window and the results can be seen in the output window. The print of the results can be seen through proc print procedure as follows-

```
Proc print;
```

```
Run;
```

Save the Result:

We can save a result and its contents to any location:

Select File Save As. The Save window opens and prompts to choose whether to save the project.

In the Save window, select a location for the project. In the File Name box, type Project file name, it will be saved .

CONCLUSION

SAS Software includes a wide range of statistical analyses, capital asset pricing model, price-earnings relationship, forecasting market performance, event studies, including analysis of variance, regression analysis, categorical data analysis, multivariate analysis, survival analysis, psychometric analysis, cluster analysis, and nonparametric analysis. Forecasting of economic variables with precision through time series data analysis is made possible by SAS. This one of-a-illustrates how to use SAS software to conduct basic analyses of data.

REFERENCES

- Boehmer, E., Broussard, J.P. and J.-P. Kallunki. 2002. Using SAS in Financial Research. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. (1990), SAS Language: Reference, Version 6,First Edition, Cary, NC: SAS Institute Inc.
- SAS Institute Inc. (1989), SAS/FSP Software: Usage and Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc.
- Prasad, R., Gupta, V.K., AK and Kole, B. Analysis of Data. Design Resources Server. Indian Agricultural Research Institute, New Delhi.

