CRIME ANALYSIS TECHNICAL ASSISTANCE

FINAL REPORT
LEAA GRANT #76-TA-04-0002

CENTER FOR CRIMINOLOGICAL
ANALYSIS AND TRAINING

July 1, 1977

INTRODUCTION

The Center for Criminological Analysis and Training (CCAT) was established in March, 1976 and began operation with a $40,245.00 grant from the Law Enforcement Assistance Administration (LEAA), Region IV.

CCAT is housed within the Criminal Justice Program of The University of Tennessee at Chattanooga (UTC). UTC is a major campus of the University of Tennessee whose state-wide system consists of five primary campuses. UTC's urban campus is located in downtown Chattanooga and serves over 6000 students. The overall faculty numbers around 300, while the criminal justice faculty numbers 5 full-time and several part-time members.

The undergraduate criminal justice program has over 250 majors. A new master's program in criminal justice will be initiated in Fall, 1977. The criminal justice program is not only involved with teaching and research, but also with the practical aspects of how the criminal justice system relates to the community as a whole. Within the criminal justice program, CCAT forms a link between the teaching aspects of the program and the practical problems involved in doing research in the "real world." The desire that the activities of the classroom be supplemented with an understanding of the real problems of research has assisted some students in learning planning, evaluation, statistics and methodology. Closely related to the teaching and research functions of the program, CCAT provides a vehicle for public service in the areas of research and training related to crime.

The technical assistance* grant awarded CCAT provided an opportunity to fulfill university public service objectives while at the same time meeting a need perceived among the eight state planning agencies (SPA) of LEAA Region IV - Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee.

Each year since 1968 the SPA's have undertaken the development of comprehensive plans for the reduction and/or control of crime in their respective states. Crime Analysis (CA) has generally been veiwed as a major component of the rational criminal justice planning model advocated both by Congress and LEAA. The LEAA M4100.1E planning guidelines define CA as "An analysis of the scope, nature and trends in crime in the state and it's subdivision in sufficient detail so that the analysis provides the data base for rational planning." With this in mind, the CCAT project was designed to assist the states in preparing and submitting acceptable crime analysis portions of their plans.

---

*Technical assistance is the providing of professional opinion and knowledge within a particular field. Within the parameters of this project technical assistance refers specifically to provision of professional opinion and knowledge to the field of crime analysis. (defined on page 4 of this report)

PROJECT OBJECTIVES

The primary objective of the grant was "to increase the capacity of State Planning Agencies (SPAs) to be responsive to the new Law Enforcement Assistance Administration (LEAA) crime analysis guidelines as set forth in M4100.1E for the 1977 planning effort."*

Other objectives of the grant were to provide technical assistance to SPA personnel in the following ways:

1. technology for building analytical capabilities (software);
2. tools for crime analysis (statistical techniques); and
3. techniques for applying the above to a body of knowledge in a manner that sufficiently meets the LEAA Comprehensive Planning guideline criteria.

The technical assistance (TA) called for in the project was of two types: proactive and reactive. It was proactive in that certain fundamentals of crime analysis were categorized and a training curricula was designed and subsequently delivered in the form of a workshop in order to provide a foundation in CA before the remainder of the technical assistance was delivered. The fundamentals covered included:

1. Data sources - what sources are available, key data elements, developing additional data sources
2. Analytical techniques - useful statistical tools, review of underlying theory, computerized application of statistical tools. These techniques were generally geared toward building competency in the areas of regression, correlation, time series, significance, etc.

---

*See appendix A for a summary of the data elements and analysis required and suggested in the M4100.1E guidelines.

3. Presentation of data - types of illustrations and choosing the most effective method of communicating data findings

The reactive portion of TA was designed to provide the states with professional opinions and advice from CCAT on a consultative basis in the following areas:

1. Selection of criminogenic data variables,
2. Methodological designs for data gathering,
3. Statistical manipulation of data,
4. Appropriate display of data findings,
5. Interpretation of data analysis, and
6. Integration of CA into the planning process.

Within the grant application each state was allocated four days of on-site technical assistance provided on an as-needed basis, with each state requesting assistance that fits their particular needs.

Another component of the project involved the development and/or modification of software packages for the SPA's using computers for analysis. This was to be delivered, again, on an as-requested basis. The SPA's were advised of this service, though none of them used it. CCAT did, however, suggest software packages already in existance, and workshop participants were exposed to one of the packages as part of the workshop curricula.

The evaluation provisions of the grant called for a pre-post test on materials presented in the workshop and a comparison of SPA comprehensive plans submitted in 1976 with those submitted in 1977 to determine the extent of change with regard to CA. The final report was to include this evaluation as well as CCAT recommendations regarding future TA implications.

PROVISION OF TECHNICAL ASSISTANCE

The proactive TA was in the form of a three-day Crime Analysis Workshop which was held on the UTC campus on April 28-30, 1976. A detailed report on this component of CCAT's technical assistance was sent to the LEAA Region IV office in July, 1976. A copy of this document is attached as Appendix B of this report. Further elaboration on the workshop will not be undertaken in this report.*

The reactive segments of the TA began after the award of the crime analysis technical assistance grant. The regional office notified each SPA of the award and the services available to them upon request from CCAT. In addition, CCAT contacted each SPA by telephone and advised of the technical assistance visits that were allocated to them. Two separate additional letters were sent to the Region IV SPA directors apprising them of CCAT services, while the Crime Analysis Newsletter published by CCAT also described available services.

The Crime Analysis Newsletter was undertaken by CCAT in October, 1976 to fulfill the grant requirement of "provision of additional CA materials to each state on a quarterly basis." During the grant period,

*In preparing for the workshop, CCAT recognized that the change in the UCR definition of larceny in 1973 was liable to cause serious problems with regard to 1977 crime analyses. The Part I UCR definition changed from defining larceny as an offense involving theft of money or articles above $50 in value to inclusion of all larcenies as Part I offenses. It was recognized that if crime analyses were conducted without adjusting the data to account for the change in definition, then the increases in the rates of larceny would be precipitous and misleading. In an attempt to forestall this problem, CCAT requested that FBI re-run Part I offense data from 1968 through 1974 by state and county utilizing the new larceny definition. They were also requested to provide a report on non-reporting agencies and clearances made by arrest by race and age of offender for Part I and Part II offenses. The ensuing reports were provided and then mailed to each SPA in early May, 1977. In all cases, except Kentucky, the re-runs were for 1968-1973. Kentucky data were available through 1974.

two issues were developed and distributed to an audience of approximately 250 including previous workshop participants, each SPA director in the USA and territories, all LEAA regional offices and other interested parties. Copies of the newsletters published under the grant are included as Appendix C of this document. As can be seen upon examination of the newsletter, its thrust is to provide information on data elements related to crime analysis, commentary on current analyses and techniques, introduction to the various analysts in Region IV, bibliographic references to current literature related to CA, and other pertinent materials. Since its development, CCAT has had numerous individual requests to be put on the newsletter mailing list.

Other CCAT technical assistance activities are summarized in Figure 1. This illustration breaks down technical assistance requests and CCAT activities by state in LEAA Region IV. With two exceptions, all requests for technical assistance were honored. Due to prior TA commitments, CCAT did not have a sufficient amount of time in one instance to be able to deliver timely training on preparing problem statements; and in the other case, the request fell beyond both the resources and the intent of the grant.

FIGURE 1

Summary of Technical Assistance Requests and Related CCAT Activities
in LEAA Region IV - 3/1976 - 6/1977

Requests/Activities	Alabama	Florida	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
On site visit(s)	X			X	X	X		X
Response to telephone queries	X	X	X	X	X	X	X	X
Response to written queries	X	X	X	X	X	X		X
Review of previous analyses to indicate areas needing improvement					X	X		
Design of Crime Analysis					X			
Construction of Data Gathering Instruments					X			
Preparing computer coding for data input					X			
Data Input for computer analysis					X			
Software selection for data analysis	X				X			
Teaching various types of statistical computation				X	X			
Selection of statistical techniques for use in analysis	X			X	X	X		X
Comparison of relative advantages among various statiscical techniques	X			X	X	X		X
Actual statistical manipulation for a segment of the crime analysis					X	X		
Selection of socio-economic-demographic variables for use in analysis	X	X		X	X	X		X
Finding Data sources	X	X		X	X	X		X

Review and comment upon draft crime analysis	X				X	X		X
Integration of CA into remainder of the comprehensive plan(s)	X			X	X	X		X
Selection of Illustrative techniques for crime analysis	X			X	X	X		X
Construction of crime analysis illustration				X	X	X		X
Providing bibliographic citations relation to CA (in addition to newsletter)	X	X	X	X	X	X		X
Providing article reprints relevant to CA	X	X	X	X	X	X		X
Recommendations for crime analysis consultants					X	X		
Interpreting data	X			X	X	X		X
Suggesting approaches to resolve data problems		X		X	X	X		X
Facilitating communications among analysts doing similar analyses	X	X	X	X	X	X		
Request to do entire crime analysis*							X	
Meeting with local planners to discuss crime analysis						X		
Evaluation design suggestions				X	X			X
Critique of evaluation designs					X			
Review of system performance analysis methodologies	X			X	X	X		X
Problem statement development training				**	X			X
Development of measurable objectives					X			
Review of draft segments of comprehensive plan other than CA					X			
CA implications in program development				X	X	X		X

*This request was beyond the purview of the TA grant. CCAT had neither the financial nor the personnel resources to respond.

**This request was not met due to an insufficient amount of lead time to prepare a training segment dealing with development of problem statements.

CONTENT ANALYSIS

The evaluation requirements, as anticipated originally under the project, involved a pre/post-test of the workshop (as mentioned earlier) and a comparative content analysis of the crime analysis portions of the plans. This analysis was to involve using LEAA Regional Office checklists for the years 1976 and 1977. After reviewing the checklists, however, it became evident that the change between the 1976 and 1977 guidelines left an insufficient number of comparable data points between the checklists for analysis.

Therefore, a manifest content analysis* (hereinafter referred to as content analysis or analysis) was conducted. Changes between the 1976 and 1977 crime analysis portions of the plans reported below, however, cannot generally be attributed to CCAT intervention. This was due primarily to the timing of the grant award with regard to the planning cycle. Though the grant was planned to begin in January, 1976, the actual award of the grant to the workshop, approximately six weeks elapsed during which time formal curricula were developed, SPAs notified of CCAT services, and other general grant activities (e.g. hiring of staff, setting up accounts, etc.) were begun.

The initial point of contact with SPA personnel was during the week of April 26, 1977 at the workshop. Since the state's comprehensive plans were scheduled for submission to LEAA at the end of June, 1976, the crime analyses were close to completion in many of the states. In several in-

---

*Manifest content analysis means assessing whether or not certain factors were present in the crime analyses.

stances (e.g. Alabama), the person(s) responsible for the crime analysis remained in their states to complete their work instead of attending the workshop.

Bearing this in mind, it should not be difficult to see that the actual impact of this first component of TA (workshop) was apt to be less than had been anticipated. Followup on-site technical assistance was available immediately following the workshop. In this period, prior to the actual submission of the 1977 plans, only three states (Alabama, North Carolina, and Tennessee) requested and subsequently received on-site assistance.

As a result of the above, the primary focus of the content analysis became the distillation of problem areas. This analysis provided CCAT staff with an over-all indication of the areas of common deficiency that could be most appropriately addressed in future workshops and in the delivery of subsequent on-site technical assistance.

An observation made during the analysis process was that the comprehensive plans, beyond some ill-defined optimum, were generally too long. As can be seen in Figure 2, between the years 1976 and 1977, plans ranged in length from 565 to 3265 pages. The plans averaged 1627 pages in length in 1976 and 1894 pages in 1977.

FIGURE 2

Length of State Comprehensive Plans in LEAA Region IV for the years 1976 and 1977*

State	1976	1977
Alabama	2095	1850
Florida	3133	2530
Georgia	1158	1573
Kentucky	2378	3265
Mississippi	1104	601
North Carolina	1827	2811
South Carolina	758	1189
Tennessee	565	1335
AVERAGE LENGTH	1627	1894

*SOURCE: State Comprehensive Plans for 1976 and 1977.

The particular focus of CCAT staff's content analysis was the crime analysis portions of the plans. The M4100.1E guidelines specified certain data elements (see Appendix A) that were required and suggested to be included in the analyses. Correlational analysis was suggested, while "straight line projection" for crime rate per population for "total crimes, person crimes and property crimes" (M4100.1E, par. 52 c (a), p. 68) was required. Since the states were charged with conducting specific types of statistical analysis, one of the areas that received close attention in the content analysis was that of the types of statistical manipulations conducted.

Figure 3 contains a comparative summary of the types of statistical analysis utilized in by Region IV states. One of the categories for summarization included whether or not the states used a computer for analysis. The CCAT staff are of the opinion that the magnitude of the analyses required and suggested under the guidelines necessitated the use of a computer.

The second area of comparison in Figure 3 regards the statistical techniques used for analysis. Regression, time series and projection (and varieties of these) are all means by which states might arrive at compliance with Chapter 3, par. 52 c (a) of the guidelines while providing potentially valuable information for planning purposes. Correlations are suggested by the guidelines in paragraph 52 d. The charge that "plans must contain a thorough, complete, total and <u>integrated</u> [CCAT emphasis] analysis of crime and the problems it causes the public and governmental agencies throughout the State" (M4100.1E, par. 52, p. 66), however, requires correlational analysis. The types of correlation useful are numerous, though the most frequently used were Spearman's Rho and Pearson's Product Moment correlations.

The mean and median, noted in Figure 3, are techniques for summarizing an array of data by providing an average. In the majority of instances, when summarizing crime data, the median average conveys a more realistic picture of the actual situation, as it is not as influenced by extremes.

Finally, Figure 3 illustrates whether or not the various states noted the limitations of statistical inference based on the data and/or statistics they used. It is accepted research practice to discuss data and analysis limitations in order to call attention to the fact that the analysis may or may not provide sufficient evidence (statistically speaking) to reach sound conclusions. In dealing with a phenomenon as inticate and multi-faceted as crime, CCAT staff felt that such a discussion was essential.

Illustrated in Figure 3, the state having the greatest degree of change between 1976 and 1977 in the number of statistical analysis factors used was Alabama. Georgia shows the second greatest degree of change while Mississippi and Tennessee changed the least. In the 1977 CA, six of the eight states in the region discussed (in some fashion) the limitations of statistical inference based on the data and/or statistics used. In view of noted weaknesses of existant crime data*, it is important that the states called it to the attention of their plan reading audiences. Five of the eight states engaged in correlational analysis. All states used percentages and rates while only half the states used median averages in 1977.

In 1976, only Tennessee used any "Other Statistical Tests," while in 1977 four more states added to their statistical repertoire (e.g. factor analyses, canonical variable analysis).

Comparative tools used for statistical analysis illustrated in Figure 3, while the tools for displaying the analysis are examined in Figure 4. A mark in the category "Tables and charts appropriately

---

*Thorough discussions of the limitations of U.C.R. data are found in the following works: Bell, Daniel. The Coming of Post Industrial Society: A Venture in Social Forecasting. N.Y.: Basic Books, 1973, p. 7. Hindilang, M. J. "The Uniform Crime Reports Revisited" Journal of Criminal Justice, Spring, 1974, pp. 1-17. Maltz, Donald D. "Crime Statistics: A Historical Perspective", Crime and Delinquency, Jan., 1977, Vol. 23, No. 1, pp. 32-40. Nettler, Gwynn. Explaining Crime. N.Y.: McGraw Hill, 1974. Price, S.E. "A Test of the Accuracy of Crime Statistics," Social Problems, Fall, 1966, pp. 214-221.

FIGURE 3

A Summary of the Crime Analysis Statistical Factors in the Comprehensive Plans of the States in LEAA Region IV - 1976/1977

Statistical Factors*	Alabama	Florida	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Computer Used for Statistical Computations	/ x	x /	/ x	/ x			/ x	x /
Statistical Techniques Used: 1. Regression	/ x	/ x	/ x					x /
2. Time Series	/ x	x /		x / x		/ x	/ x	
3. Projection	/ x	/ x	/ x	/ x		/ x		x / x
4. Correlation	/ x	/ x	/ x	/ x			/ x	
5. Percentages and Rates	x / x	x / x	/ x	x / x	x / x	x / x	x / x	x / x
6. Mean	/ x	x /	/ x	/ x	/ x		x / x	x /
7. Median	/ x		/ x				/ x	/ x
8. Other statistical tests	/ x	/ x	/ x				/ x	x / x
Contains a discussion of the limitations of statistical inference based on the data and/or statistics used	/ x	/ x		x / x	x / x		/ x	/ x

*An x indicates that the factor on the left was present in the 1976 and/or 1977 crime analysis. A blank indicates that the factor was not present

1976 → [diagonally divided box] ← 1977

constructed" means that 75% or more of the illustrations used by the state in either 1976 or 1977 were correctly prepared. This was determined by analyzing each illustration's form and tallying them as correct or incorrect.

CCAT staff are of the opinion that appropriate illustration of data presented conveys more impact than mere narration. That is, all the elements necessary for understanding the illustration must be present. Employing a variety of correctly constructed illustration techniques, then, holds the reader's attention longer.

The remainder of the categories in Figure 4 catalog the types of illustrative techniques used by the states. Frequency polygons are line graphs; bar histograms, bar graphs; pictographs, pictures representing works or ideas on a single axis; and circle graphs, are more commonly referred to as pie charts. The category "Other forms of illustration" include triangle graphs, maps, picture illustrations, etc.

Summarizing Figure 4, three of the eight states appropriately constructed their figures and tables in their 1977 crime analyses. The most frequent errors encountered in the construction of illustrations were: unlabeled axes (vertical and horizontal bounds); units on axes not identified; insufficient or no titles on illustrations; and data sources not identified. Since illustrations are often lifted from the context of the document they are contained in and used for other media purposes, each illustration should contain a sufficient amount of information to enable the viewer to understand it without the textual narration.

FIGURE 4

A Summary of Crime Analysis Illustrations in the Comprehensive Plans of the States in LEAA Region IV - 1976/1977

Crime Analysis Illustrations*	Alabama	Florida	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Tables and Charts/Figures Free from Computational Error	/ x	/ x	x / x	/	/	/	x /	x / x
Tables and Figures Appropriately Constructed	/ x	/	/	/	x / x	/	/	x / x
Illustration Techniques Used: 1. Tables	x / x	x / x	x / x	x / x	x / x	x / x	x / x	x / x
2. Frequency Polygons	x / x	x / x	/ x	/	x / x	/ x	x / x	x / x
3. Bar Histograms	x / x	x / x	/ x	/ x	x / x	x /	x / x	x / x
4. Pictographs	x /	/ x	/	/	/	/	x /	/
5. Circle graphs	x /	/	/	/ x	x /	x / x	/	x / x
6. Other Forms of Illustration	x / x	x / x	/ x	x / x	x /	x /	x / x	/

*An x indicates that the Illustration consideration on the left was present in the 1976/1977 crime analysis. A blank indicates that the point under consideration was not present.

1976 → ⧅ ← 1977

In 1977, Florida used the largest variety of CA illustrations (at least 5 different types), while Mississippi and North Carolina used the least. In most cases the 1976 and 1977 crime analyses were replete with tables, while fewer other types of illustration were used. One of the ways the crime analyses could have been shortened would have been more reliance upon illustration as a means of summary.

Figure 5 comparatively illustrates five categories that the CCAT staff considered important in crime analysis development. Considering the plethora of socio-economic-demographic data available to the states, selection of variables should relate to some particular rationale. As Alabama stated in their 1977 analysis,

> "....a decision as to what demographic variables to use had to be made. The criminal justice area, particularly where crime is concerned, has potential of working a conscientious statistician to exhaustion. For every piece of crime data available, there are virtually an infinite number of demographic variables...."

Selection of variables, relates to a number of factors: how old the data is, the potential of the data to explain part of the crime phenomena under consideration, etc. A discussion of the rationale for choice of variables, is the first category on Figure 5, was considered important.

Another important crime analysis factor found in Figure 5 is the written explanation methodology employed to analyze crime. Each state conducts its own analysis in particular ways. Being able to answer any potential questions as to how the analyses were conducted, the procedures used, the significance levels sought, etc. is accepted practice in research.

Explanations contained within the documents forestall misperceptions as to how the final products were derived.

In the area of crime analysis, it is particularly useful to cite relevant literature. Citations may be used to: point out concurrence of other studies with the analysis findings; fill in data gaps; point out factors considered important but where available data were not available to test; and to indicate that the analysts have appropriately prepared themselves by being familiar with literature in the field. Thus, the third category on Figure 5 indicates whether or not the various analyses cited relevant literature.

Provision of a narrative relevant to the formal analysis is necessary to communicate the important findings of the analysis. Insofar as narrative is not provided or it is not relevant to the analysis, the reader of the analysis will be left to ferret out significant points and interpret them for himself. Often, the reader will not go to this trouble, hence, the findings of the analysis are not communicated.

The final category on Figure 5 relates to an explanation of the limitations of the analysis. This is a necessary ingredient to assure that the persons reading the analysis will not infer what is not intended to be inferred. For instance, correlational analysis conducted between crime rate and number of families receiving Aid to Dependent Children might infer causality. Failure to explain the limitations of an analysis, particularly in the area of inference of causality, may lead to misunderstandings that are easily circumvented.

Summarizing Figure 5, we find that only Alabama and Tennessee included a rationale for their choice of socio-demographic-economic variables in 1977. In 1976, no analysis contained such a rationale.

While only 3 states explained their CA methodologies in 1976, 7 analyses (or supporting documentation thereof) contained methodological explanations in 1977.

North Carolina and Tennessee cited related literature in 1976, while only Tennessee cited related literature in 1977. Narrative related to the crime analysis was contained in 6 analyses in 1977. The limitations of analysis were clearly explained in 2 states in 1976, while in 1977, 6 states explained limitations.

If one goes on the assumption that addition of crime analysis factors between 1976 and 1977, as listed on Figure 5, indicates an improvement in the crime analysis, then Alabama exhibited the greatest degree of improvement (from 0 factors to 4 factors) in 1977. Tennessee closely follows with the addition of 3 new factors in 1977.

Figure 5

A Summary of Crime Analysis Factors in the Comprehensive Plans of the States in LEAA Region IV - 1976/1977

CRIME ANALYSIS FACTORS*	Alabama	Florida	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Rationale for choice of socio-demographic - Economic variables used	/ x							/ x
Written Explanation of methodology used	/ x	x / x	/ x	x / x	x / x		/ x	/ x
Literature Relevant to Crime Analysis Cited						x /		x / x
Narrative Relevant to Crime/Analysis	/ x	x / x	/ x	x / x	x /		x / x	x / x
Limitations of Analysis/Clearly Explained	/ x	/ x		x / x	x / x		/ x	/ x

*An x indicates that the factor on the left was present in the 1976 and/or 1977 crime analysis. A blank indicates that the factor was not present.

1976 → ⟋ ← 1977

Observations and Suggestions

The Center for Criminological Analysis and Training (CCAT) has now been operational for sixteen months. During this period, we have confirmed the opinion that technical assistance is a necessity for persons in operational agencies. If the on-site technical assistance role were extended to encompass other areas of planning and analysis, then it would meet a need, often expressed by operating agency personnel. For instance, CCAT has been called several times for on-site assistance with state and local planning and analysis problems (e.g. Memphis, Southwest Tennessee Development District, SPAs outside LEAA Region IV, and several local agencies in Florida). These requests have been beyond the financial and personnel resources of CCAT, and thus were declined.

On-site technical assistance, as mentioned earlier in the report, is designed to be reactive to various crime analysis areas of concern. What we found during the on-site crime analysis TA visits was a great need for assistance in other planning and evaluation areas. Since we were readily available, considerable use was made of our services. We cannot stress enough the importance of a responsive technical assistance component to enhance the analysis and planning effort. With the exception of a few instances, CCAT on-site visits could be scheduled at the convenience of the SPAs. Thus, when a difficult problem arose, it could be dealt with in a timely fashion. For whatever reason(s), SPAs have been reluctant to call for assistance in developmental stages of analysis and have called often only during a crisis. By being able to immediately respond to problems, we were also in a position to suggest ways of circumventing future problems. Often following a crisis oriented on-site visit, we have been called upon again to deal with other developing concerns. These may be dealt with, in part, through training. However, one of the major benefits of being on-site is that one deals with a particular area that may be peculiar to a given state, geographic area, or political situation. More time is available to deal with specifics rather than generalities.

A unique aspect of the CCAT project was that it offered crime analysis technical assistance, for the first time that was not attached to the regional office. Therefore, CCAT staff was placed in a neutral or non-threatening position vis-a-vis the SPA's. It appeared to us that, in some cases, the SPAs were less reluctant to admit their need for assistance to an independent observer than to one (e.g. an LEAA technical specialist) who was in a position to place sanctions on their plan. Thus, CCAT was available to offer only suggestions, not sanctions. CCAT staff could offer interpretation of guidelines and means by which compliance might be achieved. Beyond that, we were in a position to advocate directions in which a crime analysis might logically be expanded, but without appearing as though we were adding another requirement (as do addition's to the guidelines).

We have also found that technical assistance is particularly useful to the new analyst/planner. Often analysis/planning positions are vacated before a replacement is hired. Thus there is a need for an ongoing training and technical assistance mechanism for new planners and analysts. This was demonstrated to us by the fact that between 35-60% of the people we worked with in the SPAs during the first year of the grant are no longer employed by the same agencies. Agency turnover has left little room for continuity and building on established bases.

There also appears to be a need for an orientation to research for non-analysts. We have observed that there seems to be little understanding among non-analysts of the problems related to sound analysis and research and the amount of time and money it demands. The non-analyst's view seems to be that once data is gathered, all that has to be done is to get it into a computer and press a button. The analysis, then, is perceived to be done by the machine; and the results yield valid and reliable information. Interpretation, then, is a function of being able to read a printout. What this lack of understanding has led to is CCAT being called in on four separate technical assistance visits only to find that there were major communication problems between the analysts and planners that had to be ameliorated before any actual crime analysis TA could be delivered.

Another area that needs to be dealt with is through training and on-site assistance expanding the crime analyses so that emerging areas of concern in the criminal justice field are taken into consideration. One of the areas of concern that seems to be generally lacking is that of victimization, particularly with regard to the elderly. This concern, long discussed in the criminal justice field, emerged in the M4100.1F planning guidelines. Among the 1977 analyses, little mention was made of this (except Florida where extensive information was available via a component of a national survey done for the city of Miami). A second area that seems to be emerging in the criminal justice field is that of female criminality. Recent studies* have shown the dramatic increase of female criminality, yet only one analysis (again Florida) contained any mention of this phenomena.

If as the guidelines direct, crime analyses must form the backbone of the planning effort, then analyses lacking consideration of significant areas do not provide a sufficient amount of background for the development of necessary ameliorative or control programs. CCAT has dealt with victimization and female criminality as topics, through TA; however, the analyses have yet to systematically deal with them.

One area that could be greatly expanded in the SPA and other analysis efforts is that of sample surveying. There seems to be an extremely heavy reliance upon "whole universe" or total population data. Total population data is expensive and time consuming to collect, not to mention the expense involved in processing and analysis. Correctly designed sampling studies yeild information as valid as whole universe studies and are considerably less expensive and time consuming.

In terms of generating analyses useful for planning purposes, planners need to be able to identify particular areas where data is needed early in the planning cycle. By this means, analysts will be able to develop the data in a timely fashion. Often what we have observed is that by the time an analysis is completed, planners have seen the necessity

---

*For example, see Simon, Rita James, The Contemporary Woman and Crime. Rockville, MD: The National Institute of Mental Health, 1975. This document contains a thorough discussion of many current studies regarding female criminality.

of additional data, but there has been insufficient time for its development. Hence, planners have often been frustrated by the lack of necessary data for program development. CCAT staff have often dealt with this problem during the process of delivering on-site assistance.

At the same time, analysts need to begin the analysis process much earlier in the planning cycle. This has been observed by the LEAA Regional Office; hence, the provision for a phased submission of comprehensive plans with the crime analysis being among the components scheduled for early submission.

Within the realm of comparative analysis among "similar" states, cooperation among the states to avoid unnecessary duplication is called for. By bringing Region IV analysts together on a periodic basis to discuss their analyses and research, many useful cooperative endeavors could be suggested, encouraged, and facilitated. The CCAT staff are of the opinion that much more analysis could be conducted at no increase in cost if the Region IV states were to pool their resources; particularly in the areas of programming, questionnaire development, survey procedure development, statistical manipulations, etc.

A final area of observation regards the support mechanism for the delivery of analysts and planners to the field of criminal justice. To date, few institutions of higher learning have afforded educational opportunities for the preparation of criminal justice analysts and planners. Hence, agencies have generally resorted to hiring personnel with variety of educational backgrounds. This necessitates a great deal of "learning on the job" subsequent to employment. It also means that training courses designed for criminal justice analysts and planners, of necessity, must often deal with materials more appropriate for a college or university course (e.g. theory, statistical manipulation, research methodology, etc.) Many of these topics require a substantial investment of time to produce a knowledgable student. When these materials are dealt with in the training format, the treatment is generally scant due to the constraints of having to cover the other training material as well.

Bearing in mind the observations made previously, CCAT offers the following suggestions (not necessarily in any preferential order):

1. All crime analyses might be contracted out to a single, regional analysis center. The savings from this approach would be substantial in that fewer personnel would need to be involved in crime analysis, one (or a few) set(s) of software for CA would be necessary, there would be uniformity and continuity in the crime analyses, as well as possibilities for cross-regional comparison's and analyses.
2. SPAs might be encouraged to form a formal crime analysis consortium to pool personnel and/or financial resources. This might be coordinated through a regional university such as the University of Tennessee at Chattanooga. In this manner, new analysts entering the system would have an opportunity to work with more experienced analysts. Computer programs, methodologies, survey instruments, statistical formats, etc. could be shared. Common socio-demographic variables could be selected and analyzed in conjunction with crime, etc. An interesting fringe

benefit of this approach might well be the identification of interstate crime problems that could be addressed by two or more states in concert. This might also lead to viable suggestions with regard to guideline changes. In the long run, a considerable amount of money might be saved by this type of unified effort, not to mention the improved calibre of analysis that could be expected.

3. Technical assistance might be expanded to include state, local, and regional analysts and planners. Since the need has been expressed, it appears only logical that this extension should occur. Also, the realm of technical assistance of a project such as CCAT operates could be expanded. Often what is initially considered a crime analysis problem by the SPAs, is more specifically a planning and/or evaluation problem. Since analysis, planning and evaluation are part of a continuum, it is difficult to relate technical assistance to a portion of the problem while leaving major areas unaddressed. Expansion of technical assistance would involve some expense in terms of grant funds, but it is felt that such an expansion would be cost-beneficial in terms of the planning product.

4. A regional personnel exchange might provide experienced analysts/planners and opportunity to work with new personnel (as hired) in the various states.

5. The basic course that appear to be needed by many planners/ analysts could be developed in a multi-media presentation and sent out to SPAs as new personnel were added. This

could easily be connected to a college or university. An instructor could be supplied at intervals to assist with any problems that may have arisen in conjunction with the multi-media package, to do testing, etc.

6. Colleges and universities, particularly in captials or major urban centers, might be encouraged to add basic courses to their curricula. The curricula might be developed by one university and supplied to any school upon request. This would serve the needs of planners and analysts, and those aspiring to the positions, of basic educational tools regarding criminal justice (e.g. theory, statistics, methodology, etc.).
7. Training workshops could be provided more frequently on a larger variety of topics. Some consideration should be given to the development of training curricula on narrow topics related to crime analysis that are designed to be delivered on-site.
8. A regional and/or national summer institute for planners and analysts might be provided to allow acquisition of basic skills. This might involve intensive, on-campus sessions of about six weeks offering graduate credit.

CCAT has generally been working toward some of the suggestions contained above. The Criminal Justice program has added new curricula in the area of analysis and planning. The workshop held by CCAT allowed crime analysts perhaps their first opportunity to meet and share experiences with one another. Technical assistance has been delivered as requested to meet, often times, pressing analysis and planning problems.

Overall, CCAT feels the services they have provided over the last year have been pertinent, useful and generally well received (if calls for additional service are any indication). We feel the need for crime analysis technical assistance (as well as planning and evaluation TA) is growing and will continue to grow. Though not necessarily attributable to CCAT, we feel that the crime analyses produced in 1976 were generally better than those produced in 1977. Should this trend continue, Region IV could well lead the nation in producing quality crime analytical documents in 1978.

APPENDIX A

M4100.1E Data Elements and Analysis Required*

1. State as a whole, each substate region and urban cities and counties more than 2500,000 population

2. Total # reported offenses

3. Total # reported arrests

4. Total # reported offenses cleared by arrests

5. Number of non-UCR reporting jurisdictions

6. Summary of special surveys of crime or research completed in state or its subdivisions

7. Yearly crime rate per population for total crimes, property crimes and person crimes

8. Average proportion of increase or decrease per year and straight line projection fro subsequest year

9. Population density and crime pattern for state, region and high crime areas, crime pattern per square mile, percentage of population living in incorporated municipalities (excluding villages) and analysis of population density and crime patterns

10. Specific analyses

    A. Homicide, aggravated assault and street robberies by # of residents
    B. Rape of female residents over 12 years of age
    C. Commercial robberies and burglaries by # commercial businesses
    D. Residential burglary by # of residential units
    E. Auto thefts by # registered autos
    F. Larcenies by either amount of loss or total # of larcenies

11. Assessment of ability to perform required crime analysis with an assessment of needs in this area

---

*Analysis to be conducted on data for the most current year available.

M4100.1E DATA ELEMENTS SUGGESTED BUT NOT REQUIRED

1. Victimization survey offenses reported

2. Total person crimes by:

   A. Victim characteristics
   B. Weapon used
   C. Injury/non-injury
   D. Monetary loss
   E. Suspect characteristics
   F. Location
   G. Time

3. Property crimes by:

   A. Amount of loss
   B. Type of property stolen
   C. Recovery rate (proportion of loss recovered)
   D. Location
   E. Time

4. Homicide and rape each by:

   A. Victim characteristics
   B. Weapon used
   C. Injury
   D. Monetary Loss
   E. Suspect characteristics
   F. Location
   G. Time

5. Commercial robberies and burglaries, residential burglary, and auto theft each by:

   A. Amount of loss
   B. Type of property stolen
   C. Recovery rate
   D. Location
   E. Time

6. Socio-demographic data with a clear relationship to the crime analysis data:

   A. Census
   B. School truancy
   C. Mental health
   D. Alcoholism and Drug Abuse
   E. Educational achievement
   F. Economic trends (including unemployment)
   G. Identify which data is utilized, how it is uded, and principal sources of data

APPENDIX B

PROGRESS REPORT ON GRANT #76-TA-C4-0002

Crime Analysis Technical Assistance

CENTER FOR CRIMINOLOGICAL ANALYSIS AND TRAINING

June 30, 1976

The Center for Criminological Analysis and Training, located within the Criminal Justice Program at the University of Tennessee at Chattanooga, was awarded a grant by the Law Enforcement Assistance Administration (LEAA) Regional Office in Atlanta for the purpose of increasing the capacity of State Planning Agencies (SPA's) in Region IV to undertake crime analysis as called for under the new LEAA guidelines M4100.1E. This objective was to be realized so as to effect the preparation and integration of crime analysis into the 1977 Comprehensive Plan for criminal justice agencies from each state.

This report covers the workshop held in conjunction with the provision of technical assistance and contains observations and comments for consideration in the event of future efforts in other LEAA regions and/or with local planners.

States within LEAA Region IV include Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama and Mississippi.

TABLE OF CONTENTS

List of Tables

Introduction

The Crime Analysis Technical Assistance grant was put into operation March 4, 1976. Initially, the focal point of grant activities was toward the presentation of a crime analysis workshop so as to provide: 1) an understanding of the LEAA guidelines as they relate to crime analysis; 2) a review of the statistical techniques used in preparing a crime analysis; 3) an introduction to the computer and its statistical capabilities and 4) the integration of crime analysis into the general planning process.

The eight State Planning Agencies were notified approximately one month in advance of the workshop both by telephone and by a follow-up letter. They were informed of the sufficiency of funds to cover all transportation and housing expenses on a double occupancy basis plus limited meal allowances for up to a maximum of four representatives per SPA. Registration forms were sent out with a request that they be returned as soon as possible so as to be received before the workshop.

The Crime Analysis Workshop has been conducted, leaving on-site technical assistance as the focal point for the remainder of the grant. An evaluation of the workshop is now in order to determine its effectiveness in presenting the new guidelines, the statistical techniques used in crime analysis, and the "hands-on" approach to the computer.

One point needs to be made clear at the outset of this report: the "target" population of this grant is SPA personnel who by choice or assignment become involved in crime analysis. For this reason, only SPA reponses will be tabulated and discussed herein, with the exception of the final evaluations of the workshop of which all will be included due to the anonyminity of the responses.

This report is divided into four sections. The sections are:

1) Registration Information; 2) Pre and Post Survey Results; 3) An Evaluation of the Workshop; and 4) Comments and Recommendations.

REGISTRATION INFORMATION

A three-day crime analysis workshop was held on the campus of the University of Tennessee at Chattanooga, April 28-30, 1976. A copy of the program can be found in Appendix A. The number of official participants totaled forty-one. Table 1 breaks down this figure by jurisdiction and agency.

TABLE 1

PARTICIPATION BY JURISDICTION AND AGENCY ATTENDING CRIME ANALYSIS WORKSHOP

JURISDICTION	SPA	LEAA
Tennessee	6	
Kentucky	1	
North Carolina	3	
South Carolina	4	
Georgia	2	
Florida	3	
Alabama	2	
LEAA Region		14
LEAA (D.C.)		2
Total	25	16

Each state in the region was represented at the workshop, some moreso than others. LEAA was represented at both the regional and national level.

In that the selection of the person(s) attending the workshop was left entirely to the discretion of each SPA, pertinent information was requested from each representative at the time of registration so as to obtain an idea of the educational background and the experience of persons working in crime analysis. A questionnaire was constructed and made available to all participants for this purpose.

Roughly half of the total number of participants responded to the questionnaire. Table 2 illustrates the responses by jurisdiction and by agency. The letter "N" will be used hereafter to refer to the total number in delegation per jurisdiction. The symbol "N/R" will be used to indicate No Response from jurisdiction.

TABLE 2

RESPONDENTS TO REGISTRATION INQUIRY
BY JURISDICTION AND AGENCY

Jurisdiction	SPA	LEAA	N	% Total
Tennessee	2		6	33
Kentucky	1		1	100
North Carolina	3		3	100
South Carolina	N/R		4	0
Georgia	2		2	100
Florida	2		3	66
Alabama	1		2	50
Mississippi	1		4	25
LEAA Region		8	14	57
LEAA (D.C.)		1	2	50
Total	12	9	41	58.1

Of these responses, only one indicated prior attendance at a crime analysis workshop. Cited was a victimization conference held this past February in Atlanta, Georgia. Table 3 presents the responses per SPA as to whether they had ever previously attended a crime analysis workshop.

TABLE 3

RESPONDENTS FROM STATE PLANNING AGENCIES WHO HAVE ATTENDED A CRIME ANALYSIS WORKSHOP BEFORE

State	Yes	No	Total Response
Tennessee		2	2
Kentucky		1	1
North Carolina		3	3
South Carolina			N/R
Georgia	1	1	2
Florida		2	2
Alabama		1	1
Mississippi		1	1
Total	1	11	12

It was somewhat expected that attendance at crime analysis workshops would be minimal due to the fairly recent emergence of crime analysis as an integral component in the planning process.

The data displayed in Table 4 indicates that most respondents hold a masters degree.

TABLE 4

HIGHEST EDUCATIONAL LEVEL ATTAINED
BY RESPONDENTS FROM STATE PLANNING AGENCIES

State	BA/BS	MA/MS	Ph.D.	Other	N	Total Response
Tennessee		2			6	2
Kentucky		1			1	1
North Carolina	1	1		1(J.D.)	3	3
South Carolina					2	2
Georgia	1	1			2	2
Florida		1	1		3	2
Alabama		1			2	1
Mississippi		1			4	1
Total	2	8	1	1	25	12

The major fields of study represented were English, Biology, Mathematics, Sociology, Public Administration and Urban Planning. A degree, however, should be viewed with considerable reservation due to its failure to identify areas of academic strength and, likewise, areas of academic weakness. At best, a degree in and of itself serves as a poor indicator of individual capability to perform the job in question: i.e., crime analysis.

In recognition of this ambiguous nature of a degree, additional information was sought in specific areas of study, e.g., statistics, research methodology, and computer programming so as to gain a better understanding of the person's academic preparedness in assuming the task of crime analysis.

Statistics play a vital role in the collection and description of crime data. It is possible, for example, through the use of various levels of statistical measurements to determine the impact of crime on society, the effectiveness of the criminal justice system in dealing with crime problems, and the efficiency of the system in operation. Few would deny the importance of being in possession of this information. Toward this end, the guidelines require the use of summary statistics in the compilation of the data, i.e. averages, percentiles, and ratios. In a few instances, more sophisticated statistical techniques are recommended, i.e. correlations and regressions. Therefore, it is reasonable to expect the person charged with this responsibility to be knowledgeable in the proper use of statistics. The degree of exposure of SPA respondents to statistics is presented in Table 5.

TABLE 5

EXPOSURE OF RESPONDENTS FROM STATE PLANNING AGENCIES TO STATISTICS BY NUMBER OF FORMALIZED COURSES TAKEN

State	None	1	2	3+	N	Total Response
Tennessee	1			1	6	2
Kentucky				1	1	1
North Carolina	3				3	3
South Carolina					4	N/R
Georgia	1			1	2	2
Florida				2	3	2
Alabama		1			2	1
Mississippi			1		4	1
Total	5	1	1	5	25	12

The responses were graded by number of formalized statistics courses taken in the belief that only through formalized instruction would one be exposed to most, if not all, of the basic statistical methods used in the literature of today, and, more importantly, be exposed to both the strengths and weaknesses of the various techniques. On the job training does not offer this assurance.

Statistics can be and oftentimes are misleading because of serious methodological errors. Common oversights include the disregard of sampling techniques, the failure to validate and test questionnaires for reliability <u>before</u> administering them, and the casual assumption of relationships when the research design is without proper controls. It would seem imperative then in order to avoid these pitfalls that person(s) collecting and/or analyzing this data be aware of the necessary conditions that must be satisfied in each case. Table 6 illustrates the exposure of SPA respondents to research methodology.

TABLE 6

EXPOSURE OF RESPONDENTS FROM STATE PLANNING AGENCIES TO RESEARCH METHODOLOGY BY NUMBER OF FORMALIZED COURSES TAKEN

State	None	1	2	3+	N	Total Response
Tennessee		1		1	6	2
Kentucky			1		1	1
North Carolina	3				3	3
South Carolina					4	N/R
Georgia		1	1		2	2
Florida			1	1	3	2
Alabama		1			2	1
Mississippi		1			4	1
Total	3	4	3	2	25	12

The third area of academic inquiry was in the field of computer science. The criminal justice system in keeping up with the technology of our times is in varying stages of converting from a manual filing and records system to an automated storage and retrieval system. In doing so, the information becomes accessible only to those people able to communicate in the language of the computer. This does not mean to say that every person in the act of collecting data should be a computer programmer, but it would seem essential that they have a basic understanding of the mechanics and capabilities of computers in general so as to be able to communicate with specialists in this field. The degree of exposure of SPA respondents to computer programming is displayed in Table 7.

TABLE 7

EXPOSURE OF RESPONDENTS FROM STATE PLANNING AGENCIES TO COMPUTER PROGRAMMING BY NUMBER OF FORMALIZED COURSES TAKEN

State	None	1	2	3+	N	Total Response
Tennessee	1			1	6	2
Kentucky				1	1	1
North Carolina	3				3	3
South Carolina					4	N/R
Georgia		1	1		2	2
Florida		1	1		3	2
Alabama			1		2	1
Mississippi	1				4	1
Total	5	2	3	2	25	12

From the foregoing results one might question the capability of some to undertake crime analysis as called for under the LEAA Guidelines.

The possibility of in-service training meeting the needs of the staff cannot be overlooked. A question was therefore included in the inquiry asking for the total number of job-related in-service training hours received in the past year. Table 8 displays the responses to this question.

A great disparity in in-service training received can be seen between the states in Table 8. The fact that the training is job-related does not necessarily mean that, in turn, the training is related to crime analysis since many SPA personnel carry dual responsi-

TABLE 8

TOTAL NUMBER OF JOB-RELATED IN-SERVICE TRAINING HOURS RECEIVED BY RESPONDENTS FROM STATE PLANNING AGENCIES IN PAST YEAR

State	Hours	N	Total Response
Tennessee	0	6	2
Kentucky	80	1	1
North Carolina	0 40 20	3	3
South Carolina		4	N/R
Georgia		2	N/R
Florida	0	3	2
Alabama	12	2	1
Mississippi	10	4	1
Total	162	25	10

TABLE 9

SIZE OF SPA STAFF PREPARING CRIME ANALYSIS
BY SPA RESPONDENTS

State	# of Staff	N	Total Response
Tennessee	7	6	2
Kentucky	2	1	1
North Carolina	4	3	3
South Carolina		4	N/R
Georgia	3	2	2
Florida	2	3	2
Alabama		2	N/R
Mississippi	3	4	1
Total	21	25	11

bilities. No significance can be attached to the quantity of hours of in-service training or lack thereof until the nature of the training and the capacities developed becomes known. This information was not solicited in the questionnaire.

Another matter of interest was the size of the staff charged with the preparation of the crime analysis in each state. Table 9 presents the size of staff per SPA preparing the crime analysis.

No comparison can be made at this time between quantity and quality.

Probably the most revealing of all questions in the Registration Inquiry asked for a self-perception of needs in relation to crime analysis. This information would better enable us to meet the individual needs during and following the workshop. Table 10 displays the SPA responses to perceived needs in crime analysis.

It is noteworthy that all representatives from two states failed to respond to this question. No response in this case may indicate a lack of recognition of their needs.

TABLE 10

CRIME ANALYSIS NEEDS AS PERCEIVED BY STATE PLANNING AGENCY RESPONDENTS

State	Need
Tennessee	-- Cooperation with state and local agencies in obtaining data -- Need victimization and demographic data
Kentucky	-- To be in compliance with new LEAA Guidelines 4100.1E
North Carolina	-- No Response
South Carolina	-- No Response
Georgia	-- Determine the focus and objectives of crime analysis and special statistical techniques (e.g. ratio, multicollinearity)
Florida	-- Data sources and methods of clearly displaying data
Alabama	-- Interpretation of crime analysis
Mississippi	-- Data base construction

The area of final inquiry was experience in the field of planning and evaluation. Table 11 presents the experience of SPA respondents in planning and/or evaluation by number of years. This table clearly shows the majority of respondents to be in their beginning years of experience in planning and evaluation.

TABLE 11

EXPERIENCE OF RESPONDENTS FROM
STATE PLANNING AGENCIES IN PLANNING AND/OR
EVALUATION BY NUMBER OF YEARS

State	Less than 1	1	2	3+	N	Total Response
Tennessee	1	1			6	2
Kentucky				1	1	1
North Carolina	1			2	3	3
South Carolina					4	N/R
Georgia	1	1			2	2
Florida	1			1	3	2
Alabama	1				2	1
Mississippi		1			4	1
Total	5	3		4	25	12

On the whole, the Registration Inquiry was successful in generating background information on those in attendance at the workshop. Certain structural and procedural changes are necessary, however, for its use in the future. A few questions need to be re-worded so as to elicit specifics instead of general replies, and the information is needed far in advance of any workshop so that programs can be geared as much as possible to their particular needs.

Crime Analysis Workshop Survey

As stated earlier, the purpose of the Crime Analysis Workshop was to provide a general understanding and knowledge of the basic components of crime analysis. The extent to which this objective was met during the workshop will become evident by the nature of the calls for follow-up technical assistance and by a comparison of crime analyses as found in the 1976 and 1977 Comprehensive Plans of each state. What can however be determined at this time is the immediate impact of the workshop upon SPA personnel in attendance.

An examination of the SPA responses to a survey questionnaire administered before and after the workshop should give some indication of this effect. A pre/post evaluation design was decided upon so as to control the variables of education and experience in measuring the level of knowledge and understanding of SPA personnel in matters relating to crime analysis. Whatever the differences in a person's response to the same question before and after the workshop could then be attributed to a learning process having taken place during the workshop, be it a positive or negative learning process.

This type of research design usually matches participant to response thereby offering a more accurate account of individual change. Matching necessarily requires some form of personal identification, i.e., a symbol, a code number, or a personalized signature. It also requires participation in both surveys. These requirements could not be satisfied in this particular survey due to the feeling that any personal identification other than the state he/she represents could discourage full cooperation with the survey. Also, the voluntary nature of the survey would have to be changed to an involuntary one should matching be desired.

The survey questionnaire was constructed by the staff with the assistance of the workshop consultants. The content consisted mainly of basic definitions, concepts, and terminology that would be commonly used in crime analysis. A copy of the questionnaire can be found in Appendix C. Specific instructions were included in the beginning of the survey requesting that the person identify himself/herself by state only.

A time limit of thirty minutes was set for the completion of the questionnaire. During this period of time all staff members departed leaving each person to their honor in supplying the information requested. Again it must be stated that the survey was strictly voluntary. No attempt was made to collect the questionnaire from those not wishing to return it.

Both the pre and post returns were reviewed and graded by one staff member to ensure consistency throughout. Each question was worth approximately three points as figured on a one hundred point grading scale. Table 12 presents the number of SPA respondents and mean average score per survey by state.

TABLE 12

NUMBER OF SPA RESPONDENTS AND MEAN AVERAGE SCORE PER SURVEY BY STATE.

State	Total N	Pre - Survey N	Pre - Survey MEAN	Post - Survey N	Post - Survey MEAN
Tennessee	6	5	69.2	3	79.6
Kentucky	1	1	77	1	85
N. Carolina	3	3	67	3	76
S. Carolina	4	N/R	----	1	68
Georgia	2	2	73.6	1	75
Florida	3	3	84.6	3	89.3
Alabama	2	2	57	N/R	----
Mississippi	4	3	73	2	84
Total	25	19	71.6	14	79.5

It should be noted that the representatives from the South Carolina SPA had not yet arrived when the pre-survey was administered. They and all others arriving late were excused from taking the pre-survey.

A noticeable improvement can be seen in the post-survey mean scores. The reader is cautioned however in the comparison of pre and post average scores due to differences in N values in five out of eight states.

Table 13 breaks down the number and ercentage of correct responses to the pre and post survey by question.

The percentage of change column cannot be taken as a true indicator of change in this case because of the inability to match respondents. Again the reader is cautioned in comparing the pre and post survey results due to changing values of N.

TABLE 13

NUMBER AND PERCENTAGE OF CORRECT RESPONSES TO PRE AND POST SURVEYS

Question	Pre - Survey (N = 19) N	%	Post - Survey (N = 14) N	%	+ - % Change
1	6	32	6	42	+ 10
2	16	84	12	85	+ 1
3	6	32	7	50	+ 18
4 A	12	63	11	78	+ 15
4 B	14	74	10	71	- 3
4 C	12	63	10	71	+ 8
5	1	5	2	14	+ 9
6	7	37	8	57	+ 20
7	10	53	10	71	+ 18
8	6	32	9	64	+ 32
9 A	12	63	11	78	+ 15
9 B_1	14	74	12	85	+ 11
9 B_2	7	37	2	14	- 23
9 B_3	6	32	2	14	- 18
10	9	47	7	50	+ 3
11	17	89	14	100	+ 11
12	6	32	3	21	- 11
13 A	19	100	14	100	----
13 B	15	79	9	64	- 15
14	18	95	12	85	- 10
15.1	13	68	14	100	+ 32
15.2	6	32	9	64	+ 32
15.3	13	68	14	100	+ 32
15.4	10	53	13	92	+ 39
15.5	13	68	13	92	+ 24
15.6	19	100	14	100	----
15.7	4	21	8	57	+ 36
15.8	4	21	8	57	+ 36
15.9	9	47	11	79	+ 32
15.10	16	84	14	100	+ 16
16	6	32	3	21	- 11

Total N = 25

Twenty-two questions in the pre-survey or over seventy percent of the questionnaire had a correct response rate of less than seventy percent. Fourteen questions or forty-five percent of the questionnaire rated likewise in the post-survey. These figures seem to suggest that most respondents lack a familiarity with the basics of crime analysis.

Three questions pertaining to the LEAA guidelines on crime analysis presented difficulty in both surveys. The questions asked for: 1) the minimum data base that must be reported and analyzed for the state and its subdivisions when conducting a crime analysis; 2) the jurisdictions that must be included in a crime analysis; and 3) the naming of the Part I offenses as classified under the Uniform Crime Reporting System.

The most difficult statistical questions were: 1) the median versus the mean in reporting "averages"; 2) the value(s) of the correlation coefficient; 3) the symbol for the coefficient of determination; and 4) the symbol for the Pearson Product Moment Correlation.

Data illustrations were another problem area. Ordinal and interval level data were confused when presented in tables. Also, few recognized the merits of a semi-logarithmic graph in illustrating relative changes over a period of time.

The survey results turned out to be of limited utility due to methodological constraints. However, the direction and degree of learning during the workshop can be evidenced in the results presented.

Evaluation of the Workshop

Following the post-survey and marking the end of the Crime Analysis Workshop was a final questionnaire calling for an evaluation of the workshop. This questionnaire afforded each participant the opportunity to comment on the manner of subject presentation, the overall quality of the workshop, and whether this type of workshop would be beneficial to local and regional planners and/or to other LEAA regions.

No identification, personal or state, was required in the completion of the evaluation questionnaire and the questions were open-ended by design to encourage their candid comments. A copy of the evaluation questionnaire can be found in Appendix D.

Twenty of the forty-one participants or forty-eight percent returned an evaluation of the workshop. This does not mean that all of the questionnaires returned were complete but only that comments were offered to a few of the questions in all cases. It should also be made clear that evaluations by LEAA participants are included in this number of respondents due to the anonyminity of the responses. Bearing this in mind, the returns will be described in narrative form by topic.

Guidelines

Sixteen commented on the presentation of the LEAA guidelines as they relate to crime analysis. Fourteen of the sixteen expressed the feeling that the presentation especially the assets/liabilities analogy was helpful in their understanding of the guidelines.

A few shortcomings were pointed out in the comments. Specifically: 1) more attention should have been given to the identification of data sources; 2) more time was needed in going through the guidelines; and 3) comment from the LEAA representatives would have been appreciated.

Statistics

All twenty respondents offered comment on the manner in which statistics were presented. The presentation received high ratings despite the complexity of the subject matter and the insufficiency of time. A comment summing up the general feeling of the respondents to the session on statistics stated the material to be too difficult for those not having a background in statistics and too elementary for those having a statistics background. A suggested remedy was to divide the group and present the material at both levels.

A few comments expressed the feeling that more emphasis could have been on the practical application of statistics in the planning process.

Computer

Ten participants commented on the "hands on" approach to the computer. It was noted by some that this was their initial exposure to the computer. Nevertheless, the experience was enjoyed by all respondents.

A more formalized presentation of the computer and its capabilities was suggested. A few questioned the relevancy of the computer in preparing a crime analysis.

Interpretation

Eleven commented on this session that dealt with the writing up of the results and translating the statistical terms into language easily understood by laypersons. The presentation was given high ratings with the only comment being that more time was needed for discussions.

Information

Twelve respondents identified information he/she felt should have been incorporated in the workshop. The following areas were suggested:

- the practical aspects of crime analysis
- how to collect information from local authorities
- how to analyze data SPA's already have
- how to show relationships
- the mechanics and the art of drawing conclusions
- more component problems - police, court, corrections
- specific individual state problems
- specific data sources
- simpler methods of data presentation
- evaluation techniques
- pattern recognition techniques

In addition to the above suggestions, many cited the need for more time.

Evaluation

The over-all evaluations of the workshop were quite favorable. Eighteen of the nineteen respondents considered the two and one-half days spent at the workshop productive and worthwhile. The comments were mostly complimentary of the planning and organization that went into the staging of the workshop. The one comment not falling into this category stated that the workshop was not directly related to crime analysis but focused more on statistical methods.

Again the shortage of time was mentioned as turning the workshop into a cram session.

Local Planners

Sixteen respondents felt this type of workshop would be beneficial to local and regional planners in their state. To be meaningful though, more emphasis would need to be given to specific problems and practical applications of the information. Three expressed the opinion that the workshop was beyond the comprehension of local and regional planners. One suggested that local planners should be involved in data collection only.

LEAA

The response was unanimous when asked whether he/she would recommend this type of workshop to other LEAA regions. All nineteen respondents stated they would. A few qualified their recommendation with the conditions that the time alloted for the workshop be extended to a week and that greater emphasis be placed on practical applications.

Facilities

Only one of the twenty respondents indicated dissatisfaction with the physical arrangements of the workshop. The nature of the complaint centered around the policy of double occupancy of rooms in housing accomodations. Otherwise, the arrangements were believed to be satisfactory in meeting their needs.

A few claimed meal allowances were insufficient in covering food expenses.

Comments and Recommendations

An enduring benefit from any workshop that draws professionals together from differing geographic regions is the opportunity to share and exchange ideas with each other. This cross-fertilization was observed taking place during meals, between workshop sessions, and at social activities. The learning taking place during the entire workshop may in actuality then be greater than the survey results indicate.

One of the recurring suggestions made throughout the workshop was for the identification of specific data sources. This particular topic was covered at length in the opening presentation. It should be noted that data packages were prepared for each state. Included were Census data and selected sociodemographic variables. Also, a request was made to the FBI to rerun all Index crime data for the past five years using the new definition of larceny. Due to a backlog of similar requests, this information was not available at the workshop but has now been forwarded to the SPA's.

Unquestionably the major obstacle in the preparation and presentation of the workshop was the shortage of time. But in order to be of any assistance in the preparation of the 1977 Comprehensive Plans the pace had to be accelerated. Admittedly the pace may have been too fast in a few sessions given the complexity of the subject matter.

Overall the workshop can be considered a success. The following recommendations are made in recognition of noted weaknesses:

Recommendation 1: Selection

The purpose and contents of the workshop should be made clear to SPA directors so that it is clear who should attend the workshop.

Recommendation 2: Registration Information

Every effort should be made to have the registration information returned one month prior to the workshop.

Recommendation 3: Pre - Post Survey

The research design must be tightened before significant conclusions can be drawn from the results. Methods of unobtrusive measurement should be explored so as to be able to match respondents without jeopardizing cooperation.

Certain questions and all illustrations need re-examination to ensure clarity and accuracy in the information given.

Recommendation 4: Time

The objectives of the workshop should be reviewed in an attempt to narrow the focus thereby reducing the likelihood of "cram" sessions.

A position paper is currently in the process of development and will deal with the recommendations in detail in relationship to future workshops in crime analysis.

APPENDIX A

Name____________________

REGISTRATION INQUIRY FOR CRIME ANALYSIS WORKSHOP

Please acquaint us with your professional background and experience in criminal justice planning and evaluation? This information will better enable us to determine the manner of presenting material to you during the workshop and will help us in meeting your individual needs following the workshop.

1. Please identify the agency you represent (if a state agency please identify the state)?

2. Have you ever attended a crime analysis workshop before? If so, please cite specifics:

3. Highest educational level attained:

4. What exposure have you had to:
   Statistics

   Research Methodology

   Computer Programing

5. Estimate the total number of job-related in-service training hours that you have received in the past year?

6. How many SPA staff members prepare your state's crime analysis?

7. What do you perceive your needs to be in reference to crime analysis?

8. Briefly, describe your experience in planning and evaluation.

APPENDIX B

CRIME ANALYSIS WORKSHOP

April 28, 29 and 30, 1976

Wednesday, April 28

12:00 - 12:30	Registration, University Center - Lower Lobby
12:30 - 1:30	Luncheon, (Chattanooga Rooms A & B)
1:30 - 5:30	Workshop Detail (Signal Mountain Room)
Linda Myers	Overview of Crime Analysis and Its Implications
Ken Venters	Survey of LEAA Guideline Requirements and how to be responsive to them
Bob Catale	Data Sources What data sources are available, key data elements, and developing additional data sources
6:30 -	HAPPY HOUR(s) - Downtown Sheraton Hotel, Suite 1211

Thursday, April 29

9:00 - 12:00 Mario Perez-Reily	Analytical Techniques, (Signal Mountain Room) Introduction to the basic statistical techniques necessary to prepare a crime analysis with emphasis on regression, correlation, levels of significance, methods of sampling, time series analysis and techniques for testing hypotheses.
12:00 - 1:00	Lunch (can be purchased in the University Center)
1:00 - 2:30	Continuation of Analytical Techniques
2:45 - 5:00 Lloyd Davis	Computer Application, Hands-on application of various statistical techniques, utilization of techniques explored in the morning session.
5:00 - 6:30	Dinner (can be purchased in the University Center)
6:30 - 9:00 Roger Thompson	Continuation of Computer Applications

Friday, April 30

9:30 - 10:30	Continuation of Computer Applications
10:45 - 12:30 Nick Norwood	Presentation of Data, (Conference Room C & D) writing up findings, tracing CA through the design and development of programs responsive to problems, needs, standards and goals.
12:30 - 2:00	Luncheon, Moccasin Room
2:00 - 4:00	Continuation of morning session
4:00 - 6:00	Individualized special assistance as needed.

APPENDIX C

CRIME ANALYSIS WORKSHOP - SURVEY

State ONLY ______________________

Please do not place your name on this survey. Note your state only at the top of the first page.

1. At minimum, what data base must be reported and analyzed for the State and its subdivisions when conducting a crime analysis?

2. Crime trend analysis must cover a __________ year time span.

3. What jurisdictions must be included in a crime analysis?

4. Define the following terms as they relate to crime analysis. Cite an example of each.
   a. "Goal":

   b. "Objective":

   c. "Standard":

5. Which statistic would you use, the median or the mean, to tell [illegible] the "average" number of training hours per police agency?

Total Police Recruit Training Hours by Jurisdiction

Police Agency	Total Hours	
Minneapolis, Minn.	455	
Washington, D.C.	459	
Dallas, Texas	520	Mean 449.25
Whittier, Calif.	280	
Birmingham, Ala.	399	Median 457
Hilo, Hawaii	262	
Cleveland, Ohio	634	
Rochester, New York	585	Total 3594

6. Name the Part I offenses classified under the Uniform Crime Reporting system.

7. Differentiate between descriptive and inferential statistics.

8. Hypotheses are suppositions presumed to be true for the sake of testing. Write a null hypothesis.

9. There are three levels of statistical measurement, nominal, ordinal, and interval. Match each term with its description and the cited examples:

a. 1. _______________ has the capability of ranking numbers from high to low

2. _______________ permits the categorization of elements.

3. _______________ is composed of equal units and capable of mathematical manipulation

b. Example 1. _______________

Index Offense	Total Offenses
Murder	17,630
Assault	364,600
Rape	41,890
Robbery	385,910

Example 2. _______________

Violent Crime Factor Scores for SMSA's*

SMSA	Violent Crime
1. Baltimore, Md.	2.12
2. Austin, Texas	1.88
3. Atlanta, Georgia	.95
4. Akron, Ohio	.66
5. Buffalo, New York	.75

* Mean = 0. Standard deviation = 1.00.

Example 3. _______________

Mortality rates for Middletown 1974-75
for persons under eighteen years of age

Age at Death	Total Deaths
Birth to 1 year	57
1 - 5	26
6 - 12	16
13 - 18	10
Unknown	8
Total	117

10. The value of the correlation coefficient is always between the values of

_______________ and _______________.

11. The most reliable method for arriving at a representative sample is _______________

A. A judgmental sample
B. A quota sample
C. A random sample
D. An accidental sample

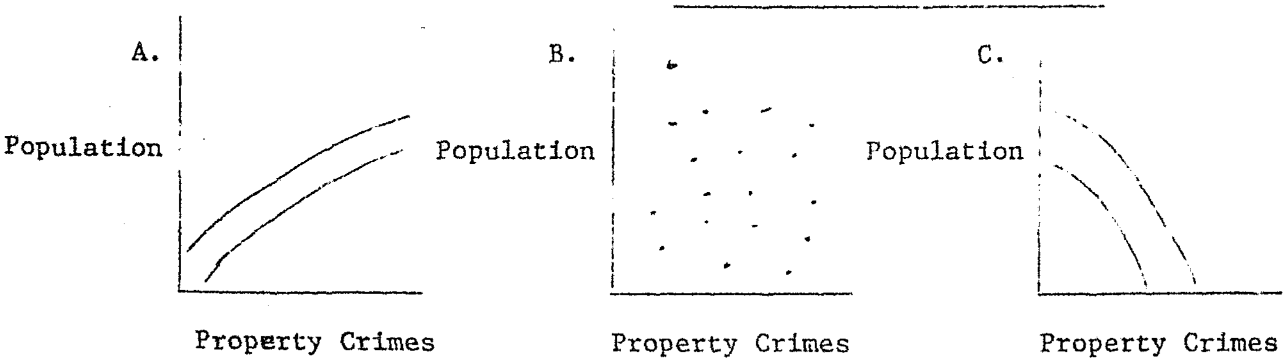
12. The most appropriate graph for illustrating relative changes over a period of time is

_______________.

A. A period graph
B. A semi-logarithmic graph
C. An arithmetic graph

13. a. Which of the three cities below shows the strongest positive association between property crimes and population? ____________

b. Which of the cities is least likely to have a strong relationship between property crime and population? ____________



14.

Black personnel in selected police departments

City	Rank in % Black pop.	Rank in % Black in police Force
Atlanta	1	3
Chicago	3	1
Louisville	4	5
New Haven	5	4
Oklahoma City	6	6
St. Louis	2	2

The above data tells us ____________________.

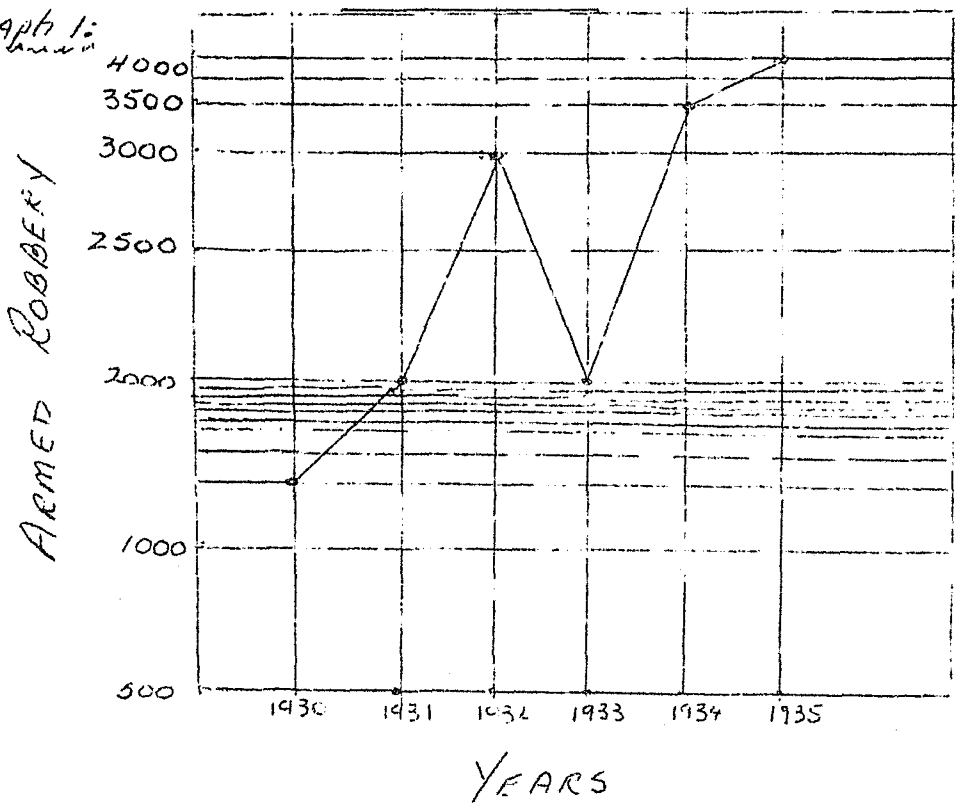
A. in cities ranking high in black population, there are fewer blacks on the police force.
B. police departments pay little attention to the hiring of blacks.
C. As the number of blacks in the general population increases, more blacks are hired by police departments.
D. the data is insufficient to allow an interpretation.

15. Match Column I with Column II.

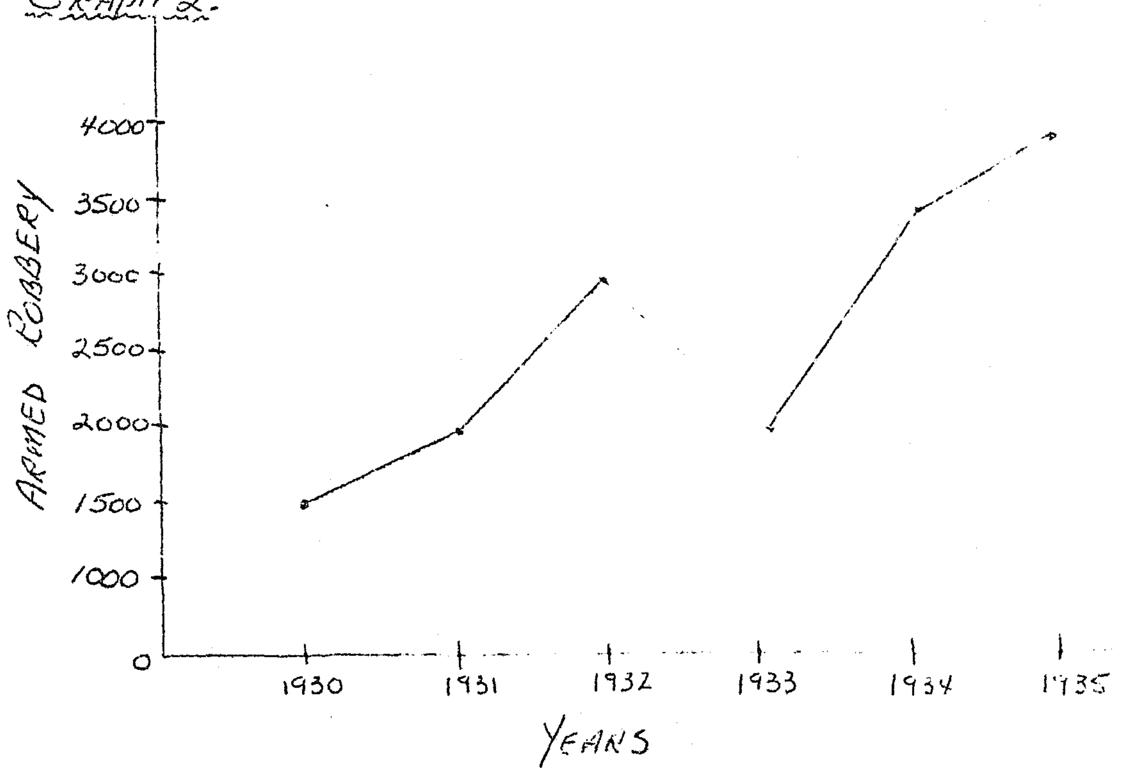
Column I	Column II
1. _____. σ or s	a. Size of Sample
2. _____. R	b. Mean
3. _____. H_o	c. Multiple Correlation Coefficient
4. _____. $\bar{X}$	d. Regressed Estimate (Predicted Value of Y)
5. _____. N	e. Standard Deviation
6. _____. $\sqrt{\ }$	f. Radical or Square Root Sign
7. _____. r^2	g. Pearson Product-Moment Correlation
8. _____. r	h. Null Hypothesis Symbol
9. _____. Y_c	i. Summation or Total of
10. _____. Σ	j. Coefficient of Determination

16. Both graphs depict the same time period. Which of the two will best convey what is happening? ______________

Graph 1:



Graph 2:



APPENDIX D

EVALUATION OF CRIME ANALYSIS WORKSHOP

Please comment on the following:

A. Subject material presented:

1. Guidelines

2. Statistical methods

3. Computer applications and "hands-on" approach

4. Interpretations

B. Was there any particular information that you feel should have been incorporated in the workshop?

C. What is your over-all evaluation of the workshop?

D. Do you feel this type of workshop would be beneficial for local and regional planners in your state?

E. Would you recommend this type of workshop to other LEAA regions?

F. Did you find the physical arrangements satisfactory for your needs?

APPENDIX E

LESSON PLAN

STATISTICAL TOOLS

I. What is/are statistics? (A theory of information/a set of tools and procedures for decision-making)

II. Thinking about human behavior

A. Hypotheses

B. Variables - Dependent and Independent

C. Measurements

D. Types of Measurements

1. Nominal

2. Ordinal

3. Interval

4. Ratio

III. Describing Phenomena

A. Averages in general

B. Measures of Central Tendency

1. Mode

2. Median

3. Mean

C. Advantages and Disadvantages

IV. Testing Hypotheses

A. The Empirical Hypothesis

B. The Null Hypothesis

C. Confidence Limits

D. Levels of Significance

V. Sampling

A. The Idea Behind Sampling

B. Types of Samples

1. Non-Probability Sampling
    a. accidental
    b. purposive or judgemental
    c. quota
2. Probability Sampling Methods
    a. Simple Random Sample
    b. Systematic Sampling
    Two - Stage Probability {
    c. Cluster Sampling (area sampling)
    d. Stratified Random Sampling

C. Determinants of Sample Size

1. Size (not very important)

*2. Variability

*3. Type of sampling procedure employed

*4. Available resources (time, money, personnel)

*5. The number of variables or characteristics under study

*6. Degree of accuracy and precision required

D. Using the Table of Random Digits and the Random Numbers Generator

E. Estimating Error

1. The standard error of the sample

Mean ($S_{\bar{X}} = \dfrac{S}{\sqrt{N - 1}}$

2. Standard Errors Associated with Different Levels of Precision

VI. Correlation and Regression

A. Parametric Statistics:

1. Simple Correlation and Regression
    a. The Scattergram
    b. Regression
    c. Correlation

d. The Coefficient of Determination

e. Tests of Significance

1. The t-test

2. The F-ratio

2. Multiple Regression and Reference to Multiple Correlation R and R^2

B. Non-Parametric, Non-Quantifiable Data

1. Spearman's Rank - Difference Correlation

2. Test of significance

VII. Time - Series

A. Characteristics of Time - Series

1. Long-term Trend

2. Cyclical Effect

3. Seasonal Effect

4. Random Variation

B. <u>What to do with Time - Series</u>

1. Use of different kinds of graphs

a. Arithmetic graph

b. Semi-logarithmic graph

(advantages of this type of graph)

2. Smoothing for random variation (smoothing methods)

3. Some Forecasting Models

a. The <u>linear model</u> only

b. The <u>auto regressive</u> model

* Important Point

Measures of Dispersion:

Measures of central tendency (averages) only tell us about the location of a representative value among a number of values but they do not provide us any more useful information. For example, while it is good to know what the mean number of violent crimes per 100,000 for a collection of SMSA's is (We have seen this problem already.), it is useful to get n idea of the amount of deviation right and left (or above and below) of the mean. Why this is useful will become apparent in a few paragraphs.

Measures of dispersion (also known as measures of variation) supply information about the amount of dispersion from the mean present in a collection of values. Let's take one more look at our problem concerning the 1974 SMSA's:

SMSA	Population	Violent Crime Rate/100,000 pop.	Property Crime Rate/ 100,000 pop.
Alabama	2,206,670	434.0	3,625.7
Arkansas	796,922	446.7	491.8
Florida	6,780,222	743.0	7,204.8
Georgia	2,770,045	568.3	4,520.9
Louisiana	2,366,031	621.2	4.504.0
Mississippi	491,719	391.1	3,516.2
South Carolina	1,332,330	559.	4,957.3
Tennessee	2,464,146	502.9	4,578.1
Mean	2,401,011	533.3875	4,174.85
s	1,827,472	107.3185	1,749.42

You would like to give the Tennessee legislators a comparison of the Tennessee SMSA to the others in terms that would be simple and meaningful. Let us see how this can be done.

The Standard Deviation

One measure of deviation (or of dispersion) about (or above and below) the mean is the Standard Deviation. Let us take population. The mean population for the 8 SMSA's was 2,401,011. We know that Tennessee had a population above or below the mean. You would have to find the difference between each state SMSA value and the mean. This is done by subtraction. This difference is represented by the term $(X-\bar{X})$. Each of these differences then is multiplied by itself (squared) to find $(X-\bar{X})^2$. All the squared differences are added and you obtain $\Sigma(X-\bar{X})^2$. This summation then is divided by N. (If the values that you have come from a sample, then you must divide by N-1). Then, the obtained quotient is placed under a radical symbol, $\sqrt{\frac{\Sigma(X-\bar{X})^2}{N}}$. Then, you find the square-root and that is your standard deviation. The standard deviation for population is 1,827,472. The standard deviation is represented by two symbols, and s. , a Greek character (known as sigma) is used to represent the standard deviation for a population of values. The letter s is used when your values come from a sample.

Now, go back and find s for violent crime and property crime.

Let us summarize our results:

Measure	Population	Violent Crime	Property Crime
Mean	2,401,011	533.3875	4,174.85
s	1,827,472	107.3185	1,749.42

So, you have calculated three standard deviations. What you have done is represented by the formula:

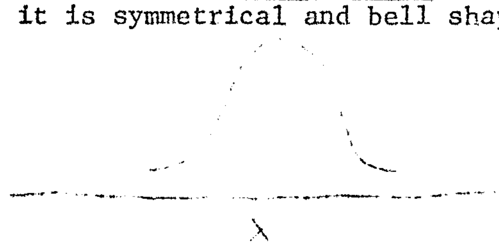
$$\sigma = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}}$$

OR

$$s = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N-1}}$$

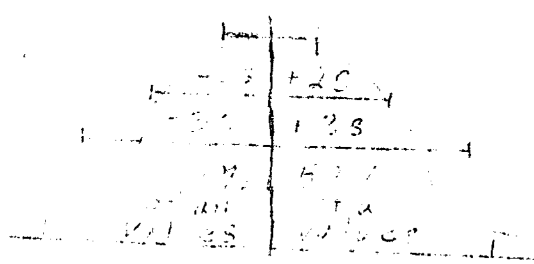
In the problem that you have just worked out you have calculated a measure which is very sensitive to extreme values and tells that the mean cannot tell - how many units to the right and left (or above and below) can the mean fluctuate. The mean can fluctuate, in the case of population, 1,827,472; in the case of violent crimes, 107.3185; and 1,749.42. in the case of property crimes.

The standard deviation is based on the theory of the normal distribution. Any group of measurements is a distribution. A normal distribution is a distribution such that it is symmetrical and bell shaped.

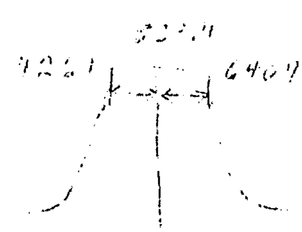


A symmetrically bell-shaped distribution.

It is also characterized by the fact that the mean divides the distribution into two halves, one of which is a mirror of the other.



Fifty percent (50%) of all the values in a distribution (if it is normal) should lie below the mean and 50% above it. If you add the standard deviation to the mean it will equal (in the case of the violent crime rate) 533.4 + 107.3 which equal 6,407. This is the area of X + s (mean plus one standard deviation), or 533.4 ←——→ 6,407. Subtract one standard deviation and now you also have the area of X - s, or 533.4 - 107.3 = 426.1 or 426.1 ←—— 533.4. Now, if we look at the normal distribution we have the



Area of plus-minus one standard deviation with 68% of the cases within that area to find the area of plus-minus 2s, double the s and add or subtract to the mean and you have the area of $\pm$ 2s. Triple the standard deviation, follow the same procedure and you have the area of plus and minus three (3) standard deviations. The area of $\pm$ 2s includes 95% of all the values in any sample (or population). The area of $\pm$ 3s contains 99.9% of all the values.

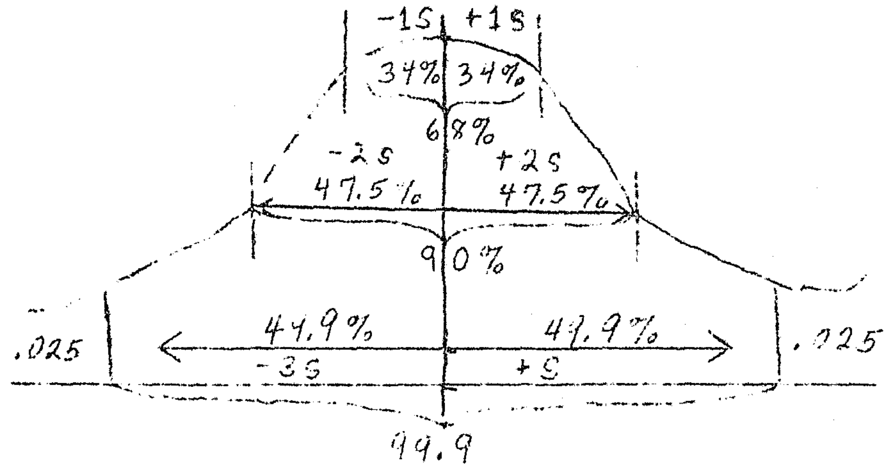


Illustration: The normal curve and its characteristics.

You can compare the distribution of values for the SMSA's with the normal curve, and if the scores don't fall as in the normal curve, then you have something to say about the SMSA's.

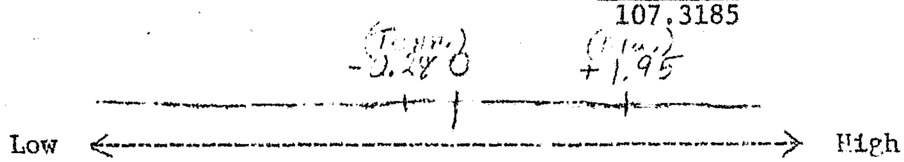
You can also take each individual SMSA, obtain the difference between its value and divide by its standard deviation and that gives you a Z - score. The formula for the Z - score (also known as the standard score) is:

(a) $\frac{X}{s}$ for a sample and (b) X for a population

On the violent crime , Tennessee has a difference (X) of -30.48 (since the value is smaller than the mean). So the z - score for Tennessee is:

$$\frac{X}{s} = \frac{-30.48}{107.3185} = -0.28$$

What this means is that on a scale using the mean as the zero-point, Tennessee would look like this in comparison to Florida $\frac{+209.613}{107.3185}$:



Where would each of the other states fall in such a scale? Find the z - scores in each case, in the case of the violent crime rate and place each state on the corresponding position. What use could you make of this device?

There are other uses of the standard deviation that could be helpful in other types of problems. Would you like to know about them?

Rank-Order Correlation

Up till now, we have been running (or learning to run) statistical tests that have two distinguishing characteristics: (1) they fit interval level data, and (2) are parametric. You are already familiar with levels of measurement, thus understand what an interval measurement is. Numerical descriptive measures are measures that locate the center and describe the spread of the distribution. In statistical tests, such as product-moment correlation and regression, we apply tests to sample data for the purposes of saying something (making generalizations) about the population from which the sample comes. Parametric statistics require that (a) data come from a random sample, and (b) that the sample represents a normally distributed population.

Whether or not a population is normally distributed is always questionable. Non-parametric statistics get us out of this dilemma because they are not based on any assumptions about means or standard deviations, nor about whether the sample data come from a normally distributed population. Non-parametric statistics are also called distribution free statistics. Data that have been measured as either ordinal or nominal lend themselves to analyses with the use of non-parametric statistics. One such technique is Spearman's Rho (or R_s), also known as the rank-difference coefficient of correlation. It is a very easy measure to compute and it is especially useful in dealing with small samples or populations (where N is less than 50). They are also appropriate when variables cannot be quantified satisfactorily. Let us test the hypothesis that high unemployment tends to associate with high incidence of delinquency. Let us say that you are able to get information for census tracts areas of a large metropolitan city. The city has 22 census tracts and you are able to collect the following information:

(an example of social area analysis)
See next page.

The formula for $R_s = \frac{1 - 6 D^2}{N3 - N}$

where ΣD^2 = the sum of the squared-rank differences

N = number of cases (size of the sample)

with our data:

$$R_s = (1) - \frac{6(203)}{223 - 22} = 1 - \frac{1218}{10648 - 22}$$

$$= 1 - \frac{1218}{10626}$$

$$= 1 - .1146245$$

$$= .88$$

Table RD. Relationship between Percentage Unemployed and Percentage of Car Theft Offenders by Census Tracts

Census Tract#	(X) Percentage Male Unemployed	(Y) Percentage Car Theft Offenders	(X) Rank	(Y) Rank	D	D^2
1	10	7				
2	22	15				
3	18	10				
4	3	2				
5	5	6				
6	21	12				
7	6	5				
8	2	1				
9	1	2				
10	7	4				
11	3	7				
12	2	3				
13	10	11				
14	15	12				
15	15	9				
16	9	6				
17	6	2				
18	4	3				
19	8	7				
20	11	12				
21	13	10				
22	14	9				

Our coefficient tells us that the association is positive and strong. These results tend to suggest that the relationship between male unemployment and auto theft offenders is verified. In order to determine the significance of this association, we go to table G: Critical Values of Rho. Since we have 22 cases, we go down the n column. We wish to test a one-tail Null Hypothesis that census tracts with high male unemployment do not associate significantly with high incidences of auto thefts. We locate 22, and go across where it reads: "level of significance for a one-tailed test, .05". The expected value of R_s .359. Since our observed R_s is larger (higher) than the expected value, we reject the null hypothesis and any rejection of the null hypothesis validates the empirical hypothesis. The interpretations of these results are basically the same as for the product-moment correlation coefficient (r).

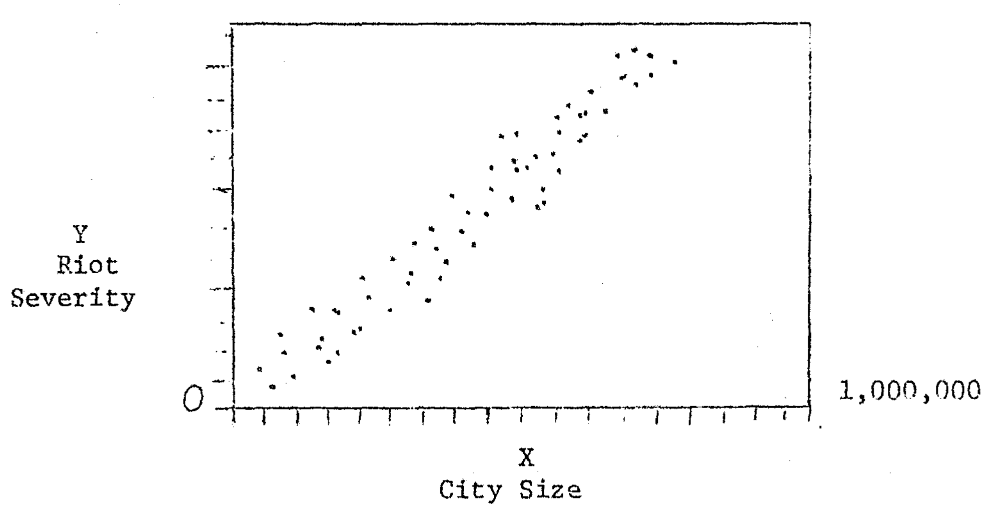
Regression and Correlation:

Let's take a case - predicting the number of vehicle units that a police department will need for a given fiscal year. The following methods could be suggested to predict the number of vehicle units that the city of Memphis, Tennessee, would need for 1974. In order to predict (really, what we are doing is estimating) we must think in terms of a relationship: we can estimate on the basis of the past relationship between city area and the number of vehicle units. For the city of Memphis, Tennessee, the data will look like this:

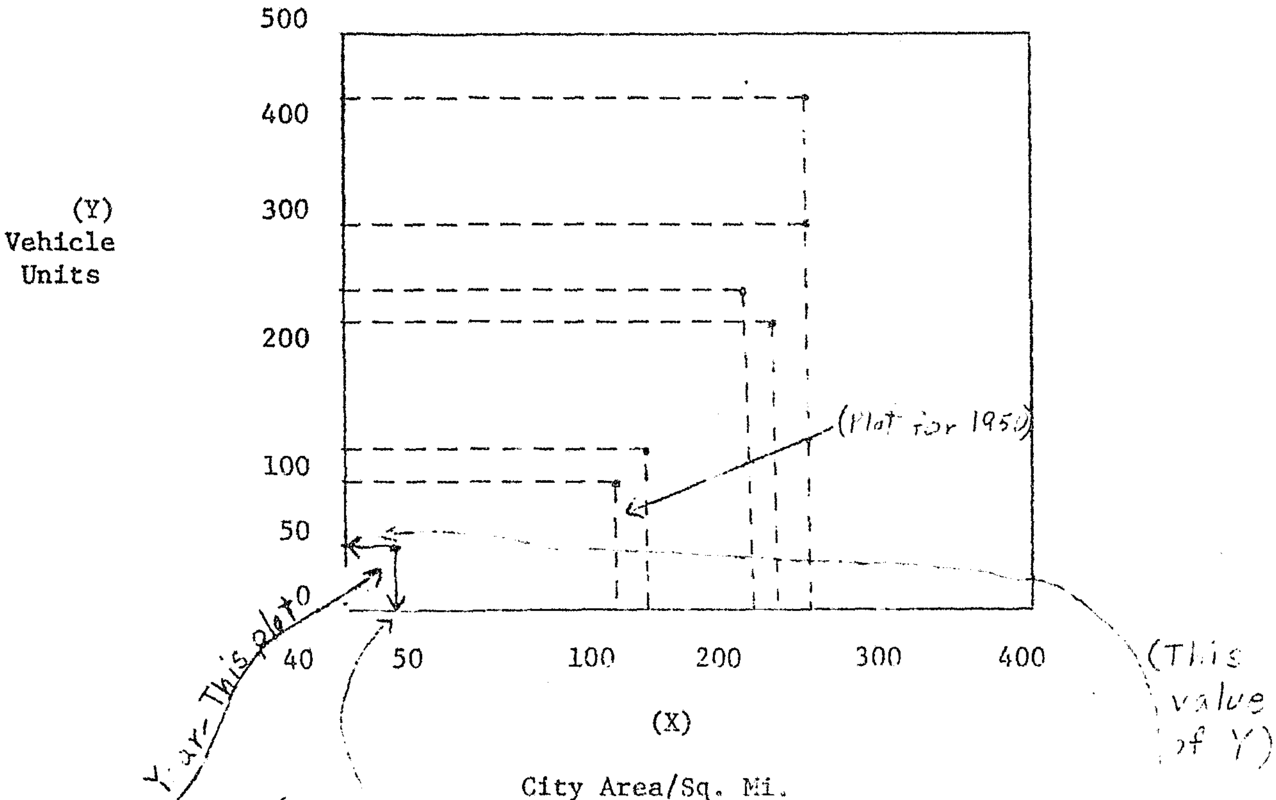
Year	(X) City Area (sq. miles)	(Y) Vehicle Units
1940	48.50	50
1950	115.80	83
1960	140.70	123
1970	234.65	252
1971	235.02	219
1972	251.22	298
1973	251.22	444
1974	274.58	?

The problem here is to predict (or estimate) how many vehicle units will be needed in 1974. Let us try the regression method of estimating or predicting. If you notice the column containing the values for city area has been labeled X and the column for the number of vehicle units has the label Y assigned to it. So, once more let us examine the above data. The year for which we are trying to estimate is 1974 (column Y). The letter XX represents the independent variable and Y the dependent variable. In this problem the hypothesis is that there is a relationship between area per square mile and the number of vehicle units required to police a city. As area per square mile increases, the number of auto units increases.

Before any numerical computations are performed, you must get a picture of the relationship under examination. This is done with a scattergram. (another word is scatterdiagram). A scattergram is constructed by plotting coordinates on a graph. These coordinates are points representing X and Y values in the scattergram.

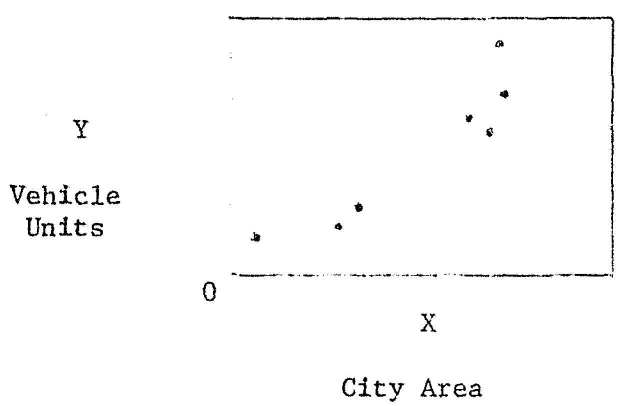


Let us plot the coordinates for our data:



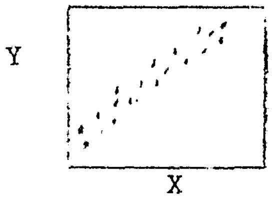
Year	City Area (X)	Vehicle Units (Y)
1940	48.50	50
1950	115.80	

The first coordinate we plot is for the year 1940 (X = 48.50, Y = 50). We locate these values on the respective scales or axes and place a dot on the place where weak lines drawn at right angles from the scales intersect. This plotting is done for each of the years. You can see that the scattergram shows a relationship that looks like this:



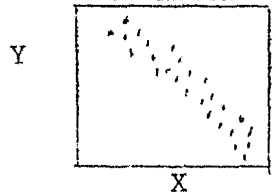
Whenever scattergrams are used, any of the following basic patterns may be produced:

(A)



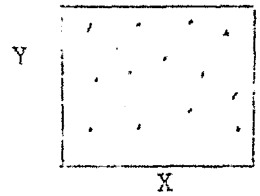
Positive Association

(B)
Rectilinear



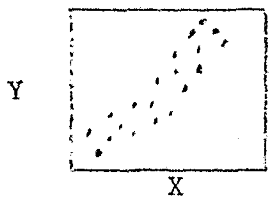
Negative Association

(C)

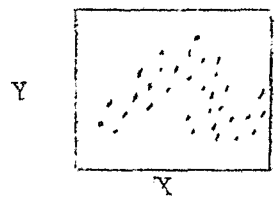


No Association (Scatter)

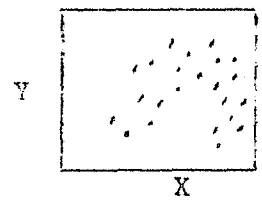
(D)



(E)



(F)



Curvilinear

Our scattergram looks like model A.

When we look for relationships through the use of a scattergram we look for linear patterns (preferably rectilinear). Our relationship looks <u>rectilinear</u> and it seems positive: as X increases Y also increases. <u>Bivariate linear regression</u> is the statistical technique usually employed in estimating the values of a variable Y when we know the values of the variable XX. We call this the estimated regression of Y on X (Yc_{yx}). The formula used to estimate Yc is:

$$Yc = a + bX$$

where (1) $a = \frac{\Sigma Y - b\Sigma X}{N}$

and (2) $b = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}$

In order to be able to estimate (or predict), we must manipulate our lata. First, we must find b, then a, and then we are able to estimate. So, let's do to the data all that the formulas require.

4

X	X^2	Y	Y^2	XY	
48.50	2352.25	50	2500	2425.00	
115.80	13409.64	83	6889	9611.40	N = 7
140.70	19796.49	123	15129	17306.10	
234.65	55060.62	252	63504	59131.80	(X)2 =
235.02	55234.40	219	47961	51469.38	1,631,009.9
251.22	63111.49	298	88804	74863.56	
251.22	63111.49	444	197136	111541.68	
1277.11	272076.38	1469		326348.92	

$$(Y)^2 = Y^2 = 341{,}923$$

Now we have all the information that we need to carry out the formulas. First, find b.

$$b = \frac{N\Sigma XY - (\Sigma X)\ (\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2} = \frac{(7)\ (326348.92) - (1277.11)\ (1469)}{(7)\ (272076.38) - (1631009.9)}$$

$$b = \frac{2{,}284{,}442.4 - 1{,}876{,}074.5}{1{,}904{,}534.6 - 1{,}631{,}009.9} = \frac{408{,}367.9}{273{,}524.7}$$

$$b = 1.49$$

Next, find a.

$$a = \frac{\Sigma Y - b\Sigma X}{N} = \frac{1469 - (1.49 \times 1277.11)}{7}$$

$$a = \frac{1469 - (1902.89) = 433.8939}{7} = -61.98$$

$$a = -61.98$$

Now we can start predicting (or estimating). As you perhaps noticed, for 1974 we need to estimate the number of vehicles. The area (X) was 274.59. So let us use the formula.

$$Yc = a + bX$$

$$= -61.98 - (1.49 \times 274.59)$$

$$Yc = 347.159 \text{ vehicle units}$$

Let's do another Yc. For the year 1940, the area (X) was 48.50 square miles. So we compute:

$$Yc = (-61.98) - (1.49 \times 48.50)$$

$$Yc = (-61.98) - (72.265)$$

$$= 10.56 \text{ auto units}$$

As you go on, you will see that the predicted values (Yc) differ from the actual number of auto units for each of the years. Next to each value of Y you can plot the corresponding estimate. Then it is possible to connect each of the estimates with a straight line. This straight line is called the line of regression. If we take every difference between Yc's and Y's and square them $(Yc-Y)^2$, then we can add them ($(Yc-Y)^2$) and once we have added them we can proceed to find the <u>standard</u>:

X	Y	Yc*	Yc-Y	$(Yc-Y)^2$	
48.50	50	10	-40	1600	
115.80	83	110	27	729	
140.70	123	148	25	625	
234.65	252	288	36	1296	
235.02	219	288	69	4761	
251.22	298	312	14	196	
251.22	444	312	-132	17424	
274.59	(480)	347 *projected for 1974		26631	$(Yc-Y)^2$

<u>Error of the regression</u> (S(Yc)

The formula for the S(Yc) is:

$$S(Yc) = \frac{(Yc-Y)^2}{N} = \frac{26631}{7}$$

$$S(Yc) = 61.68 \text{ or } 62$$

This <u>standard error of the regression</u> tells you how accurate or reliable the regression predictions are. As indicated by the S(Yc), the predictions are going to fluctuate an average of 62 auto units above and below the line of regression (the area of plus and minus S(Yc). Notice that two original values lie outside standard error limits. Those two values could be considered out of line. Though we are not going to find that explanation here, it must be pointed out that what is happening might be due to some other variable that tended to produce the unusually high and low auto units for those two years (444 and 219).

In our original regression formula the value <u>a</u> tells us what the starting value of Y would be if X were equal to zero. The value b tells us how many values Y will increase for every unit increase in XX. In simple terms, b tells how many auto units will be required for every square mile increase.

The term <u>correlation</u> is very often likened to regression. The regression line (or slope) that you have just learned to compute serves to get estimates on the basis of the relationship between two variables, but it has several weaknesses. First, there are two possible lines of regression, Y on X (the one just learned), and X on Y (the one we are not going to do). Second, the b value makes sense only in terms of the units in which the data are measured. A change in units, say from feet to inches, or from square miles to square kilometers, will change the b value. Also, if we wish to compare the effects of two independent variables upon a dependent variable (if we also wanted to study the effect of auto accidents on the number of police auto units), we could not do so if one is expressed in terms of square miles and the other in terms of auto accidents per 10,000 automobiles. We would have to turn to the standard deviation to know whether the path of the dots on the scattergram is is a narrow one, and reliable, or broad and unreliable, This, too, is expressed in terms of the original values, and consequently would not make for adequate comparisons.

Correlation deals with these weaknesses by producing a single measure that has the quality of varying between 1 for a perfect positive association, down through 0 for no association, to -1 for perfect negative association. The product-moment, or Pearsonian coefficient of correlation, symbolized by r, is the statistic that expresses each variable in terms of what is called a standardized deviation. This makes it possible to compare the strength of relationships between variables expressed in all sorts of units (auto units, square miles, auto accident rates, population, etc.), just as long as we have enough values to establish a viable mean and a standard deviation.

Let us look once more at the diagrams on page 3 to see the connection between regression and correlation. If the cluster of points (plots) is tight against the diagonal (the regression line), the correlation is high; if it is football shaped, the correlation is moderate; and if it is widely circular or scattered, then there is no correlation (r = 0). If the diagonal runs from lower left to upper right, the correlation is positive, and if the points run diagonally from the upper left to the lower right, the correlation is negative.

The coefficient of correlation (r) is used as an index of association applicable to any pair of variables. This mea-s that we can tell, if we go back to the relationship between square miles and numbers of police auto units, how closely the two variables are related in the case of Memphis. Though we must perform most of the calculations necessary for the regression before finding the correlation coefficient, it is customary (in most analyses) to report only correlations.

The formula for r is:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

We can use the information already obtained when we computed the regression "estimates."

$$r = \frac{2,284,442.4 - 1,876,074.5}{1,904,534.6 - 1,631,009.9 \quad 2,393,461 \quad -2,157,961}$$

$$r = .875 \text{ or } .88$$

This resulting indicates the existence of a strong association between city area in square miles and the number of auto units available to the Memphis police department. Once you have calculated the coefficient of correlation we must determine how much of the variation (or of increases and decreases) in the number of police auto units could be explained by a knowledge of the area of a city (in square miles). We take the obtained coefficient ($r = .88$) and square it. The resulting measure is known as the coefficient of determination (r^2). The r^2 is $(.88)^2$ which is equal to .77 or 77% of the variations in auto units can be explained (or predicted) with our knowledge of the area of the city of Memphis.

Correlations of less than = .30 are regarded to be of very little importance, because when they are squared they explain very little. An $r = .30$ only explains .09 (or 9%) of a phenomenon. While r tells us about the amount of relationship between two variables, r^2 tells us how much we may improve our estimates of the value of any case on the dependent variable by knowing its value on the independent variable. A a conclusion to any analysis of this kind (Regression-Correlation), we must also compute an estimate of the significance of r.

This is fairly simple. We only have to consider the number of cases (N). When the number of observations are small (in the case of our problem, seven years), a test of significance known as t is used:

$$t = r\sqrt{\frac{N-2}{1-r^2}} = .88\sqrt{\frac{7-2}{1-.77}}$$

$$t = .88\sqrt{\frac{.5}{.23}} = .88\sqrt{21.739} = .88\ (4.66)$$

$$t = 4.10$$

With d.f.= N - 2

d.f.= 5

We test the Null Hypotheses (Ho) that the relationship between area and number of auto units is not statistically significant at the .05 level of significance ($p = .05$). We go to Table 2 and on the first column we locate d.f. = 5, then we go across to the values under column heading 0.05 and we find that the expected value of te = 2.571. Since our obtained (to) is 4.10, this value being greater than te, we reject the Null Hypothesis. We conclude that the observed strong relationship between city area and number of auto units is a significant one. The results are also applicable to the regression estimates.

Regression and correlation can also be used to predict population, crime rates, specific types of offenses, and various other types of phenomena. These techniques are currently being employed in attempts to understand better various types of crime patterns. An example from a recent publication will illustrate the use of this technique in this context (Harris, The Geography of Crime and Justice, p. 21).

Table 2.5

Intercorrelations by States between Reported Index Offenses

	Murder	Rape	Robbery	Assault	Burglary	Larceny	Auto Theft
Murder	1.00	.45	.28	.70	.12	.03	.01
Rape		1.00	.64	.65	.62	.63	.40
Robbery			1.00	.62	.69	.61	.62
Assault				1.00	.47	.37	.27
Burglary					1.00	.88	.75
Larceny						1.00	.65
Auto Theft							1.00

Problem: Let us take two different groups of states and compare them in terms of the relationship between their total crime index and specific offenses:

Region Northeastern States	Rate/100,000 Total Crime Indes	Murder	Rape	Larceny	Auto Theft
1. Connecticut	4,407.0	3.3	11.2	2,275.0	549.9
2. Delaware	5,949.6	10.3	17.3	3,459.9	528.8
3. New York	4,857.1	10.6	28.4	2,054.4	567.5
4. New Hampshire	3,143.9	3.5	8.4	1,973.0	259.0
5. New Jerseay	4,771.7	6.8	19.7	2,392.2	546.7
6. Maine	3,600.2	2.9	8.7	1,946.0	197.7

Region		Total Crime Index	Murder	Rape	Larceny	Auto Theft
North Central States						
7.	Ohio	4,223.4	8.9	23.9	2,285.1	402.4
8.	Michigan	6,519.6	13.0	37.1	3,331.6	634.7
9.	Illinois	5,184.3	11.8	27.7	2,761.5	531.4
10.	Wisconsin	3,641.1	3.0	11.3	2,417.8	246.1
11.	Iowa	3,413.7	1.9	10.1	2,282.1	219.1
12.	Indiana	4,336.9	8.0	23.5	2,396.0	393.0
13.	Kansas	4,300.4	6.9	19.7	2,516.9	238.6
Southern States						
14.	Alabama	3,000.1	15.0	22.7	1,308.7	260.6
15.	Florida	7,387.3	14.7	36.0	3,939.5	482.9
16.	Tennessee	3,659.1	13.4	25.7	1,575.6	346.7
17.	Georgia	3,912.4	17.8	27.1	1,660.8	347.1
18.	Louisiana	3,816.4	16.0	25.2	1,956.2	341.7
19.	North Carolina	3,511.2	11.7	15.5	1,647.1	188.3
20.	Arkansas	3,300.7	11.2	23.9	1,745.3	163.9
Western States						
21.	California	6,846.8	9.5	40.6	3,525.5	638.8
22.	Colorado	6,165.8	6.0	36.5	3,354.2	538.6
23.	Oregon	6,344.7	5.6	32.3	3,665.5	468.7
24.	Washington	6,009.1	5.1	29.0	3,484.8	405.5
25.	Nevada	7,827.1	14.8	45.2	4,086.0	606.1
26.	Arizona	8,221.7	9.6	37.5	4,518.6	602.3
27.	New Mexico	5,212.9	11.3	34.8	2,873.4	306.1

Source: Uniform Crime Reports, 1974, pp. 62-72

Instructions:

1. Take all 27 states. Assume that this is a random sample (N = 27). Using each state as an observation, correlate each variable against every other. Prepare a correlation table. Order the correlations in the order they appear on the above data table.

2. Take each region and run correlations of each variable against each other. How do the correlations come out (a) by regions, (b) nationwide?

3. Is there any way that you could describe each region in terms of discernible crime patterns?

TABLE V—*Continued*

5 per cent (Roman) and 1 per cent (**Boldface**) critical values of F

df_2 \ df_1	1	2	3	4	5	6	7	8	10	14	20	40	100	∞
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.07	1.95	1.84	1.69	1.59	1.51
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.80**	**2.56**	**2.37**	**2.11**	**1.94**	**1.81**
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.06	1.94	1.82	1.68	1.57	1.49
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.77**	**2.54**	**2.35**	**2.08**	**1.91**	**1.78**
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.05	1.92	1.81	1.66	1.56	1.48
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.75**	**2.52**	**2.32**	**2.06**	**1.88**	**1.75**
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.04	1.91	1.80	1.65	1.54	1.46
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.73**	**2.50**	**2.30**	**2.04**	**1.86**	**1.72**
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.03	1.90	1.79	1.64	1.53	1.45
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.71**	**2.48**	**2.28**	**2.02**	**1.84**	**1.70**
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.02	1.90	1.78	1.63	1.52	1.44
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.70**	**2.46**	**2.26**	**2.00**	**1.82**	**1.68**
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.00	1.88	1.76	1.61	1.50	1.41
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.66**	**2.43**	**2.23**	**1.96**	**1.78**	**1.64**
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	1.99	1.86	1.75	1.59	1.48	1.39
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.63**	**2.40**	**2.20**	**1.93**	**1.74**	**1.60**
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	1.98	1.85	1.73	1.57	1.46	1.37
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.61**	**2.37**	**2.18**	**1.90**	**1.71**	**1.56**
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	1.97	1.84	1.72	1.56	1.45	1.35
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.59**	**2.35**	**2.15**	**1.88**	**1.69**	**1.53**
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.95	1.82	1.70	1.54	1.42	1.32
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.55**	**2.32**	**2.11**	**1.84**	**1.65**	**1.49**
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.92	1.79	1.68	1.51	1.39	1.28
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.51**	**2.26**	**2.06**	**1.79**	**1.59**	**1.43**
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.90	1.77	1.65	1.49	1.36	1.25
	6.84	**478**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.47**	**2.23**	**2.03**	**1.75**	**1.54**	**1.37**
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.89	1.76	1.64	1.47	1.34	1.22
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2 76**	**2.62**	**2.44**	**2.20**	**2.00**	**1.72**	**1.51**	**1.33**
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.87	1.74	1.62	1.45	1.32	1.19
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.41**	**2.17**	**1.97**	**1.69**	**1.48**	**1.28**
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.85	1.72	1.60	1.42	1.28	1.13
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.37**	**2.12**	**1.92**	**1.64**	**1.42**	**1.19**
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.83	1.69	1.57	1.40	1.24	1.00
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.32**	**2.07**	**1.87**	**1.59**	**1.36**	**1.00**

TABLE J CRITICAL VALUES OF T

The symbol T denotes the smaller sum of ranks associated with differences that are all of the same sign. For any given N (number of ranked differences), the obtained T is significant at a given level if it is equal to or less than the value shown in the table.

	Level of significance for one-tailed test					Level of significance for one-tailed test			
	.05	.025	.01	.005		.05	.025	.01	.005
	Level of significance for two-tailed test					Level of significance for two-tailed test			
N	.10	.05	.02	.01	N	.10	.05	.02	.01
5	0	--	--	--	28	130	116	101	91
6	2	0	--	--	29	140	126	110	100
7	3	2	0	--	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	302	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

(Slight discrepancies will be found between the critical values appearing in the table above and in Table 2 of the 1964 revision of F. Wilcoxon, and R.A. Wilcox, Some Rapid Approximate Statistical Procedures, New York, Lederle Laboratories, 1964. The disparity reflects the latter's policy of selecting the critical value nearest a given significance level, occasionally overstepping that level. For example, for N = 8,

the probability of a T of 3 = 0.0390 (two-tail)

and

the probability of a T of 4 = 0.0546 (two-tail).

Wilcoxon and Wilcox select a T of 4 as the critical value at the 0.05 level of significance (two-tail), whereas Table J reflects a more conservative policy by setting a T of 3 as the critical value at this level.)

TABLE G CRITICAL VALUES OF r_{rho}

	Level of significance for one-tailed test			
	.05	.025	.01	.005
	Level of significance for two-tailed test			
n*	.10	.05	.02	.01
5	.900	1.000	1.000	--
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.683	.783	.833
10	.564	.648	.746	.794
12	.506	.591	.712	.777
14	.456	.544	.645	.715
16	.425	.506	.601	.665
18	.399	.475	.564	.625
20	.377	.450	.534	.591
22	.359	.428	.508	.562
24	.343	.409	.485	.537
26	.329	.392	.465	.515
28	.317	.377	.448	.496
30	.306	.364	.432	.478

*n = number of pairs

TABLE B RANDOM DIGITS

Row number										
00000	10097	32533	76520	13586	[illegible]	[illegible]	80959	[illegible]	[illegible]	[illegible]
00001	37542	04805	64894	74296	24805	24037	20636	10402	[illegible]	[illegible]
00002	08422	68953	19645	09303	23209	02560	15953	34764	[illegible]	[illegible]
00003	99019	02529	[illegible]	[illegible]	38311	31165	88676	74397	[illegible]	[illegible]
00004	12807	99970	[illegible]	35147	64032	36653	93951	[illegible]	[illegible]	[illegible]
00005	66065	74717	[illegible]	[illegible]	[illegible]	[illegible]	65813	39885	[illegible]	[illegible]
00006	31060	10805	45571	82406	35303	[illegible]	86799	07439	[illegible]	[illegible]
00007	85269	77602	02051	65692	68665	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
00008	63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	35808	34282	[illegible]	[illegible]	[illegible]	[illegible]
00010	98520	17767	14905	[illegible]	22109	40558	[illegible]	[illegible]	[illegible]	[illegible]
00011	11805	05431	39808	[illegible]	50725	68248	[illegible]	[illegible]	[illegible]	[illegible]
00012	83452	99634	06288	98033	13746	70078	18475	40610	[illegible]	77817
00013	88685	40200	86507	58401	36766	67951	[illegible]	76493	[illegible]	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	[illegible]	34113
00015	65481	17674	17468	50950	58047	76974	[illegible]	57186	40218	16544
00016	80124	35635	17727	08015	45318	22374	[illegible]	78253	14385	[illegible]
00017	74350	99817	77402	77214	43236	00210	[illegible]	64237	[illegible]	[illegible]
00018	69916	26803	66252	29148	36936	[illegible]	76621	[illegible]	[illegible]	56418
00019	09893	20505	14225	68514	46427	56788	[illegible]	78822	[illegible]	[illegible]
00020	91499	14523	68479	27686	46162	83554	94750	89923	[illegible]	20048
00021	80336	94598	26940	36858	[illegible]	34135	53140	[illegible]	[illegible]	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	[illegible]
00023	12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
00024	63606	49329	16505	34484	40219	52563	43651	77082	[illegible]	[illegible]
00025	61196	90446	26457	47774	51924	33729	65394	59593	[illegible]	60527
00026	15474	45266	95270	79953	59367	[illegible]	82396	10118	33211	59466
00027	94557	28573	67897	54387	54622	44431	91190	42592	92927	45973
00028	42481	16213	97344	08721	16868	48767	[illegible]	12059	25701	[illegible]
00029	23523	78317	73208	89837	68935	91416	[illegible]	29663	05522	82562
00030	04493	52494	75246	[illegible]	45862	51025	[illegible]	79335	65337	[illegible]
00031	00549	[illegible]	64051	88159	96119	63896	54692	82391	23287	[illegible]
00032	35963	[illegible]	26898	09354	33351	35462	77974	50024	90103	[illegible]
00033	59808	08391	45427	26842	83609	49700	13021	24892	78565	[illegible]
00034	46058	85236	01390	92286	77281	44077	[illegible]	[illegible]	[illegible]	[illegible]
00035	32179	00597	87379	25241	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
00036	69234	61406	20117	45204	[illegible]	[illegible]	18743	92423	97118	96338
00037	[illegible]	[illegible]	01758	75379	[illegible]	[illegible]	66674	36806	[illegible]	[illegible]
00038	[illegible]	[illegible]	19476	07246	[illegible]	[illegible]	[illegible]	90033	[illegible]	[illegible]
00039	94864	31994	36168	10851	34888	[illegible]	[illegible]	35456	[illegible]	[illegible]
00040	98086	24826	45240	28404	44999	08896	[illegible]	73407	[illegible]	[illegible]
00041	33185	16232	41941	50949	89435	48581	88695	41994	37548	[illegible]
00042	80951	00406	96382	70774	[illegible]	23387	[illegible]	25298	94624	[illegible]
00043	[illegible]	49140	71961	28296	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
00044	[illegible]	[illegible]	98145	[illegible]	31010	[illegible]	[illegible]	61427	[illegible]	[illegible]
00045	74029	43902	77557	32270	97790	17119	[illegible]	58021	80814	[illegible]
00046	54178	[illegible]	80993	37143	[illegible]	12969	[illegible]	19255	[illegible]	[illegible]
00047	11664	49883	52079	84827	59381	71539	[illegible]	[illegible]	[illegible]	[illegible]
00048	48324	77928	31249	64710	02295	36870	32307	57546	[illegible]	[illegible]
00049	69074	94138	87637	91976	[illegible]	04401	10518	21615	[illegible]	[illegible]
00050	09188	20097	32825	[illegible]	[illegible]	86304	[illegible]	[illegible]	[illegible]	[illegible]
00051	90045	85497	51981	[illegible]	[illegible]	[illegible]	91870	76150	[illegible]	[illegible]
00052	73189	50207	[illegible]	[illegible]	[illegible]	[illegible]	27124	67018	[illegible]	[illegible]
00053	[illegible]	[illegible]	20971	87749	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
00054	54016	44056	66281	31003	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
00055	08358	69910	78542	42785	[illegible]	58873	[illegible]	[illegible]	[illegible]	[illegible]
00056	28306	03264	81333	[illegible]	40510	07893	[illegible]	[illegible]	94119	[illegible]
00057	53840	86233	81594	13628	51215	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
00058	[illegible]	53741	61613	62269	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
00059	89415	[illegible]	00397	58391	[illegible]	[illegible]	48949	72306	94541	[illegible]

TABLE V.

5 per cent (Roman) and 1 per cent (**Boldface**) critical values of F

df_2 \ df_1	1	2	3	4	5	6	7	8	10	14	20	40	100	∞
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.78	8.71	8.66	8.60	8.56	8.53
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.23**	**26.92**	**26.69**	**26.41**	**26.23**	**26.12**
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.87	5.80	5.71	5.66	5.63
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.54**	**14.24**	**14.02**	**13.74**	**13.57**	**13.46**
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.64	4.56	4.46	4.40	4.36
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.27**	**10.05**	**9.77**	**9.55**	**9.29**	**9.13**	**9.02**
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	3.96	3.87	3.77	3.71	3.67
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.87**	**7.60**	**7.39**	**7.14**	**6.99**	**6.88**
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.63	3.52	3.44	3.34	3.28	3.23
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.62**	**6.35**	**6.15**	**5.90**	**5.75**	**5.65**
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.34	3.23	3.15	3.05	2.98	2.93
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.82**	**5.56**	**5.36**	**5.11**	**4.96**	**4.86**
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.13	3.02	2.93	2.82	2.76	2.71
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.26**	**5.00**	**4.80**	**4.56**	**4.41**	**4.31**
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.97	2.86	2.77	2.67	2.59	2.54
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.21**	**5.06**	**4.85**	**4.60**	**4.41**	**4.17**	**4.01**	**3.91**
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.86	2.74	2.65	2.53	2.45	2.40
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.54**	**4.29**	**4.10**	**3.86**	**3.70**	**3.60**
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.76	2.64	2.54	2.42	2.35	2.30
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.30**	**4.05**	**3.86**	**3.61**	**3.46**	**3.36**
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.67	2.55	2.46	2.34	2.26	2.21
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.10**	**3.85**	**3.67**	**3.42**	**3.27**	**3.16**
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.60	2.48	2.39	2.27	2.19	2.13
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**3.94**	**3.70**	**3.51**	**3.26**	**3.11**	**3.00**
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.55	2.43	2.33	2.21	2.12	2.07
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.80**	**3.56**	**3.36**	**3.12**	**2.97**	**2.87**
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.37	2.28	2.16	2.07	2.01
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.69**	**3.45**	**3.25**	**3.01**	**2.86**	**2.75**
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.45	2.33	2.23	2.11	2.02	1.96
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.59**	**3.35**	**3.16**	**2.92**	**2.76**	**2.65**
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.29	2.19	2.07	1.98	1.92
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.85**	**3.71**	**3.51**	**3.27**	**3.07**	**2.83**	[illegible]	**2.57**

Table V is abridged from Table 10.5.3 of Snedecor, *Statistical Methods*, 5th edition 1956, Iowa State College Press, Ames, Iowa, by permission of the author and publisher

TABLE V—*Continued*

5 per cent (Roman) and 1 per cent (**Boldface**) critical values of F

df_2 \ df_1	1	2	3	4	5	6	7	8	10	14	20	40	100	∞
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.38	2.26	2.15	2.02	1.94	1.88
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.43**	**3.19**	**3.00**	**2.76**	**2.60**	**2.49**
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.35	2.23	2.12	1.99	1.90	1.84
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.71**	**3.56**	**3.37**	**3.13**	**2.94**	**2.69**	**2.53**	**2.42**
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.32	2.20	2.09	1.96	1.87	1.81
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.65**	**3.51**	**3.31**	**3.07**	**2.88**	**2.63**	**2.47**	**2.36**
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.30	2.18	2.07	1.93	1.84	1.78
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.26**	**3.02**	**2.83**	**2.58**	**2.42**	**2.31**
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.28	2.14	2.04	1.91	1.82	1.76
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.21**	**2.97**	**2.78**	**2.53**	**2.37**	**2.26**
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.26	2.13	2.02	1.89	1.80	1.73
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.17**	**2.93**	**2.74**	**2.49**	**2.33**	**2.21**
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.24	2.11	2.00	1.87	1.77	1.71
	7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.46**	**3.32**	**3.13**	**2.89**	**2.70**	**2.45**	**2.29**	**2.17**
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.22	2.10	1.99	1.85	1.76	1.69
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.09**	**2.86**	**2.66**	**2.41**	**2.25**	**2.13**
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.20	2.08	1.97	1.84	1.74	1.67
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.06**	**2.83**	**2.63**	**2.38**	**2.21**	**2.10**
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.19	2.06	1.96	1.81	1.72	1.65
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.03**	**2.80**	**2.60**	**2.35**	**2.18**	**2.06**
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.18	2.05	1.94	1.90	1.71	1.64
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.00**	**2.77**	**2.57**	**2.32**	**2.15**	**2.03**
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.16	2.04	1.93	1.79	1.69	1.62
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**2.98**	**2.74**	**2.55**	**2.29**	**2.13**	**2.01**
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.14	2.02	1.91	1.76	1.67	1.59
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**2.94**	**2.70**	**2.51**	**2.25**	**2.08**	**1.96**
34	[illegible]	3.28	[illegible]	[illegible]	2.49	2.38	2.30	2.23	2.12	2.00	1.89	1.74	1.64	1.57
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.89**	**2.66**	**2.47**	**2.21**	**2.04**	**1.91**
36	4.11	[illegible]	2.86	[illegible]	2.48	[illegible]	2.28	2.21	2.10	1.98	1.87	1.72	1.62	[illegible]
	7.39	**5.25**	**4.38**	[illegible]	**3.58**	[illegible]	**3.18**	**3.04**	**2.86**	**2.62**	**2.43**	**2.17**	**2.00**	**1.87**
38	[illegible]	3.25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	7.35	[illegible]	**4.34**	**3.86**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TABLE V—*Continued*

5 per cent (Roman) and 1 per cent (Boldface) critical values of F

df_2 \ df_1	1	2	3	4	5	6	7	8	10	14	20	40	100	∞
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.07	1.95	1.84	1.69	1.59	1.51
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.80**	**2.56**	**2.37**	**2.11**	**1.94**	**1.81**
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.06	1.94	1.82	1.68	1.57	1.49
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.77**	**2.54**	**2.35**	**2.08**	**1.91**	**1.78**
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.05	1.92	1.81	1.66	1.56	1.48
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.75**	**2.52**	**2.32**	**2.06**	**1.88**	**1.75**
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.04	1.91	1.80	1.65	1.54	1.46
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.73**	**2.50**	**2.30**	**2.04**	**1.86**	**1.72**
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.03	1.90	1.79	1.64	1.53	1.45
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.71**	**2.48**	**2.28**	**2.02**	**1.84**	**1.70**
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.02	1.90	1.78	1.63	1.52	1.44
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.70**	**2.46**	**2.26**	**2.00**	**1.82**	**1.68**
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.00	1.88	1.76	1.61	1.50	1.41
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.66**	**2.43**	**2.23**	**1.96**	**1.78**	**1.64**
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	1.99	1.86	1.75	1.59	1.48	1.39
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.63**	**2.40**	**2.20**	**1.93**	**1.74**	**1.60**
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	1.98	1.85	1.73	1.57	1.46	1.37
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.61**	**2.37**	**2.18**	**1.90**	**1.71**	**1.56**
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	1.97	1.84	1.72	1.56	1.45	1.35
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.59**	**2.35**	**2.15**	**1.88**	**1.69**	**1.53**
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.95	1.82	1.70	1.54	1.42	1.32
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.55**	**2.32**	**2.11**	**1.84**	**1.65**	**1.49**
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.92	1.79	1.68	1.51	1.39	1.28
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.51**	**2.26**	**2.06**	**1.79**	**1.59**	**1.43**
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.90	1.77	1.65	1.49	1.36	1.25
	6.84	**478**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.47**	**2.23**	**2.03**	**1.75**	**1.54**	**1.37**
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.89	1.76	1.64	1.47	1.34	1.22
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.44**	**2.20**	**2.00**	**1.72**	**1.51**	**1.33**
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.87	1.74	1.62	1.45	1.32	1.19
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.41**	**2.17**	**1.97**	**1.69**	**1.48**	**1.28**
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.85	1.72	1.60	1.42	1.28	1.13
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.37**	**2.12**	**1.92**	**1.64**	**1.42**	**1.19**
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.83	1.69	1.57	1.40	1.24	1.00
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.32**	**2.07**	**1.87**	**1.59**	**1.36**	**1.00**

TABLE J CRITICAL VALUES OF *T*

The symbol T denotes the smaller sum of ranks associated with differences that are all of the same sign. For any given N (number of ranked differences), the obtained T is significant at a given level if it is equal to or less than the value shown in the table.

N	.05	.025	.01	.005	N	.05	.025	.01	.005
	Level of significance for one-tailed test					Level of significance for one-tailed test			
	Level of significance for two-tailed test					Level of significance for two-tailed test			
	.10	.05	.02	.01		.10	.05	.02	.01
5	0	--	--	--	28	130	116	101	91
6	2	0	--	--	29	140	126	110	100
7	3	2	0	--	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	302	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

(Slight discrepancies will be found between the critical values appearing in the table above and in Table 2 of the 1964 revision of F. Wilcoxon, and R.A. Wilcox, *Some Rapid Approximate Statistical Procedures*, New York, Lederle Laboratories, 1964. The disparity reflects the latter's policy of selecting the critical value nearest a given significance level, occasionally overstepping that level. For example, for N = 8,

the probability of a T of 3 = 0.0390 (two-tail)

and

the probability of a T of 4 = 0.0546 (two-tail).

Wilcoxon and Wilcox select a T of 4 as the critical value at the 0.05 level of significance (two-tail), whereas Table J reflects a more conservative policy by setting a T of 3 as the critical value at this level.)

TABLE G CRITICAL VALUES OF r_{rho}

n*	.05	.025	.01	.005
	Level of significance for one-tailed test			
	Level of significance for two-tailed test			
	.10	.05	.02	.01
5	.900	1.000	1.000	--
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.683	.783	.833
10	.564	.648	.746	.794
12	.506	.591	.712	.777
14	.456	.544	.645	.715
16	.425	.506	.601	.665
18	.399	.475	.564	.625
20	.377	.450	.534	.591
22	.359	.428	.508	.562
24	.343	.409	.485	.537
26	.329	.392	.465	.515
28	.317	.377	.448	.496
30	.306	.364	.432	.478

*n = number of pairs

CONTINUED

1 OF 2

1. Introduction
   - Computing
   - Types of computing
   - UTC's experience

2. Terminals
   - Power on
   - Keyboard
   - Special Keys
   - CTRL, BREAK, RETURN, LINEFEED
   - CTRL H or ^H - deletes last character
   - CTRL X or ^X - deletes last line
   - Use of return key

3. Errors
   - CHATANN^c3^cH^cHTANNOOGA

     or

     CHATANN^cX

     CHATTANOOGA

4. Logging in
   - Sequence (Return) (Linefeed)

     PLEASE LOG IN

     REL-U109,CONTED (Return)

     News

     READY

5. EXE-$GAMES (Return)
   - ABAGEL
   - LANDER
   - BLACK
   - Note -- Press (Break) [illegible]

6. When finished
   - Type BYE - (Return)

Data

1970 Census Data per Capita Income and State Expenditures on Education for Nine States

State	Per Capita Income X	Public Education per Student Expenditure Y
	$	$
Arkansas	2520	534
California	4272	922
Colorado	3568	695
Michigan	3944	842
Mississippi	2194	476
North Carolina	2890	609
New York	4421	1237
Rhode Island	3779	904
South Dakota	3051	657

Problem

1. Compute correlation between X and Y.
2. Compute regression equation $Y = a + bY$
3. Using the file CRIMEB.A103

```
RDIM   for 193 observations on 6 variables.
ENTER  file CRIMEB.A103.
CORR   RACE and JAIL.
REGR   BNDAMT dependent variable
       against OFFENS independent variable.
COEF   to view coefficients.
```

A USERS GUIDE TO IDA -
INTERACTIVE DATA ANALYSIS

by
Dr. Leila Pratt
Assistant Professor
School of Business

Published by
UTC - Academic Computing
March 15, 1976

After you have successfully logged in you need to obtain the program named IDA (Interactive Data Analysis.) To do this you type "GET-" followed by "$IDA" and a carriage return. Therefore the first thing you would type after you have logged in is:

GET-$IDA

Now you have to be able to put IDA to work for you. To do this you would type RUN followed by return:

RUN

Now you are under the control of IDA and only IDA commands are valid. HP system commands will no longer be accepted until you get out of IDA.

SAMPLE SESSION - Entering Data Using IDA

The following is a sample session that introduces you to all you need to know in order to obtain information about IDA and to enter data using IDA.

IDA is a self-documenting program. That is, the program itself will provide you with all the instructions you need in order to use its various capabilities. Below you will find a brief description of some of the more important documentation you need in order to use IDA and the IDA commands you use to get it.

After you type
GET-$IDA
RUN

the program will answer you with a greeting either GOOD MORNING, GOOD AFTERNOON OR GOOD EVENING followed by NEW COMMAND CHANGES IN 'NEWS' and NEED HELP?

If you type Y or YES after this question, you will obtain much of the basic information you need to run IDA. Included in this is the following:

1. That there are three prompt levels to IDA. Level 1 prompts are very detailed, level 2 prompts are less detailed, and level 3 prompts are at times cryptic. Prompt level 1 is in effect unless you change it by issuing the IDA command CHGP. The first time you run the program I strongly suggest that you answer YES to the NEED HELP? question.

2. That there are two IDA commands that you may use to obtain information about all other IDA commands. To get a list of IDA commands by function you can type COMM. To get an explanation of all the commands in a group, such as Data Definition, you would type INFO. I strongly urge you to make extensive use of INFO since it can provide you with a short concise definition of each command or only one command in a group as well as general comments about IDA.

3. That you can enter data into IDA by using the IDA command ENTER. Normally you can enter a maximum of 100 rows (observations) and a maximum of 19 columns (variables) of data in an IDA data matrix. But this can be changed by using the command RDIM to redimension to more rows (a maximum of 563) but only at the expense of fewer columns (a minimum of 1.)

4. That the symbol ">" is IDA's command readiness symbol.

5. That you can get out of IDA and back under the control of the HP system by issuing the IDA command QUIT.

Now let's look at how we would enter data using IDA. I will underline what I typed into the terminal.

HEL-F999,XXX

CRE-DUTT1,1

Here I used the HP system command "CRE-" to open a file that I named "DUTT1." This file will be used later to store data. The 1 after the comma means that the file will be one record in length, that is, it will store up to 128 numbers (not just digits.)

GET-$IDA
RUN

Now you are under IDA control and can no longer use HP system commands.

DEC 17 1975 10:00 P.M.
GOOD EVENING
NEW COMMAND CHANGES IN 'NEWS'
NEED HELP? N

I answered no here since I am familiar with the program. Once you have answered yes to this question and either studied or kept the response, there is no need to answer yes again. Notice that Y and N are satisfactory abbreviations for YES and NO respectively.

> ENTER

This is the command used to enter data using IDA.

WANT EXPLANATION ? N

* MODE OF INPUT : FROM 'FILE' OR 'TERMINAL'? T

This indicates that I am going to enter the data from the terminal.

* SAMPLE SIZE (N) = 20

Here I tell the program that I have twenty rows of data.

* NO. OF VARIABLES (K) = 2

Here I tell the program that I have two columns of data.

ENTER ELEMENTS OF MATRIX BY ROWS BELOW:

```
ROW 1  :  ?   230.5,249.6
ROW 2  :  ?   232.8,255.7
ROW 3  :  ?   239.4
```

Note the comma between each entry only two numbers per row. Always hit carriage return after each row.

```
?? 263.3
```

Here I hit the carriage return before I input the second variable. IDA came back with "??" to ask me for the second number which I typed in.

```
ROW 4  :  ?   250.8,275.4
ROW 5  :  ?   255.7,728.3
ROW 6  :  ?   274.2,296.7
ROW 7  :  ?   281.4,309.3
ROW 8  :  ?   288.2,315.8
ROW 9  :  ?   290.1,318.8
ROW 10 :  ?   370.3,333.0
ROW 11 :  ?   316.1,340.2
ROW 12 :  ?   322.5,350.7
ROW 13 :  ?   338.4,367.3
ROW 14 :  ?   353.3,381.3
ROW 15 :  ?   373.7,407.9
ROW 16 :  ?   397.7,435.0
ROW 17 :  ?   418.1,458.9
ROW 18 :  ?   430.3,477.7
ROW 19 :  ?   452.6,497.6
ROW 20 :  ?   466.0,509.4
```

Here I typed 728.3 instead of 278.3 but I hit carriage return before I noticed it. Therefore I will have to change it later.

DATA MATRIX NOW DEFINED FOR 20
ROWS (OBSERVATIONS) BY 2 COLUMNS
(VARIABLES)
COMPUTING MEANS   STD. DEVS. (S)...

> CHGO IDA66.A101 IDA66.A101 WANT EXPLANATION ? N	*This command allows me to change the mistyped data item from 728.3 to 278.3.*

GIVE ROW NUMBER I, COLUMN NUMBER J,
AND THE NEW DATA VALUE X(I,J) :
I, J, X(I,J) = ? 5,2,278.3

UPDATING MEAN(S), STD.DEV.(S)...

> PRTS

Here I have the data matrix printed again to make sure I have corrected the error.

WANT EXPLANATION ? N

TO PRINT SUBMATRIX BETWEEN ROWS I1 AND I2 INCLUSIVE,
GIVE 2 ROW NUMBERS, SEPARATED BY A COMMA, FOR I1 AND I2, OR
FIRST ROW, LAST ROW = ? 1,5

I ask it only to print the first five rows since row five is where the error occurred.

* DO VARIABLES TO BE PRINTED OCCUPY
CONSECUTIVE COLUMNS OF DATA MATRIX ? Y
GIVE 2 COL. NUMBERS, SEPARATED BY A COMMA, FOR:

FIRST COL., LAST COL. = ? 1,2

ROW	CONS	DI
* 1 *	230.50000	249.60000
* 2 *	232.79999	255.70001
* 3 *	239.39999	263.29999
* 4 *	250.79999	275.40002
* 5 *	255.70001	278.29999

> FSAV IDA19.A101 IDA10.A101 WANT EXPLANATION ? N	*This command asks IDA to save the data matrix I have just entered.*

* NAME OF OUTPUT FILE TO SAVE DATA MATRIX IS DUTT1
WANT ENTIRE ACTIVE DATA MATRIX SAVED ? Y

Note I saved the data in the file that I opened at the beginning of the session before I went into IDA.

THE FOLLOWING ARE NOW SAVED IN 'DUTT1' :
20 , 2 , AND THE 20 BY 2 DATA MATRIX
NAMES:
CONS
DI
YOU MAY ADD EXTRA TEXT, ONE LINE AT A TIME, TO DESCRIBE 'DUTT1'.

YOU MAY GIVE EACH VARIABLE A 1 TO 6 CHARACTER NAME.
YOU ARE STRONGLY URGED TO SUPPLY NAMES FOR YOUR VARIABLES.
IF YOU DO, SEQUENCE PLOTS, TABLES, ETC. WILL BE LABELED.

WANT TO SUPPLY NAME(S) ? Y
PLEASE GIVE NAME OF:

VAR. 1 ? CONS
VAR. 2 ? DI

> PRTS
IDA25.A101
IDA25.A101
WANT EXPLANATION ? N

This causes the matrix of data I have just input using the terminal to be printed out.

TO PRINT SUBMATRIX BETWEEN ROWS I1 and I2 INCLUSIVE, GIVE 2 ROW NUMBERS, SEPARATED BY A COMMA, FOR I1 AND I2, OR FIRST ROW, LAST ROW = ?1,20

* DO VARIABLES TO BE PRINTED OCCUPY
CONSECUTIVE COLUMNS OF DATA MATRIX ? YES
GIVE 2 COL. NUMBERS, SEPARATED BY A COMMA, FOR:

FIRST COL., LAST COL. = ? 1, 2

ROW	CONS	DI
* 1 *	230.50000	249.60000
* 2 *	232.79999	255.70001
* 3 *	239.39999	263.29999
* 4 *	250.79999	275.40002
* 5 *	255.70001	728.30004
* 6 *	274.20001	296.70001
* 7 *	281.40002	309.29998
* 8 *	288.20001	315.79999
* 9 *	290.09997	318.79999
* 10 *	307.29999	333.00000
* 11 *	316.09997	340.20001
* 12 *	322.50000	350.70001
* 13 *	338.40002	367.29998
* 14 *	353.29999	381.29999
* 15 *	373.70001	407.90002
* 16 *	397.70001	435.00000
* 17 *	418.09997	458.90002
* 18 *	430.29998	477.70001
* 19 *	452.59997	497.59997
* 20 *	466.00000	509.40002

> PAUS

This command tells IDA to pause for a while to give me a chance to examine the data printed above. I note that I have mistyped the second variable in the fifth row.

```
WANT TO ADD SOME TEXT ? Y

WHEN DONE, USE 'CARRIAGE RETURN' KEY ONLY.

INPUT LINE OF LESS THAN 73  CHARACTERS.
DATA ON PERSONAL CONSUMPTION AND DISPOSABLE INCOME FOR

INPUT LINE OF LESS THAN 73  CHARACTERS.
1950-1969 FROM DUTTA ECONOMETRIC METHODS PROBLEM 3 CHAPTER 2 PAGE 68

INPUT LINE OF LESS THAN 73  CHARACTERS.

TEXT SAVED AT END OF 'DUTT1'.
```

Now the data are in a file which is in my own library. The data will stay in that file until I either save other data in DUTT1 or I kill the file.

```
> FILE
IDA27.A101
IDA27.A101
NAME OF FILE TO PRINT IS ? DUTT1
FIRST ELEMENT OF 'DUTT1' IS 20
SECOND ELEMENT IN 'DUTT1' IS 2
'DUTT1' PROBABLY HAS DATA MATRIX OF 20   ROWS AND 2   COLUMNS.
HOW MANY ROWS DO YOU WANT TO PRINT ? 10

ROW  1 :    230.5      249.6
ROW  2 :    232.8      255.7
ROW  3 :    239.4      263.3
ROW  4 :    250.8      275.4
ROW  5 :    255.7      278.3
ROW  6 :    274.2      296.7
ROW  7 :    281.4      309.3
ROW  8 :    288.2      315.8
ROW  9 :    290.1      318.8
ROW 10 :    307.3      333

*  WANT TO CHECK NAMES OR TEXT IN 'DUTT1' ? Y

NAMES:
CONS
DI
TEXT:
DATA ON PERSONAL CONSUMPTION AND DISPOSABLE INCOME FOR
1950-1969 FROM DUTTA ECONOMETRIC METHODS PROBLEM 3 CHAPTER 2 PAGE 68
END OF FILE CONTENTS.
```

Now let's look at how you enter data from a file instead of a terminal.

```
> ENTER
IDA23.A101
IDA23.A101
WANT EXPLANATION ? N
```

```
FURTHER USE OF 'ENTE' WILL ERASE
DATA NOW IN FIRST 20  ROWS, 2     COLUMNS.
WANT TO CONTINUE WITH 'ENTE' ? Y

*  MODE OF INPUT   :  FROM 'FILE' OR 'TERMINAL' ? F
```

Here I tell the computer that I want to enter data from a file.

```
NAME OF INPUT FILE IS ? DUTT1

*  ARE THE FIRST TWO ELEMENTS OF YOUR DATA FILE
   VALUES FOR N AND K (SIZE OF DATA MATRIX)? Y
DATA MATRIX NOW DEFINED FOR 20 ROWS (OBSERVATIONS)
BY 2 COLUMNS (VARIABLES)
COMPUTING MEANS & STD.DEVS.(S)...
```

NOTE: If the file was created by IDA you must answer yes to this question.

```
PROVISIONAL NAME OF:
VAR.     1  IS CONS
VAR.     2  IS DI

WANT TO SUPPLY NEW NAME(S) ? N

> PRTS
IDA25.A101
IDA25.A101
WANT EXPLANATION ? N

TO PRINT SUBMATRIX BETWEEN ROWS I1 AND I2 INCLUSIVE,
GIVE 2 ROW NUMBERS, SEPARATED BY A COMMA, FOR I1 AND I2, OR
FIRST ROW, LAST ROW = ? 18,20
```

Here I ask only for the last three rows of the data matrix.

```
*  DO VARIABLES TO BE PRINTED OCCUPY
   CONSECUTIVE COLUMNS OF DATA MATRIX ? Y
GIVE 2 COL. NUMBERS, SEPARATED BY A COMMA, FOR:

FIRST COL., LAST COL. = ? 1,2

   ROW          CONS          DI

*  18   *     430.29998    477.70001
*  19   *     452.59997    497.59997
*  20   *     466.00000    509.40002

> QUIT
DONE
```

The command QUIT gets me out of IDA and back under HP control. In order to log off, I must now type the HP command

BYE.

USING IDA TO OBTAIN SUMMARY STATISTICS

IDA has four commands that will allow you to obtain summary statistics on your data before you do any transformations or analysis. These four commands will allow you to obtain the mean and standard deviation of the variables in your data set, the correlation matrix, the partial correlation matrix between any variables you choose and the variance-covariance matrix. Below you will find an actual sample session using these IDA commands and the actual output you will obtain.

```
> ENTER
IDA23.A101
IDA23.A101
WANT EXPLANATION? N

*MODE OF INPUT:  FROM 'FILE' OR 'TERMINAL'? F

NAME OF INPUT FILE IS? DUTT2
```

Here I entered data from a file I had created earlier named DUTT2.

```
*  ARE THE FIRST TWO ELEMENTS OF YOUR DATA FILE
   VALUES FOR N AND K (SIZE OF DATA MATRIX)? Y
DATA MATRIX NOW DEFINED FOR 21  ROWS (OBSERVATIONS) BY 3
COLUMNS (VARIABLES)
COMPUTING MEANS & STD. DEVS.(S)...

PROVISIONAL NAME OF:
VAR.  1  IS  Y
VAR.  2  IS X1
VAR.  3  IS X2

WANT TO SUPPLY NEW NAME(S)? N

> MEAN
IDA27.A101
IDA27.A101

VARIABLE          MEAN            STD. D.V.

   Y         5.91714E+01      8.80523E+00
  X1         6.52000E+01      1.11618E+01
  X2         5.17333E+01      8.93769 E+00

BASED ON 21  ACTIVE ROWS.
```

This command will give you the mean and the standard deviation of every variable contained in the data file entered above.

```
> CORR
IDA28.A101
IDA28.A101

UPDATING CORR. MATRIX...IDA29.A101       500
```

This command will give you the correlation matrix for all the variables in the data file entered above or for any combination of the variables entered above.

HOW MANY VARIABLES? 3
GIVE COLUMN NUMBERS
OF VARIABLES, SEPARATED BY COMMAS: ? 1, 2, 3
* NUMBER OF DECIMAL PLACES WANTED IN THE PRINTOUT? 5

I chose to use all three variables.

I also chose to have five decimal places in my output.

	Y	X1	X2
Y	1.00000		
X1	0.99213	1.00000	
X2	0.70273	0.71764	1.00000

> PARC
IDA36.A101
IDA36.A101
LET N1 = NO. OF VARIABLES IN PARTIAL CORR. MATRIX,
AND N2 = NO. OF GIVEN VARIABLES FOR EACH PARTIAL.
GIVE VALUES N1, N2, SEPARATED BY A COMMA:
? 2,1
COL. NOS. OF FIRST 2 VARIABLES:
? 1,2
COL. NOS. OF SECOND 1 VARIABLES:
? 3
NUMBER OF DECIMAL PLACES
WANTED IN THE PRINTOUT? 5

This command allows you to obtain the partial correlation matrix holding one variable or a set of variables constant. In the following I obtain the partial correlation between Y and X1 holding X2 constant, Y and X2 holding X1 constant, and X1 and X2 holding Y constant. In all cases I requested five (5) decimal places in my printout.

PARTIAL CORRELATION MATRIX GIVEN VARIABLE(S):
X2

Y	1.00000	
X1	0.98457	1.00000

> PARC
IDA36.A101
IDA36.A101
LET N1 = NO. OF VARIABLES IN PARTIAL CORR. MATRIX,
AND N2 = NO. OF GIVEN VARIABLES FOR EACH PARTIAL.
GIVE VALUES N1, N2, SEPARATED BY A COMMA:
? 2,1
COL. NOS. OF FIRST 2 VARIABLES:
? 1,3
COL. NOS. OF SECOND 1 VARIABLES:
? 2
NUMBER OF DECIMAL PLACES
WANTED IN THE PRINTOUT? 5

PARTIAL CORRELATION MATRIX GIVEN VARIABLE(S):
X1

Y	1.00000	
X2	-0.10624	1.00000

> PARC
IDA36.A101
IDA36.A101
LET N1 - NO. OF VARIABLES IN PARTIAL CORR. MATRIX,

```
AND N2 = NO. OF GIVEN VARIABLES FOR EACH PARTIAL.
GIVE VALUES N1, N2, SEPARATED BY A COMMA:
? 2,1
COL. NOS. OF FIRST 2 VARIABLES:
? 2,3
COL. NOS. OF SECOND 1 VARIABLES:
? 1
NUMBER OF DECIMAL PLACES
WANTED IN THE PRINTOUT? 5

PARTIAL CORRELATION MATRIX GIVEN VARIABLE(S):
Y

X1        1.00000
X2        0.22947    1.00000

> COVA
IDA28.A101
IDA28.A101

HOW MANY VARIABLES? 3
GIVE COLUMN NUMBERS
OF VARIABLES, SEPARATED BY COMMAS: ? 1,2,3

          Y              X1             X2
Y        7.75322E+01
X1       9.75085E+01    1.24585E+02
X2       5.53040E+01    7.15925E+01    7.98823E+01
```

This command allows you to obtain the variance - covariance matrix of your data or of a subset of your data. Again I chose to use all three variables in my data set.

In addition to the above you may use IDA to obtain frequency distributions, histograms and normal cumulative probability plots of your data. As an example, I will obtain each of these for the variable Y in my data set above.

```
> FREQ
IDA39.A101
IDA39.A101
*  GIVE NAME OR COLUMN NUMBER FOR THE VARIABLE TO BE TABULATED: Y

MIN. OBS. = 42.4          MAX. OBS. = 74.6
MEAN = 59.1714         STD. DEV. = 8.80523
SAMPLE SIZE = 21

*  GIVE MIDPOINT OF A CENTRAL INTERVAL OF TABLE: 60
*  GIVE WIDTH OF EACH CLASS INTERVAL OF TABLE: 4

FREQUENCY DISTRIBUTION OF 'Y

  CLASS                                   CUMULATIVE
MIDPOINT           PERCENTAGE             PERCENTAGE

44.0E+00              9.5                    9.5
48.0E+00              4.8                   14.3
```

This command allows you to obtain a frequency distribution of the variable.

You are asked to supply both the midpoint of the central interval and the width of each class interval.

52.0E+00	14.3	28.6
56.0E+00	9.5	38.1
60.0E+00	19.0	57.1
64.0E+00	19.0	76.2
68.0E+00	14.3	90.5
72.0E+00	4.8	95.2
76.0E+00	4.8	100.0
TOTAL	100.0	
(NUMBER)	(21)	

```
> HIST
IDA39.A101
IDA39.A101
```

This command will give you a histogram plot of your data.

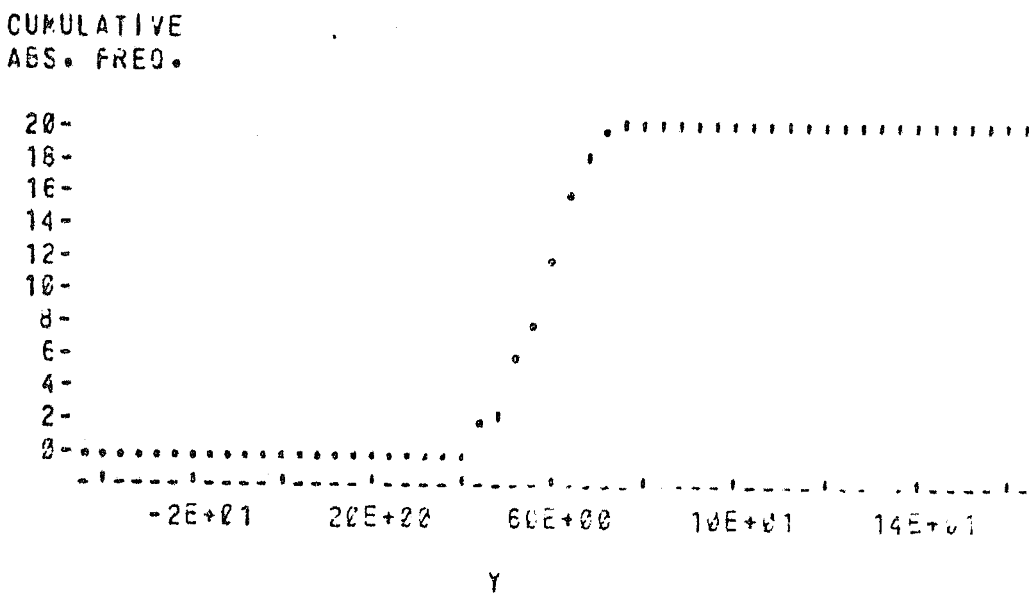
```
*  GIVE NAME OR COLUMN NUMBER FOR THE VARIABLE TO BE PLOTTED : Y

MIN. OBS. = 42.4           MAX. OBS. = 74.6
MEAN = 59.1714          STD. DEV. = 8.80523
SAMPLE SIZE = 21

*  GIVE MIDPOINT OF A CENTRAL INTERVAL OF HISTOGRAM: 60
*  GIVE WIDTH OF EACH CLASS INTERVAL OF HISTOGRAM: 4
```

Again you are asked to supply the midpoint of the central interval and the width of the class interval.

You may also obtain a plot of the cumulative probability distribution if you desire one.

```
HISTOGRAM

ABS. FREQ.

   -                             **
   -                           *  ***
   -                         *  *****
   -                         *********
 -!----!----!----!----!----!----!----!----!----!----!-
   -200E-01  2000E-02  6000E-02  1000E-01  1400E-01

                             Y

MEAN = 59.1714
STD. DEV. = 8.80523
SAMPLE SIZE = 21
```

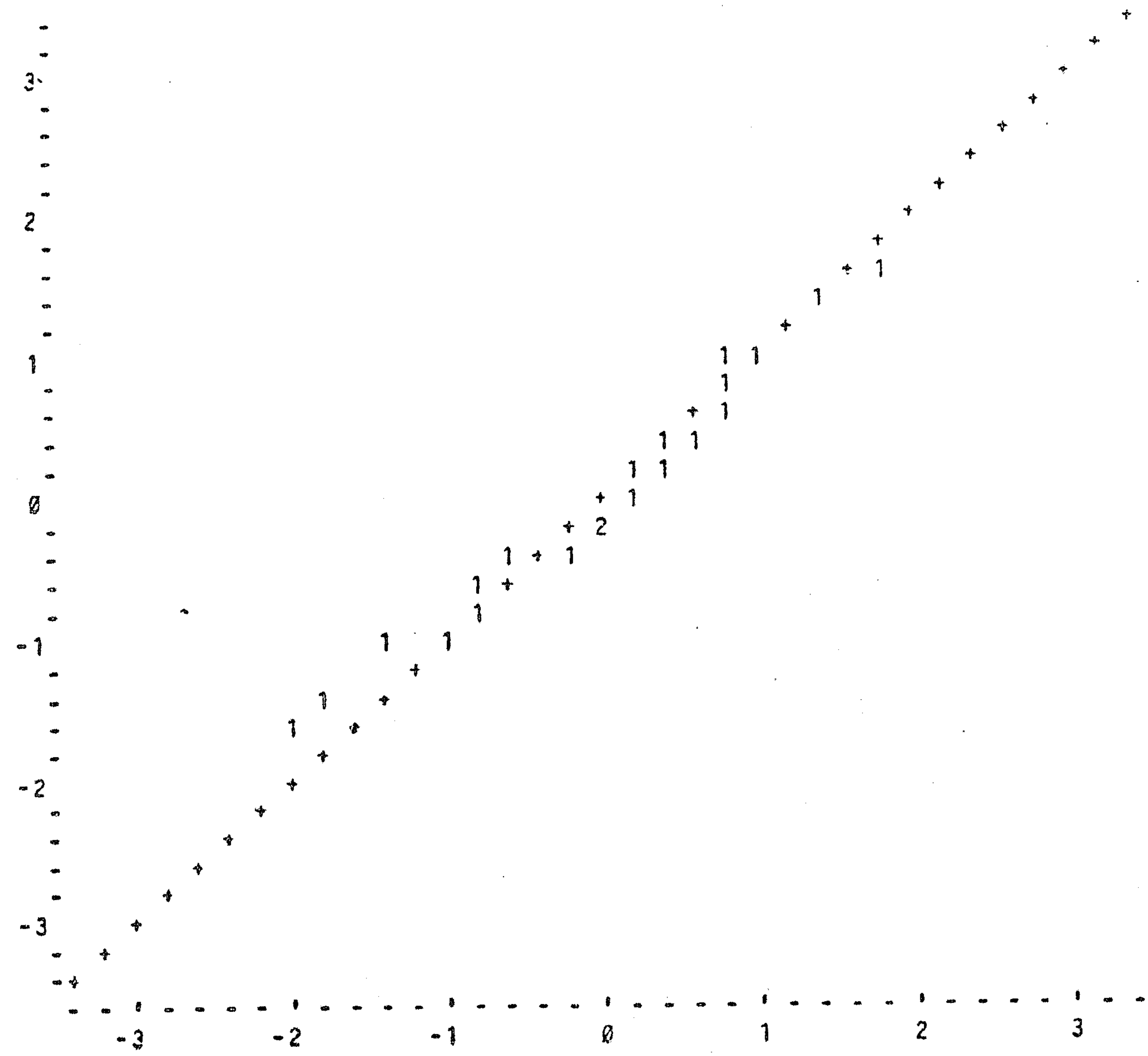
WANT CDF ALSO?Y



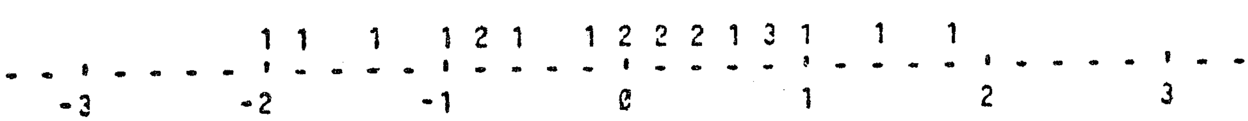
NORM
IDA39.A101
IDA39.A101
* GIVE NAME OR COLUMN NUMBER FOR THE VARIABLE TO BE PLOTTED: Y

This command obtains the normal cumulative probability plot of the variable.

NORMAL CUMULATIVE PROBABILITY PLOT OF Y



FREQUENCY DISTRIBUTION



MEAN = 59.1714
STD.DEV. = 9.80523
SKEWNESS = -.316854
KURTOSIS = -.8795
STUDENTIZED RANGE = 3.65692
SAMPLE SIZE = 21

DATA TRANSFORMATIONS USING IDA

There are several commands that will allow you to transform your data before you analyze it. The following are just a few of the commands that I have found most useful.

```
>TRAN
IDA24.A101
IDA24.A101
WANT EXPLANATION:N
```

*This command allows you to transform several variables using the same or different transformations. In other words you could do all your transformations by issuing this one command. If you want a list of the possible transformations you may answer "?" to the question "*TRANS.?"*

```
WHEN DONE, USE CARRIAGE RETURN ONLY AFTER '* TRANS. ? '.

*  TRANS.  ?  SUBV
*  IN WHICH COLUMN WILL THE TRANSFORMED
VARIABLE BE PLACED?  GIVE NUMBER : 3
COLUMN 3 = (COLUMN I) - (COLUMN J)
GIVE VALUES I, J, SEPARATED BY A COMMA : ? 2,1
NAME FOR COL. 3   ?C
```

This command allows me to subtract one column (variable) of data from another column. You supply the column to have the new variable placed in as well as the two columns to be subtracted. I chose to put the transformation in a new column but you could have replaced an old column with the transformed data. You also are allowed to name the new variable.

```
*  TRANS. ? ABSO
*  IN WHICH COLUMN WILL THE TRANSFORMED
VARIABLE BE PLACED?  GIVE NUMBER : 4
*  COLUMN TO BE TRANSFORMED :3
NAME FOR COL. 4   ?D
```

This command takes the absolute value of a column of data.

```
*  TRANS.  ?  POWE
*  IN WHICH COLUMN WILL THE TRANSFORMED
VARIABLE BE PLACED:  GIVE NUMBER : 5
*  COLUMN TO BE TRANSFORMED :4
*  VALUE OF POWER = 2
NAME FOR COL. 5   ?E
```

This command raises the variable to a power. Here I squared the variable. If I would have responded -1 to the question "VALUE OF POWER=", I would have taken the reciprocal, if I had responded .5 I would have taken the square root.

```
*  TRANS.  ?

UPDATING MEAN(S), STD. DEV.(S)...

CURRENT NAME OF:
VAR.  3  IS  C
VAR.  4  IS  D
VAR.  5  IS  E

WANT TO CHANGE ANY NAME(S) ?N
```

```
> ADDV
IDA24.A101
IDA24.A101
WANT EXPLANATION ?N
```

This command allows you to add to columns of data, specify the column to place them in, and give the new variable a name.

```
*  IN WHICH COLUMN WILL THE TRANSFORMED
VARIABLE BE PLACED:  GIVE NUMBER : 6
COLUMN 6 = (COLUMN I) = (COLUMN J)
GIVE VALUES I,J, SEPARATED BY A COMMA : ?1,2

UPDATING MEAN(S), STD. DEV.(S)...
PLEASE GIVE NAME(S) OF NEW VARIABLE(S):

VAR.  6 ?F

> DIVI
IDA24.A101
IDA24.A101
WANT EXPLANATION ?N
```

This command allows you to divide one column of data by another, specify the column to put the new data in and give the new variable a name.

```
*  IN WHICH COLUMN WILL THE TRANSFORMED
VARIABLE BE PLACED:  GIVE NUMBER : 7
COLUMN 7 = (COLUMN I) / (COLUMN J)
GIVE VALUES I,J, SEPARATED BY A COMMA : ?1,2

UPDATING MEAN(S), STD. DEV.(S)...
PLEASE GIVE NAME(S) OF NEW VARIABLE(S):

VAR. 7 ?G

> LOG1
IDA24.A101
IDA24.A101
WANT EXPLANATION ?N
```

This takes the log to the base 10 of a column of data. If you wanted to take the natural log you would use the command LOGE.

```
*  IN WHICH COLUMN WILL THE TRANSFORMED
VARIABLE BE PLACED:  GIVE NUMBER : 8
*  COLUMN TO BE TRANSFORMED :1

UPDATING MEAN(S), STD. DEV.(S)...
PLEASE GIVE NAME(S) OF NEW VARIABLE(S):

VAR. 8 ?H

> STAN
IDA64.A101
IDA64.A101
WANT EXPLANATION ?N
```

This command standardizes the data. That is, the mean of the column is subtracted from each observation and that value (the deviation) is then divided by the standard deviation.

```
*  IN WHICH COLUMN WILL THE TRANSFORMED
VARIABLE BE PLACED:  GIVE NUMBER : 9
*  COLUMN TO BE TRANSFORMED :1
PLEASE GIVE NAME FOR COL. 9   ?I
```

```
> CATE
IDA17.A101
IDA17.A101
WANT EXPLANATION ?N
```

This command allows you to create a categorical value from values in a specified column. It is very useful in creating dummy variable to use in regression analysis.

```
* GIVE NUMBER OR NAME
FOR THE COLUMN TO BE USED FOR CREATING A CATEGORICAL VARIABLE : 1

* GIVE NUMBER OF DATA COL. TO PUT CATEGORICAL VARIABLE IN:10
THERE ARE 5  ACTIVE OBSERVATIONS IN COL. 1    (A     )  :
MIN. = 5     MAX. = 20     MEAN = 12     STD. DEV. = 5.70088
HOW MANY CATEGORIES DO YOU WANT ?2
```

Here I specified I wanted to use column one to create the categorical variable and that I want to place the created variable in column 10.

Here I specified I wanted two categories.

```
LOWER LIMIT OF FIRST CATEGORY IS 5
PLEASE GIVE, FOR CATEGORY :
  1     , CATEGORY'S VALUE, UPPER LIMIT OF CATEGORY : ?0,5
UPPER LIMIT OF CATEGORY 2  IS 20
PLEASE GIVE VALUE OF CATEGORY 2  ?1

UPDATING MEAN(S), STD. DEV.(S)...
PLEASE GIVE NAME(S) OF NEW VARIABLE(S):

VAR.  10 ?J
```

By giving the first category a value of 0 and setting the upper limit equal to the lower limit, I caused every value in column 1 that was a 5 to be set equal to 0 in column 10. By giving the value of 1 to the second category I caused every value that was not a 5 in column 1 to be a 1 in column 10. Thus I have created a dummy variable that takes on the value of 0 when A = 5 and takes on the value of 1 otherwise.

```
> PRTS
IDA25.A101
IDA25.A101
WANT EXPLANATION ?N
```

Here I simply had the original data and all the transformations printed.

```
TO PRINT SUBMATRIX BETWEEN ROW  11 AND 12 INCLUSIVE,
GIVE 2 ROW NUMBERS, SPEARATED BY A COMMAS, FOR 11 AND 12, OR
FIRST ROW, LAST ROW = ?1,5

* DO VARIABLES TO BE PRINTED OCCUPY
  CONSECUTIVE COLUMNS OF DATA MATRIX ? Y
GIVE 2 COL. NUMBERS, SEPARATED BY A COMMA, FOR:

FIRST COL., LAST COL. = ?1,10
```

ROW	A	B	C	D	E
* 1 *	20.00000	10.00000	-10.00000	10.00000	100.00000
* 2 *	15.00000	5.00000	-10.00000	10.00000	100.00000
* 3 *	10.00000	10.00000	0.00000	0.00000	0.00000
* 4 *	5.00000	3.00000	-2.00000	2.00000	4.00000
* 5 *	10.00000	2.00000	-8.00000	8.00000	64.00000

ROW					
* 1 *	30.00000	2.00000	1.30103	1.40329	1.00000
* 2 *	20.00000	3.00000	1.17609	0.52623	1.00000
* 3 *	20.00000	1.00000	1.00000	-0.35082	1.00000
* 4 *	8.00000	1.66667	0.69897	-1.22788	0.00000
* 5 *	12.00000	5.00000	1.00000	-0.35082	1.00000

Using IDA for Regression Analysis

IDA may be used to do several types of data analysis. The following is an example of how IDA may be used in performing regression analysis.

```
GET-$IDA
RUN
IDA

IDA01.A101
IDA01.A101
DEC 19 1975        9:12 A.M.

GOOD MORNING.
NEW COMMAND CHANGES IN 'NEWS'
NEED HELP?N

> ENTER
IDA23.A101
IDA23.A101
WANT EXPLANATION?N
```

In this case I used the command ENTER to enter data from a file named KEM2 that I had created earlier and saved.

```
* MODE OF INPUT:  FROM 'FILE' OR 'TERMINAL'?F

NAME OF INPUT FILE IS? KEM2

* ARE THE FIRST TWO ELEMENTS OF YOUR DATA FILE
  VALUES FOR N AND K (SIZE OF DATA MATRIX)? Y
DATA MATRIX NOW DEFINED FOR 12  ROWS (OBSERVATIONS) BY 3
COLUMNS (VARIABLES)
COMPUTING MEANS & SID.DEVS.(S)...

PROVISIONAL NAME OF:
VAR.  1 IS QUANT
VAR.  2 IS PRICE
VAR.  3 IS ADV

WANT TO SUPPLY NEW NAME(S)?N

> CHGP

* LEVEL = 3
```

Here I changed the command to prompt level 3 since I am now familiar with the IDA command structure. I suggest that the first time you run an IDA command you use the command level 1 to take advantage of the additional explanations available at this level. (Prompt level 1 is in effect unless changed.)

```
> FILE
  FILE NAME IS? KEM2
'KEM2' PROBABLY HAS DATA MATRIX OF 12 ROWS AND 3 COLUMNS.
HOW MANY ROWS DO YOU WANT TO PRINT?4

ROW 1:  55  100  5.5
ROW 2:  70   90  6.3
ROW 3:  90   80  7.2
ROW 4: 100   70  7.

*  CHECK NAMES. TEXT? Y

NAMES:
QUANT
PRICE
ADV
TEXT:
DATA FROM KMENTA PAGE 353
END OF FILE CONTENTS.
```

```
> REGR
IDA29.A101
IDA29.A101
*  DEP. VAR. = QUANT
HOW MANY INDEP. VAR.?2
INDEP. VAR.  1 = ?PRICE
INDEP. VAR.  2 = ?ADV
```

This command performs the actual regression calculations but does not give the results. The following commands must be used to get the results.

```
UPDATING CORR. MATRIX...COMPUTING REGRESSION...
ANALYZING RESIDUALS...CHECKING AUTO CORRELATIONS...

IDA.A101          150
```

```
> COEF
IDA34.A101
IDA34.A101
```

This gives the standard deviation of the B coefficients, the standard error of the B coefficient, and the t values.

VARIABLE	B(STD.V)	B	STD.ERROR(B)	T
PRICE	-0.7816	-.1308E+01	1.2937E-01	-10.110
ADV	0.3123	1.1246E+01	2.7844E+00	4.039
CONSTANT	0	1.1616E+02	2.4646E+01	4.713

```
> SUMM
IDA34.A101
IDA34.A101
```

This gives the adjusted and unadjusted R and R^2 and the standard deviation of the residuals.

	MULTIPLE R	R-SQUARE
UNADJUSTED	0.9801	0.9606
ADJUSTED	0.9756	0.9518

STD. DEV. OF RESIDUALS = 5.25454

> SEPR
IDA33.A101
IDA33.A101
USE DATA MATRIX?Y
HOW MANY Y'S?12

This command will print out the Y value and the standard error of the fitted and predicted Y values.

	Y	S.E.(FITTED)	S.E.PRED.(Y)
ROW:1	4.7222E+01	3.8939E+00	6.5401E+00
ROW:2	6.9297E+01	2.6933E+00	5.9046E+00
ROW:2 3	9.2497E+01	2.7795E+00	5.9444E+00
ROW:4	1.0333E+02	1.7261E+00	5.5308E+00
ROW:5	9.5455E+01	1.8887E+00	5.5837E+00
ROW:6	1.0726E+02	2.3526E+00	5.7571E+00
ROW:7	8.7583E+01	3.4282E+00	6.2740F+00
ROW:8	1.1155E+02	1.8522E+00	5.5714E+00
ROW:9	1.2203E+02	2.4337E+00	5.7908E+00
ROW:10	1.1528E+02	1.8823E+00	5.5815E+00
ROW:11	1.2463E+02	2.2625E+00	5.7209E+00
ROW:12	1.2386E+02	3.2921E+00	6.2007E+00

> ANOV
IDA34.A101
IDA34.A101

This command performs an analysis of variance on the regression.

SOURCE	SS	DF	MS	F
REGRESSION	6.05151E+03	2	3.02575E+03	109.59
RESIDUALS	2.48491E+02	9	2.76102E+01	
TOTAL	6.30000E+03	11	5.72727E+02	

> BCOV
IDA33.A101
IDA33.A101

This command gives the variance and covariance matrix for the B coefficients in the regression.

PRICE	1.67368E-02	
ADV	1.86068E-01	7.75285E+00
CONST.	-.24190E+01	-.65001E+02

6.07410E+02

> BCOR
IDA33.A101
IDA33.A101
DECIMALS = ?5

This command gives the correlation matrix for the B coefficients in the regression.

PRICE	1.00000	
ADV	0.51654	1.00000
CONST.	-0.75868	-0.94722

> FPRF
IDA25.A101
IDA25.A101

This command gives a formated print out of the fitted values of the regression, the residual values of the regression and one other variable if desired.

WANT FITTED VALUES ONLY?N
TO PRINT ONE COL. OF DATA PLUS
'FITTED' AND 'RESIDUAL' VALUES, GIVE
* COLUMN OF DATA MATRIX TO BE PRINTED:1

Here I chose to have the original value of Y printed as my additional variable.

* FMT = #,DDDD.DDD

Here I specified the format to be used in printing the data.

ROW			QUANT	FITTED	RESIDU
*	1	*	55.000	47.222	7.778
*	2	*	70.000	69.297	0.703
*	3	*	90.000	92.497	-2.497
*	4	*	100.000	103.327	-3.327
*	5	*	90.000	95.455	-5.455
*	6	*	105.000	107.263	-2.263
*	7	*	80.000	87.583	-7.583
*	8	*	110.000	111.553	-1.553
*	9	*	125.000	122.029	2.971
*	10	*	115.000	115.281	-0.281
*	11	*	130.000	124.632	5.368
*	12	*	130.000	123.861	6.139
MEAN			100.000	100.000	0.000
S.DEV.			23.932	23.455	5.255

The format must be preceded by #,. The D's following the "." indicate the number of decimal places you want in your print out. You may have a maximum of five.

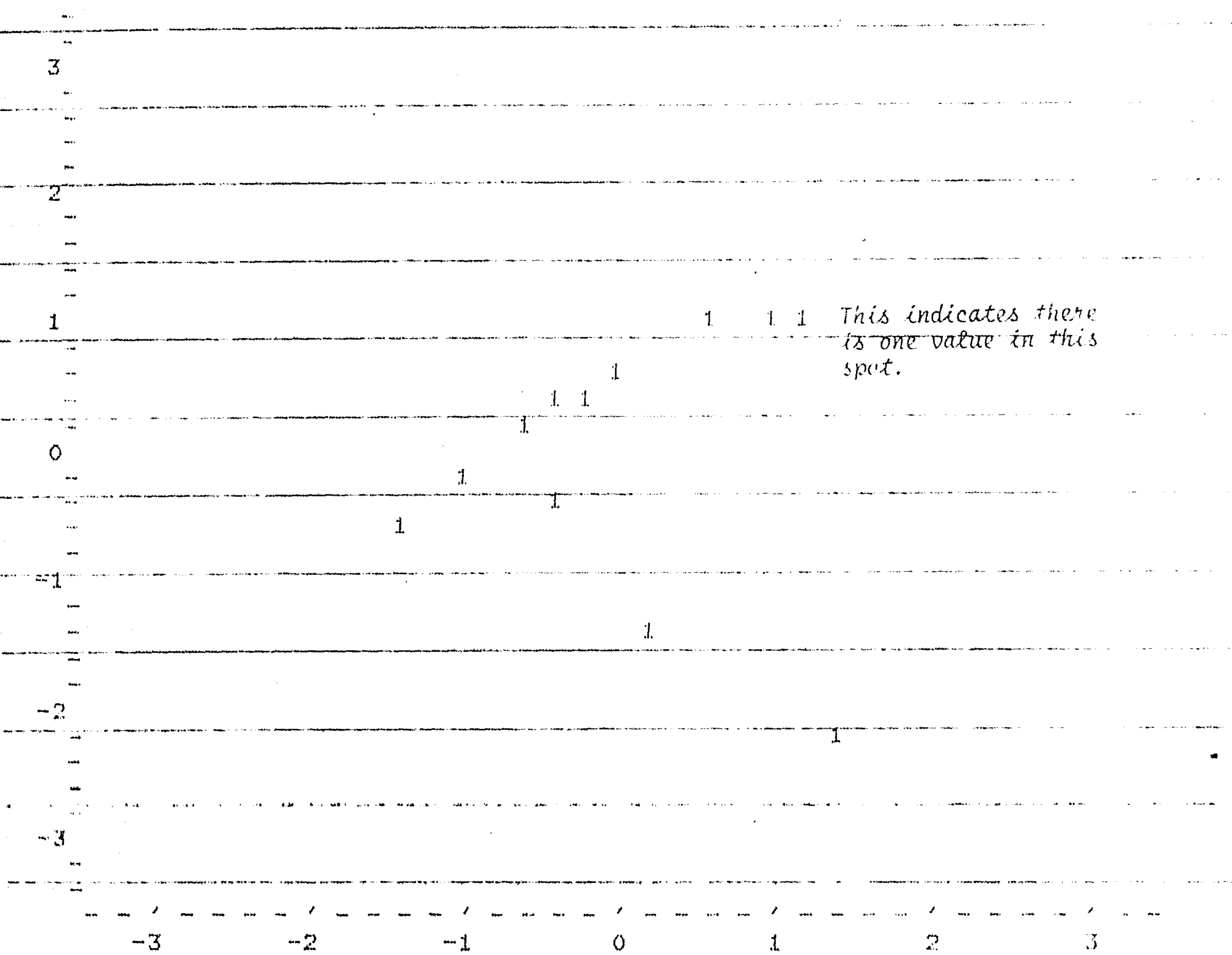
> SCATT
IDA02.A101
IDA02.A101
* VERT. VAR. : FITTED
* HORIZ. VAR. : RESIDU

This command allows me to obtain a plot of the standardized values of any two variables.

* ALL ROWS ? Y

Here I chose to plot the fitted values of Y and residual values of Y.

SCATTER PLOT OF 12 STANDARDIZED VALUES OF 'FITTED' VS. 'RESIDU'



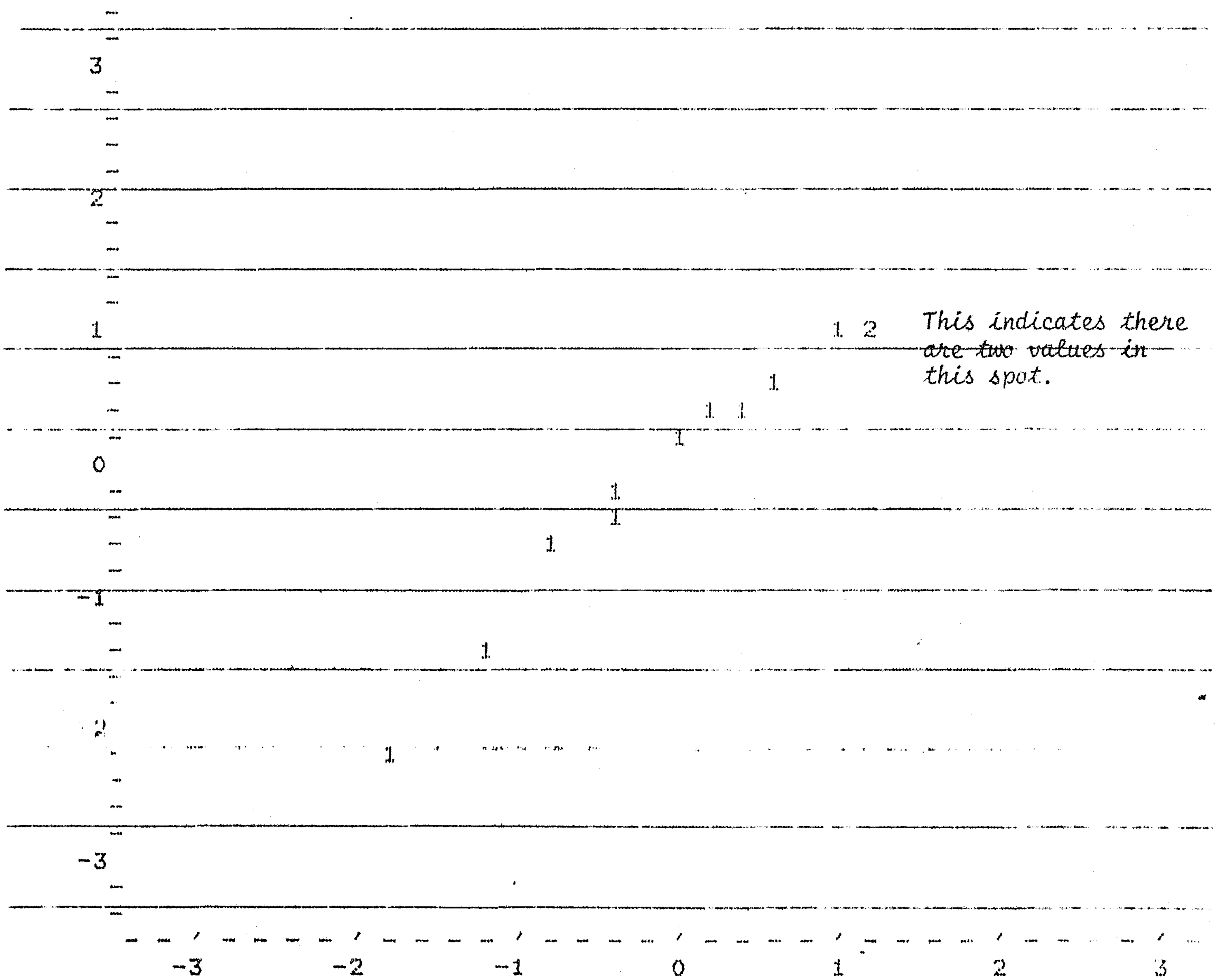
	MEAN	STD. DEV.
VERT. VAR.	100.	23.455
HORIZ. VAR.	0	5.25452

SAMPLE SIZE = 12

> SCATT
IDA02.A101
IDA02.A101
* VERT. VAR. : FITTED
* HORIZ. VAR. : QUANT
* ALL ROWS ? Y

Here I chose to plot the fitted Y values and the original Y values.

SCATTER PLOT OF 12 STANDARDIZED VALUES OF 'FITTED' VS. 'QUANT '



This indicates there ~~are two values in~~ this spot.

	MEAN	STD. DEV.
VERT. VAR.	100.	23.455
HORIZ. VAR.	100	23.9317

SAMPLE SIZE = 12

> QUIT

Here I get out of IDA and back under HP system control.

Using IDA to Perform Cross-Tabs

```
  ENTER
WANT EXPLANATION : N
*  MODE OF INPUT : FROM 'FILE' OR 'TERMINAL' ? F

NAME OF INPUT FILE IS ?  KEM2

*  ARE THE FIRST TWO ELEMENTS OF YOUR DATA FILE
   VALUES FOR N AND K (SIZE OF DATA MATRIX)? Y
DATA MATRIX NOW DEFINED FOR 12 ROWS (OBSERVATIONS) BY 3
COLUMNS (VARIABLES)
COMPUTING MEANS & STD. DEV.(S)...

PROVISIONAL NAME OF:
VAR.  1 IS QUANT
VAR.  2 IS PRICE
VAR.  3 IS ADV

WANT TO SUPPLY NEW NAME(S) ? N

  CTAB
WANT EXPLANATION ? N
*  GIVE NAME OR COLUMN NUMBER FOR THE VARIABLE TO BE TABULATED
ON THE VERTICAL AXIS : QUANT
*  GIVE NAME OR COLUMN NUMBER FOR THE VARIABLE TO BE TABULATED
ON THE HORIZONTAL AXIS : PRICE

NEED TO DEFINE CLASS INTERVALS FOR VERTICAL OR HORIZONTAL
VARIABLES? Y

VERTICAL VARIABLE:
MIN. OBS. = 55    MAX. OBS. = 130
MEAN = 100    STD. DEV. = 23.9317
SAMPLE SIZE = 12

*  GIVE MIDPOINT OF A CENTRAL INTERVAL OF VARIABLE:

CHOOSE A CONVENIENT NUMBER NEAR THE MEAN AND/OR ONE NOT
TOO FAR FROM HALFWAY BETWEEN THE MIN. OBS. AND MAX. OBS.

*  MIDPOINT:  100
*  GIVE WIDTH OF EACH CLASS INTERVAL OF VARIABLE:

USE YOUR COMMON SENSE, OR TRY A 'ROUND NUMBER' ABOUT HALF
THE STD. DEV.

*  WIDTH OF INTERVAL: 12
HORIZONTAL VARIABLE:
MIN. OBS. = 50    MAX. OBS. = 100
MEAN = 70    STD. DEV. = 14.3019
SAMPLE SIZE = 12
```

Here I am entering data from an existing file.

Here I am invoking CTAB.

I am naming row and column variables.

Here I define class intervals in case data. This is done when data has values in excess of 9 for AY or 21 for AX variable.

These are the prompts to determine class interval size for X variables.

These are the prompts to determine class interval size for Y variable.

* GIVE MIDPOINT OF A CENTRAL INTERVAL OF VARIABLE: 70
* GIVE WIDTH OF EACH CLASS INTERVAL OF VARIABLE: 7

JOINT DISTRIBUTION OF QUANT VERSUS PRICE

	PRICE							
QUANT	2	3	4	5	6	8	9	TOTAL
7	0	0	0	0	0	0	1	1
9	0	0	0	1	0	1	0	2
10	0	0	0	1	1	0	0	2
11	0	0	0	2	0	0	0	2
12	0	0	2	0	0	0	0	2
13	0	0	1	0	0	0	0	1
14	1	1	0	0	0	0	0	2
TOTAL	1	1	3	4	1	1	1	12

This is the produced CROSS TABS TABLE.

WANT PRINTOUT OF KEY FOR CATEGORY LABELS? Y

Here I answer the PROMPT FOR CATEGORY LABELS associated with the class intervals.

VERTICAL VARIABLE:

CATEGORY	MIDPOINT
7	5.20000E+01
9	7.60000E+01
10	8.80000E+01
11	1.00000E+02
12	1.12000E+02
13	1.24000E+02
14	1.36000E+02

This gives the coded X VARIABLE CLASS INTERVAL MIDPOINT.

HORIZONTAL VARIABLE:

CATEGORY	MIDPOINT
2	4.90000E+01
3	5.60000E+01
4	6.30000E+01
5	7.00000E+01
6	7.70000E+01
8	9.10000E+01
9	9.80000E+01

This gives the coded Y VARIABLE CLASS INTERVAL MIDPOINT.

THE EXTREME CATEGORIES ARE OPEN-ENDED

WANT CONDITIONAL DISTRIBUTION (IN PERCENTAGE FORM) OF THE VERTICAL VARIABLE GIVEN THE HORIZONTAL VARIABLE?Y

Here I answer the PROMPT FOR A CONDITIONAL DISTRIBUTION.

CONDITIONAL DISTRIBUTION OF QUANT GIVEN PRICE

				PRICE				
QUANT	2	3	4	5	6	8	9	TOTAL
7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	8.3
9	0.0	0.0	0.0	25.0	0.0	100.0	0.0	16.7
10	0.0	0.0	0.0	25.0	100.0	0.0	0.0	16.7
11	0.0	0.0	0.0	50.0	0.0	0.0	0.0	16.7
12	0.0	0.0	66.7	0.0	0.0	0.0	0.0	16.7
13	0.0	0.0	33.3	0.0	0.0	0.0	0.0	8.3
14	100.0	100.0	0.0	0.0	0.0	0.0	0.0	16.7
TOTAL	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
NUMBER	1	1	3	4	1	1	1	12

I receive the CONDITIONAL DISTRIBUTION TABLE.

INDEX TO IDA COMMANDS

DATA DEFINITION :

```
ENTE    TO ENTER DATA FROM FILE, TAPE, OR TERMINAL
ENTS    TO ENTER SELECTED DATA FROM AN 'IDA-SERIAL' DATA FILE
ENRA    TO ENTER A SAMPLE OF DATA FROM AN 'IDA-RANDOM' DATA FILE
EOBR    TO ENTER A SAMPLE OF DATA FROM AN 'IDA-RANDOM' DATA FILE
        USING A LIST OF OBSERVATION NUMBERS IN AN IDA
        DATA COLUMN
EOBS    TO ENTER SELECTED DATA FROM AN 'IDA-SERIAL' DATA FILE
        USING A LIST OF OBSERVATION NUMBERS IN AN IDA DATA COL.
```

NOTE: YOU MAY USE A MODIFIED VERSION OF '$END' AS AN IDA 'NEWC' TO ENTER DATA, USING YOUR OWN DATA IN DATA STATEMENTS.

```
INDX    TO CREATE AN INDEX VECTOR (SUCH AS 1,2,...,N)
        IN A COLUMN OF THE DATA MATRIX.  (USEFUL FOR KEEPING
        TRACK OF ORIGINAL ORDER IF 'MSOR' IS USED.
RAND    TO GENERATE RANDOM DATA FROM SOME MODEL AND ENTER
        IT IN ONE COLUMN OF THE DATA MATRIX
STAR    TO DEFINE AN INITIAL DATA MATRIX OF ZEROES.

CRSP    TO ENTER DATA, OR APPEND OR REPLACE A COLUMN WITH
        DATA, FROM ONE OF THE $CRSP FILES
DRID    TO ENTER DATA, OR APPEND OR REPLACE A COLUMN WITH
        DATA, FROM ONE OF THE $DRI FILES
COMP    TO ENTER DATA, OR APPEND OR REPLACE A COLUMN WITH
        DATA, FROM THE $COMP FILE
```

NOTE: SINCE 'DRID' AND 'COMP' ARE NOT YET IMPLEMENTED IN IDA, PLEASE USE '$DRI' AND '$COM' AS IDA 'NEWCS'.

```
FSAV    TO SAVE DATA MATRIX ON FILE TO BE USED LATER. (NOTE:
        FILE MUST HAVE BEEN OPENED ALREADY)
MOVF    TO MOVE THE FITTED VALUES FROM THE CURRENT REGRESSION
        TO A COLUMN OF THE DATA MATRIX
MOVR    TO MOVE THE RESIDUALS FROM THE CURRENT REGRESSION TO
        A COLUMN OF THE DATA MATRIX
```

DATA EDITING :

APPB TO APPEND A BLOCK OF ROWS TO THE DATA MATRIX FROM A DATA FILE OR DIRECTLY FROM THE TERMINAL

APPO TO APPEND AN OBSERVATION VECTOR TO THE DATA MATRIX. YOU MAY USE THIS TO ADD A ROW TO THE EXISTING DATA MATRIX OR TO CHANGE A ROW IN IT

APPS TO APPEND A SUBMATRIX TO OR CHANGE A SUBMATRIX IN THE DATA MATRIX OR THE SCRATCH PAD DIRECTLY FROM THE TERMINAL OR USING A DATA FILE.

ENTS TO APPEND DATA TO OR REPLACE DATA IN THE DATA MATRIX WITH SELECTED DATA FROM A DATA FILE. (FIRST TWO ELE-MENTS MUST BE THE NUMBER OF ROWS AND THE NUMBER OF COLUMNS IN THE FILE DATA MATRIX.)

APPV TO APPEND A VARIABLE (COLUMN) TO OR REPLACE A COLUMN IN THE DATA MATRIX

CHGO TO CHANGE THE VALUE OF A SINGLE ENTRY IN THE DATA MATRIX

DELB TO DELETE A BLOCK OF OBSERVATIONS FROM THE DATA MATRIX. YOU CAN RECOVER THE DELETED BLOCK BY THE COMMAND 'RETB' OR 'RECO'

DELO TO DELETE AN OBSERVATION VECTOR FROM THE DATA MATRIX. DELETED VECTOR CAN BE RETRIEVED BY 'RETO' OR 'RECO'

DELV TO DELETE A VARIABLE (COL.) VECTOR FROM THE DATA MATRIX. DELETED VECTOR CANNOT BE RETRIEVED. (USE 'FSAV' OR 'MOVE' INSTEAD OF 'DELV' IF YOU WISH TO USE THAT COLUMN'S DATA LATER.)

SELR TO DELETE ROWS FOR WHICH VALUES IN A SPECIFIED COL. ARE OUTSIDE A SPECIFIED RANGE OF VALUES.

RECO TO RECOUP ALL THE DELETED OBSERVATIONS

RETB TO RETRIEVE A BLOCK OF DELETED OBSERVATIONS

RETO TO RETRIEVE A DELETED ROW OF OBSERVATIONS

NOTE: YOU CANNOT RETRIEVE A ROW MADE INACTIVE BY THE USE OF 'DIFF' OR 'LAGO' UNLESS YOU REPLACE THE COLUMN WITH ANOTHER VARIABLE WITH FEWER UNDEFINED OBSERVATIONS OR YOU REMOVE THAT COLUMN BY USING 'DELV' OR 'MOVE'.

MOVE TO MOVE A DATA COL. TO THE SCRATCH PAD OR A SCRATCH COL. TO THE DATA MATRIX

MSOR SORTS ONE VARIABLE (COLUMN) IN ASCENDING ORDER AND ALL OTHER COLUMNS ACCOMPANY IT. RESULTS PLACED IN SAME COLUMNS

PSOR PAIRED SORT OF ONE VARIABLE (COLUMN) AND ACCOMPANYING VARIABLE (COLUMN) INTO TWO OTHER COLUMNS

SORT SORTS THE VALUES OF ONE VARIABLE (COLUMN) INTO ASCENDING ORDER AND PLACES RESULTS IN ANOTHER COLUMN

DATA DISPLAY (PRINT) :

FILE TO PRINT OUT ONE OR MORE ROWS OF A DATA MATRIX STORED TO A FILE. THIS ALLOWS YOU TO TAKE A LOOK AT THE DATA BEFORE DECIDING WHETHER IT'S THE MATRIX YOU WANT TO ENTER.

FPRF FORMATTED PRINT OF FITTED VALUES (IN REGRESSION)

FPRO FORMATTED PRINT OF AN OBSERVATION (VECTOR)

FPRR FORMATTED PRINT OF RESIDUALS (IN REGRESSION)

FPRS FORMATTED PRINT OF A SUBMATRIX OF DATA OR SCRATCH VALUES

FPRV FORMATTED PRINT OF A VARIABLE (COLUMN)

NOTE: IN THE ABOVE FIVE COMMANDS, THE USER WILL BE ASKED TO SUPPLY THE FORMAT FOR PRINTING. (A MODIFIED VERSION OF '$PRT' MAY ALSO BE USED AS AN IDA 'NEWC' FOR SPECIAL FORMATS.)

NAME TO LIST THE NAMES OF THE VARIABLES (IF YOU SUPPLIED THEM.) TO BE USED WHEN YOU HAVE FORGOTTEN WHICH VARIABLE IS IN WHICH COLUMN OF THE DATA MATRIX. THE COMMAND WILL CAUSE THE FIRST ACTIVE ROW OF THE DATA MATRIX TO BE PRINTED IF NO NAMES HAVE BEEN GIVEN TO THE VARIABLES.

PRTF PRINT FITTED VALUES

PRTO PRINT OBSERVATION

PRTR PRINT RESIDUALS

PRTS PRINT SUBMATRIX OF DATA OR SCRATCH VALUES

PRTV PRINT VARIABLE

THE COMMANDS BEGINNING WITH 'PRT' WILL AUTOMATICALLY GIVE VALUES IN THE FORM DDDDD.DDDDD, UP TO FIVE VALUES PER LINE. IF ANY OF YOUR DATA VALUES IS GREATER THAN 99999, YOU SHOULD USE THE CORRESPONDING 'FPR' COMMANDS, SUPPLYING THE FORMAT YOU CHOOSE. BECAUSE OF FLOATING POINT CONVERSION OF NUMBERS, YOU MAY GET GARBAGE FOR CERTAIN TRAILING DIGITS WHEN 'PRT' COMMANDS ARE USED. FOR EXAMPLE, THE NUMBER 12345 IS PRINTED AS 1234.99989 BECAUSE THE MACHINE DOES NOT CARRY AN EXACT REPRESENTATION OF 12345.

WHEN YOU GIVE A FORMAT FOR PRINT, THE SAME FORMAT MUST BE APPLIED TO ALL OF THE VARIABLES; THAT IS, YOU DO NOT HAVE THE OPTION OF SPECIFYING DIFFERENT FORMATS FOR DIFFERENT VARIABLES AS CAN BE DONE IN 'FORTRAN'. FOR EXAMPLE, IF A ROW OF DATA CONSISTS OF
1.2, 2.3456, 3500
THE 'FPR' COMMANDS WILL NOT ENABLE YOU TO PRINT IT AS

```
  1.2     2.3456    3500.
```

IF YOU USE THE FORMAT #,4D.4D,2X YOU WILL GET:

```
   1.2000      2.3456   3500.0000
```

WHICH IS NOT MUCH DIFFERENT FROM THE FORMAT YOU WOULD HAVE OBTAINED BY ONE OF THE 'PRT' COMMANDS. THE 'FPR' COMMANDS ARE USEFUL WHEN ALL THE VARIABLES ARE ROUGHLY COMPARABLE IN MAGNITUDE, OR WHEN ALL THE DATA VALUES ARE INTEGERS.

NOTE: YOU COULD HAVE USED A MODIFIED '$PRT' AS A 'NEWC' TO PRINT:

```
  1.2     2.3456    3500
```

DATA DISPLAY (PLOT) :

CTAB	TWO-WAY TABLES OF FREQUENCIES
FREQ	TABLE OF RELATIVE FREQUENCIES
HIST	HISTOGRAM OF ABSOLUTE FREQUENCIES
NORM	NORMAL PROBABILITY PLOT
PLTS	TO PLOT A (STANDARDIZED) VARIABLE IN SEQUENCE
MPLS	MULTIPLE SEQUENCE PLOTS OF 1 TO 3 TIME SERIES
RVSF	A TINY PLOT OF RESIDUALS VERSUS FITTED VALUES FOR A QUICK LOOK. FOR DETAILS, USE
SCAT	TO SCATTER PLOT ANY VARIABLE VERSUS ANY OTHER. VARIABLES 'FITTED' AND 'RESIDU' ARE ALWAYS AVAILABLE AFTER A REGRESSION

TRANSFORMATIONS :

ABSO	ABSOLUTE VALUE
ADDC	ADD A CONSTANT TO A COLUMN
MULC	MULTIPLY A COLUMN OF THE DATA MATRIX BY A CONSTANT
ADDV	ADD TWO COLUMNS OF DATA MATRIX
SUBV	SUBTRACT ONE COLUMN OF DATA MATRIX FROM ANOTHER
MULV	MULTIPLY TWO COLUMNS OF THE DATA MATRIX
DIVI	DIVIDE ONE COLUMN OF DATA MATRIX BY ANOTHER

CATE	CREATE CATEGORICAL VALUE FROM VALUES IN SPECIFIED COLUMN
DIFF	DIFFERENCING TRANSFORMATION LET J BE THE COLUMN TO PLACE THE TRANSFORMED VARIABLE, I BE THE VARIABLE TO BE TRANSFORMED, AND K BE THE GAP FOR DIFFERENCING. THEN X(L,J) = X(L,I) - X(L-K,I), L=K+1,...
LAGG	LAG TRANSFORMATION X(L,J) = X(L-K,I), L=K+1,...

NOTE: THE FIRST K ROWS OF THE ACTIVE DATA MATRIX BECOME INACTIVE IN 'DIFF' OR 'LAGG' AND CANNOT BE RETRIEVED FOR FURTHER ANALYSIS UNLESS COL.J IS REPLACED OR DELETED.

LOGE	NATURAL LOG (LN) TRANSFORMATION
LOG1	COMMON LOG (BASE 10) TRANSFORMATION
EXPO	EXPONENTIAL TRANSFORMATION
HVAR	HOMOGENEOUS VARIANCE TEST TRANSFORMATION
POWE	POWER TRANSFORMATION. NOTE VALUE OF POWER = -1 FOR RECIPROCAL TRANSFORMATION; .5 FOR SQUARE ROOT TRANSFORMATION, ETC.
RANK	ASSIGNS RANKS TO THE OBSERVATIONS (ROWS) OF A VARIABLE (COLUMN) AND PLACES THE RANKS IN ANOTHER COLUMN
STAN	STANDARDIZATION TRANSFORMATION--SUBTRACT MEAN FROM EACH OBSERVATION, DIVIDE THE DEVIATION BY THE STANDARD DEVIATION
TRAN	MULTIPLE TRANSFORMATIONS BEFORE UPDATING

SUMMARY STATISTICS :

CORR	CORRELATION MATRIX OF VARIABLES
COVA	COVARIANCE MATRIX OF VARIABLES
MEAN	MEANS AND STANDARD DEVIATIONS OF VARIABLES
PARC	PARTIAL CORRELATION MATRIX OF ONE SET OF VARIABLES GIVEN ANOTHER SET OF VARIABLES

ONE SAMPLE STATISTICS :

AUTO AUTOCORRELATION (BOX-JENKINS ESTIMATES)
DURB DURBIN-WATSON STATISTIC (FOR RESIDUALS ONLY)
RUNS EXPECTED AND OBSERVED NUMBER OF RUNS ABOVE AND BELOW THE MEAN. NORMAL APPROXIMATION
SERC SERIAL CORRELATION (MAXIMUM LIKELIHOOD ESTIMATE OF AUTOCORRELATION)

DATA ANALYSIS:

WANT COMMANDS FOR TABULAR ANALYSIS ? Y

TABULAR ANALYSIS :

FREQ TABLE OF RELATIVE FREQUENCIES OF A VARIABLE
CTAB TWO-WAY TABLES OF FREQUENCIES
MTAB TWO-WAY TABLE OF MEANS

WANT COMMANDS FOR REGRESSION ANALYSIS ? Y

REGRESSION ANALYSIS :

1. SIMPLE OR MULTIPLE REGRESSION
REGR ORDINARY REGRESSION
WLSR WEIGHTED LEAST SQUARES

2. FOR SELECTING INDEPENDENT VARIABLES
BACK BACKWARD SELECTION PROCEDURE (AUTOMATIC)
FORW FORWARD SELECTION PROCEDURE (AUTOMATIC)
STEP STEPWISE PROCEDURE (USER TO SPECIFY STEPS)
SWEE SWEEP OPERATION. USED TO DELETE A VARIABLE FROM OR REPLACE A VARIABLE WHICH HAS BEEN PREVIOUSLY REMOVED FROM THE REGRESSION EQUATION
ALLS TO PERFORM REGRESSIONS USING ALL POSSIBLE SUBSETS OF A SET OF INDEPENDENT VARIABLES
SUBS TO REGRESS THE DEPENDENT VARIABLE ON ALL POSSIBLE COMBINATIONS OF A GIVEN SIZE OF A SET OF INDEPENDENT VARIABLES

3. FOR PRINTING REGRESSION RESULTS :
ANOV ANALYSIS OF VARIANCE TABLE
BCOR CORRELATION MATRIX OF REGRESSION COEFFICIENTS
BCOV COVARIANCE MATRIX OF REGRESSION COEFFICIENTS
COEF REGRESSION COEFFICIENTS, STANDARD ERRORS, T
SUMM SUMMARY STATISTICS -- MULTIPLE R, STANDARD ERROR OF RESIDUALS, ETC.

4. FOR EXAMINATION OF RESIDUALS :
AUTO TO COMPUTE AUTOCORRELATION COEFFICIENTS (BOX-JENKINS ESTIMATES)
DURB DURBIN-WATSON STATISTIC
NORM TO OBTAIN NORMAL PROBABILITY PLOT OF RESIDUALS
PLTC TO PLOT CONFIDENCE BAND OF FITTED VALUES
PLTS TO PLOT SEQUENCE OF RESIDUALS
RVSF MINIPLOT OF RESIDUALS VERSUS FITTED VALUES
RUNS RUNS TEST FOR RESIDUALS
SAMP TO PERFORM REGRESSION USING RANDOM SUBSAMPLES OF DATA. FOR ERROR ANALYSIS
~~SCAT PLOT OF RESIDUALS VERSUS FITTED VALUES~~
SEPR TO COMPUTE STANDARD ERRORS OF PREDICTED VALUES

WANT COMMANDS FOR TIME-SERIES ANALYSIS ? ?

TIME SERIES ANALYSIS :

```
AUTO   TO COMPUTE AUTOCORRELATION COEFFICIENTS
       (BOX-JENKINS ESTIMATES)
BOXJ   BOX-JENKINS ESTIMATION OF ARIMA MODELS BY NON-
       LINEAR LEAST-SQUARES WITH BACK FORECASTING.
       ALSO POINT FORECASTS TO FUTURE PERIODS.
CROS   CROSS CORRELATION FUNCTION BETWEEN TWO SERIES
PACF   PARTIAL AUTOCORRELATION FUNCTION ESTIMATE
MPLS   PLOT SEQUENCE OF 1 TO 3 TIME SERIES
SPEC   SPECTRAL ANALYSIS
STAR   TO DEFINE A DATA MATRIX OF ZEROES IN THE FIRST
       N ROWS AND FIRST K COLUMNS
```

MISCELLANEOUS COMMANDS :

```
CALC   A CALCULATOR FOR ARITHMETIC OPERATIONS
CRFI   TO CREATE OR MODIFY A FILE WITH A DATA MATRIX TOO LARGE
       TO ENTER ALL AT ONCE IN THE MAXIMUM SIZE IDA DATA MATRIX
CRTS   TO ALTER OUTPUT FOR CATHODE RAY TERMINAL USE
CHGP   TO CHANGE THE LEVEL OF PROMPTS
CTAB   CROSS TABULATION
MTAB   2-WAY TABLE OF MEANS
EXPL   TO EXPLAIN INDIVIDUAL COMMAND WORDS
FSAV   TO SAVE DATA MATRIX ON FILE TO BE USED LATER. (FILE MUST
       HAVE BEEN OPENED ALREADY.)
GAUS   TO COMPUTE PROBABILITIES OR DENSITIES OF A NORMAL VARIABLE
INFO   TO OBTAIN GROUPS OF COMMANDS IN DIFFERENT CATEGORIES
COMM   TO GET A LIST OF COMMAND WORDS GROUPED BY FUNCTION
LIST   TO GET AN ALPHABETICAL LIST OF ALL THE IDA COMMAND WORDS
MISS   TO DECLARE A NUMBER REPRESENTING MISSING VALUES
NEWC   TO DEFINE A NEW COMMAND NAME
NEWS   TO PRINT NEWS ABOUT $IDA
PAUS   TO PAUSE AT THE COMMAND LEVEL.  OTHERWISE IDA
       WILL ASK YOU IF YOU NEED HELP IF NO COMMAND IS
       ISSUED WITHIN ONE MINUTE
PSAM   TO SELECT A PROBABILITY SAMPLE:  SIMPLE RANDOM,
       STRATIFIED OR SYSTEMATIC
QUIT   TO EXIT FROM IDA TO HP SYSTEM CONTROL
RDIM   TO REDIMENSION MAX. SIZE OF DATA MATRIX. (YOU MAY SET
       ASIDE SOME OF THE COLUMNS IN A SCRATCH PAD.)
SPAD   TO SET ASIDE UNUSED DATA COLUMNS IN A SCRATCH PAD

>
```

DATA FILES CONTAINED IN IDA

The following is a list of data files that are contained in IDA and a brief description of the files. Any of these files may be obtained by using the IDA command ENTER and responding to the question "NAME OF INPUT FILE IS?" with the file name followed by ".A103". For example, if you wanted the IDA file ATT you would do the following

```
> ENTER
* MODE OF INPUT:  FROM 'FILE' OR 'TERMINAL?"  F
  NAME OF INPUT FILE IS?  ATT.A103
```

ATT This file contains returns by month for AT&T common stock for 84 months from January, 1961, through December, 1967, in column one, and column two contains the corresponding return data for the same months for the New York Stock Exchange Arithmetic Index.

BELOIL This file has 32 observations and 3 variables. The variable names are PROD, LAB, and CAP. The data collected by Broeckx, D'Hooge, Goovaerts, and Van Den Broeck of Belgian Oil.

BLUME This is a matrix that consists of one variable, monthly sales of Winnebego Motorhomes, starting with November, 1966, and concluding with February, 1972.

BROWNL This file contains six variables and 33 observations. The data is on the percent reduction in blood sugar after injection of insulin into rabbits. The source of the data is K. A. Brownlee, _Statistical Theory and Methodology in Science and Engineering_, Wiley 1960, first edition, page 517.

DJ This file contains data on the monthly closings of the Dow-Jones Industrial Index from August, 1968, through July, 1972.

ECDAT1

This file has quarterly observations on various economic variables from the first quarter 1947 through the fourth quarter of 1970. The variables are:

(1) Unadjusted GNP in billions of dollars expressed as a quarterly rate.
(2) Money supply (M1) without seasonal adjustment. The data are means of daily data in billions of dollars.
(3) Federal government purchases of goods and services unadjusted quarterly rate in billions of dollars.
(4) Standard and Poor's Composite Index of Stock Prices (means of daily data).
(5) Man-days idle due to strikes and lock-outs in millions unadjusted seasonally.
(6) GNP implicit price deflators 1957-1959 = 1
(7) Time period identification (e.g. 1950 second quarter is designated 502).

HALSEY

This file contains data on degree days in column one and the year in column 2. The data was recorded by the National Weather Service at Midway Airport Chicago for 1932-1971.

IBM

This file contains data on the monthly returns on IBM common stock from January, 1961, through December, 1967, in column one and data for the same period for the New York Stock Exchange Arithmetic Index in column two.

IQ

This is a set of data that was simulated by the computer. It consists of 100 numbers that are similar to drawings from a normal distribution with mean 100 and standard deviation of 15.

ISAACS

This file contains weekly data from the week ending January 12, 1952 to the week ending May 31, 1952, for a commercial laundry. Column one contains data on calls made (picks and deliveries), column two contains data on sales.

LMICH

This file contains data on the mean monthly levels for July of each year from 1875 through 1972 of Lake Huron. The levels are expressed as elevations in feet above the mean water level at Father Point, Quebec. The level is contained in column one, the year in column two.

MCDFRM

This file gives 53 paired observations on two methods of measuring iron content in crushed blast furnace slag. Column one is the measurement by chemical test at the laboratory, column two is the measurement by an alternative magnetic test.

MICH

This file contains data on Michigan unemployment. It contains observations on 83 counties in Michigan on the following variables:

(1) Designated for funding: 1 = yes 0 = no
(2) Funded? 1 = yes 0 = no
(3) Unemployment percentage 1966
(4) Unemployment percentage 1970
(5) Labor Force 1966
(6) Labor Force 1970
(7) Dollar funding for public works
(8) Dollar funding for technical assistance
(9) Dollar funding for business loans

MICH1

This file contains data on Michigan unemployment. It contains observations on 83 counties in Michigan on the following variables:

(1) Special Federal Funding: 1 = yes 0 = no
(2) Percentage unemployment in county 1966
(3) Percentage unemployment in county 1970.

MMIND

This file contains data on management salaries and their job evaluation. There are five variables and 67 observations. Column one contains data on know-how (KNOWHO), column two contains data on problem-solving (PROBSO), column three contains data on accountability (ACCTBL), column four is the total of the scores made on each of the three tests: Know-How, Problem Solving and Accountability. The fifth column contains a variable that classifies the person being evaluated (CLASS). The variable has the following coding scheme:

0 Top Management
2 Major Division 1
3 Major Division 2
4 Major Division 3
5 Engineering
7 Administration
10 Sales

ROCKY	This file contains a set of readings of "hardness" of 100 steel coils produced by a steel mill in the Chicago area.
ROCKY2	This file also contains a set of readings of "hardness" of 100 steel coils produced by a steel mill in the Chicago area. It differs from ROCKY only in the fact that these data were collected at a later date.
SALES1	This file consists of actual monthly sales in thousands of dollars in column one and judgemental forecasts of sales made by the sales manager of the division in column two. There are 24 months of observations.
TJT66	This file contains data on soilage percentages for a particular kind of bottle cap in 56 consecutive production runs.
VELMON	This file contains data on the velocity of money collected by Friedman and Schwatz for the American economy from 1891 to 1961. Column one contains the velocity figure, column two the year.

APPENDIX C

THE UNIVERSITY OF TENNESSEE AT CHATTANOOGA
CENTER FOR CRIMINOLOGICAL ANALYSIS AND TRAINING
CHATTANOOGA, TENNESSEE 37401

CRIME ANALYSIS NEWSLETTER

Editor: Bob Catale May 1977

AN INVITATION

The staff of Crime Analysis Newsletter invites our readers to contribute articles and reviews dealing with any aspect of crime analysis. Ideally, articles should run between 500 and 750 words. Topics may fall into such categories as, but certainly not restricted to, statitical analysis, evaluation techniques, planning theory, methodological and research designs, and criminological theory. Book and article reviews should be of lesser length.

COMPUTERS AND CRIME ANALYSIS by Joe Maloney

Computers have drastically simplified the art of crime analysis. In the past, the analyst had to have lots of time, deep knowledge of statistics and relatively little data. With the advent of the computer, these factors have become somewhat reversed.

When crime analysis was first emphasized, most analysts operated by hand. They had to obtain the data, sort it into relative categories, and then perform all the mathematical analysis using pencil and paper. If any sizeable amount of data, 50 or 100 observations for example, were to be studied and this data had only relatively few variables, say 10, it might take hundreds of hours to examine each group observations, make some statistical tests (using complicated and time consuming mathematics) and reach conclusions, some of which might force the analyst to reclassify the data and retest. Assuming it would take a week to sort and classify the data, it could take another half a week to do the math correctly. When the process had to go through several iterations, months could be spent in relatively simple analysis.

Today the amount of time spent can be cut to days through the use of the modern computer. This time reduction is possible because once the data is loaded into the computer and one of the standard computer packages is applied, the analyst can perform several statistical tests in a matter of hours just by slightly modifying instructions to the computer. The "big black box" works at such high speeds (billionths of a second in some cases) that it is possible for the analyst to see the results of one round of tests, draw hypotheses or conclusions, design a new test of his conclusions and run those tests, receiving the second (and even third or fourth) round of results the same day. Thus, the computer will reclassify the data and calculate the mathematics in seconds, where as the analyst might have taken days or weeks to do it by hand.

To manually perform statistical operations, in the past, the analyst not only had to know what a coefficient of correlation was and what it meant but also how to group the data properly. In addition, he had to know what manipulations had to be done to the data, why those steps were necessary and how to do each step, plus actually performing the oftentimes intricate calculations himself. At times it seemed only those with advanced statistics degrees and education could perform simple correlations. Today, the analyst needs only to know what a coefficient of correlation looks like and what it means. The computer will do all the rest. It is still very helpful if the analyst is thoroughly familiar with statistics, but it is no longer essential. With the development of software packages such as BMC, SPSS, STP andSAS the analyst needs to know only how to put the data in the computer and how to read the printout. The computer does the rest.

The computer has simplified analysis. Large amounts of complicated data can now efficiently be dealt with. In the past, only a few variables could reasonably be analyzed. Too many factors or variables made analysis impractical. With the use of a low number of factors, tests were always subject to question because of factors omitted. Inter-relationships, whether dependent or independent, had to be ignored and/or placed low on the priority list. With the computer, many more factors can be included in the analysis. Tests can now be performed that were impractical or impossible before. The computer has simplified analysis because relationships that had to be ignored in the past now can be brought to light.

In the beginning of this article, I referred to crime analysis as an art. This is because while the computer has greatly simplified analysis, final analysis is still imperfect and subject to several interpretations. A piece of sculpture made of garbage may be an excellent work of modern art, but it is still garbage. A crime analysis made from insufficient or bad data may be an excellent analysis, but will be worthless. The acronym that applies is GIGO--Garbage In-Garbage Out. This principal applies to crime analysis, computers, and other art forms. The artist/crime analyst must still be familiar with his media/data. He must still be familiar with the tools of his trade (brusiery points, canvases or computers) although he does not need to know how to make the tools. And, of utmost importance, the artist/analyst must be able to look at the product and decide if it makes sense.

This article has not addressed itself directly to the ability of the artist/analyst to understand his work. The computer does not generate creativeity or enhance insight. It must be remembered that worthless data run through the best analysis stystem will desolve into nothing more than a good analysis of bad information, i.e., it will still be worthless. Computers do not make analysis better, just faster.

JOSEPH P. MALONEY is Systems Development Planner for the Kentucky Department of Justice. He has been employed in a wide range of planning, analysis and system development positions with the Kentucky SPA since 1972. He holds an M.S. in Community Development Planning from the Institute of Community Development, University of Louisville.

CCAT UPDATE

The second CCAT Crime Analysis Workshop was held on March 14-18, 1977, at the Sheraton Hotel in downtown Chattanooga. Special computer-oriented seminars were held at the University of Tennessee at Chattanooga. Participants totaling nearly 20 came from several states in the region.

Areas covered in this workshop included statistical interpretations and analysis, handled by Mike Devine of Alabama's Statitical Analysis Center and Mario Perez-Reilly of Middle Tennessee State University, and data sources and presentation , handled by Keith Harries of Oklahoma State University. Also included were LEAA Guidelines Review and a Literature Review conducted by UTC personnel and others in such diverse areas as formation of goals and objectives, questionnaire design and development, new computer packages, and effective graphic data display.

Interspaced throughout the five days were times set aside so that conference participants could have private meetings with the conference consultants regarding any special problems they had.

Under the same LEAA Grant that provided funds for the crime analysis workshop,CCAT personnel devoted themselves to providing technical assistance (TA). This technical assistance for crime analysis was provided to the State Planning Agencies of Kentucky, North Carolina, Florida, Missisissippi, and Tennessee.

BIBLIOGRAPHY UPDATE

CRIMINAL JUSTICE PLANNING: A PRACTICAL APPROACH, Michael E. O'Neill, Ronald F. Bykowski and Robert S. Blair. Justice Systems Development, Inc., 1976.

EXPLAINING CRIME, Gwynn Nettler, McGraw-Hill Book Company, 1974. A fine introduction to the problems surrounding crime analysis.

CITIES AND CRIME: A GEOGRAPHIC MODEL, Keith D. Harries in Criminology. Volume 14, No. 3, pages 369-386; November 1976. Five measures of serious crime and 25 socio-economic variables are presented as a basis for interurban crime analysis.

ON BEHALF OF A MORATORIUM ON PRISON CONSTRUCTION, William G. Nagel in Crime and Delinquency. Volume 23, No. 2, pages 154-172, April 1977. Though initially a re-evaluation of the author's views on prison construction the article contains valuable analysis and data information applicable to crime analysis and planning.

ECONOMIC IMPLICATIONS OF STANDARDS AFFECTING CORRECTIONAL PROGRAMS, Neil M. Singer in Crime and Delinquency. Vol. 23, No. 2, pages 180-195, April 1977. An interesting examination of the realities of correctional reform.

CALENDAR OF UPCOMING EVENTS

The Center for Criminological Analysis and Training will hold its second Crime Analysis Workshop & Conference of this year in Chattanooga, TN, August 22-26, 1977. The workshop/conference is sponsored under a grant from The Law Enforcement Assistance Administration, US Department of Justice. The program (see this issue page 3, CCAT Update) is designed basically for those engaged in crime analysis and planning in state planning agencies, but others are welcome. For full information contact: Ms. Linda Myers, Director, Center for Criminological Analysis & Training, The University of Tennessee at Chattanooga, Fletcher Hall, Chattanooga, TN 37401. Telephone (615) 755-4135.

****
****
****

Crime Analysis Newsletter was prepared under Grant No. 77-TA-04-0001 from the Law Enforcement Assistance Administration of the US Department of Justice.

THE CENTER FOR CRIMINOLOGICAL ANALYSIS & TRAINING
The University of Tennessee at Chattanooga
Chattanooga, Tennessee 37401

ADDRESS
CORRECTION
REQUESTED

THE UNIVERSITY OF TENNESSEE AT CHATTANOOGA
CENTER FOR CRIMINOLOGICAL ANALYSIS AND TRAINING
CHATTANOOGA, TENNESSEE 37401

CRIME ANALYSIS NEWSLETTER

Editor: Bob Catale

January 1977

CCAT's 1977 CRIME ANALYSIS CONFERENCE WORKSHOPS

Plans are underway for CCAT's 1977 Crime Analysis Conference Workshops. The first of the workshops is planned to be held the week of March 14-18, 1977. The second workshop is planned for August 15-19, 1977. The aim of the conference workshops is primarily to help build greater capacity in LEAA Region IV among analysts/planners to meet the problems associated with crime analysis and state comprehensive plans. Site for the conference workshops will be the University of Tennessee at Chattanooga. Though the site for the workshops is the same as last year's, the length of the workshops has increased in order to present a wider range of topics. Both conference workshops will deal with essentially the same materials. Areas to be covered during the week long workshops will include: statistical interpretations and analysis, LEAA guidelines, data sources and presentation, methodology, criminological theory related to crime analysis (CA), and the role of the computer in CA.

The various segments of the conference workshops will be conducted by CCAT staff members and several other specialists in crime analysis. Among those included will be Keith D. Harries of Oklahoma State University, Michael DeVine of the Alabama Statistical Analysis Center, Mario Perez-Reilly of Middle Tennessee State University, and George Datesman of LEAA's Washington, D. C. office of Criminal Justice Education and Planning.

Participants in the conference workshops will be SPA analysts and planners responsible for crime analysis, LEAA Region IV state representatives and others interested in crime analysis. In addition to participants invited from Region IV, there are a limited number of openings for other persons who might wish to attend. Such parties may write to Ms. Linda Myers, Director, Center for Criminological Analysis and Training, The University of Tennessee at Chattanooga, Chattanooga, Tennessee 37401. There will be a registration and materials fee for those not specifically invited. The fees do not include food, lodging or travel expenses.

CCAT's 1977 Conference Workshops are funded under a continuation grant from the Law Enforcement Assistance Administration administered through LEAA Region IV in Atlanta.

FINDING A FOUNDATION FOR YOU: CAUGHT IN A PINCH?
by Linda L. Myers

Need money? Are you caught in the financial crunch? Do shifting priorities leave analysis, research and evaluation wanting for money in your area? If so, you may be interested in information on alternate sources of funding in the United States. Have you considered private foundations? One single source of foundation information may be sufficient for your needs: Taft Information System: A Method for Keeping Current on Foundations, Foundation Reporter. This is published periodically by Taft Products, Inc., 1000 Vermont Avenue, NW, Washington, DC 20005. Telephone (202) 347-0788. This is an expensive service - around $200 per year, but most Research and Development Offices on college and university campuses already subscribe to the service and may allow you to use their materials.

Volumes are published in an 8-1/2" x 11" format and are approximately 1-3/4" thick. Within the September 1976 volume, 391 major American foundations are reported on. Each report conveys the assets of the foundation, quantity of grants made annually, geographic focus and location of awards, particular fields of interest to the foundation, contact persons, funding cycle and grant format. Examples of the most current grants awarded are set forth by category, e.g., Arts and Humanities, Education, Health and Social/Welfare.

In perusing the Taft document, one finds that a substantial number of the foundations listed provide funds for research and education. Each foundation is described in order to provide a person preparing a proposal a sufficient amount of information and assistance.

An additional source of foundation information is The Foundation Directory. This is distributed by the Columbia University Press. Again this resource might be easily located in a University Research and Development Office.

The focus of the directory is state and regional. It provides only a minimal amount of information and includes foundations "established to maintain or aid social, educational, charitable, religious, or other activities serving the common welfare." The listings include foundations that award more than $25,000 per year and/or possessed assets of $500,000 or more. Information is contained, state-by-state, on 5,454 foundations existing in 1971. Each listing contains the name and address of contacts, donors, purpose, activities and financial status.

Linda L. Myers is Assistant Professor of Human Services and Director of the Center for Criminological Analysis and Training at the University of Tennessee at Chattanooga. She has worked as a state planner doing crime analysis and teacher courses in planning and evaluation in criminal justice.

AN INVITATION--

The staff of CA Newsletter's invite our readers to contribute articles and reviews dealing with any aspect of crime analysis. Ideally, articles should run between 500 and 750 words. Topics may fall into such categories as, but certainly not restricted to, statistical analysis, evaluation techniques, planning theory, methodological and research designs, and criminological theory. Book and article reviews should - be of lesser length.

BIBLIOGRAPHY UPDATE

The goal of this feature will be to "update" the bibliographical knowledge of those involved with crime analysis by noting recent books and articles. Future issues will include reviews of some of the more important literature in crime analysis.

CRIMINAL JUSTICE PLANNING: A PRACTICAL APPROACH, Michael E. O'Neill, Ronald F. Bykowski and Robert S. Blair. Justice Systems Development, Inc., 1976.

EXPLAINING CRIME, Gwynn Nettler, McGraw-Hill Book Company, 1974. A fine introduction to the problems surrounding crime analysis.

CITIES AND CRIME: A GEOGRAPHIC MODEL, Keith D. Harries in Criminology. Volume 14, No. 3, pages 369-386; November, 1976. Five measures of serious crime and 25 socio-economic variables are presented as a basis for interurban crime analysis.

CALENDAR OF UPCOMING EVENTS

The National Conference of Criminal Justice Evaluation will be held in Washington, DC, February 22-24, 1977, in the Sheraton-Park Hotel. The conference is sponsored by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, US Department of Justice. Conference Chairman will be Arnold S. Trebach. Registration fee is $10.00. Contact: Koba Associates, Inc., 2001 S Street, NW, Washington, DC 20009. Telephone (202) 265-9114.

Courses in Crime Analysis are being sponsored by the Theorem Institute. Plans call for holding courses in Dallas, Texas, February 16-18, 1977 in Washington, DC March 16-18, 1977, in Las Vegas, Nevada, May 18-20, 1977, and in San Francisco, California, June 22-24, 1977. Contact: Michael E. O'Neill, Theorem Institute, 1737 North First Street, Suite 590, San Jose, CA 95112. Telephone (408) 294-1427.

Progress in Criminal Justice: By Whose Standards? will be the theme of a fourth national symposium sponsored by the National Clearinghouse for Criminal Justice Planning and Architecture of the University of Illinois and the Law Enforcement Assistance Administration. Associate sponsors include the American Institute of Architects and the American Probation and Parole Associations, Inc. Dates for the symposium are set for April 6-9, 1977 and it will be held at the Fairmont Hotel in New Orleans, LA. Tuition will be charged. Information can be obtained from Elmer F. Edwards, 116 Illini Hall, Champaign, Illinois 61820.

Crime Analysis Newsletter was prepared under Grant No. 76-TA-04-0002 from the Law Enforcement Assistance Administration of the US Department of Justice.

The University of Tennessee at Chattanooga is a primary campus of The University of Tennessee, Dr. Edward J. Boling, President, UT, Dr. James E. Drinnon, Jr., Chancellor, UTC.

THE CENTER FOR CRIMINOLOGICAL ANALYSIS & TRAINING
The University of Tennessee at Chattanooga
Chattanooga, Tennessee 37401

THE UNIVERSITY OF TENNESSEE AT CHATTANOOGA
CENTER FOR CRIMINOLOGICAL ANALYSIS AND TRAINING
CHATTANOOGA, TENNESSEE 37401

CRIME ANALYSIS NEWSLETTER

Editor: Bob Catale

Vol. 1, No. 1, October 19

HELLO!

The Center for Criminological Analysis and Training (CCAT) was established in March 1976 and is housed within the Criminal Justice Program of The University of Tennessee at Chattanooga. The primary purpose of CCAT is to foster sound analysis in all areas of criminal justice research and planning. In its first year of activities, CCAT prepared and presented a Crime Analysis Workshop and delivered technical assistance among the eight states in LEAA Region IV. New workshops and seminars are being planned for the coming year.

Currently, CCAT has funds available to provide technical assistance to State Planning Agencies in Region IV. Specific areas where CCAT is available to assist include, but are not limited to, overall general guideline compliance for crime analysis, the addressing of special conditions, selecting of data for comparisons and predictions, developing research designs and methodologies for crime analysis, and data display techniques.

In addition, CCAT personnel are available as consultants for crime analysis, time permitting, outside of LEAA Region IV.

BOUNDARY IMPACTS IN CRIME ANALYSIS by Keith D. Harries

Problems in urban crime analysis that are frequently overlooked - or underestimated - relate to the impact of changes in city boundaries, or the effects of socioeconomic changes occurring just beyond these boundaries. First, the impact of boundary changes.

Annexation to city territory is commonest in regions outside the eastern U.S. The latter tends to be "locked in" to current city boundaries because cities are relatively old and have become surrounded by other incorporated areas. Outside the east, however, more territory is available for expansion, which in some cases has taken the form of consolidation of the city with the county. Consider the impact of annexation on crime statistics. Typically, the annexed area will be a white suburb, often affluent, and may include essentially rural areas. These new areas are likely to be targets of property crime - particularly burglary - since they have historically lacked police patrol and yet are attractive by virtue of their wealth and the opportunities associated with that wealth. Street crimes of violence are less likely, since the new areas are probably distant from central city problems and will often lack the retail establishments, including banks, that would be robbery targets in more developed districts. Thus population-specific rates of Index crime would, over time (holding all else constant) be expected to dip in the violent categories but possibly increase in the property columns. Since annexation often precedes the development of major subdivisions, the nature of those new areas of construction should also be considered. Does the design offer "defensible space"? What will be the age/sex composition of the incoming population? Will the population consist of migrants from another state or nation? If so, what are the culture traits of the group with respect to possible criminal activity?

Second, what is the effect of socioeconomic change beyond existing boundaries? The structure of our local government tends to encourage a rather myopic view of their jurisdictions by police and other agencies. And yet what is going on outside the city limits may have a profound impact on the city itself. Rural non-farm population is very varied demographically, ranging from affluent mansions to mobil home parks with sharply differing levels of amenity. Given the mobility that our citizens have, the characteristics of this outlying population are worth looking at rather carefully - from both sides of the city line. The sheriff may have to deal with the crime displaced out of the city by successful urban enforcement, and vice versa.

In summary, crime trends over time may be profoundly affected by numerous factors, two of which are changes in city boundaries and in "exurban" population characteristics. Rigorous crime analysis should take these problems into account.

About the author...

Keith D. Harries is Associate Professor of Geography at Oklahome State University. Recent works by Professor Harries important to those involved in crime analysis include The Geography of Crime and Justice (1974) and "A Crime Based Analysis and Classification of 729 American Cities" in Social Indicators Research (1976).

PEOPLE IN CRIME ANALYSIS

Our objective with this feature is to encourage and facilitate better communications in Region IV by providing brief profiles of a few of the people doing crime analysis - their work, their backgrounds, and their special interests.

MICHAEL DEVINE, born under the astrological sign of Sagitarius, has been working since 1975 at the Statistical Analysis Center (SAC) of Alabama's Criminal Justice Information Center in Montgomery. Mike graduated from both the B.A.(1970) and M.S.(1971) programs in psychology at Auburn University. He has primary responsibilities at SAC for coordination of data collection and supervision of analysis and interpretation of data for Alabama's comprehensive planning process. Additional duties include development of resource allocation systems for the state's criminal justice modules, administration of the Alabama Uniform Crime Reports section of the Alabama Criminal Justice Information System (ACJIS), and evaluation of law enforcement data systems effectiveness. Areas of interest, evidenced by publications, have included evaluation of criminal behavior and measurement and prediction studies of criminal behavior and recidivism. Before assuming his present position at SAC, Mike was a research project director at both the University of Alabama and Auburn University.

Current activities include working through a flow simulation of a sample of the Alabama Courts System, development of both a mapping/modeling system for use in the criminal justice area and of a research project to define the "real" crime rate in Alabama.

BIBLIOGRAPHY UPDATE

The goal of this feature will be to "update" the bibliographical knowledge of those involved with crime analysis by noting recent books and articles. Future issues will include reviews of some of the more important literature in crime analysis.

EXPLAINING URBAN CRIME RATES. Sheldon Danziger in Criminology. Vol.14, No.2, pages 291-296; August, 1976.

SAMPLE SURVEYS OF THE VICTIMS OF CRIME. Wesley G. Skogan in Review of Public Data Use. Vol.4, No.1, pages 23-28; January, 1976.

SINUSOIDAL PATTERN ANALYSIS IN CRIMINAL INCIDENCE. Jerry Banks and David Vatz in Criminology. Vol.14, No.2, pages 251-258; August, 1976.

HOMICIDE TRENDS IN ATLANTA, R.S. Munford, R.S. Kazer, R.A. Feldman, and R.R. Stivers in Criminology. Vol.14, No.2, pages 213-232; August, 1976.

CALENDAR OF UPCOMING EVENTS

OCTOBER

The Sixth National Forum on Volunteers in Criminal Justice will be held in Atlanta, GA from Oct. 17-20. For information write: Volunteers in Criminal Justice, 800 Peachtree St., Room 312, Atlanta, GA 30308.

NOVEMBER

The American Society of Criminology Annual Meeting will be held in Tucson, AZ Nov. 4-7. Contact June Morrison, School of Police Administration, University of Arizona at Tucson 85712.

The 1976 Conference of the Society of Police and Criminal Psychology is scheduled for Nov. 18-20 at Gulf Park Campus of the University of Southern Mississippi. Contact: Dr. Mickey Braswell, Box 221, Southern Station, Hattiesburg, MS 39401.

Indiana University has published its 1976-'77 training calendar. For information contact: Donald P. Weir, Trng. Spec., Center for Criminal Justice Trng., Indiana University, Harrison Bldg., Suite 502, 143 Market St., Indianapolis, IN 46204, (317) 264-4607.

Crime Analysis Newsletter was prepared under Grant No. 76-TA-04-0002 from the Law Enforcement Assistance Administration of the U. S. Department of Justice.

The University of Tennessee at Chattanooga is a primary campus of The University of Tennessee, Dr. Edward J. Boling, President, UT, Dr. James E. Drinnon, Jr., Chancellor, UTC.

THE CENTER FOR CRIMINOLOGICAL ANALYSIS & TRAINING
The University of Tennessee at Chattanooga
Chattanooga, Tennessee 37401

Non-Profit Organization
U. S. POSTAGE
PAID
CHATTANOOGA, TENN.
Permit No. 26

END