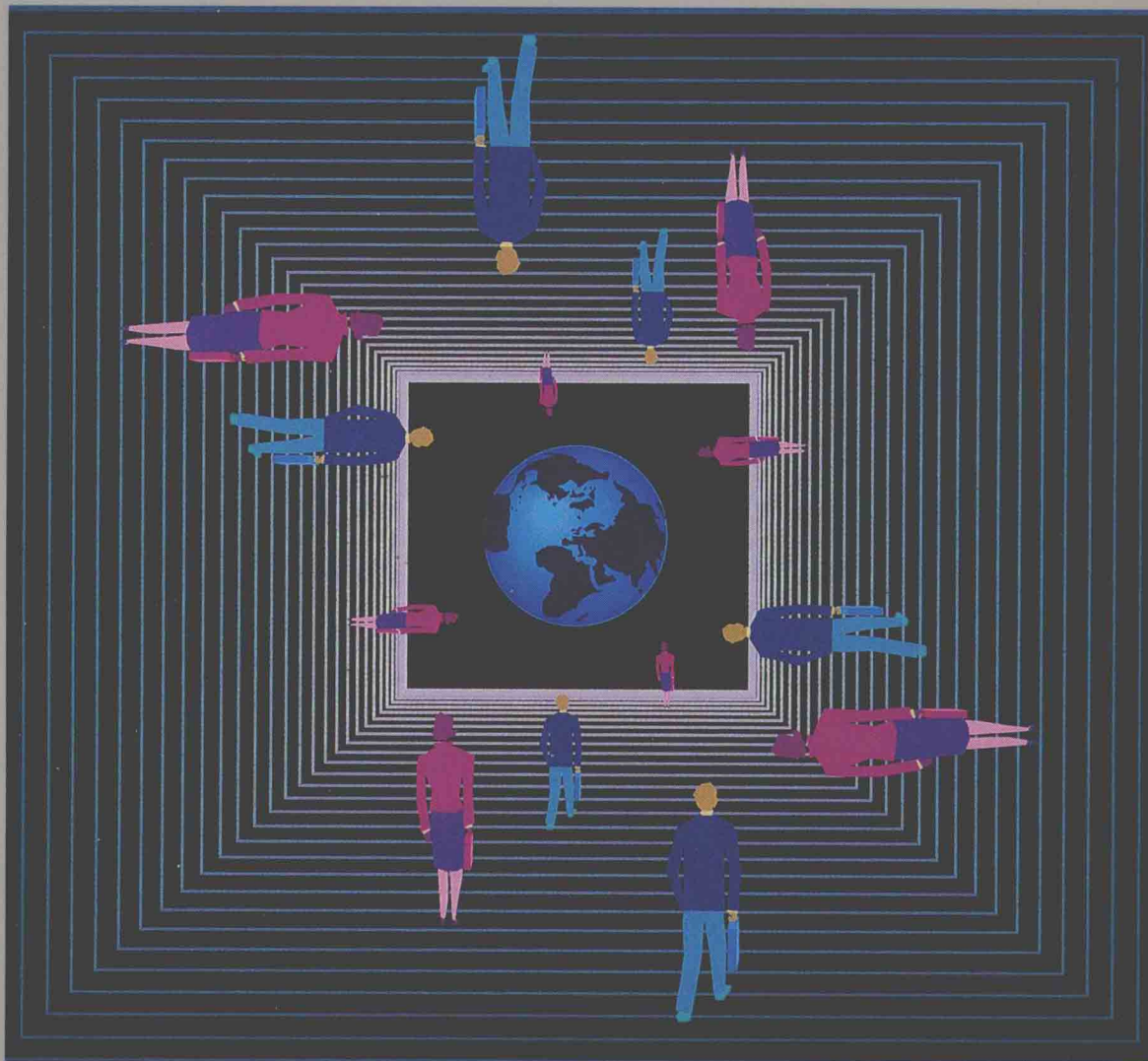# ISP

INTERACTIVE STATISTICAL PROGRAMS

SPYROS MAKRIDAKIS
ROBERT L. WINKLER

ISP

INTERACTIVE STATISTICAL PROGRAMS

SPYROS MAKRIDAKIS
INSEAD

ROBERT L. WINKLER
Duke University

Preface

This book is designed to be used in conjunction with a set of interactive statistical programs, called ISP, which has been developed to improve and facilitate the teaching and learning of basic statistics. The ISP software and this book are intended for use in an introductory statistics course. Alternatively, they can be used for self-paced individual learning of statistics, as a "refresher course" for those who studied statistics some time ago, and as a convenient computational tool and accompanying reference manual for users of statistics. Introductory statistics courses are seldom among the more popular courses in the curriculum, and we strongly believe that students of statistics can profit from using this book and the ISP software as an integral part of an introductory statistics course.

A very important aspect of ISP and this book is that *no* previous background is needed in either statistics or computers, and the mathematics needed is also minimal (say, the equivalent of the prerequisite for a very low-level basic statistics course). The software is extremely easy to use, is completely interactive and user-friendly, and requires absolutely no programming. ISP facilitates the learning of statistics by removing the computational burden of statistics and enabling students to concentrate on statistical concepts. Moreover, provisions for working with real data (including data chosen by the instructor or student), sampling from data sets, conducting simulations, and experimenting with games of chance help to make statistics more interesting and to give students a better understanding of the nature of uncertainty, sampling variability, and statistical modeling.

The book begins with an introduction to ISP which describes various aspects of the programs and should enable a user to get started with ISP. The rest of the book is organized by statistical concepts. First, it is divided into seven parts, covering descriptive statistics, probability and probability distributions, sampling and sampling distributions, estimation and hypothesis testing, comparing two or more populations, regression and correlation, and forecasting. Each part is then divided into chapters. A chapter begins with a brief introduction of the concepts of interest, gives a description of the ISP commands that are useful for these concepts, provides a solved example to illustrate both the statistical concepts and the ISP commands, and includes a few exercises to help you use ISP and learn the concepts.

For information about purchasing the ISP programs, write to

> Lincoln Systems Corporation
> P.O. Box 391
> Westford MA 01886

or call   (617) 692-3910.

The programs are written in FORTRAN and are currently available for VAX computers and IBM-compatible personal computers. The micro version of ISP consists of three modules which are available separately or as a package:

> Main Module—Data handling, Statistical spreadsheet, Transformations, Distributions, Basic statistical functions, Games
>
> Training Module—Descriptions of statistical concepts, Exercises
>
> Correlation/Regression Module—Correlation, Regression, Forecasting

The micro version will run on any IBM-compatible personal computer with 256kb RAM, two 320kb floppy disk drives or one floppy and one hard disk, PC or MS DOS (version 2.0 or greater), and a monochrome or color monitor. The VAX version of ISP covers all of the material in the three micro modules and runs on any Digital VAX with tape drive. For information on possible adaptations for other computers, write or call Lincoln Systems Corporation.

We are grateful to all of the people who provided help and encouragement in bringing this work to completion. Paule Villain did much of the original programming for ISP, and Tim Davidson and Mike Harde were involved in adaptations. Students at INSEAD used ISP in various stages of development and provided valuable feedback, as did some colleagues who taught ISP. INSEAD and Indiana University also provided important support to facilitate the development of ISP and the writing of this book. Finally, Dick Fenton gave helpful advice during the development stage and David Farr did an excellent job of shepherding the manuscript through the production process.

We hope that this book and the ISP software will make the learning of statistics more interesting and more enjoyable and that it will help provide students with a better understanding and appreciation of statistics.

SPYROS MAKRIDAKIS
Fontainebleau, France

ROBERT L. WINKLER
Durham, North Carolina

Contents

Introduction to ISP: Interactive Statistical Programs

ISP is a set of Interactive Statistical Programs to help you learn statistics more effectively and with less effort by using a computer terminal or a microcomputer. ISP allows you to start applying from the very beginning the various statistical methods and concepts you are learning. It takes full advantage of today's computer capabilities to facilitate learning by simplifying the data handling and computational chores. The ISP training modules have been developed for people having little or no mathematical background. As you gain experience, you will soon find that ISP's extensive collection of statistical functions is not only valuable pedagogically but extremely useful in a wide variety of practical applications of statistics.

The purpose of this introductory chapter is to give you an idea of the nature and scope of ISP. It should also provide enough information about some basic commands for entering and manipulating data to enable you to start using ISP. Occasionally, a technical term such as "mean" or "standard deviation" is used in this chapter. If you do not understand these terms, don't worry. They will be explained and illustrated later in this book.

THE ISP COMMAND LANGUAGE AND THE MENU-DRIVEN APPROACH

ISP responds to various commands, which are easy to learn and which instruct the computer to do a variety of tasks. A list of these commands can be found in the *ISP Quick Reference Guide*, or you can see them on the screen by typing NAMES after ⟨n⟩, the command prompt. The various commands are grouped in categories (teaching MODULES, STATISTICS, GAMES, TRANSFORMATIONS, etc.), which can be listed at the terminal by entering the first one or two letters of each group of commands. For

instance, the letter G entered after ⟨n⟩ will provide you with a list of all the games-related commands.

GAMES and SIMULATIONS

GCARD:	Simulates a deck of playing cards
GCOIN:	Simulates the tossing of coins
GCOMB:	Combinations of M elements taken K at a time
GDIE :	Simulates the throwing of dice
GLOTO:	Simulates the game of lotto
GMARB:	Simulates sampling marbles from an urn
GPERM:	Permutations of M elements taken K at a time
GROUL:	Simulates the game of roulette
GSAMP:	Selects a sample from a population of your choice
GTACK:	Simulates the tossing of a thumbtack

In the menu-driven version of ISP, the programs are even easier to use because you are prompted by a menu, which is a list of alternatives. By entering a letter or moving a cursor on the screen, you choose a command or group of commands. For instance, when you first enter ISP, the main menu, listing the major groups of commands, is shown on the screen.

Interactive Statistical Programs

ENTER THE LETTER OF YOUR CHOICE OR MOVE CURSOR TO YOUR CHOICE

A Introduction to statistics and ISP
B Games and simulations
C Transformations
D Basic statistics
E Calculations
F Distributions
G Plotting and printing
H Regression and analysis of variance
I Forecasting
J Enter, edit, store, and list data/files; modify level/language
X Exit

If you choose the group of commands called BASIC STATISTICS, a menu showing all of the commands in this group will be displayed.

Basic Statistics

ENTER THE LETTER OF YOUR CHOICE OR MOVE CURSOR TO YOUR CHOICE

A Compute the covariance and correlation of two variables
B Compute point and interval estimates

C Compute expected values
D Test hypotheses
E Compute the skewness and kurtosis
F Compute the mean
G Compute the median
H Compute the mode
I Compute the range
J Compute the standard error of the mean
K Compute the variance and standard deviation
L Bayesian statistics
M Summary statistics
R Return to main menu
X Exit

Note that the options include returning to the main menu or exiting from the program.

Occasionally you may find that you have inadvertently chosen the wrong command or answered some question inappropriately while using the command. Many of the commands ask you which column of your data matrix you would like to work on. If you respond with a "0" (zero), the program will exit from the command and let you choose a new command or repeat the old command. As noted above, options such as returning to the main menu or exiting from the programs are available in the menus. To end your session with ISP, you can choose the Exit option or you can type BYE, EXIT, STOP, END, or QUIT.

SELF TEACHING MODULES

There are more than 40 teaching modules available in ISP, including descriptive statistics, probability, sampling, estimation, hypothesis testing, comparison of populations, regression and correlation, and forecasting. These modules explain major statistical concepts and tell you how such concepts can be applied or illustrated through the various ISP commands. In order to use a teaching module you simply need to type its name after ⟨n⟩. Refer to the *ISP Quick Reference Guide* for a complete list of the teaching modules; alternatively, type M in ISP to get a list of the modules. To use any of the teaching modules, type the command (such as MDA3) after ⟨n⟩.

⟨n⟩ MDA3

***** Measures of Location *****

The mean, median, and mode are three commonly used measures of location. They tell us something about the "typical" member of a data set. The mean is found by adding up all of the members of a data set and then dividing this sum by how many numbers there are in the set. This average is called the arithmetic mean and is widely used in statistics. The command SMEAN can be used to compute the mean of a set of numbers. Alternatively, the command PRINT provides a list of the data set with the sum and the mean.

If the values of a data set are ordered from the smallest to the largest, the median is the middle number in this ordered list of numbers. The command SMEDI can be used to find the median of a set of numbers. Alternatively, the command TSORC can be used to order the data from the smallest to the largest. The median is then the middle value.

The mode is the number that appears most often in a set of data. This number can be found by using the command SMODE. Alternately, the command TSORC can be used. Since the numbers are sorted, those that have the same value will be next to each other and can be counted easily to find the value that occurs with the highest frequency.

Here is an example of the use of the commands SMEAN, SMEDI, and SMODE for a set of 230 ages.

⟨n⟩ SMEAN
　　***** Computes the Mean *****
Which column(s) do you want to work on ? AGES
Mean of your 230 numbers for:
— Column 1 (AGES) :   27.839130

⟨n⟩ SMEDI
　　***** Computes the Median *****
Which column(s) do you want to work on ? AGES
Median of your 230 numbers for:
—Column 1 (AGES) :   28.000000

⟨n⟩ SMODE
　　***** Computes the Mode *****
Which column(s) do you want to work on ? AGES
Mode(s) of your 230 numbers for:
—Column 1 (AGES) :   28.000000

To get some practice with these commands, take a file from those that are stored in the computer (type LISTF for a list of the available files), or enter your own data set, and use SMEAN, SMEDI, and SMODE. Also, try to find the mean, median, and mode without using these commands. (The commands PRINT and TSORC will be helpful here.) By using different data sets and looking at their histograms as well as the three measures of location, you can see how the shape of the histogram affects the mean, median, and mode.

HELP STATEMENTS

With some ISP commands, you are asked if you want help in using the command. A yes answer will tell the computer to provide a brief "HELP" statement concerning the command. Furthermore, even when the computer does not ask if you want assistance, in most instances you can get some clarification or a "HELP" statement by pressing the carriage return key instead of answering a question posed by the computer. The "HELP" statements make it easier to use ISP without having to refer to this book repeatedly during a session with ISP. The book does, however, provide explanations of statistical concepts and examples of the use of ISP commands. The ISP programs are very easy to use, and the book is a useful supplement to the programs.

In terms of assistance, ISP can be used at two levels. While prompts such as "Do you need help?" are often welcome to a beginning user of ISP, a more experienced user may find that having to read and respond to such questions is irritating. Thus, you are given a choice of levels: "beginner" or "experienced user." To change from one level to the other, simply use the command CHANG.

CHOICE OF LANGUAGE

ISP is available not just in English, but also in other languages such as French. Some versions of the ISP programs include more than one language. For such versions, you will be asked which language you would like to use when you first use the programs. If you want to change the language at any time, the command CHANG can be used to change language as well as level.

SAVING RESULTS

Once you have finished with your ISP session, type BYE, EXIT, STOP, END, or QUIT. This will end your session with ISP. However, remember that nothing is lost while using ISP (unless the computer breaks down or you break out of ISP). Everything important appearing on the screen is kept in the file named RESULT.DAT, which is automatically stored in the memory of the computer until you use ISP again. Every time ISP is used, the old contents of RESULT.DAT are destroyed and a new version is created. If you do not want this to happen and you want to keep the contents of the RESULT.DAT file, rename it before using ISP again.

EXERCISES TO REINFORCE LEARNING

There are plenty of training exercises available in ISP that allow you to reinforce learning and to get practical experience using the various statistical concepts. You can use the exercises by typing MEXER after ⟨n⟩. An indication of how well you have done on these exercises will be given to you after you finish each session.

GAMES TO ILLUSTRATE PROBABILITY AND UNCERTAINTY

A good deal of statistics involves learning about probabilities and understanding uncertainty. This learning can be facilitated by the various games of chance available in ISP. For a list of the games available, see the *ISP Quick Reference Guide*, or type G after ⟨n⟩. In order to play any of the games, type its name after the ⟨n⟩ prompt.

In order to use the modules, the exercises, or the games, you do not need to enter any of the data on your own. Any required information is supplied in the programs. You only need to answer the various questions asked by the programs. These will be adequate to get results or to achieve some learning objective. However, if you do want to work with your own numbers you will first have to enter your data in the computer. (See the discussion of datafile operations later in this chapter.)

COMPUTING WITH ISP AS A HAND CALCULATOR

In some cases you might wish to do arithmetic operations on data. For instance, you might want to do some calculations such as

$$(35 + 4(29.3)/6)^2 + \sqrt{29 + 126} \quad .$$

You can find the value of this expression by using ISP's CALC command.

⟨n⟩ CALC
>   * * * Hand Calculator * * *

Enter the mathematical expression to be computed :
(35 + 4*29.3/6)^2 + (29 + 126)^.5

The result of your computations is : 2986.33430

Do you have more computations ? N

If you have some simple calculations to do, you can use any of the C (computing) commands, which allow you to use ISP as a powerful hand calculator. For instance, squaring five numbers and adding the results can be done as follows:

⟨n⟩ CSXX

>   ***** Sums n Squared Numbers *****

Enter the n numbers :
12.3 25 32.12 67 25.5

The sum of your   5 squared numbers is 6947.2344

The computer does not store these data or remember them for subsequent uses.

ISP AS A STATISTICAL SPREADSHEET

ISP provides the statistics student or practitioner the ability to create a multiple-column worksheet, using typical transformations on a column-by-column basis. The ISP transformations include arithmetic operations, differencing, standardizing, and sorting. Here is an example:

⟨n⟩ TCC

***** Squares all elements of a Column *****

Observ.	Value	(This is now column no: 2)
1	729.00000	
2	625.00000	
3	625.00000	
4	1024.0000	
5	900.00000	
6	841.00000	
.	.	
.	.	
.	.	
225	841.00000	
226	841.00000	
227	625.00000	
228	1369.0000	
229	1156.0000	
230	841.00000	
Sum	180113.00	
n =	230	
Mean	783.10000	

Give a title for variable no. 2: XSQU

⟨n⟩ PRINT

* * * To Print Data * * *

Do you want to print:
1 . The first 8 and last 8 observations
2 . Rows specified by you
3 . Successive rows
4 . All rows
Enter a number from 1 to 4: 1

Observ.	AGES	XSQU
* 1*	27.000	729.000
* 2*	25.000	625.000
* 3*	25.000	625.000

*	4*	32.000	1024.000
*	5*	30.000	900.000
*	6*	29.000	841.000
*	7*	26.000	676.000
*	8*	28.000	784.000
		.	.
		.	.
	.	.	.
*	223*	34.000	1156.000
*	224*	34.000	1156.000
*	225*	29.000	841.000
*	226*	29.000	841.000
*	227*	25.000	625.000
*	228*	37.000	1369.000
*	229*	34.000	1156.000
*	230*	29.000	841.000
Sum		6403.0000	180113.0000
n =		230	230
Mean		27.8391	783.1000

ISP'S STATISTICAL FUNCTIONS

Many commands are provided to allow you to use all of the concepts learned in the teaching modules on the statistical spreadsheet data. These include correlation, regression, confidence intervals, hypothesis testing, and many descriptive statistics (mean, mode, median, standard deviation, standard error, range, variance, skewness, kurtosis, etc.). In addition, several important probability distributions are provided.

PLOTTING AND PRINTING DATA

Various ISP commands are available to print or plot your data at the terminal. They range from listing data for editing purposes (using PRINT and EDIT) to illustrative graphics (such as histograms, scatter plots, time series plots, and plots of data and theoretical distributions). For instance, to print your data, type PRINT after ⟨n⟩.

⟨n⟩ PRINT

* * * To Print Data * * *

Observ.	X1	X2	Y
* 1*	12.000	30.000	100.500
* 2*	15.000	43.000	124.000
* 3*	14.000	35.000	114.300

* 4*	22.000	63.000	205.000
* 5*	19.500	38.700	127.100
* 6*	33.200	52.000	301.000
* 7*	26.000	75.000	251.000
* 8*	25.000	70.000	209.000
* 9*	44.000	131.000	438.000
* 10*	29.000	73.000	333.700
Sum	239.7000	610.7000	2203.6000
n =	10	10	10
Mean	23.9700	61.0700	220.3600

DATAFILE OPERATIONS

There are many practical applications that require data to be used in several different statistical operations. For instance, you might want to calculate the mean and the variance of the data, and later on you might decide that you want the median and the mode of the same data. It would be impractical and time consuming if you had to input your data anew each time you wanted to compute different statistical measures. To avoid unnecessary effort you can enter your data once and keep them in the memory of the computer until you finish your session. In order to do this you must enter your data by the ISP commands ENTER or RENTE. The command ENTER can be used as follows to create a one-variable datafile:

⟨n⟩ ENTER

* * * To Enter a Data Set * * *

Need help ? N

Do you want to input your data:
1 . From the terminal
2 . From a file

Enter 1 or 2: 1

Enter n numbers:
123,105.23 88 129 100,120.1234 99.23,120,115 104

Observ.	Value	(This is now column no: 1)
1	123.00000	
2	105.23000	
3	88.000000	
4	129.00000	
5	100.00000	
6	120.12340	
7	99.230000	
8	120.00000	
9	115.00000	
10	104.00000	
Sum	1103.5834	
$n =$	10	
Mean	110.35834	

Give a title for variable no. 1: X

When the numbers are entered, they must be separated by either a space or a comma so that the computer knows when one number ends and another begins. In the example, spaces are used in some instances and commas in other instances.

The command RENTE can be used as follows to create a datafile containing several variables.

⟨n⟩ RENTE

* * * To Enter a Multivariate Data Set * * *

Need help ? N

Do you want to input your data:
1 . From the terminal
2 . From a file

Enter 1 or 2: 1

Number of observations ? 10

Number of variables ? 3

For observation   1 Enter   3 variables
12 30 100.5

For observation   2 Enter   3 variables
15 43 124

For observation   3 Enter   3 variables
14 35 114.3

For observation   4 Enter   3 variables
22 63 205