

A NOTE ON THE DEVELOPMENT OF A LOCAL STATISTICAL PACKAGE FOR SMALL COMPUTERS¹

by

Rudy H. Tan & Ceferino S. Sincioco²

Introduction

Statistical packages like the BMDP, SPSS, SAS, GENSTAT, MINITAB, etc., have relieved applied statisticians of the tedium of computer programming. Furthermore, these packages have permitted researchers from other fields to employ advanced statistical techniques with minimum assistance from statisticians. Those who use statistical packages are primarily interested in statistics and computing as tools of research and have very little interest in computer programming. Generally, they want to undertake a complete and searching analysis of their data as quickly as possible, and then spend more time interpreting the results.

Although there are many computer installations in the Philippines, only a few have statistical packages. The reasons include the high rental cost and large core storage requirement of these packages. Thus, the cost of using a statistical package like SPSS or SAS in any of the commercial computers in Metro Manila is still very expensive. The absence of any statistical package in the three computer installations in the University of the Philippines at Diliman, Quezon City, has prevented the Statistical Center from offering and developing computer-based data analysis courses for its students and researchers from other fields who are interested in applied statistics.

¹Paper presented at the 1981 Philippine Statistical Association Annual Conference held at the Central Bank of the Philippines, Manila.

²Professor and Associate Professor, respectively, of the Statistical Center, University of the Philippines, Diliman, Quezon City.

This paper announces the local development of a statistical software package, called STATPACK, under a project funded by the National Research Council of the Philippines. STATPACK would first be used in the Statistical Center starting next semester (November, 1981) in courses like exploratory data analysis, multivariate analysis, time series analysis and forecasting, and statistical simulation and modelling. Eventually, it would be made available to institutions and individuals at very low cost compared to other imported statistical packages. Details about this package are discussed in the next section.

STATPACK Design

STATPACK is a comprehensive statistical software package designed specifically for small (mini and micro) computers. The reference computer is the School of Economics PANAFACOM U-SERIES 300 with 32 to 35 K-bytes user's memory. The programming design is modular, and presently it consists of the following eight modules:

- ST-00 General purpose and linkage module
- ST-01 Descriptive statistics module
- ST-02 Contingency table module
- ST-03 Correlation and regression module
- ST-04 Multivariate statistics module
- ST-05 Experimental design module
- ST-06 Time series and forecasting module
- ST-07 Statistical simulation module

Modules 01 to 07 can be compiled separately and each one has its own complete set of utility subroutines. Two or more modules can be interfaced using the linkage module. Also, a user can write his own program and interface it with any of the modules. The number of modules that can be compiled together, however, will depend on the size of the main storage capacity of the computer. Modular programming design may not be practical for large computers because some compilers impose large overhead costs on separately compiled modules. However, this approach appears to be advantageous for small computers. Because of the limited storage capacity

of small computers, modules are compiled only when they are needed by the users. The compilation cost is usually offset by the cheaper CPU or clock time of small computers.

STATPACK is written in FORTRAN IV and should run on different machines with only slight modifications. It makes use of single precision and currently, the arrays are dimensioned to handle up to 20 variables and 200 observations per variable. The programs can be redimensioned to handle larger problems, depending on the storage capacity of the computer.

The command structure of STATPACK will be patterned after MINITAB. For example, to run Box and Jenkins' time series analysis using the School of Economics computer, the Time Series Analysis and Forecasting Module is compiled together with the parameter cards to produce outputs for identification, preliminary estimation, final estimation, and diagnostic checking (see Appendix A).

STATPACK outputs are free of unwanted statistics and can readily be incorporated in a report. Variable labelling and data print-out are at the user's option.

STATPACK Programs

STATPACK will initially consist of the following statistical programs:

- ST-01.1 Descriptive Statistical Analysis (summary statistics from ungrouped data)
- ST-01.2 Descriptive Statistical Analysis (summary statistics from grouped data)
- ST-01.3 Descriptive Statistical Analysis (grouping of raw data into a frequency distribution)
- ST-02.1 Construction and Analysis of Two-Way Contingency Table
- ST-02.2 Construction and Analysis of Three-Way Contingency Table
- ST-03.1 Simple Correlation Analysis
- ST-03.2 Partial and Multiple Correlation Analysis

- ST-03.3 Simple Linear Regression Analysis (one value of Y for a given X)
- ST-03.4 Simple Linear Regression Analysis (Several values of Y for a given X)
- ST-03.5 Simple Linear Regression Analysis (X and Y are measured with error)
- ST-03.6 Curvilinear Regression Analysis (Logarithmic)
- ST-03.7 Curvilinear Regression Analysis (Power)
- ST-03.8 Curvilinear Regression Analysis (Exponential)
- ST-03.9 Curvilinear Regression Analysis (Polynomial)
- ST-03.10 Nonlinear Regression Analysis (Fitting of $Y = \beta_0 + \beta_1 \beta_2^x + \epsilon$)
- ST-03.11 Nonlinear Regression Analysis (Fitting of $Y = \beta_0 + \beta_1 x \beta_2 + \epsilon$)
- ST-03.12 Multiple Linear Regression Analysis
- ST-04.1 Principal Components Analysis
- ST-04.2 Factor Analysis (orthogonal factors with rotation by varimax or quartimax method)
- ST-04.3 Cluster Analysis (single linkage cluster)
- ST-04.4 Multivariate analysis of variance (one-way classification)
- ST-04.5 Canonical correlation analysis
- ST-04.6 Multiple discriminant analysis
- ST-05.1 Completely randomized design
- ST-05.2 Randomized Complete Block Design
- ST-05.3 Latin Square Design
- ST-05.4 Factorial Experiment (Two or three factors in CRD or RCBD)
- ST-05.5 Balanced Incomplete-Block Design
- ST-05.6 Probit Analysis
- ST-06.1 Holt's Nonseasonal Time Series Analysis
- ST-06.2 Winter's Seasonal Time Series Analysis
- ST-06.3 Box and Jenkins' Time Series Analysis
- ST-06.4 Spectral Analysis
- ST-07.1 Normally Distributed Random Numbers
- ST-07.2 Uniformly Distributed Random Numbers
- ST-07.3 Binomially Distributed Random Numbers

ST-07.4 Poisson Distributed Random Numbers

ST-07.5 Erlang Distributed Random Numbers

STATPACK Future

Developing a statistical software package takes a long time and requires large financial resources. The limited funding received from NRCPC for this project will terminate by the end of November, 1981. However, the authors are optimistic that financial assistance can be obtained from other sources in order to continue this project for at least two more years. Due to some constraints, no support of whatever form has been received from the U.P. Statistical Center, which is the principal beneficiary of this project.

Since STATPACK has been designed for small computers, a commercial version in BASIC (Beginner's All Purpose Symbolic Instruction Code) will have to be developed for WANG, APPLE, TRS-80, SHARP, and other microcomputers. Furthermore, there is a need to have additional modules for univariate statistical tests, data editing and quality control, file manipulation, discrete multivariate analysis, general linear models, response surface methodology, etc. Also, some of the existing modules may have to be expanded to include, for example, spline regression analysis, stepwise regression analysis, hierarchical and non-hierarchical clustering techniques, factor analysis by the method of maximum likelihood, and a general multivariate analysis of variance.

The future of STATPACK will lie in its portability and flexibility. A user will be able to run a module at a computer installation with the cheapest computer time. Thus, even if the U.P. Computer Center has already a SPSS, it may still be practical for a student to use STATPACK at the School of Economics or Transport Engineering where the queue is short and the computer time is relatively cheaper. Also, special modules can easily be designed for specific applications and integrated with the existing package.

APPENDIX A

Example of STATPACK Control Cards for Running
Box and Jenkins' Time Series Analysis Program

```
//JOB RHTAN XXXXXXXX
//4430
PROG 3
DATA 11F5.3
IDEN 144 1 1 12 1.0000 0.0000 1
PEST 0 0 112 0
FEST 0 0 112 0
DIAG 1 1
STOP
/END
//
```

Parameter/Option Cards

Card 3	Col. 1-4	PROG	[Program card]
			[1 - Holt's nonseasonal time series analysis]
			[2 - Winter's seasonal time series analysis]
	Col. 6-7	CODE	[3 - Box and Jenkins' time series analysis]
			[4 - Spectral analysis]
Card 4	Col. 1-4	DATA	[Data format card]
	6-12	FORM	[F-format specification]
Card 5	Col. 1-4	IDEN	[Identification card]
	6-10	NOBS	[Number of observations]
	11-15	NRDF	[Number of regular differencing]
	16-20	NSDF	[Number of seasonal differencing]
	21-25	OSDF	[Order of seasonal differencing]
	26-30	DMU1	} Value of μ_1 }
	31-35	DMU2	

		36-37	PLOT	[1 – Plot auto and partial auto- corrections 0 – Do not plot auto and partial autocorrelations]
Card 6	Col.	1-4	PEST	[Preliminary estimation card]
		6-7	ORAR	[Order of regular autoregressive term]
		8-9	OSAR	[Order of seasonal autoregressive term]
		10-11	ORMA	[Order of regular moving-average term]
		12-13	OSMA	[Order of seasonal moving-average term]
		14-15	MEAN	[1 – Working series is to be centered 0 – Working series is not to be centered]
Card 7	Col.	1-4	FEST	[Final estimation card]
		6-7		
		8-9		
		10-11	(Same as for PEST)	
		12-13		
		14-15		
Card 8	Col.	1-4	DIAG	[Diagnostic checking card]
		6-7	PROB	[1 – Plot residuals on probability chart 0 – Omit residual analysis]
Card 9	Col.	1-4	STOP	[Termination card]