

STATISTICAL TECHNIQUES FOR MANAGERS

Harry V. Roberts

Benjamin F. King

Draft textbook prepared for use in QMB 6603, *Data Analysis for Managers*, a course in the MBA Program, College of Business, Florida Atlantic University.

This textbook is designed to be used with the software package, *SPSS 12.0 Student Version for Windows*, created by SPSS, Inc., Chicago, IL, and distributed by Prentice-Hall. *SPSS 12.0* runs under Microsoft Windows 98, 2000, XP, Me, or NT 4.0. It requires a Pentium Class processor, 128 MB RAM, 220 MB of drive space, CD-Rom drive, and an SVGA monitor.

This material may not be quoted or reproduced without permission from the authors. Inquiries should be addressed to *bfking@adelphia.net*.

Copyright 2004

Harry V. Roberts

Benjamin F. King

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND OVERVIEW		1-1
1.	Background	1-1
2.	Quality Management	1-1
3.	Quality Management and Traditional Management	1-3
4.	Statistics and Quality	1-5
5.	Installing <i>SPSS 12.0 (Student Version)</i> and Getting Underway	1-7
CHAPTER 2: DATA ANALYSIS AND STATISTICAL CONTROL		2-1
1.	Introduction	2-1
	"Lightning Data Sets"	2-1
	The Role of <i>SPSS</i>	2-2
2.	A Simple Target Game to Illustrate Statistical Control	2-2
3.	Preliminary Entry and Examination of Data	2-6
4.	The Control Chart	2-8
5.	Statistical Control and Visual Checks	2-11
6.	Runs Check for Statistical Control	2-13
	"Expected Number of Runs"	2-15
	Comparison of Actual and Expected	2-16
	"Significance"	2-16
	The 0.05 Rule of Thumb	2-17
7.	Distribution of the Data	2-17
	An Important Caution	2-19
8.	Normal Distribution	2-19
9.	Why Does the Normal Distribution Sometimes Provide a Smoothed Description of Sample Histograms?	2-24
10.	Checking for Normality of Distribution	2-24
	The Normal Probability Plot	2-28
11.	Statistical Control, Process Capability, and Tampering	2-32
	Process Capability	2-33
	Tampering	2-34
	A Common and Disastrous Form of Tampering	2-39
	"Hot Streaks and Slumps"	2-42
12.	"Outliers" and "Special" or "Sporadic" Causes	2-43
13.	Two Further Applications	2-46
	Timing the Passage of Ten Seconds	2-46
14.	Checking for Bias	2-53
	Target Game	2-53
	Timing 10 Seconds	2-57
	Guessing 10 Seconds	2-58
Appendix: FURTHER REMARKS ON THE RUNS TEST	2-59	

CHAPTER 3: SIMPLE EXTENSIONS OF CONTROL CHARTS	3-1
1. Introduction: Counting Data and Other Complications	3-1
Statistical Control but Nonnormal Histogram	3-2
Not in Statistical Control	3-2
Counting Data	3-3
Conditions Leading to the Poisson Distribution	3-8
Another Application: "Non-Words" in Public Speaking	3-9
2. Data Arising as Time Intervals: the Exponential Distribution	3-11
A Quality Management Application of Cycle Times	3-17
3. Analysis of Defect Rates	3-20
The "Hot Hand"	3-23
4. Checking Stock Market Information	3-24
An Alternative, and Appropriate, Analysis	3-27
5. A New Wrinkle: Trend	3-30
The Mortality Data Reconsidered	3-32
6. Another New Wrinkle: Periodic Effects	3-32
7. A Final New Wrinkle: Intervention Effects	3-37
Appendix: RANDOM WALKS	3-41

CHAPTER 4: INTRODUCTION TO REGRESSION	4-1
1. Fitting Trends by Regression: the Running Example	4-1
The Basic Regression Command:	4-6
The Key Result: the Fitted Line	4-7
Making the Plot Look Pretty	4-8
The Save Feature in <i>SPSS</i> Regression	4-12
Geometric Interpretation of Residuals	4-15
The Principle of Fitting the Line	4-16
Diagnostic Checking	4-16
Standardized Residuals for Checking Normality	4-18
Interpretation of Trend	4-19
Diagnostic Checking for Linearity	4-19
Regression and Prediction	4-21
2. Trend Fitting by Regression: the Intensive Care Example	4-23
Statistical Significance of the Apparent Trend	4-27
"Significance" or "p-value"	4-29
Significance of Trend in the Running Data	4-30
Practical Interpretation of "Significance"	4-30
3. Fitting Periodic Effects by Regression: the Ishikawa data	4-31
"Indicator" Independent Variables for Periodic Effects	4-31
Diagnostic Checking	4-36
4. Fitting Intervention Effects by Regression: the Putting Data	4-38
5. Introduction to Multiple Regression: Allowance for Severity in the Intensive Care Application	4-44
Multiple Regression of MORTRATE on both TIME and SEVERITY	4-46

Analysis of Variance	4-47
Graphical Display of Multiple Regression	4-48
Diagnostic Checking of Multiple Regression	4-49
Relationship between Independent Variables	4-55
Effects of Correlation between Independent Variables	4-60
APPENDIX: MORE ABOUT PREDICTED VALUES	4-61
CHAPTER 5: INTERVENTION ANALYSIS AND RANDOMIZED EXPERIMENTATION	5-1
1. Introduction	5-1
2. Stepwise Regression and Variable Selection: Medical Audit	5-2
Systematic Selection of Independent Variables Using	
STEPWISE REGRESSION	5-11
An Alternative Approach to Modeling	5-20
Introducing the Medical Audit as an Explicit Indicator Variable	5-21
Diagnostic Checking	5-23
Visualizing the Fit	5-25
Systematic Strategies for Regression Analysis	5-29
3. A Randomized Experiment	5-31
The Role of Randomization	5-39
Summary	5-42
4. Randomized Pairs	5-43
Simplicity and Usefulness of Randomized Pairs	5-48
5. Intervention, Experimentation, and Statistical Control	5-48
CHAPTER 5 APPENDIX 1: GRAPHIC DEPICTION OF NEAR-MULTICOLLINEARITY	5-50
CHAPTER 5 APPENDIX 2: TWO SAMPLE t-TESTS	5-53
CHAPTER 6: AUTOREGRESSION	6-1
1. Introduction	6-1
2. A Chemical Reactor Process	6-2
Have We Overlooked Anything?	6-14
3. A More Complicated Application of Autoregression: History of a Diet	6-15
What Lies Behind the Regression Model?	6-25
4. Introduction of Information on Calories (C) to Improve the Diet Regression Model	6-25
R-SQUARE	6-32
R-SQ(ADJ) versus UNADJUSTED R-SQUARE	6-32
5. Study of Calories Consumed (C); Lessons Learned	6-36
6. Time-Series Regression Strategy and Hazards	6-39
"Rash Regression"	6-40

CHAPTER 6 APPENDIX: FURTHER REMARKS ON R-SQ AND R-SQ(ADJ). . . . 6-44

CHAPTER 7: CROSS-SECTIONAL DATA ANALYSIS AND REGRESSION	7-1
1. Introduction	7-1
2. Special Causes and Pareto Analysis	7-2
Operator Errors	7-2
Library Book Borrowing	7-5
Air Delays in October, 1987	7-7
Performance Differences among Baseball Teams	7-8
3. Simple Cross-Sectional Regressions	7-10
Sales Proposals and Sales	7-10
Selling Time and Sales	7-14
Transformation for Nonconstant Variance	7-21
Lightning Data Set: Predictability of Professional Football	7-24
4. Paired-sample Experiment for Cross-Sectional Data: BOYSHOES	7-27
5. Car Gas Mileage, Studied by Cross-Sectional Regression	7-29
Time-Series Checks Are Not Applicable	7-38
6. Successive Cross Sections and Performance Assessment	7-39

PREFACE TO CHAPTER 8: WHY DO WE DO FACTORIAL EXPERIMENTS?

CHAPTER 8: TWO-LEVEL EXPERIMENTS WITH SEVERAL VARIABLES	8-1
1. Introduction	8-1
2. The Pilot Plant Experiment: Full Factorial Design with Replication	8-2
A Visual Explanation of Interaction	8-10
3. Two-level Factorial without Replication: Running Experiment	8-16
Optional Computer Preliminaries	8-16
Actual Data Analysis	8-20
Analysis of Both Blocks: All 16 Observations	8-23
4. A Fractional Factorial Experiment: Top Spinning	8-27
5. A Pool Experiment: 2² Factorial with Blocking	8-33

CHAPTER 1: INTRODUCTION AND OVERVIEW

1. Background

This textbook is an outgrowth of a series of textbooks and notes written by Harry V. Roberts for courses in the MBA and the executive programs at the Graduate School of Business of the University of Chicago. The last course that Roberts, now retired, taught at Chicago evolved from an Executive MBA statistics course called "Statistics for Managers", a course that stressed practical data analysis, with informal, mostly non-mathematical, treatment of statistical ideas. Over the years Roberts kept introducing quality management ideas to motivate data analysis. For example, control charts were introduced at the very start, as a natural way of getting into statistics. Since statistics and quality management turned out to be remarkably synergistic, it was not necessary to drop any of the statistics. Hence "Statistics for Managers" gradually evolved into "Statistics and Quality Management".¹

The present textbook, **Statistical Techniques for Managers (STM)**, rather than adopting a total orientation toward the ideas of quality management, is designed for a course that is closer to the conventional required course in statistics for a typical MBA program. It does not, however, abandon the quality and continuous improvement approach to business problems, and, as the reader will see, there are frequent references to those important concepts. Although we try to maintain the quality management spirit that is present in Roberts' other texts, we also intend, as these notes evolve, to discuss matters of data analysis that are not so readily associated with statistical process control as it is taught, say, in a course in operations management, but which are deemed important topics for an introductory statistics course by faculty in other areas in a school of business such as accounting, marketing, and organizational behavior. Examples planned for the future are the chi-square procedure for dealing with frequency counts in contingency tables, and the discussion of nonresponse in sample surveys.

2. Quality Management

A good short definition of quality management is **doing the right things right the first time**. But that definition doesn't tell you what the right things are or how you should go about doing them right. Consider then this more explicit and only slightly longer definition:

Continually serve customers better and more economically, using scientific method and teamwork, and focusing on removal of all forms of waste.

The term "process" is key to quality management: **a process is a well-defined series of**

¹Roberts' earlier book, *Data Analysis for Managers with Minitab (DAFM)*, Second Edition, Boyd & Frasier, 1991, was used in "Statistics for Managers" at the University of Chicago and in "Data Analysis for Managers" at Florida Atlantic University. The course materials that Roberts last used at Chicago resemble **DAFM** in some ways but they place much greater emphasis on the non-statistical side of quality management.

steps by which inputs are converted to outputs, either goods or services. Continual improvement of work processes (by small or large increments) is guided by the two simple goals implicit in the phrase "serve customers better and more economically".

- "Serving customers better" means adding more value, increasing customer satisfaction. **(This serves as an implicit definition of the word "quality" itself: anything that serves customers better, as customers see it, is an improvement of quality.)**
- "Serving customers more economically" means reducing cost to provide any given customer service.

"Scientific method and teamwork" are the two main strategies for serving customers better and more economically. Scientific method entails the application of ideas of systematic learning -- mainly the domain of statistics. Teamwork entails processes by which people can cooperate in doing their work -- processes that owe much to research in the behavioral sciences.

"Focusing on removal of all forms of waste" provides the focus for improvement efforts. The meaning of waste is clear intuitively, but there is more to waste than is apparent at first glance. Some waste is obvious, but even in well-managed organizations, waste -- even substantial waste -- can be pervasive without making its presence obvious.

The working assumption of quality management is that further improvements -- increased customer satisfaction, lower costs -- are always possible, no matter how long and hard you have been working to improve, and that never-ending improvement is necessary for long-term survival of most organizations.

Opportunities for improvement are recognized primarily by continuing questioning of the effectiveness of the processes by which organizational work is done. This questioning leads to:

- Orientation of all efforts towards delighting customers and removing waste in organizational processes.
- Stress on team effort at all levels inside and outside the organization, including cooperative efforts with suppliers and customers.
- Use of data and scientific reasoning to guide improvement efforts, and to hold the gains from past improvements.

The foundation of quality management is philosophical: the scientific method in general and statistics in particular. Quality management includes systems, methods, and tools. The systems permit change; the philosophy stays the same. Quality management is anchored in values that stress the dignity of the individual and the power of community action.

Quality management is a total system approach and an integral part of high-level strategy.

It works horizontally across functions and departments, involves all employees, top to bottom, and extends backwards and forwards to include the supply chain and the customer chain. It stresses learning and adaptation to continual change as keys to organizational survival.

The ideas of quality management pose a profound challenge to us all: they say that, no matter what we have done up to the present moment, we must be prepared to find that we can do substantially better. This is gratifying in the sense that improvement is always gratifying, but it is a bit discouraging because it also suggests that our past accomplishments will always seem inadequate in the light of present knowledge. We can never rest on our laurels.

3. Quality Management and Traditional Management

Quality management leads to many management practices that are very different from traditional practices. The new practices are often so appealing that many people will insist that they have been following them all along. They will say that quality management is just common sense. In a broad interpretation of "common sense", this may be true, but, if so, common sense has often been missing in traditional management practice.

For example, traditional management thinking has emphasized optimization within currently fixed constraints, and has often taken "currently fixed" to mean "immutable". Traditional management thinking says that improved quality necessarily increases costs, because, given traditional work processes, we can improve quality only by more intensive inspection to weed out defective output, and added inspection obviously adds to cost.

Quality management stresses the importance of going beyond optimization within the currently fixed constraints to aim at ever-moving improvement targets by relaxing or eliminating the current constraints. We continually try to improve current processes so that less defective output will be created in the first place; less rework or scrapping of defective output will be needed; and less inspection will be necessary. Less defective output will reach the inspectors who remain, and costs will go down. Hence we can obtain higher quality for lower cost.

We are thus not confined to deciding whether improvements in quality attributable to increased inspection are worth the costs of the increased inspection. This is narrow "tradeoff thinking": how much quality improvement can be gained in exchange for an additional expenditure on inspection? By looking beyond this immediate tradeoff, we can explore the possibility of a broad tradeoff: **investment in the improvement of processes**, which can lead both to lower cost and higher quality.

In short, process improvement may be an excellent internal investment. Whenever possible, quality management tries to sidestep painful narrow tradeoffs by investment in improved processes, and there are likely to be many opportunities to do so. These opportunities exist because of the prevalence of widespread waste in many organizations.

Here are additional examples of divergences between quality management and traditional management:

- Under many circumstances, choosing suppliers by competitive bidding based on lowest price has been discredited. There are potential advantages in cooperative rather than adversarial relations with suppliers, and these may more than offset the market discipline imposed by competitive bidding.
- Unlike the older programs of "quality control", quality assurance and improvement are not tasks for professional specialists alone; they concern everyone in an organization. Upper management usually initiates quality management, but everyone must be involved, not only in special team improvement projects but in applying quality ideas to everyday work.
- Older "quality control" concepts, such as "acceptable quality level" (AQL) and "average outgoing quality limit" (AOQL), can have a crippling effect on quality improvement. This can happen because AQL and AOQL are minimum standards. Minimum standards -- however high they initially may seem -- can discourage further improvement once they are met. Quality improvement, by contrast, is to be never-ending.
- Cutting employment -- downsizing -- to cut costs is a traditional response to economic adversity. Downsizing is inherently crude and demoralizing; it removes organizational bone and muscle along with organizational fat. Under quality management, the first approach is to reduce errors and wastes, and this reduction leads to reduced costs, and there need be no downsizing so long as employees can be switched to other tasks. Normal attrition can then work to achieve the needed economies.
- Quality management also continually calls attention to the possibility of increasing revenues by increasing customer satisfaction, which can be done not only by improving present products and services but by searching for new products and new markets. Moreover, employees whose current jobs in production or middle management become unnecessary can often be transferred to sales and marketing, with the aim of expanding demand for company products.
- Creativity in working around problems, expediting shipments, and setting up inventories for emergencies can impede the continuing long-term improvement that comes from removing the flaws in business processes that create crises. For example, in quality management, the slogan, "If it's not broken, don't fix it", changes to, "If it's not broken, it's a good time to think about developing better preventive maintenance practices".
- In many circumstances, aiming at long production runs and large batch sizes is not a wise strategy. It may be better to acquire the ability to change production setups quickly so that several product variations can be produced every day or week, thus more closely matching production and market demand. Being able to respond rapidly to demand allows companies to avoid the traditional marketing emphasis on trying to persuade customers to buy things

they don't want.

- Middle managers and staff specialists need to emphasize training, coaching, and facilitation rather than simply giving or passing on orders. Exhortations to "do better" or to "try harder" seldom lead to substantial and sustained improvement; specific changes in work practices and processes must be made.

If you are confused about the claims of proponents and critics of quality management, use the above examples as test cases: in each of these clashes between the quality management approach and the traditional principles of management, which principle seems more reasonable to you?

Some people say that quality management is good but that it is not new. It is true that most of the numerous component ideas of quality management have been around a long time.² However, successful **synthesis** of these components can lead to organizations that look very different from traditional organizations. One suggestive name for such organizations is "learning organization".

Finally, there is an ongoing displacement of traditional management principles by their conflicting counterparts from quality management, often without conscious awareness or attribution. Quality management is becoming the new traditional management.

4. Statistics and Quality

Statistics has been described as "a body of methods for making wise decisions in the face of uncertainty".³ An important key to the improvement of a process is understanding the nature of that process and the characteristics of other processes that are related to the process of primary interest. For example, suppose that our principal aim is to improve the sales of our company's product, while at the same time reducing waste and enhancing customer satisfaction. One aspect of the process that we face is the number of units sold each month (day, week, year -- whatever). The exact value of that number, before it happens and before it is measured, is subject to uncertainty. It makes sense that reducing the uncertainty associated with future sales via some statistical estimation method could result in considerable savings in the cost of the company's operations. We call that activity of reducing uncertainty "forecasting"-- only one of the many activities that constitute the field of statistics. Further statistical investigation might lead us outside of the immediate area of company data analysis-- we might, for example, hire a marketing research firm to select a sample of our customers and potential buyers to elicit their opinions on ways in which service could be improved. Although potentially very useful, the results of such a survey are subject to the uncertainty caused by sampling and imperfect measurement. Other economic consultants would

²One exception is the concept of just-in-time production in the production of a variety of products or services; this concept was invented by Toyota well after World War II, in the process of implementing quality management. It has revolutionized manufacturing world wide, and its ideas are beginning to find application to services.

³Roberts, H.V., and Wallis, W.A., *Statistics: A New Approach*, Glencoe: The Free Press, 1956.

perhaps use statistical techniques to study the relationships between **fundamental factors** such as Gross National Product and the performance of our firm with the aim of better understanding the mechanism behind our success or failure in the business arena. Company engineers might perform **controlled experiments** to determine whether proposed new manufacturing methods will be effective or not. A psychologist at the local university, funded by tax-deductible contributions from our firm, might investigate the different personality traits that seem to be **statistically related** to preferences for one type of warranty over another. You can see that statistics, a legitimate field of scientific endeavor, is much more than just keeping track of the yardage for Monday night football or providing the data that are found in almanacs.⁴

Unfortunately, statistics as a college subject can be badly taught and, consequently, very dull.⁵ We hope to improve upon this all too traditional situation by means of the following:

- We encourage students to work through the material in **STM** step by step. You will see that the text in the following chapters is richly enlaced with output from the statistical package, **SPSS**. You should have ready access to this software and you should replicate the analyses as you read. That means that you should read the text with hands close to a computer keyboard. Only when you really get the hang of data analysis, by virtue of a lot of practice, can you safely distance yourself from that keyboard.
- Many of the examples, and those assigned for homework, are real data from real-life statistical situations, available on the data diskette that accompanies this book. Other good sources of data, however, are “**lightning data sets**”—Roberts’ name for interesting and meaningful data sets that you can collect very quickly, often in a few minutes. We present many lightning data sets in the following chapters and give some general guidance on how to go about gathering your own lightning data sets.
- We avoid wherever possible the kinds of formulas that are often found in even the most introductory statistics textbooks. If we do indulge in more mathematical explanations of statistical concepts we present most of that material as technical notes that may be skipped without serious damage to the learning process.
- We believe that real learning comes from collecting and analyzing your own data, preferably data that you find understandable, interesting, and relevant to some problem in your own life, either working or personal. You will find that the software package, **SPSS**, makes such personal projects very easy to accomplish.

⁴ “A statistician is somebody who is pretty good with figures, but who doesn’t have the personality to be an accountant.” *Anonymous*. (With apologies to our CPA friends.)

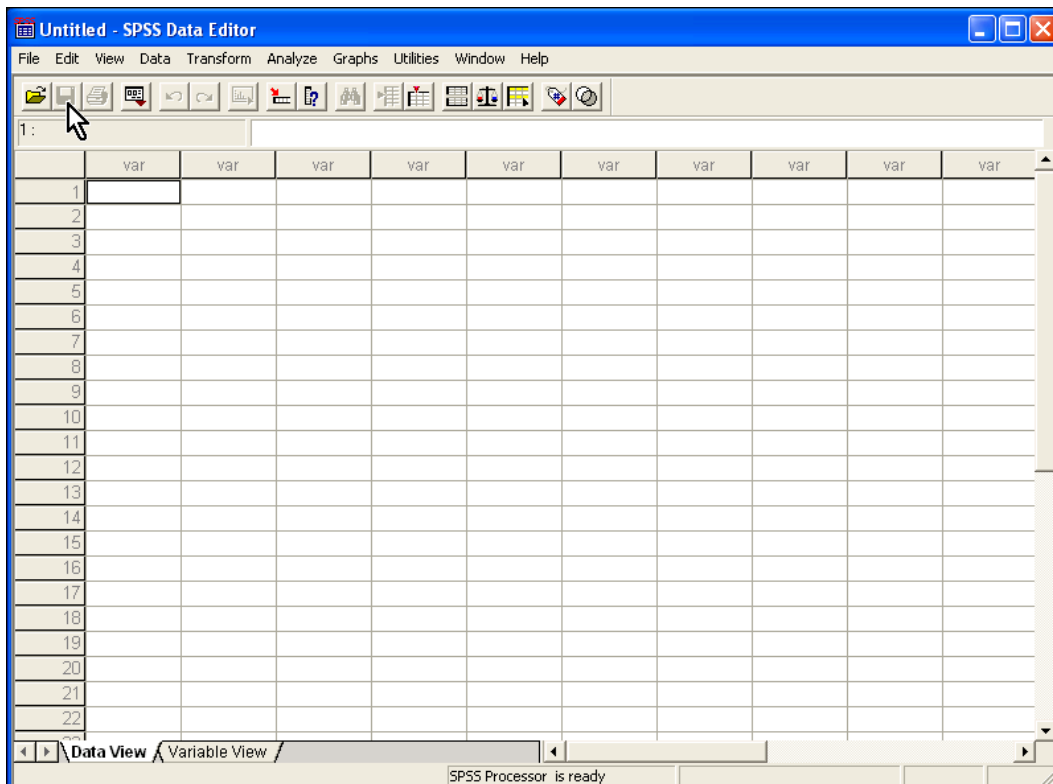
⁵ “If this were my last day on earth I would rather spend it in my statistics class. It would seem so much longer” *Anonymous*. (That fellow *Anonymous* sure gets around.)

5. Installing *SPSS* and Getting Underway

For the first few lectures at least you should bring the manual for *SPSS* to class along with this textbook. We encourage you to install *SPSS* as soon as you can and start experimenting with it even before the first class session. After successfully installing *SPSS* you should read the manual, even if you do not understand everything at the first reading.⁶ You should also work through as much of the tutorial as possible. This will help you navigate about in the program. Don't worry if things are a bit confusing at first. We shall tell you about, and demonstrate with data, all that you need to know about *SPSS* for this course.

You will see that the *SPSS* program contains many sets of data that can be used in the tutorial and elsewhere to illustrate various statistical techniques. For this course, however, we require certain additional data files. These will be made available to you on an FTP site as soon as the course begins.

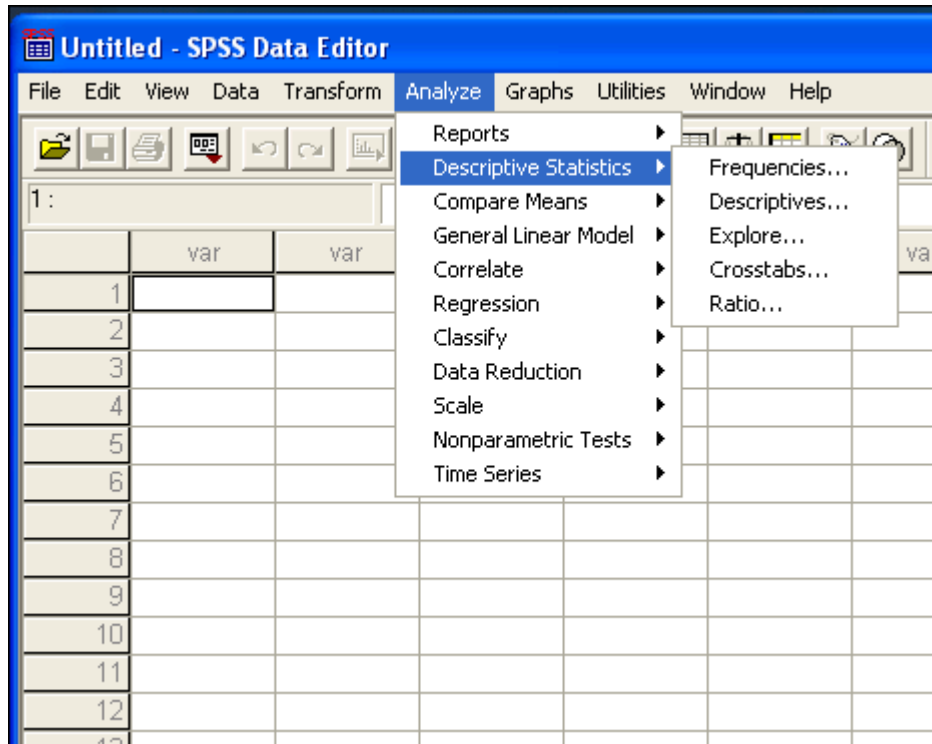
When you initiate your first *SPSS* session the following screen will appear:



Placing the mouse arrow on one of the choices listed on the menu bar just below the title bar that says **Untitled – SPSS Data Editor** and left-clicking will open a menu. For example, if your choice is the

⁶ Not all of the procedures discussed in the manual entitled *SPSS 12.0 Brief Guide* are available in the student version that you have received. An example is the topic “Working with Syntax” in Chapter 10. The statistical analytical tools that are available, however, are sufficient for this course.

sixth from the left, **Analyze**, the following topics menu appears:



Note that under the topic **Descriptive Statistics** there are further choices in a second menu.

Using **SPSS** may seem daunting at first, especially if you have never seen an interactive statistical package before, but all of the necessary details are described in the **SPSS 12.0 Brief Guide**, which you must read diligently at the beginning of this course, or by the help function in the program, and in the subsequent chapters of this set of notes. After a few trials at working with **SPSS** you will see how extremely user-friendly it is. In a short time you will hardly ever have to refer to the manual when analyzing real data.