To the Graduate Council:

I am submitting herewith a thesis written by Jonathan G. Overly entitled "Characterization and Statistical Analysis of Industrial Data Compiled by the University of Tennessee Industrial Assessment Center." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science.

Richard Jendrucko, Major Professor

We have read this thesis and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and Dean of The Graduate School

CHARACTERIZATION AND STATISTICAL ANALYSIS OF INDUSTRIAL DATA COMPILED BY THE UNIVERSITY OF TENNESSEE INDUSTRIAL ASSESSMENT CENTER

A Thesis
Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Jonathan Gregory Overly
May 1997

Acknowledgments

There are several people to whom I owe a great deal for making my time at the University of Tennessee so prosperous. Primarily, and most directly, I would like to give thanks to Dr. Richard J. Jendrucko. His undying interest in helping me improve my ways and means as a growing engineer is inestimable. My tenure under his tutelage as Center Director for the University of Tennessee Industrial Assessment Center has proven invaluable in more ways than are scriptable here. Additionally, as the remaining members of my Thesis Committee, I gratefully appreciate the inputs, advice and assistance of Dr. Esteban E. Walker and Dr. William S. Johnson.

I would like to also thank all the professors of the Engineering Science and Environmental Engineering departments for their support and encouragement. I have gained so much by selecting these University of Tennessee Engineering departments as my engineering home.

Additionally, much appreciation must be given to my mother and father for their perpetual motivation in furthering myself in life. To them both I owe an unpayable debt for their indefatigable excellence in parenting.

Lastly, I wish to thank my extraordinary wife, Nancy. She has quietly provided unyielding emotional support to me in my pursuits of greater things and in attaining the next level of personal and professional success.

Abstract

In this study, a large quantity of industrial manufacturers' energy consumption and waste generation data compiled by the University of Tennessee Industrial Assessment Center was analyzed for inherent trends and relationships. This effort first required the amassing of the energy and waste data generated through over 20 years of service to regional manufacturing facilities, then organization of the amassed data into an onsite database. Previous efforts to compile this type of data for statistical analysis provided some preliminary insights, but the earlier investigatory analyses did not reach the depth of analysis this work undertook in categorizing and analyzing different types of information as well as examining the data for relationships, including nonlinear functions. Specifically, the objectives of this work were to 1) create the database to facilitate easier access to and analysis of the UTIAC-generated data and 2) perform several different types of analyses in search of trends and relationships within the data.

The characterization analyses performed on the energy and waste data revealed some notable information and trends. The two most interesting included the following. The first is that the implementation rate of the recommended measures (both energy and waste related) varies moderately from state to state in which the assessments were performed, with Tennessee and South Carolina clients reporting an implementation rate of approximately 50% and Kentucky, North Carolina and Virginia clients reporting an implementation rate of approximately 60%. The second is that the implementation rate for all recommendations increases significantly and

almost linearly with an increase in the number of energy recommendations made per report up to 9 (from 0% at 1 recommendation per report to about 60% at 9).

The simple linear, curvilinear and multiple linear regression analyses performed on the data were limited to a sampling of the analyses that could be done, simply because of the sheer quantity of program data amassed. However, five investigatory questions were propounded with the overriding theme of which factors most strongly relate to plant energy consumption, waste generation and recommendation implementation rate. In order as posed, the results yielded the following.

The analysis of question #1 revealed a strong, positive relationship for all SIC groups between annual sales and annual electrical consumption, with five specific SIC groups also showing moderate to highly significant, similar relationships. Question #2's analysis exposed a modest, curvilinear relationship for all SIC groups between annual sales and the UTIAC-recommended annual energy cost savings, along with several, modest SIC-specific linear and curvilinear relationships between these variables. The investigation of question #3 revealed a mediocre, positive relationship between annual production hours and annual energy consumption for all industry types, with the same results for six specific SIC groups. Question #4's analysis showed that there is practically no relationship between the number of recommendations made per report and the total UTIAC-recommended energy conservation or waste generation quantities, outside of some modest correlations within three specific SIC groups. Lastly, the investigation of question #5 exposed that, for all industrial manufacturing types as well as all the specific ones, there is almost no association between implementation rate and either the capital cost of implementation or payback period.

Most certainly, these findings show that there are significant, inherent relationships within the various types of the industrial data compiled, and the thorough analysis of this data can provide valuable insights into the factors affecting plant energy consumption, waste generation and recommendation implementation. In addition, the results can be used to prioritize future program work for maximum client and program benefit.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	Analysis of Variance
AR	Assessment Recommendation
ARC	Assessment Recommendation Code
ASCII	American Standard Code for Information Interchange
BTU	British Thermal Unit
DOE	Department Of Energy
EADC	Energy Analysis and Diagnostic Center
ECO	Energy Conservation Opportunity
HVAC	Heating, Ventilation and Air Conditioning
IAC	Industrial Assessment Center
MB	Megabyte
MLR	Multiple Linear Regression
OIPEA	Office of Industrial Productivity and Energy Assessment (Rutgers University)
OIT	Office of Industrial Technologies (U.S. DOE)
OMB	Office of Management and Budget
QA/QC	Quality Assurance/Quality Control
SCR	Simple Curvilinear Regression
SIC	Standard Industrial Classification
SLR	Simple Linear Regression
UTIAC	University of Tennessee Industrial Assessment Center
UTK	University of Tennessee at Knoxville
WWW	World Wide Web

Chapter I

Introduction and Background

Energy Analysis and Diagnostic Center/Industrial Assessment Center

Program History

In 1976, the Energy Analysis & Diagnostic Center (EADC) program was instituted as a component of the National Energy Strategy under the aegis of the U.S. Federal Energy Administration (which evolved into the Department of Energy (DOE) in 1977). As a major DOE energy conservation initiative administered by the Office of Industrial Technologies (OIT), the focus of the program has been to utilize teams of university engineering faculty and students to provide no-cost energy (and more recently waste and productivity) assessments to small and medium-sized industrial manufacturers located within the geographical region centered by each participating university. The teams traveled to plant locations and spent one day in the manufacturing facility collecting data pertaining to energy use in plant equipment. Following the site visit, the team prepared a report of findings and recommendations. The recommended measures to save energy, or "Energy Conservation Opportunities" (ECOs) specified how the facility could save energy and money associated with their facility's energy consumption. The recommendations included an engineering-based discussion of operating conditions, the recommended action to be taken, anticipated savings, implementation costs and a simple payback analysis. Completed reports were mailed to the facility, with an anonymous copy mailed to a field management center. These field management centers were established to provide a link to the DOE sponsor and to maintain program quality.

Following the broadening of the overall DOE mission in 1992, the EADC program was renamed the Industrial Assessment Center (IAC) program and the focus was expanded to include waste minimization as well as energy conservation. The participating universities have expanded in number from three schools in 1976 to 30 universities since 1994, serving manufacturers in 42 of the contiguous 48 states. In 1995, the program experienced another focus expansion adding productivity enhancement to the energy and waste assessment focus. Figure I-1 shows the general operating areas of the 30 current IACs (actual service area layout varies from Center to Center; stars in Figure I-1 indicate the location of the two field management agencies). At the current rate of 30 assessments per year per center, 900 manufacturers per year across the U.S. are receiving industrial assessments by IAC program Centers. The entire IAC program has through 1996 served almost 7,000 manufacturing facilities, with the UTIAC program having served over 600 clients alone, or about 8.5% of the total due to its participation from program initiation.

For participation in the IAC program, each client plant must be a "manufacturing" facility. A facility is classified as a "manufacturer" if they have a Standard Industrial Classification (SIC) code in the range of 2000 to 3999. The first two digits of these codes are termed groups, and codes are grouped by similar products manufactured. Table I-1 lists the 20 two-digit manufacturing SIC groups and their respective industry types. SIC codes were developed by the U.S. Office of Management and Budget (OMB) and

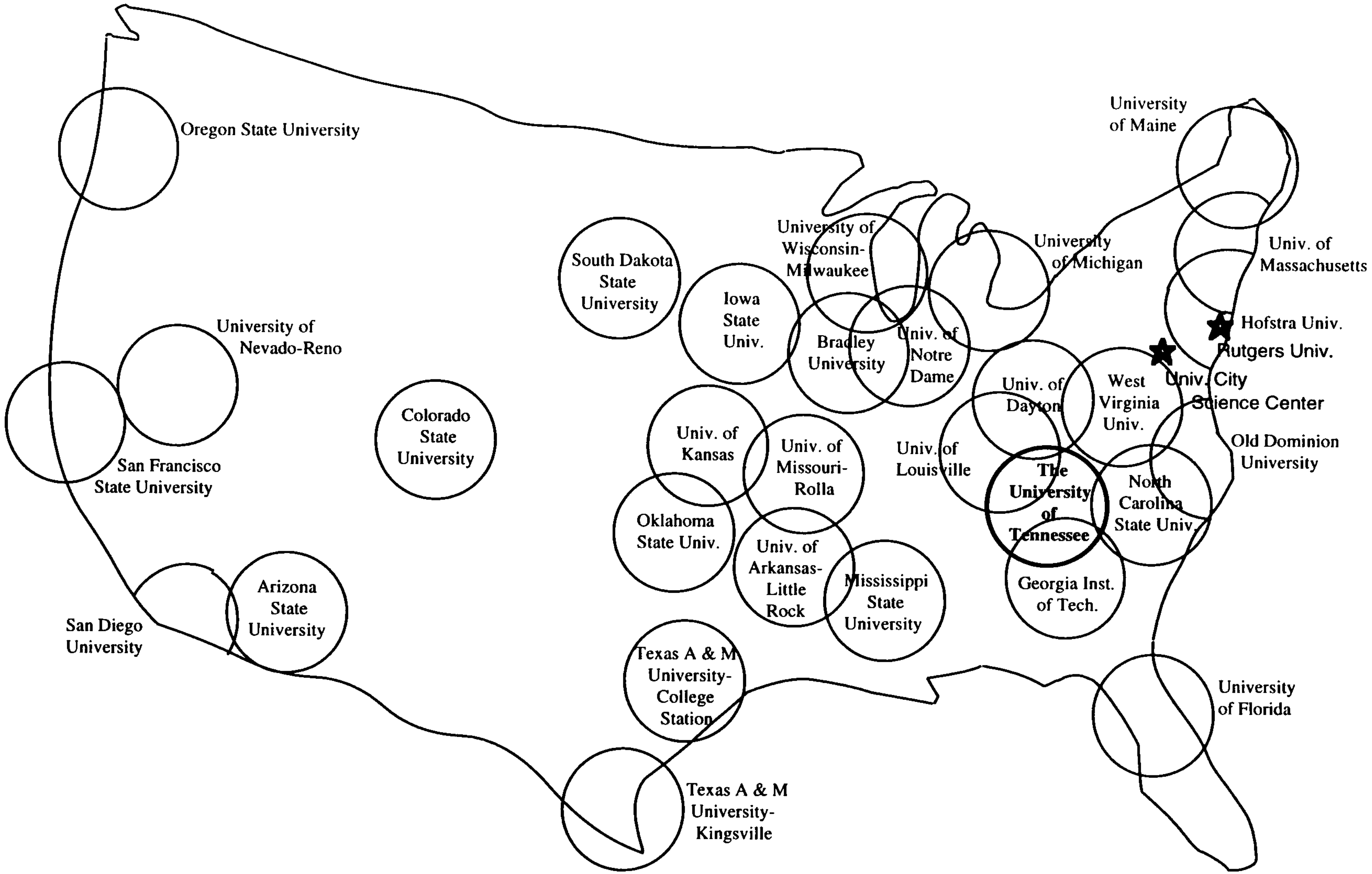


Figure I-1. BASIC GEOGRAPHICAL SERVICE AREAS OF THE 30 CURRENT IACS

Table I-1. STANDARD INDUSTRIAL CLASSIFICATION (SIC) GROUPS FOR MANUFACTURERS

SIC Group	Description
20	Food and kindred products
21	Tobacco products
22	Textile mill products
23	Apparel and other textile products
24	Lumber and wood products
25	Furniture and fixtures
26	Paper and allied products
27	Printing and publishing
28	Chemicals and allied products
29	Petroleum and coal products
30	Rubber and miscellaneous plastic products
31	Leather and leather products
32	Stone, clay and glass products
33	Primary metal industries
34	Fabricated metal products
35	Industrial machinery and equipment
36	Electronic and other electric equipment
37	Transportation equipment
38	Instruments and related products
39	Miscellaneous manufacturing

range from numerical categories 0000 to 9999 (Table A-1, p. 109, lists all the SIC groupings).

From its beginning, the IAC program has limited its service to small and medium-sized manufacturing facilities based on the premise that larger manufacturing plants generally have the personnel and financial resources to either 1) employ personnel with full-time responsibility for either or both energy conservation, waste minimization or productivity enhancement, or 2) be able to afford the services of outside consultants. To be eligible for program participation, prospective clients must meet three of the following four criteria:

1) Gross annual sales less than $75 million (for the assessed site only)
2) Fewer than 500 employees
3) Total annual energy costs below $1.75 million
4) Lack of in-house technical staff with expertise and major responsibility for energy management, waste minimization and/or productivity enhancement

Archived Data

When the earliest UT EADC audits were performed in 1976, the computer revolution was just dawning on the United States. At that time, hard copies (paper bound editions) of client reports were prepared on typewriters and photocopies were retained for future reference in responding to client questions (if any) and to serve as a base for future reports containing similar recommendations. In the early 1980s, as desktop computers started becoming an asset to every office and university in the U.S., more and more businesses and institutions of higher education began utilizing them for document preparation and editing, as well as for their data storage and retrieving capabilities.

The UT EADC began utilizing computers in the early 1980s to generate reports utilizing word processing programs. Document files such as questionnaires, client reports, client characteristic sheets and the history of student involvement were maintained. Reports were stored on computer after about report #300 (January 1987). The utilization of desktop computers for report generation improved program efficiency substantially in that the common content of reports did not have to be reproduced; a standard format could be saved in computer memory and used repeatedly to generate the 'skeleton' of subsequent reports. As the standard content of reports evolved, older report versions were purged from computer memory for a lack of necessity to retain them. However, the UT EADC began retaining all issued reports in computer format after about report #400 (June 1990). As an example of the type of reporting information cataloged in reports today, Table I-2 on the following page reveals the sections currently included in assessment reports issued to clients.

In 1992, Rutgers University (Piscataway, New Jersey) became a field management center for the eastern U.S. EADCs. (The western U.S. EADCs are managed by the University City Science Center, a nonprofit research and development agency based in Philadelphia, Pennsylvania.) Two of Rutgers University's directives and requirements as a new field management center were to 1) establish a computerized database containing the majority of newly-generated EADC report information and 2) transfer all of the data

Table I-2. OUTLINE OF CURRENT UTIAC REPORT CONTENTS

Section	Section Name
I.	SUMMARY OF FINDINGS AND RECOMMENDATIONS
A.	Energy Utilization
B.	Waste Generation
C.	Productivity Status
D.	Assessment Recommendations Summary
II.	PERTINENT ABBREVIATIONS & CONVERSION FACTORS
III.	PLANT BACKGROUND
A.	Facility Characteristics
B.	Plant Building Characteristics
C.	Plant Layout
D.	Overview of Manufacturing Process
E.	Process Flowsheet
F.	Previously Implemented Measures
IV.	PLANT ENERGY USE
A.	Major Plant Production and Facility Equipment
B.	Summary of Patterns Energy Use and Costs
C.	Annual Energy Use Breakdown by Energy Type and Equipment Class
V.	PLANT WASTE GENERATION
A.	Summary of Waste Generation and Costs
B.	Breakdown of Waste Streams and Associated Costs
C.	Summary Assessment of Major Waste Related Issues
VI.	PLANT PRODUCTIVITY
A.	Summary of Relevant Metrics
B.	Overview Status of Operations Productivity
VII.	ASSESSMENT RECOMMENDATIONS
VIII.	ADDITIONAL MEASURES TO CONSIDER FOR ENERGY CONSERVATION, WASTE MINIMIZATION AND PRODUCTIVITY ENHANCEMENT
IX.	GUIDANCE FOR FINANCING RECOMMENDED MEASURES
X.	SELECTED REFERENCES

compiled by the previous sole field management agency (University City Science Center) pertaining to previous reports into the database. It was the design to build this database to make the information available to governmental and outside groups for their use, maintaining clients' confidentiality through the omission of some facility and operational statistics pertaining to each assessment performed.

The database that was created is actually composed of two components. The first component is a database of assessment-specific data, for which there is one value for each assessment or facility. Examples are number of employees, plant total floorspace area and facility total annual electricity costs. There are a few exceptions to this "one value" rule, namely SIC code, for which there is occasionally more than one specified for a given facility, and assessment date, for which a few facilities require two sequential visits for completion of an assessment. The second component is a database of recommendation-specific data. For each recommendation made for any facility there are multiple items of information that are recorded, for example, estimated total annual savings in dollars and the total implementation cost (for recommended measures). Tables A-2 (p. 110) and A-3 (p. 112) present a complete set of the data contained in the two Rutgers databases, the assessment-specific database and the recommendation-specific database, respectively.

The Rutgers database is currently accessible through the WWW on the Internet (WWW address <http://oipea-www.rutgers.edu/>). The address is actually the homepage for the Office of Industrial Productivity & Energy Assessment (OIPEA) based at Rutgers, with the databases contained therein at

<http://128.6.70.23/html_docs/dbase.html>. Both databases can be downloaded in several formats, depending on the type of computer and computer software utilized. Additionally, there are files explaining more about exactly how the two databases are constructed and formatted.

Purposes of UTIAC Program Database Development and Objectives of Subsequent Systematic Analysis

There are two main purposes for any individual IAC university creating its own database: to allow for the input and cataloging of additional data for future analytical purposes and to facilitate easier access to the information for the purposes of analysis. Pertaining to the former reason, Rutgers controls the specific nature of information inputted into their database (per the DOE contract), and thereby is accessible through their databases on the WWW. By generating a UTIAC-specific database, more information can be retained to expand the possibilities of future analytical analysis. Pertaining to the latter reason, the need to download more recent upgrades of the university-specific portion of Rutgers database as more university-specific information is added to it, is reduced. Having a database onsite in Knoxville will remove the time-consuming task of downloads. (Considering the current size of the databases, a total of 17 MB, this is significant!)

One other reason for a university-specific database is that the database can be built on a daily basis, as reports are completed and mailed to clients and as client feedback about recommendation implementation becomes available. In a constantly up-to-date state, the database is more valuable as a tool for statistical or categorical analysis.

Specifically, the long-term collection of industrial manufacturing data provides a wealth of information pertaining to small and medium-sized manufacturing facility operations and their specific energy-using and (more recently) waste-generating habits or patterns. The systematic analysis of this data will provide insights into some of the relationships inherent to different kinds of manufacturers (within different SIC codes) as well as characteristics that are not specific to different SICs. These relationships are revealed and explored in Chapter IV.

With the wealth of information available from an onsite, comprehensive database, data relationships can be explored and the resulting information utilized to help in many ways, including, but not limited to the following:

1) Reveal the types of manufacturers that are prone to higher energy consumption and waste generation in the geographical region served by the UTIAC
2) Facilitate selection of future clients likely to be associated with high energy and waste reduction potential
3) Illuminate the industry-specific characteristics affecting the rate of implementation of the recommended measures
4) Reveal linkages between different types of industries relative to types of energy consumed, wastes generated or recommendations implemented (of interest to national and state-level planners)

As examples, these highlight the potential usefulness of a UTIAC database.

Chapter II

UTIAC Database Design and Development

Principles of Database Development

Webster's Ninth New Collegiate Dictionary [Webster's, 1989, p. 325] defines a database as "A collection of data organized especially for rapid search and retrieval." Groliers Encyclopedia CD-ROM [Groliers, 1994] states the following about databases:

> "A database is a large, complex list of facts and information--usually containing text and numbers, and possibly also images, sounds, and video or film clips. Typical databases include telephone directories, airline flight guides, and bibliographic references, but databases have been used to store every conceivable kind of knowledge. Databases are usually, but not always, constructed on computers; an elaborate paper filing system is a sort of database. Databases are distinguished from mere lists by one essential feature of their design: upon request, a specific group of disparate facts can be extracted from the full collection."

Therefore, depending on the complexity of the data being organized, the underlying database design should be relatively simple. Most are comprised of several to many different formats through which the data can be organized, sorted, viewed or printed. More complex databases may have multiple components which combine to form the whole database. Often, these more complex databases utilize 'fields' to relate similar or conjunctive data between components.

When setting up a computer database, designers have multiple software packages from which to choose. Some currently available database software programs include FileMaker by Claris (Santa Clara, California) and Access by Microsoft (Seattle, Washington). Each of these is designed to facilitate convenient organization and use of data. However, depending upon the complexity of the database design, simpler programs can be utilized. For the purposes of the UTIAC's data storage and retrieval, Microsoft Excel was chosen as the utilized database software for its ready availability and ease of use. In addition, all project office employees are already familiar with its operation, eliminating the need for user training.

Considering the type of data to be contained within the UTIAC database, it was required to select a particular database organization. The data to be compiled in the UTIAC database consisted of two main types of data, the same as in Rutgers' overall program database. As indicated in Chapter I, the first type of data is data for which there is only one value per assessment or plant, and the second type is data relevant to recommendations made to facility managers. Therefore, after considering the more complex database nature that would have been required to facilitate having only one component containing all relevant information, it was decided to construct two database components. The purpose of this two-component database was to simplify data organization and retrieval, thus easing operator data input and data manipulation for statistical analysis.

In the UTIAC database, the first type of data is termed 'plant data,' while the second type is referred to as 'recommendation data.'

The UTIAC Database Data Compilation Method

To begin construction of the UTIAC database, the kinds of data that would be organized into the two component databases had to be specified. Thus, a listing of all the types of data to be included was generated. Tables A-4 (p. 114) and A-5 (p. 117) list the types of data that were selected to be included in the plant database and recommendation database, respectively.

Following specification of the data to be collected and included in the UTIAC database, a data compilation method had to be specified. It was decided to utilize in-house data initially, on the reasoning that any information previously transferred offsite might have been modified in an undesirable way. Therefore, the database was built with the most recent information first, working backwards in time to the first assessments performed by UTK in 1976. Even though the UTIAC's computerized reports dated back to approximately report number 400, the Center had only been organizing major report and facility data into a database-like format since report number 507 (first report of the 1994 performance year). Thus, the data was already somewhat organized for approximately the 100 most recent reports, and was consequently assembled into one document and reorganized and formatted to become the UTIAC database plant data component.

Before proceeding with the inputting of the remaining plant-specific information, it was determined that there were two avenues for attaining this information: gather it from the written documents cataloged at the UTIAC or utilize Rutgers' assessment-specific database to download and reformat the available information. After searching the Rutgers database for a listing of the UTIAC data contained therein, it was determined that Rutgers' assessment-specific database contained about 70% of the data desired for the UTIAC database. Thus, this data was downloaded from Rutgers and reformatted as required for the new onsite database.

For the approximate 30% of the data not included in the Rutgers database, individual reports had to be manually reviewed to extract necessary information.

To build the second UTIAC database component (recommendation data), the same approach was taken. The more recent UTIAC data stored onsite was first used to develop the general content organization and to build the initial portion, with Rutgers' second main database component, the recommendation-specific database, being accessed to obtain earlier data. When both UTIAC databases had been filled with data from previous UTIAC work from Rutgers' databases, that data was randomly checked for errors by reviewing various reports and verifying the contents. Fewer than 5 errors were found and corrected.

Comments on the Design and Utility of the Developed Database

With both of the UTIAC database components complete, data manipulation for characterization and statistical analysis has been straight forward. With the uncomplicated layout (column and row data) and differentiation between types of data, the data was easily manipulated or copied into statistical or other software programs for analysis. Variables could be compared or analyzed with very little time spent transferring data between formats.

Tables II-1 and II-2 illustrate the organization of portions of the developed UTIAC plant and recommendation databases, respectively.

Table II-1. EXAMPLE PORTION OF PLANT DATABASE DATA*

	BASIC INFO																		
Ass. #	Assessment Date	Ass. Yr	Ass. Sea-son	Ass. State	Prn-cpl Ass.	# of Studnt Astnts	SIC Code	SIC Group	Annual Sales	# of emps	Plant sq. ft.	Annual Prod. Hours	# of NRG Rec.	# of NRG Rec. Imp.	# of WST Rec.	# of WST Rec. Imp.	# of TTL Rec.	# of TTL Rec. Imp.
									MM$		k sq. ft.							
500	4/27/93	93	g	T	W	1	3743	37	25.0	120	40	6240	7	5			7	5
501	6/2/93	93	s	K	W	1	3999	39	199.0	250	600	6240	6	5			6	5
502	6/15/93	93	s	T	J	1	2499	24	15.0	167	50	2125	8	2			8	2
503	7/15/93	93	s	T	W	1	3553	35	18.0	160	250	2000	9	9			9	9
504	6/17/93	93	s	K	J	1	3432	34	57.0	320	162	6000	14	8			14	8
505	6/21/93	93	s	T	W	1	3432	34	27.0	220	88	6240	8	5			8	5
506	6/22/93	93	s	K	W	1	2328	23	7.0	137	29	2000	8	0			8	0
507	11/23/93	93	f	K	W	1	2851	28	300.0	225	241	7488	9	2			9	2
508	10/21/93	93	f	T	W	1	3448	34	45.0	275	190	5616	5	2			5	2
509	12/2/93	93	w	A	J	2	2330	23	31.0	478	83	6552	5	5			5	5
510	11/4/93	93	f	K	W	1	3251	32	10.0	78	88	6864	6	3			6	3
511	12/17/93	93	w	N	W	1	2284	22	30.0	350	235	8400	8	0			8	0
512	1/7/94	94	w	T	J	3	3498	34	17.0	184	53	2601	6	1			6	1
513	1/11/94	94	w	N	J	1	3620	36	20.0	300	65	3875	11	4			11	4
514	1/28/94	94	w	T	J	1	2361	23	150.0	635	350	2400	5	4			5	4
515	5/27/94	94	g	T	J	2	3531	35	300.0	400	360	7200	6	3			6	3
516	6/9/94	94	s	T	J	1	3069	30	44.0	560	190	6000	9	8			9	8
517	5/12/94	94	g	T	W	2	3079	30	13.0	75	165	4463	6	2			6	2
518	3/2/94	94	g	S	W	1	3566	35	54.0	300	206	8568	8	3			8	3
519	3/2/94	94	g	T	J	1	2834	28	250.0	433	190	6000	4	0			4	0
520	6/22/94	94	s	T	J	1	2861	28	30.0	105	300	7488	7	0			7	0
521	6/28/94	94	s	T	W	1	3111	31	4.0	58	40	4208	6	3			6	3
522	5/31/94	94	g	T	W	2	3086	30	7.0	100	100	7200	6	5			6	5
523	6/6/94	94	s	T	W	1	3086	30	10.0	104	163	6864	8	4			8	4
524	5/24/94	94	g	T	W	2	2361	23	30.0	230	50	2938	6	4			6	4
525	5/21/94	94	g	K	W	1	2891	28	70.0	216	204	6240	9	7			9	7
526	5/16/94	94	g	T	W	1	3089	30	20.0	102	92	6240	7	3			7	3
527i	11/29/93	93	f	T	J	2	3471	34	6.0	120	90	4290	10	8	5	2	15	10
528i	1/6/94	94	w	T	J	4	3714	37	14.0	100	98	4680	7	0	4	0	11	0
529i	3/10/94	94	g	T	J	3	3325	33	5.0	75	20	4940	5	2	5	2	10	4
530i	3/17/94	94	g	T	J	3	2511	25	20.0	400	200	4313	9	1	4	1	13	2

* Abbreviations utilized in this table and in plant database are explained in Table A-6.

Table II-2. EXAMPLE PORTION OF RECOMMENDATION DATABASE DATA*

Audit #	AR #	App. Code	ARC Type	ARC	Imp. Date	Imp. Status	Imp. Cost	Recom. Conserved	Rec. Cost Saved	Imp.? (Y or N)	SIC Group	ARC Group	Primary Resource Info Src Cd	Primary Resource Info Cnsrvd	Primary Resource Info Savd	Secondary Resource Info Src Cd	Secondary Resource Info Cnsrvd	Secondary Resource Info Savd
500	1	3	2	7144	6/1/95	4	5042	290	4608	Y	37	71	E1	290	4608		0	0
500	2	3	2	7261	6/1/95	4	1460	129	1890	Y	37	72	E1	129	1890		0	0
500	3	2	2	4211		5	0	0	0	N	37	42	E1	44	645		0	0
500	4	2	2	4314	6/1/95	4	0	35	513	Y	37	43	E1	35	513		0	0
500	5	3	2	7142	6/1/95	4	242	25	366	Y	37	71	E1	25	366		0	0
500	6	2	2	4221		5	440	0	0	N	37	42	E1	17	249		0	0
500	7	2	2	6212	5/1/95	4	560	4	59	Y	37	62	E1	4	59		0	0
501	1	3	2	7261	4/28/94	3	6930	3193	15769	Y	39	72	E2	3134	15043	E1	59	726
501	2	3	2	7144	4/1/95	4	43092	1180	17524	Y	39	71	E1	1180	17524		0	0
501	3	3	2	2123	8/1/93	1	580	317	1522	Y	39	21	E2	317	1522		0	0
501	4	3	2	7142	4/1/95	4	2457	148	1822	Y	39	71	E1	148	1822		0	0
501	5	1	2	4111	8/1/93	1	170	31	382	Y	39	41	E1	31	382		0	0
501	6	2	2	4211		5	0	0	0	N	39	42	E1	23	283		0	0
502	1	1	2	2212		5	13400	0	0	N	24	22	E9	4972	691	E1	243	5698
502	2	1	2	2131		5	321	0	0	N	24	21	E10	1338	186	E2	40	248
502	3	1	2	4111		5	1980	0	0	N	24	41	E1	134	3141		0	0
502	4	1	2	6218		5	0	0	0	N	24	62	E1	95	2227		0	0
502	5	3	2	7111	8/1/93	1	0	94	2204	Y	24	71	E1	94	2204		0	0
502	6	3	2	7142	8/1/93	1	750	87	2039	Y	24	71	E1	87	2039		0	0
502	7	1	2	2122		5	65	0	0	N	24	21	E1	44	273		0	0
502	8	3	2	3211		5	19400	0	0	N	24	32	E1	0	16070		0	0
503	1	1	2	2513	6/1/94	3	1308	789	4418	Y	35	25	E2	789	4418		0	0
503	2	3	2	7111	1/1/94	3	480	305	5703	Y	35	71	E1	305	5703		0	0
503	3	1	2	4111	6/1/94	3	4628	299	5346	Y	35	41	E1	299	5346		0	0
503	4	3	2	7413	1/1/94	3	2640	279	1562	Y	35	74	E1	279	1562		0	0
503	5	2	2	2312	1/1/94	3	1320	178	3182	Y	35	23	E1	178	3182		0	0
503	6	2	2	4211	9/1/93	1	0	135	2413	Y	35	42	E1	135	2413		0	0
503	7	2	2	4314	1/1/95	4	560	83	1484	Y	35	43	E1	83	1484		0	0
503	8	3	2	7142	9/1/93	1	80	3	54	Y	35	71	E1	3	54		0	0
503	9	1	2	3123	9/1/93	1	0	0	5116	Y	35	31	E1	0	5116		0	0
504	1	3	2	7261	12/1/94	3	3000	3304	9192	Y	34	72	E2	3304	9192	E1	21	246
504	2	1	2	2511	4/1/95	4	1125	1260	8307	Y	34	25	E2	741	2223	E1	519	6084
504	3	3	2	7144		5	45191	0	0	N	34	71	E1	1166	15927		0	0
504	4	1	2	2135	8/1/93	1	1350	736	2208	Y	34	21	E2	736	2208		0	0
504	5	2	2	1233	11/1/93	1	1500	734	2202	Y	34	12	E2	734	2202		0	0
504	6	1	2	2511	1/1/95	4	770	628	7362	Y	34	25	E1	628	7362		0	0
504	7	3	2	3121		5	9600	0	0	N	34	31	E1	609	7140		0	0

* Abbreviations utilized in this table and in recommendations database are explained in Table A-6.

Chapter III
Statistical Analysis Methodology

Prior Analysis of Center Data

The large volume of information assimilated through the production of 596 program assessments over a twenty year period (through September, 1996) has led to numerous, previous limited analyses of different types. The field management centers have generated the majority of material documenting relationships among the characteristics of small and medium-sized manufacturing facilities. The main focus usually encompasses either basic characterizations or implementation information, covering such aspects as relationships within SIC codes about implementation or ways to improve implementation through inventive strategies. Listed in the References are some of the known works generated by the field management agencies (designated with an asterisk (*) at then end of such reference).

One particular recent document published by Rutgers University pertaining to all previous EADC/IAC work and entitled, "Analysis of Energy-efficient Investment Decisions by Small and Medium-sized Manufacturers," [Woodruff, et al., pp. 51-54, 1996] highlighted some of the more important conclusions revealed from almost twenty years of EADC/IAC program operation, and are as follows:

- The majority of small and medium-sized manufacturing industries served fall into SIC groups 20, 30, 34 and 35, or food and kindred products, rubber and miscellaneous plastics, metal fabrication and industrial and commercial machinery and computer equipment, respectively.

- The recommendations made most often are related to building lighting, heating and air conditioning, the building envelope, heat recovery and containment and motors and air compressors.
- The recommendations most likely to be implemented are related to boilers and steam systems, motors and compressors and lighting; those least likely are cogeneration, drying and heating and heat treatment.
- Facilities typically save on the order of 10% of their energy bills for the recommendations implemented and pay for those recommendations within two years.
- The only SIC group that has never had an assessment performed by any program Center is SIC group 21, the tobacco industry.
- As a result of the assessment, facilities tend to adopt additional energy-efficiency measures that were not part of the assessment recommendations, a sign of heightened energy awareness.
- $5,000 was the border over which recommendations initial investment was considered to be a "large investment."
- A payback period of two years or less is typically required for implementation.
- No specific relationship exists between plant size or energy costs and willingness to invest in recommendations with longer payback periods.
- Since the IACs operate without a 'profit motive,' the information the facilities receive is objective and unbiased by vendors or suppliers. Therefore, the IACs fill an important niche in the manufacturing environment by providing this unbiased information at no cost to plants that otherwise might not have the resources to identify potential energy-saving projects or to evaluate them to the point where a decision on implementation can be made.

Additionally, Thomas [p. iii, 1994], a past UTIAC graduate student, investigated the existence of linear relationships between energy consumption and waste generation within a few specific SIC groups (34, 36 and 37). The first main result of this work was that an overall proportional relationship exists for clients between energy use and waste generation. The second was that even though there appeared not to be a relationship between hazardous waste generation and energy consumption, when the quantity of hazardous waste generated was normalized to facility size (floorspace), a significant relationship with energy use was unveiled. This previous thesis demonstrated the overall importance of considering the linkages between industrial energy use and waste generation, which formed the basis for further analysis.

Selection of Statistical Software

In the statistical marketplace, there are many different software packages available for use, and the growth in the number of software-generating companies has followed the growth of computer use in the last 20 years. Some are designed to work directly with typical desktop operating systems such as Windows 3.1 and 95 and Macintosh OS, while others are setup to work through operating systems such as UNIX and VAX. Additionally, some were developed to be used in specific applications and some where developed for broad-based statistical usage. Table A-7 (p. 121) lists some of the major statistical software packages currently available in the marketplace. Of the software packages listed, the best known is SAS, which operates in several different types of computer system environments. "Most companies that are looking to hire statistics-related employees have one prerequisite related to knowledge of software packages, and that is knowing how to use SAS statistical software" [Walker, 1996]. SAS is widely considered to be the standard by which other statistical software packages are measured. SAS performs practically every possible type of statistical analysis that could be needed; however, the system's graphing capabilities are somewhat limited, utilizing standard ASCII characters in the generation of graphs and figures.

SAS statistical software is produced by The SAS Institute (Cary, North Carolina), which also produces JMP statistical software. The SAS Institute's main JMP web page (<www.sas.com/jmp>) describes JMP software [SAS, 1997] in the following way:

> "JMP (Statistical Discovery) software is a highly interactive environment for statistical visualization and exploratory data analysis (for Microsoft Windows and Apple Macintosh users). JMP presents statistics in an easily understood, graphical environment. Data tables are presented clearly in spreadsheet form and are dynamically linked to related graphs. JMP also features integrated capabilities for quality improvement and design of experiments, offering five types of classical designs for estimating the effect of one or more factors on a dependent variable."

Thus, because 1) the makers of SAS produce JMP, 2) JMP has excellent analytical and graphing capabilities and 3) JMP was readily available to the author, this software was chosen for the statistical regression analyses performed in this thesis.

Nature of the UTIAC Database Data

The data contained within the UTIAC database is of four main types: general plant, energy consumption, waste generation and recommendation implementation. Less than 10% of the data is waste generation related information because of the limited time frame (1993-present) during which waste generation issues were dealt with by the UTIAC. Due to the fact that all of the data was obtained through site visits to regional manufacturing facilities for the purposes of developing recommendations for those facilities, the type of data is labeled "observational" data. All data utilized for statistical analysis is classified as either observational or "experimental," where the latter is defined as data collected specifically through experimentation. Thus, since the data contained within the UTIAC database was gathered for purposes other than specific experimental analysis, this data is classified as observational data.

Additionally, data must be divided into the categories quantitative data (continuous) or qualitative data (categorical), where the former represents any value in a continuous series of data and the latter represents a value that fits into one of a group of defined categories. The UTIAC database contains some of each, with the majority of the data being quantitative in nature. Examples of each type are listed below in Table III-1:

Table III-1. EXAMPLES OF QUANTITATIVE AND QUALITATIVE DATA TYPES

Quantitative Data Examples	Qualitative Data Examples
# of employees	Assessment state
Annual natural gas consumption	SIC code
Recommendation implementation cost	Implementation status (Yes or No)

Statistical Methods Utilized for Characterization and Analysis

In common statistical analysis, *sample* data are taken from certain *populations* of interest to search for relationships or information about some type of group. The *population* is a large data set that contains information of interest to the analyst. The *sample* is a portion of that population's data that is extracted for analysis. An example of these types of data might be, for a population, all the manufacturers of plastic injection-molded food containers, and for a sample, the much smaller portion of that group from which one might be able to contact and ask a few questions about their processes (e.g., operational hours, production rates, equipment age). This extraction process leads to the ever important facet in statistical analysis of *randomization*. Many statistical theories are based on the premise that the sample data to be analyzed is randomly chosen from the population of interest. Random sampling of data populations is described as "the most common type of sampling procedure ..... that gives every different sample of fixed size in the population an equal probability (chance) of selection." [Mendenhall, and Sincich, 1996, p. 7]

As pertains to the data compiled into the UTIAC database, it has previously been noted that the type of data is observational. As is typical with observational data, analysis must be tempered somewhat because of the data nature, meaning it is much more difficult to establish cause-and-effect relationships in obtained results. Utilizing experimental data in statistical analysis, the experiment is usually setup so that after the experiment has been designed and independent variable settings decided, "the experimental units are randomly assigned in such a way that each combination of the independent variables has an equal chance of receiving experimental units with unusually high (or low) readings." [Mendenhall, and Sincich, 1996, p. 348] Utilizing this assignment procedure, the analyst can infer that when two sample means are significantly different, this was due to the way the experiment was setup, thus justifying inference about a cause-and-effect relationship. When utilizing observational data in statistical analysis, a statistically significant relationship does not justify the same allowable inferences because the extent of data randomization is not known. The existence of significant relationships in observational data does, however, reveal the contribution of some certain variable 'X' to another variable 'Y'. Therefore, even though we cannot say that, through some of the relationships found in the UTIAC data, one or more independent variables causes a certain type of affect on a different dependent variable, we can say that one or more independent variables contributes to the prediction of a different dependent variable.

For both the characterization and statistical regression analysis parts of this work, Apple Macintosh computer systems were utilized, including a desktop Performa 578 and a PowerBook 145B. For the characterization part of this thesis, various types of graphs were utilized, mostly from DeltaGraph Pro (Monterey, California) software version 3.1, a computer graphing software. This being the simpler part of the analyses, the data to be reorganized was copied from the database in Excel either directly into DeltaGraph, or into JMP for data manipulation and then into DeltaGraph, and manipulated to obtain the appropriate type of information displays, as are presented in Chapter IV. For the regression analyses performed, the statistical computer software JMP version 3.0 was utilized. All of the tabular and graphical information shown was generated in JMP. All results obtained from graphs and output tables are explained immediately preceding or following the graph or table. The regression analyses results are presented in Chapter IV.

Due to the more complex nature of the statistical regression analyses versus the characterizations, some basic analytical methodology information is included in the following text. The example shown is from a preliminary analysis of the first question posed in the Chapter IV regression study, and is meant to help describe some of the more basic information being conveyed. As the results of more detailed analyses are presented in Chapter IV, additional explanation is given as appropriate.

The statistical regression analysis part of this study was performed in attempts to reveal significant relationships in the types of specific of data collected by the UTIAC over the last 20 years. Data types include: small and medium-sized manufacturing facility general plant information, energy consumption and waste generation data and energy and waste related recommendation and recommendation implementation data. The statistical analytical techniques utilized were mainly Simple Linear and Multiple Linear Regression (SLR/MLR). SLR is utilized when investigating the level of correlation between one dependent variable (y) and one independent variable (x). MLR is utilized when investigating the level of correlation between one dependent variable (y) and more than one independent variable (x_1, x_2, x_3, and so on). An example of SLR is shown in Figure III-1 (p. 28) and reveals information about the correlation between facility square footage (under roof - in this case, the dependent variable y) and annual sales (the independent variable x).

Regression analysis normally involves the fitting of a regression line to the data and the subsequent testing of the regression line "fit," or strength of the correlation between the two variables being analyzed. SLR and MLR normally involve the use of the standard least squares method of statistical analysis, where the analysis begins with a scatterplot of the data (simple plotting of data values on a regular linear x and y coordinate graph). The least squares method focuses on developing a line that minimizes the difference between the predicted value ($\hat{y}_i$) and the observed value (y_i) of the dependent variable for a given x. [Montgomery, 1991, p. 125] The Sum of Squares values for each of the variables is SS_{xx} and SS_{yy}, and the sum of squares of the interaction term is SS_{xy}. These terms are defined as follows:

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{(Equation \#1)}$$

$$SS_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \text{(Equation \#2)}$$

$$SS_{xy} = \sum_{i=1}^{n}y_i(x_i - \bar{x}) \qquad \text{(Equation \#3)}$$

These equations are used to develop the best regression straight line fit to the data through the use of two parameter estimates, $\hat{\beta}_1$ which is the slope of the fitted regression line and $\hat{\beta}_0$ which is the y-intercept for the fitted regression line. The equations utilized to generate these values and an example equation using them (the least squares line) are presented next:

$$\hat{\beta}_1 = \frac{SSxy}{SSxx} \qquad \text{(Equation \#4)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \qquad \text{(Equation \#5)}$$

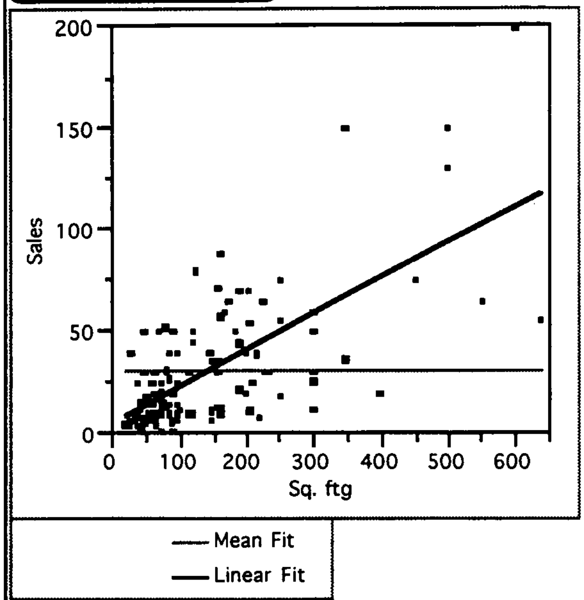
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad \text{(Equation \#6)}$$

The equation generated for the example data shown in Figure III-1 is presented just below the Linear Fit box, and indicates that the least squares method best fit calculates that $\hat{\beta}_1$ equals 0.1749629 and $\hat{\beta}_0$ equals 5.8342209. These values can also be found in the Parameter Estimates table under the Estimate column. The diagonal line shown on the scatterplot is the least squares line and the horizontal line shown is the mean fit (simply the mean value of all the y values).

After the equation of the line is developed, it can be tested for its fit to the data, or strength of fit. This interpretation of the linear fit is an indicator of the strength of the correlation between the two variables, and is found by analyzing the variation within the SS_{yy} value. This value is also called SST, or the total sum of squares, which is a measure of the combination of the following two terms: SSR (sum of squares regression; variation in data explained by regression line) and SSE (sum of squares error; variation in data not explained by regression line). Stated differently, the purpose of the best least squares method regression line is to minimize the sum of squares of the error (SSE). These terms are calculated from the following equations:

$$\text{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad \text{(Equation \#7)}$$

Sales By Sq. ftg



Mean Fit

Mean	31.48319
Std Dev [RMSE]	31.70514
Std Error	2.906405
SSE	118615.5

Linear Fit

Sales = 5.83422 + 0.17496 Sq. ftg

Summary of Fit

RSquare	$\boxed{0.459538}$
RSquare Adj	0.454919
Root Mean Square Error	23.40778
Mean of Response	31.48319
Observations (or Sum Wgts)	119

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	54508.32	54508.3	99.4815
Error	117	64107.14	547.9	Prob>F
C Total	118	118615.47		$\boxed{<.0001}$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>ltl
Intercept	$\boxed{5.8342209}$	3.349238	1.74	0.0841
Sq. ftg	$\boxed{0.1749629}$	0.017542	9.97	$\boxed{<.0001}$

Figure III-1. JMP OUTPUT FROM EXAMPLE DATA ANALYSIS

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \text{(Equation \#8)}$$

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = SS_{yy} = SSR + SSE \qquad \text{(Equation \#9)}$$

The level of regression correlation is relative to several different values calculated from these numbers, which are utilized in the example problem in the Analysis of Variance (ANOVA) table under the Sum of Squares column. The software package (JMP in this case) automatically calculates all these values and outputs the tables of results (as shown in Figure III-1).

The strength of the regression fit is specified in terms of several numbers listed in the tables shown in Figure III-1, but is usually compared by the following two main factors: the R square value (R^2) in the Summary of Fit table and the Prob>F value (p value) in the ANOVA table. The R^2 value, or Coefficient of Determination, always has a value between 0 and 1, and is an indication of model utility, or the "proportion of variation of y explained by the linear association with x." [Walker, 1996, p. 13] This value is obtained utilizing the following equation:

$$R^2 = \frac{SSR}{SST} \qquad \text{(Equation \#10)}$$

where as can be seen from the equation, as the proportion of error explained by the regression line (SSR) increases for any particular model, the R^2 value increases, thus revealing that the model has a greater overall utility.

The p value from the ANOVA table is calculated from the means square terms and F Ratio term. The mean square terms, Mean Square Regression (MSR) and Mean Square Error (MSE), are estimates of the sample variance for each component. The F Ratio is the ratio of two variances and is described by an F probability distribution (assuming the response variable is an independent, normal variable). [Thomas, 1994, p. 26] The equations that are used to calculate these values are:

$$\text{MSR} = \frac{\text{SSR}}{\text{degrees of freedom in the model}} \qquad \text{(Equation \#11)}$$

$$\text{MSE} = \frac{\text{SSE}}{\text{degrees of freedom in the error term}} \qquad \text{(Equation \#12)}$$

$$\text{F Ratio} = \frac{\text{MSR}}{\text{MSE}} \qquad \text{(Equation \#13)}$$

The calculated F Ratio value is compared to an appropriate F distribution value, where it is compared to tabular data utilizing three quantities: the desired alpha (α) or threshold level of significance (usually chosen to be 0.05) and the number of degrees of freedom for the model and error terms (indexing terms). When the p value is less than the chosen level of significance, this indicates that the fitted regression line is "statistically significant" at explaining variation within the model.

Figure III-1 illustrates the results from this analysis. Of the two values noted for comparison, the R^2 value indicates that the regression line fitted explains approximately 46% of the variation within the data (a rather high response considering the observational nature of the data), and the p value from the ANOVA table indicates that the model is statistically significant at explaining the variation within the data (<0.0001). Therefore, to summarize and interpret this example JMP output, the following could be said of the linear fit: The results from this analysis support the conclusion that there is a statistically significant, strong, positive correlation between the facility size in thousands of square feet and the facility's annual sales in million of dollars. This correlation is limited to the range of square footage of less than about 600,000 ft^2, and to annual sales of less than 200 million dollars. Additionally, the correlation indicates that with every 10,000 ft^2 increase in square footage under roof, a increase in annual sales of approximately $7.6 million could be expected.

Some additional information should be noted here about the analysis in general and the particular example used here. In general, it should be stated that the least squares method of statistical analysis gives even weighting to each observation within a model. On occasion, other choices for model analytical tools may be appropriate, however, this is usually the first method used to perform initial data analysis. Also, for all regression analyses, several computer programs had to be utilized to make the JMP output tables and graphs fit within the pages shown in this text. Through that process, the individual observations (points) as shown in the scatterplots were altered in size somewhat, being enlarged either vertically or horizontally. Therefore, care should be exercised when interpreting the data directly from these pages.

Pertaining to the example analysis, the analysis was performed due to the lack of plant size data in terms of square footage. Two of the questions posed in Chapter IV are relative to plant size, and normally, the best variable to utilize would be the direct measurement of floorspace. However, facility square footage data was not recorded for many years of program operation, reducing the quantity of data analyzable with that parameter. In particular, there are only 128 of over 600 assessments for which plant square footage data was recorded, and nine of those could not be used for other reasons (i.e., other missing data). Both number of employees and annual sales data was recorded for almost 500 of over 600 assessments, greatly increasing the quantity of data analyzable. For the correlation of non-SIC-specific data, this is enough data for statistical analysis. However, for SIC-specific analysis, the utilization of facility square footage data would have reduced the average number of data values to compare to approximately 7, too few for this type of analysis.

Additionally, the example data shown was the best fit of all the combinations of variables analyzed, including the SLR of plant square footage to number of employees, annual sales and production hours, and MLR of square footage to number of employees and annual sales, number of employees and production hours, annual sales and production hours and all three combined. From analysis of each of the SLR results, the annual sales/square footage fit had several influential observations and outliers (extreme values found in regression analysis; these will be further explained with graphical illustrations later in Chapter IV), all for companies with annual sales of over $200 million. Removing these influential observations improved the fit to that shown in Figure III-1.

Chapter IV

Analyses, Interpretations and Discussion

Characterizations

Of all the data amassed from the almost 600 completed UTIAC reports, many characterizations could be made revealing and defining interesting information about small and medium-sized manufacturers as well as all the work the UTIAC has done over a 20 year period. Because of this variation, the characterizations performed here are divided into the following relative categories: general UTIAC work background, SIC codes, assessment recommendation codes (ARCs) and recommendation implementation. Each of the characterizations within these relative groupings conveys different types of information.

BACKGROUND CHARACTERIZATIONS

In being one of the first Centers to provide regional manufacturers with no-cost assessments, the UTIAC has generated 596 assessment reports as of the end of the UTIAC 1995 performance year. Figure IV-1 reveals the number of assessments performed sorted by plant size reported in total floorspace in units of square feet. From Figure IV-1, it can be seen that just under one-third of the total number of assessments performed were for plants under 100,000 ft^2 in size. Additionally, the trend is that as plant size increases, the overall number of assessments performed decreases, excluding the above 500,000 ft^2 range. Through this change in trend at the larger plant size-end of the scale, an interesting point is to be noted. Typically each year the UTIAC receives a small number of requests from companies of larger size to perform an

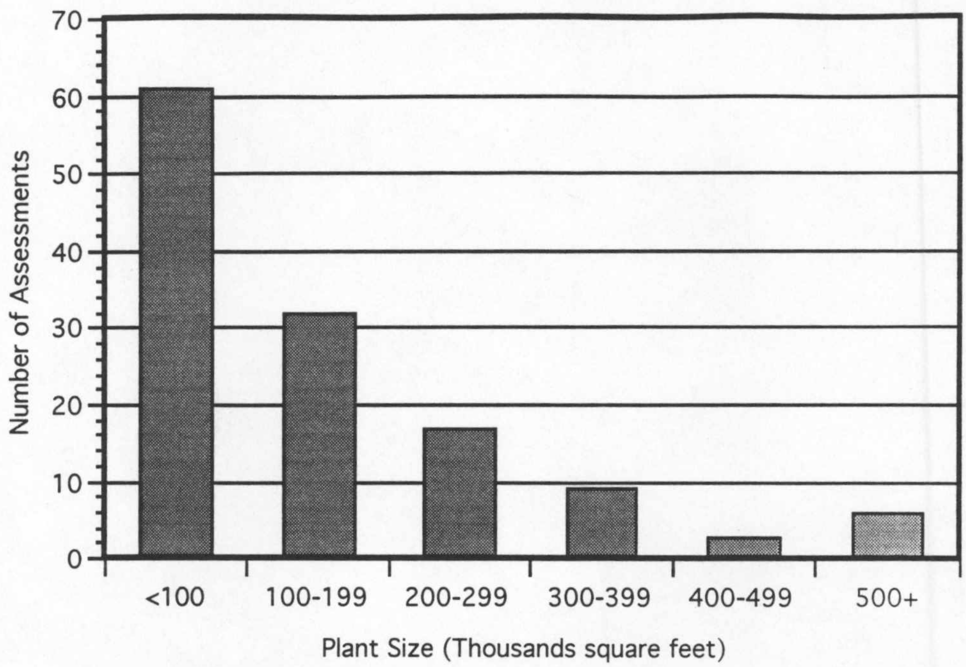


Figure IV-1. NUMBER OF ASSESSMENTS PERFORMED BY PLANT SIZE

assessment. Meanwhile, it is normal UTIAC protocol for the student employees to select facilities that would like to receive an assessment, therefore, the majority of facilities assessed were selected by the UTIAC to receive an assessment. The notable point is that the larger facilities receiving assessments were most likely those requesting our services rather than responding to a solicitation. More than likely, the facilities in this range were considered "exceptions," in that they did not meet three of the four qualifying criteria. In these cases, the UTIAC had to obtain permission from its field management center to perform the assessments for the larger facilities.

Another interesting point to make about Figure IV-1 is that the overall number of assessments depicted is only 128. (The totals will vary from graph to graph depending on the specific nature of the information being graphically depicted.) Most likely, this is due to the fact that as the IAC program has grown and matured through the years, it has developed a broader scope and thus more information has been required to investigate newer avenues of savings for the facilities. An example of a type of recommendation where square footage might have been important is in determining if a facility has an appropriate amount of space cooling or heating equipment. This is usually determined by associating the class of work being performed in a particular area of the plant (e.g., office area, manufacturing area) with that part of the facility's floorspace in square feet. Specifically, the UTIAC did not begin gathering plant square footage information until report number 475.

The next two figures, Figure IV-2 and Figure IV-3, depict the number of assessments distributed by the facility's number of employees and annual sales, respectively. These relationships follow the same general trend as in Figure IV-1 with decreasing assessment numbers associated with plants having a larger number of employees and higher annual sales. A notable point pertaining to these figures is that in Figure IV-3, the $50-$100 million annual sales range is uniquely large compared to its neighbors, and is most likely contributed to the "lumping" of several categories into one range.

When the original three EADCs where established in 1977, the geographical service area of each Center was defined by a radius from the university of 150 miles. For the Knoxville location, this included out-of-state work in North and South Carolina, Virginia and Kentucky. After several years, this lead to

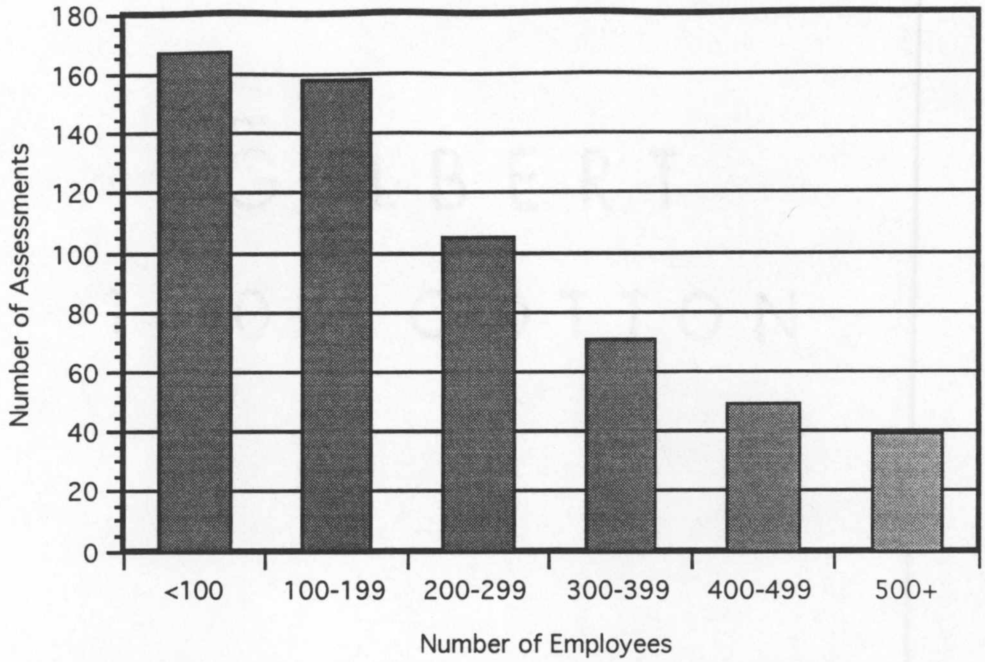


Figure IV-2. NUMBER OF ASSESSMENTS PERFORMED BY NUMBER OF EMPLOYEES

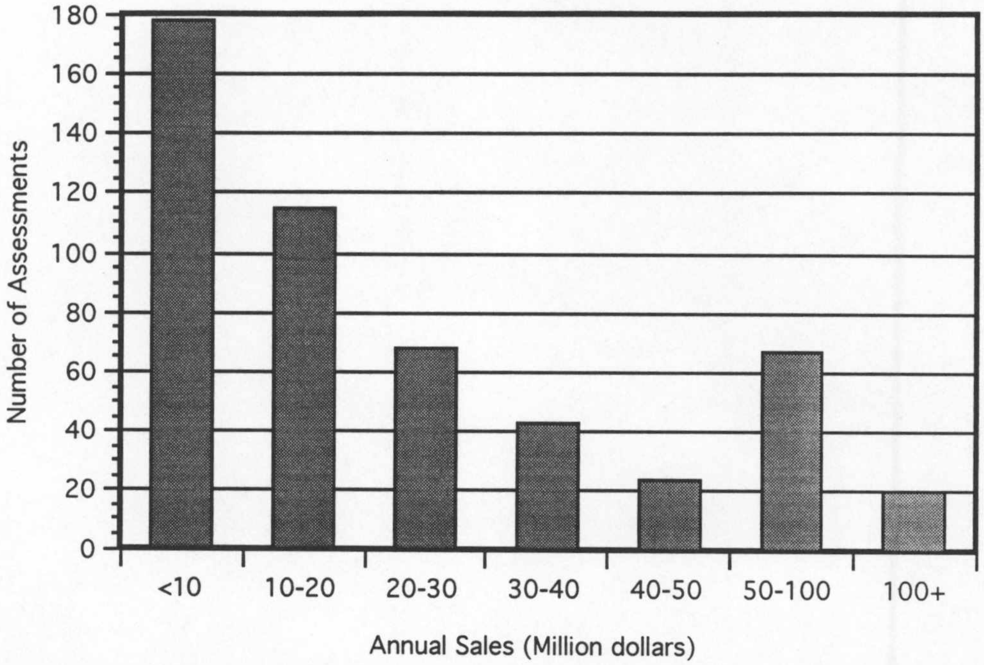


Figure IV-3. NUMBER OF ASSESSMENTS PERFORMED BY ANNUAL SALES occasional confusion between Centers because of overlapping operating areas. In or around 1994, the DOE directed the field managers to reallocate the Centers' service areas as a basis of specific nonoverlapping counties within states served by the program. Only under special circumstances would one Center operate outside of the specified counties. This decision lead to some rather deformed operational 'circles' around the Centers, but eliminated earlier confusion. As of the end of fiscal year 1995, Figure IV-4 reveals the percentage of the total number of UTIAC assessments performed within each state (and the actual number in parenthesis). Additionally, Figure IV-5 shows the county breakdown utilized for the five-state region served by the UTIAC, the University of Tennessee is located in Knoxville at the star symbol.

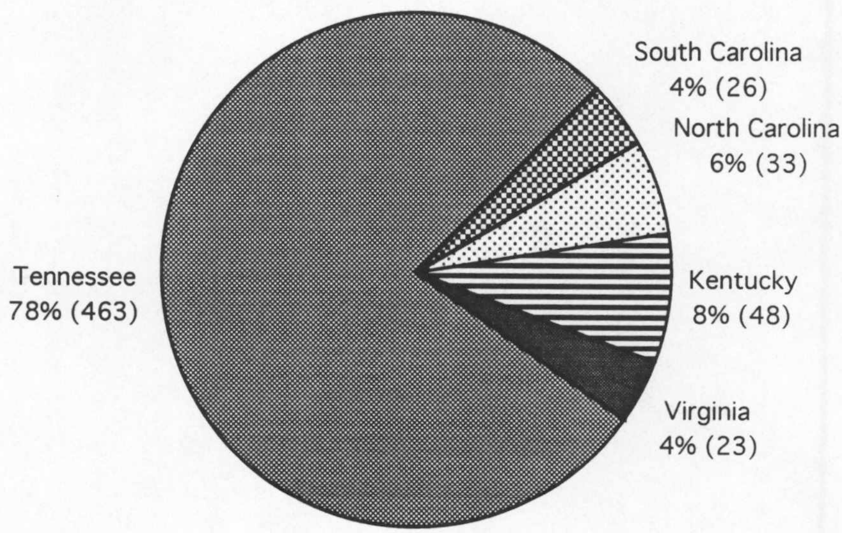


Figure IV-4. DISTRIBUTION OF ASSESSMENTS PERFORMED BY PERFORMANCE AREA (STATE)

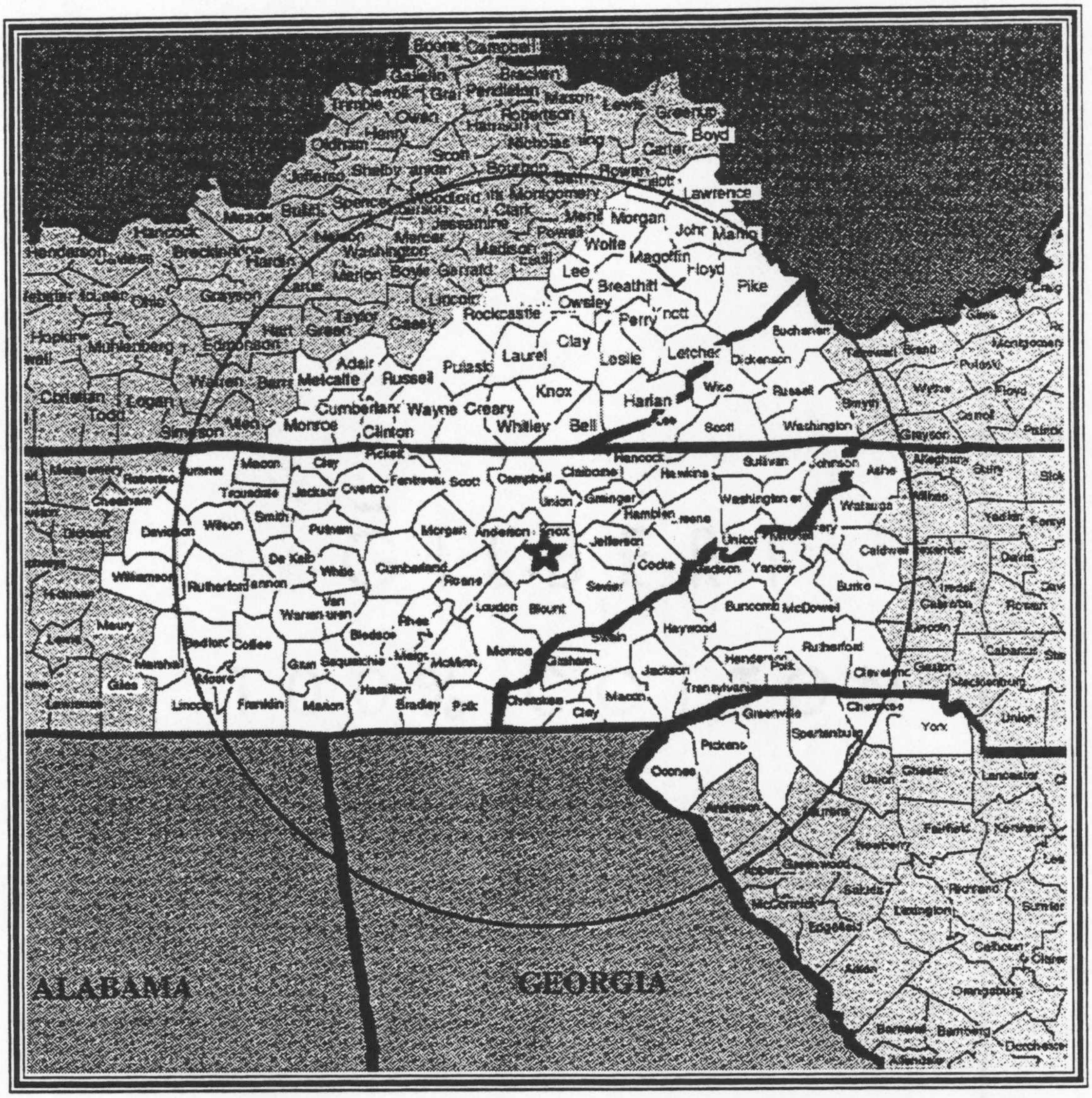


Figure IV-5. FIVE-STATE SERVICE AREA FOR THE UTIAC

SIC CHARACTERIZATIONS

As described in Chapter I, the Office of Management and Budget, a division of the Executive Office of the President of the United States, developed an organizational system by which all industries could be classified, "to facilitate statistical economic analysis and reporting of the state of the U.S. economy based on enterprises engaged in production, trade and service." [Wandell, 1996] This numbering system is termed the Standard Industrial Classification (SIC) system. As part of the original focus of the EADC program, assessments were meant to serve manufacturers, as is listed in the SIC manual, and covers SIC groups 20 through 39 (as listed in Table I-1, p. 4).

Through 596 UTIAC assessments, Figure IV-6 reveals the breakdown of assessments performed within each SIC group and state. Because of the more centralized focus of each IAC for many years around its home university location, it is not surprising that the number of assessments performed in Tennessee greatly outweighs those performed in the other four states. Additionally, it can be seen that the SIC groups served most frequently within each state are similar (SIC groups 20, 22, 23 and 34 through 36). This fact tends to lend validity to the premise that these SIC groups are most likely the smaller manufacturers in the UTIAC operating area (reference Figures IV-1 through IV-3).

To further characterize the work done by the UTIAC, the average plant size, number of employees and annual sales were all plotted against SIC groups in Figure IV-7. Owing to the use of a similar abscissa, only trends should be compared in this figure, not scale. Some general trends across the three

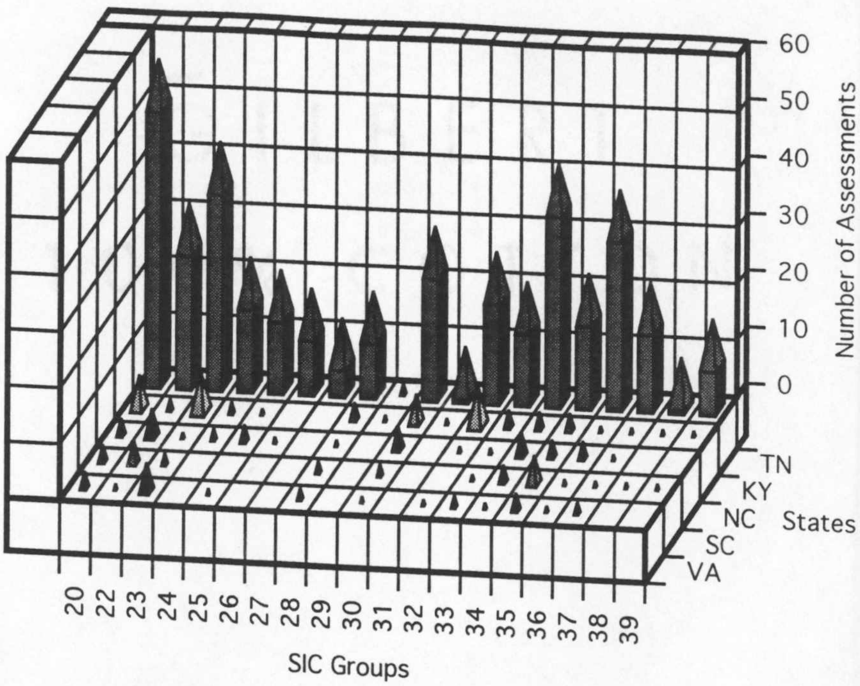


Figure IV-6. NUMBER OF ASSESSMENTS PERFORMED BY SIC GROUP AND STATE

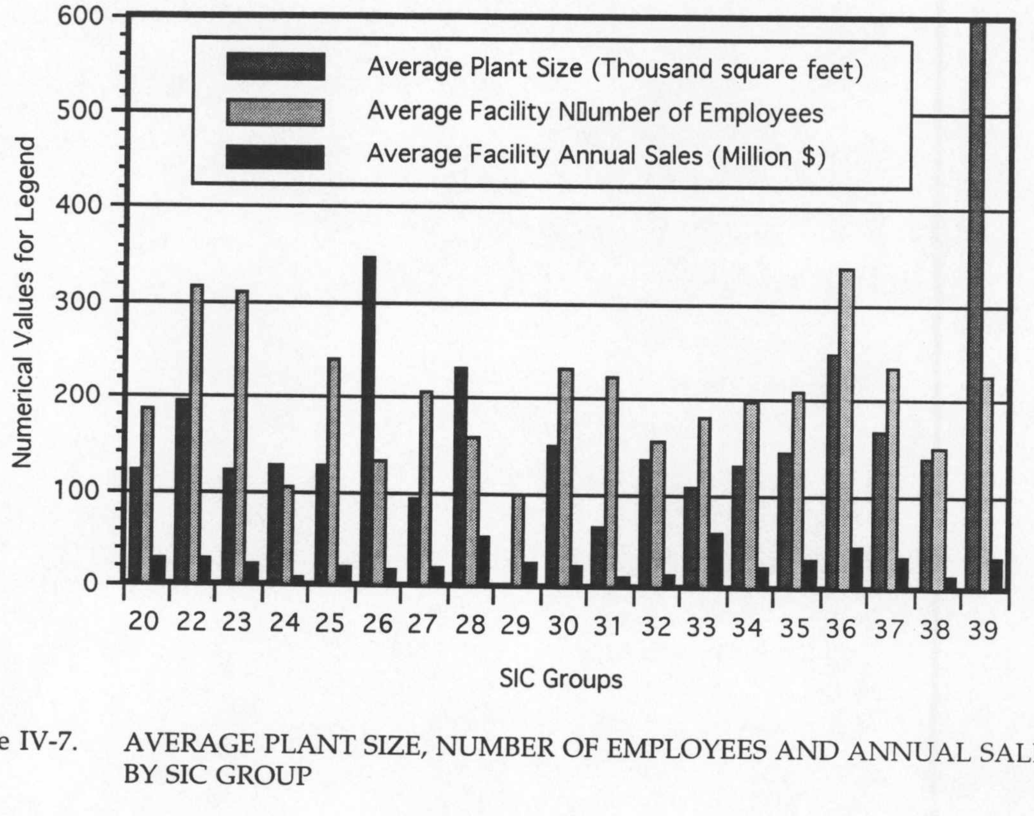


Figure IV-7. AVERAGE PLANT SIZE, NUMBER OF EMPLOYEES AND ANNUAL SALES BY SIC GROUP categories can be seen, for example, in the SIC range 34 through 38. Through these SIC groups, as one category increases or decreases, the other two follow suit to some degree. However, there is also discontinuity across the three categories, as in SIC groups 24 through 28. The discontinuity reveals the lack of relationship throughout *all* SIC groups served for these categories, yet the visible trends reveal the possible relationships between some particular SIC groups. Additionally, it should be noted that some of the categories within specific SIC groups are averages of few actual values due to the small number of assessments performed within those SIC groups (reference Figure IV-6). An excellent example is the average plant size for SIC group 39; there was only one facility assessed in that SIC group from which the plant square footage was obtained. Because of the number of facilities for which plant square footage was not obtained (over half of 600 assessments), this value as an average is erroneous. However, after noting this occurrence, there are only four SICs in the average number of employees and average annual sales categories for which there were fewer than ten values to be averaged, and only five in the average plant size category for which there were fewer than five.

An additional interesting characterization by SIC group reflected in Figure IV-8 represents the number of recommendations made per assessment in the energy and waste categories averaged for all assessments. For the energy category, the range fluctuates little, varying from about 4.3 to about 6.3, with an overall average of approximately 5.3. For waste recommendations, there exists greater variation, with a minimum of zero (due to lack of any assessments involving waste analysis within three SIC groups) and a

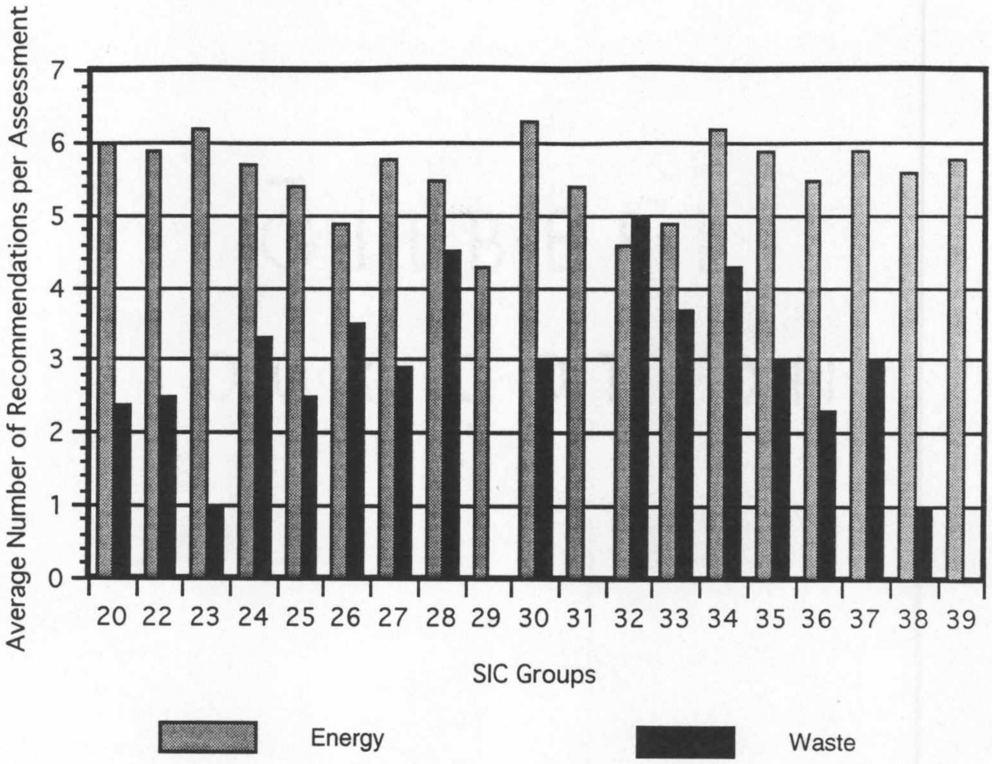


Figure IV-8. AVERAGE NUMBER OF ENERGY AND WASTE RECOMMENDATIONS PER ASSESSMENT BY SIC GROUP

maximum of 5. Broken down by total number of assessments performed for each category, all 596 assessments performed through the end of fiscal year 1996 had energy analysis components, whereas only 51 assessments had waste analysis components. More specifically, the average number of assessments within each SIC group averaged to calculate the numbers utilized in the figure was 31 for energy, and only three for waste.

Lastly, to reveal more about the relationship between the number of assessments performed and number of recommendations made within individual SIC groups, Figure IV-9 shows the total number of assessments and total number of recommendations made within each SIC group. It is

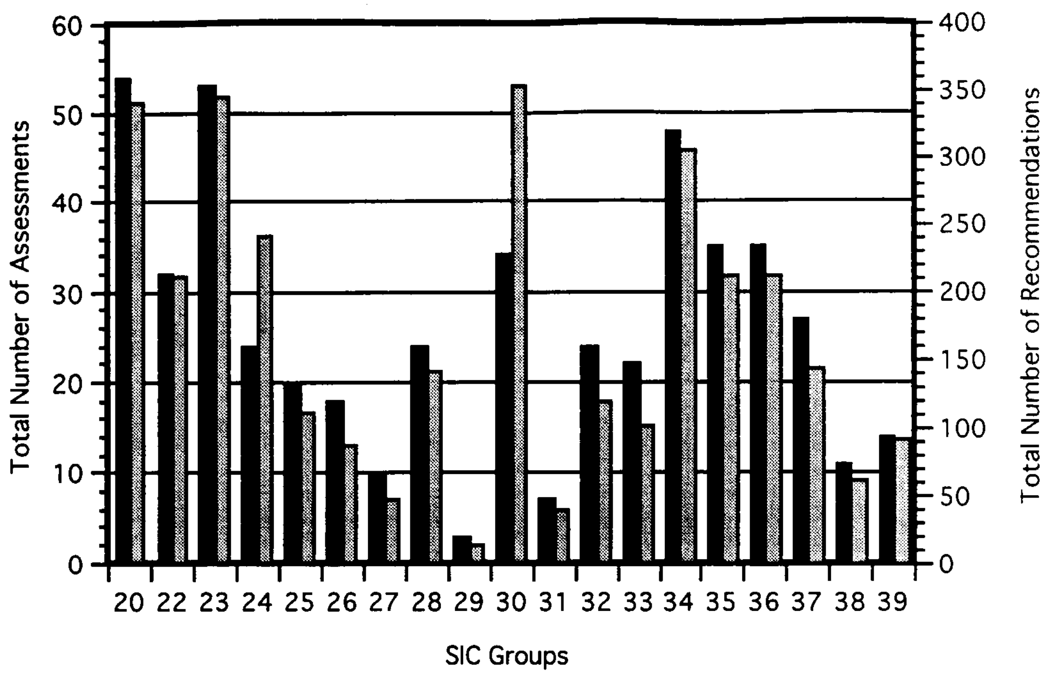


Figure IV-9. NUMBER OF ASSESSMENTS AND RECOMMENDATIONS BY SIC GROUP

apparent that there is not much variation across the manufacturing SIC range except for two specific SIC groups: 24 (lumber and wood products) and 30 (rubber and miscellaneous plastic products). For these two specific SIC groups, the ratio of recommendations to assessments is approximately 1.5:1 versus about 1:1 for the remaining SIC groups. (For a tabular listing of the information displayed in Figure IV-9, see Table A-8, p. 122.)

ARC CHARACTERIZATIONS

In 1993, the OIPEA developed the Assessment Recommendation Code (ARC) system for the purpose of "systematic categorization of assessment recommendations .... in[to] the program database." [Muller et al., 1995, p. 4] These codes organized the recommendations that were being made by the 30

IACs into broad categories and specific groups covering all types of recommendations pertaining to systems and equipment found in manufacturing facilities. Soon after its development, the IACs utilized the codes to categorize each recommendation reported to Rutgers for incorporation into the program database. As with most large organizational systems, there have been several modifications over time involving the reorganization of some classifications and the addition of new ones. The current version is 6.0. This version contains major categories for energy management, waste minimization/pollution prevention and productivity enhancement recommendations, and are identified as 'recommendation types' by the first numeral in the code being 2, 3 or 4, respectively.

The ARC codes are 6-digit codes in the format (1.2345.6) where the first digit lumps the recommendations into one of the above mentioned major types (energy, waste or productivity). The next four digits are similar to SIC codes in that the first two digits group recommendations by similar types of systems and equipment, and the last two digits break those groups into individual recommendations pertaining to particular types of systems and equipment. The sixth and last digit in the series reveals the application code, or what level of production importance the recommendation relates to, which varies between 1 and 4. Table A-9 (p. 123) reveals more specifics about the application codes. All of the major ARC categories and subcategories and their two-digit codes are presented in Table IV-1. Additionally, the number of recommendations made to facilities by the UTIAC within the subcategories is shown in Table IV-1. As can be seen from this table, the total number of

Table IV-1. ARC CATEGORIES AND SUBCATEGORIES, INCLUDING NUMBER OF RECOMMENDATIONS MADE BY THE UTIAC WITHIN EACH ARC SUBCATEGORY

ARC Category	Subcategory	2 digit code	Number of recommendations* 596	Number of recommendations* 566
Energy Management				
Combustion systems	Direct-fired operations	11	21	20
	Boilers	12	124	120
	Fuel switching	13	21	19
Thermal systems	Steam	21	331	326
	Heating	22	6	6
	Heat treatment	23	1	1
	Heat recovery	24	131	125
	Heat containment	25	264	256
	Cooling	26	10	10
	Drying	27	0	0
Electrical power	Demand mgmt.	31	37	32
	Power factor	32	34	33
	Generation of power	33	0	0
	Cogeneration	34	9	8
	Transmission	35	0	0
Motor systems	Motors	41	221	207
	Air compressors	42	474	450
	Other equipment	43	109	99
Industrial design	Systems	51	12	11
Operations	Maintenance	61	37	37
	Equipment control	62	289	280
Buildings and grounds	Lighting	71	509	480
	Space conditioning	72	265	249
	Ventilation	73	21	21
	Building envelope	74	89	86
Ancillary costs	Administrative	81	12	9
	Shipping, distribution and transportation	82	0	0
Alternative energy usage	General	91	1	1
Total number of recommendations:			3,028	2,886

ARC Category	Subcategory	2 digit code	Number of recommendations* 596	Number of recommendations* 566
Waste Minimization/Pollution Prevention				
Operations	Procedures	11	30	16
	Waste stream contamination	12	2	1
	CAD/CAM	13	1	1
Equipment	General	21	8	4
Post generation treatment/ minimization	General	31	4	2
Water use	General	41	9	4
Recycling	Liquid waste	51	5	5
	Solid waste	52	5	2
	Other materials	53	36	15
Waste disposal	General	61	19	6
Maintenance	Cleaning/ degreasing	71	7	7
	Spillage	72	5	4
	Other	73	0	0
Raw materials	Solvents	81	8	7
	Other solutions	82	2	1
	Solids	83	6	4
Total number of recommendations:			147	79

ARC Category	Subcategory	2 digit code	Number of recommendations* 596	Number of recommendations* 566
Direct Productivity Enhancement				
Manufacturing enhancements	Bottleneck reduction	11	0	0
	Defect reduction	12	1	0
Purchasing	Raw materials	21	2	0
	Ancillary materials	22	0	0
	Capital	23	0	0
Inventory	Just in Time	31	0	0
	Other inventory controls	32	0	0
Labor optimization	Scheduling	41	0	0
	Practices/ procedures	42	0	0
	Training	43	3	0
	Automation	44	0	0
Space utilization	Floor layout	51	0	0
	Rental space	52	0	0
Reduction of downtime	Maintenance	61	0	0
	Quick change	62	0	0
	Power conditioning	63	0	0
	Alarms	64	0	0
	Other equipment	65	0	0
Management practices	Total Quality Management	71	0	0
	Certifications	72	0	0
Other administrative savings	Taxes	81	0	0
	Fees	82	0	0
Total number of recommendations:			6	0

*The column heading numbers shown reveal to which assessment the numbers in that column are indicative. This is separated because assessments 567-596 have not have implementation information gathered on them yet.

recommendations made for energy conservation (3,028) vastly outweighs the total for waste minimization (147) or productivity enhancement (6). Due to the limited information pertaining to the latter two classes of recommendations, only energy recommendations will be addressed in the following ARC-related data.

Figure IV-10 presents the total number of energy recommendations made within the major ARC categories, while Figure IV-11 reveals the number of recommendations made within the ARC subcategories. Notice in Figure IV-11 that the vertical lines within the graph separate the various major categories from one another. From Figure IV-10, it can clearly be seen that the majority of recommendations made by the UTIAC have been relative to three major categories: thermal systems (21 - 27), motor systems (41 - 43) and

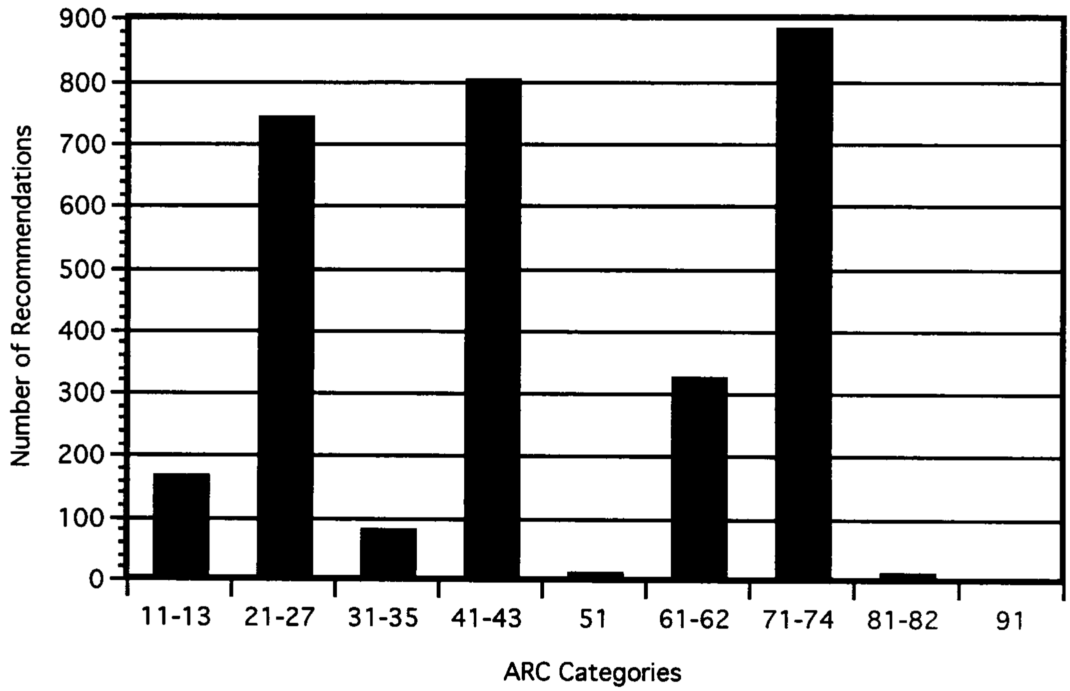


Figure IV-10. NUMBER OF ENERGY RECOMMENDATIONS BY ARC CATEGORIES

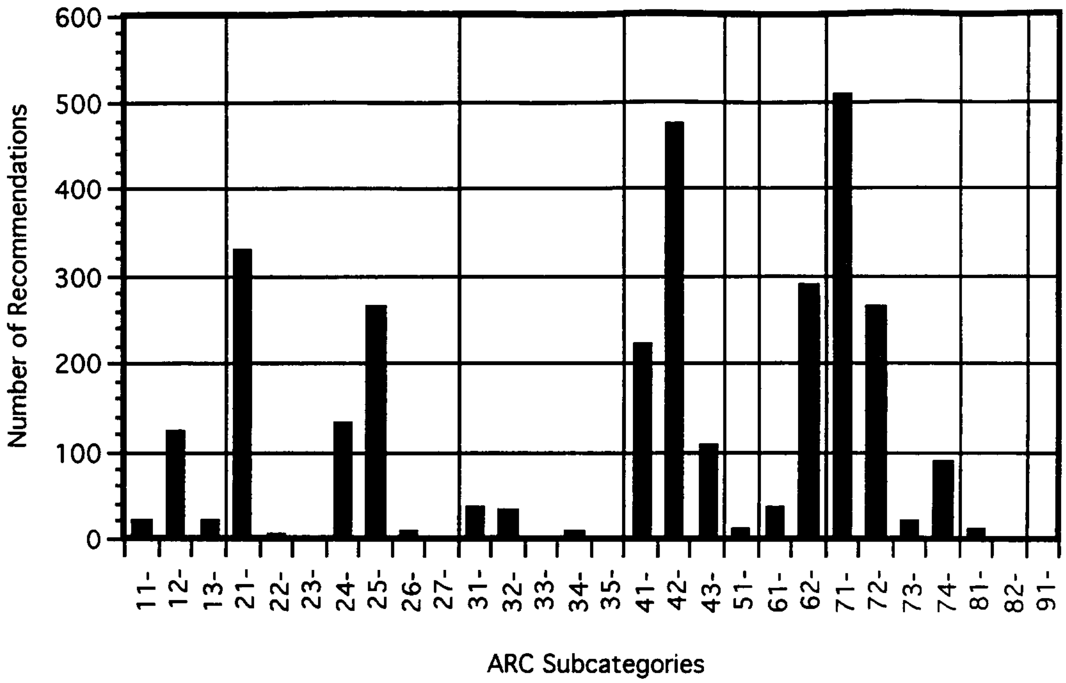


Figure IV-11. NUMBER OF ENERGY RECOMMENDATIONS BY ARC SUBCATEGORIES

buildings and grounds (71 - 74). These three categories of recommendations make up just over 80% of all the recommendations made pertaining to energy conservation (2,431 out of 3,028 recommendations). Undoubtedly, with reference to Table IV-1, these three categories contain the equipment types most likely to be found in the majority of manufacturing facilities, and include such common types of systems as steam supply and distribution systems, air compressors, lighting and space conditioning equipment. Additionally, comparing the distribution of number of recommendations in Figure IV-11 to the distribution in Figure IV-10, one can see that those specific systems and equipment (as subcategories) are those about which the most recommendations have been made on a long-term historical basis.

Additionally, even though there are no figures for the waste and productivity-related recommendations versus the ARC categories and subcategories, some meaningful observations can still be made from Table IV-1 data. Over 55% of the total number of waste-related recommendations were made in one of the following subcategories: operating procedures (30), recycling of miscellaneous materials (39) and waste disposal (19). For productivity recommendations, the major issues addressed by the IAC-recommended measures (through report number 596) are within the employee training (43) and raw material purchasing (21) ARC subcategories.

RECOMMENDATION IMPLEMENTATION CHARACTERIZATIONS

Each IAC is contractually required to contact the companies served within 6 months to a year after the assessment to obtain data on implementation of the recommendations. Historically, for all of the IACs considered collectively, the implementation rate for all recommendations made is about 50%. For the UTIAC, the implementation rate has been of almost exactly the same magnitude, calculated to be 52% through 566 reports.

To begin investigating implementation characterizations, the rates of recommendation implementation were graphed against four general categories (state in which assessment was performed, annual sales, annual production hours and number of energy recommendations in a report) and are shown if Figures IV-12, IV-13, IV-14 and IV-15, respectively. In Figure IV-12, the most notable feature is that of there being far more recommendations for Tennessee clients than for the other states. This is due to 1) the exclusive enlistment of Tennessee only firms in the early program years and 2) the

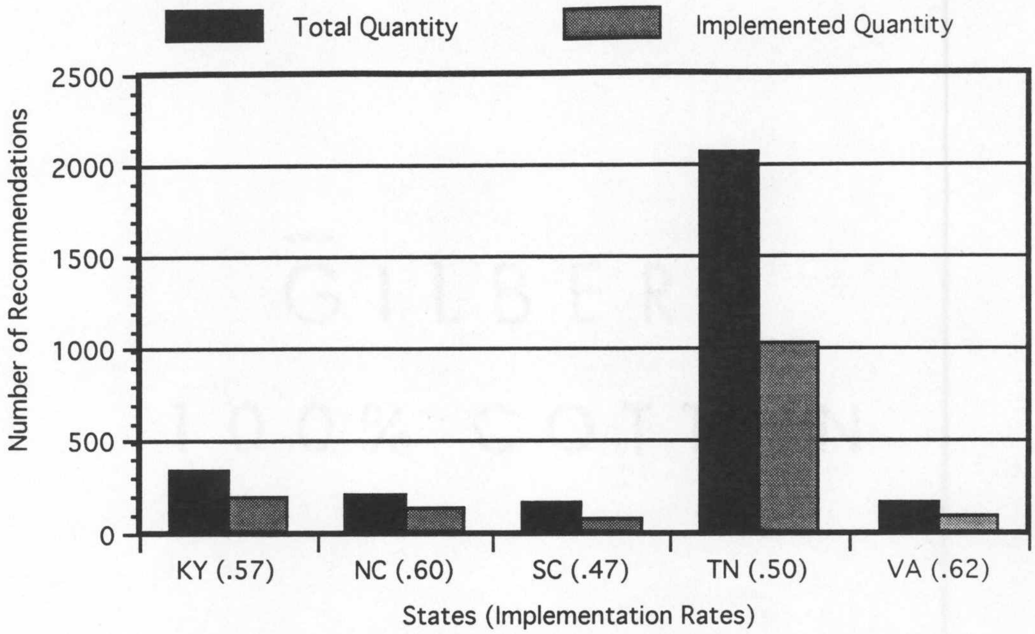


Figure IV-12. TOTAL AND IMPLEMENTED NUMBER OF RECOMMENDATIONS BY STATE

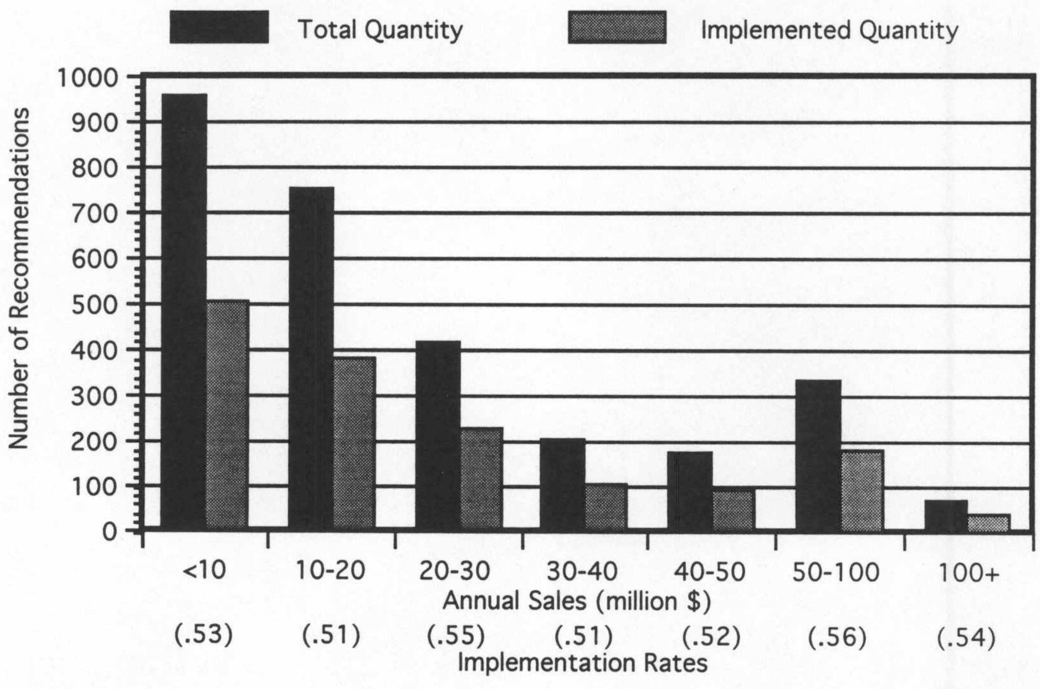


Figure IV-13. TOTAL AND IMPLEMENTED NUMBER OF RECOMMENDATIONS BY ANNUAL SALES relatively large Tennessee geographical area as a portion of he more recently defined several-state service area. The more important characteristic evident in this figure, however, is that for clients in three of the four states outside of Tennessee that the UTIAC serves (Kentucky, North Carolina and Virginia), the implementation rate is about 10% higher than Tennessee's clients, on average. Only South Carolina clients lag Tennessee clients with a reported 47% implementation rate.

Figure IV-13 shows the variation in implementation rate with increasing annual sales. The implementation rate range does not vary substantially with respect to annual sales (only 5%), and there is no significantly differentiating characteristic in this figure. Referencing Figure IV-3, one can also see that there is not a significant difference in the number of recommendations made with respect to the number of assessments performed within the various annual sales brackets.

Figure IV-14 illustrates the implementation rate variation within different annual production hours for facilities served by the UTIAC. The most significant point to make about this figure is that above 6,000 hours of annual manufacturing operations time, the implementation rate is between 54 and 58%. Below 6,000 hours, the categories with the greatest number of recommendations (and thus greater weight when dealing with statistical significance) have smaller implementation rates; the two categories with the highest implementation rates have only 46 total recommendations between them (below 2,000 hours). Thus, this data indicates that future work with manufacturing facilities of greater than 6,000 annual operating hours could

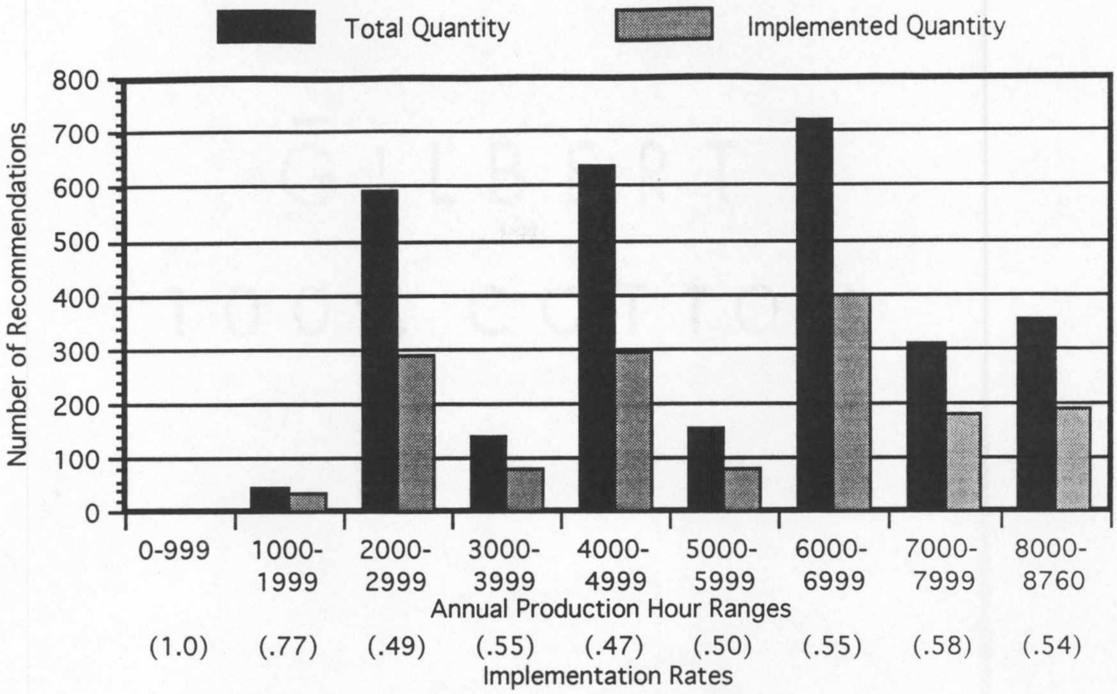


Figure IV-14. TOTAL AND IMPLEMENTED NUMBER OF RECOMMENDATIONS BY ANNUAL PRODUCTION HOURS

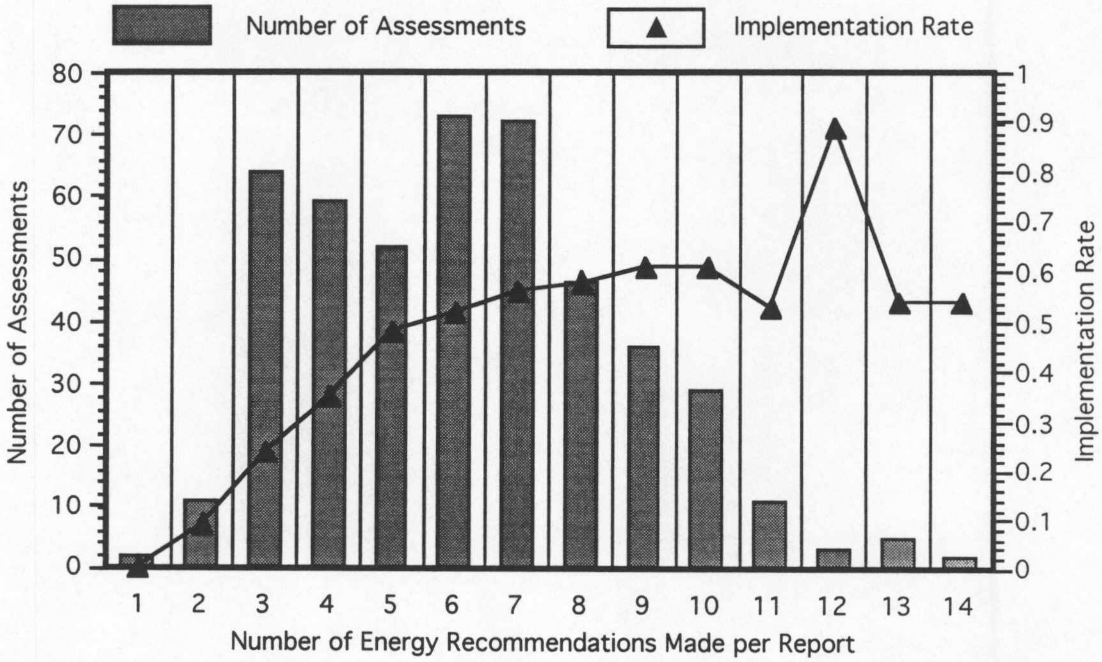


Figure IV-15. NUMBER OF ASSESSMENTS AND OVERALL IMPLEMENTATION RATE BY NUMBER OF ENERGY RECOMMENDATIONS MADE PER ASSESSMENT possibly, over time, increase the UTIAC's implementation rates and thus, the overall program benefit. It is also noted that a likely contributing factor to this effect is the increase in savings with operational hours which usually results in shorter payback periods, and thus enhanced implementation rates.

The last of these general implementation characterizations is Figure IV-15, which is presented as a combination. This graph depicts the number of reports and the implementation rate within each X-axis category (a particular number of recommendations made per report). The general trend of increasing implementation rate with an increase in the number of recommendations made per report does not have an obvious explanation, other than the fact that in its history, the quantity of recommendations made per report by the UTIAC has increased slowly over the years (known only because of the extensive amount of time the author has spent with the data). However, the reason the number of assessments part of the graph was included was to shed further light on the relationship between the implementation rate and number of recommendations. It can clearly be seen that as the implementation rate increases, the total number of assessments performed in each category decreases, and thus there are fewer actual data values over which the implementation rate is being averaged.

Nevertheless, the most interesting point to make about Figure IV-15 is the distribution of implementation rates. They increase steadily from 1 to 9 recommendations per report, and then declines slightly and levels off, 12 recommendations per report notwithstanding. The visible trend within the majority of this data (not including the extreme value) indicates that, in general, as the number of recommendations increases, so does their implementation rate.

Figure IV-16 presents the recommended measure implementation rate as it relates to the varying SIC groups served by the UTIAC. This figure reveals significant fluctuation or implementation rates within SIC groups. Excluding the three lowest implementation rates for printing and publishing (SIC 27, 13%), paper and allied products (SIC 26, 26%) and electronic and other electric equipment (SIC 36, 36%), the SIC groups have implementation rates which vary between 40% and 64%. The SIC groups served by the UTIAC each with greater than 200 total recommendations all have implementation rates of greater than 50%, suggesting that implementation rates appear to reach and remain above 50% as the total number of recommendations made within

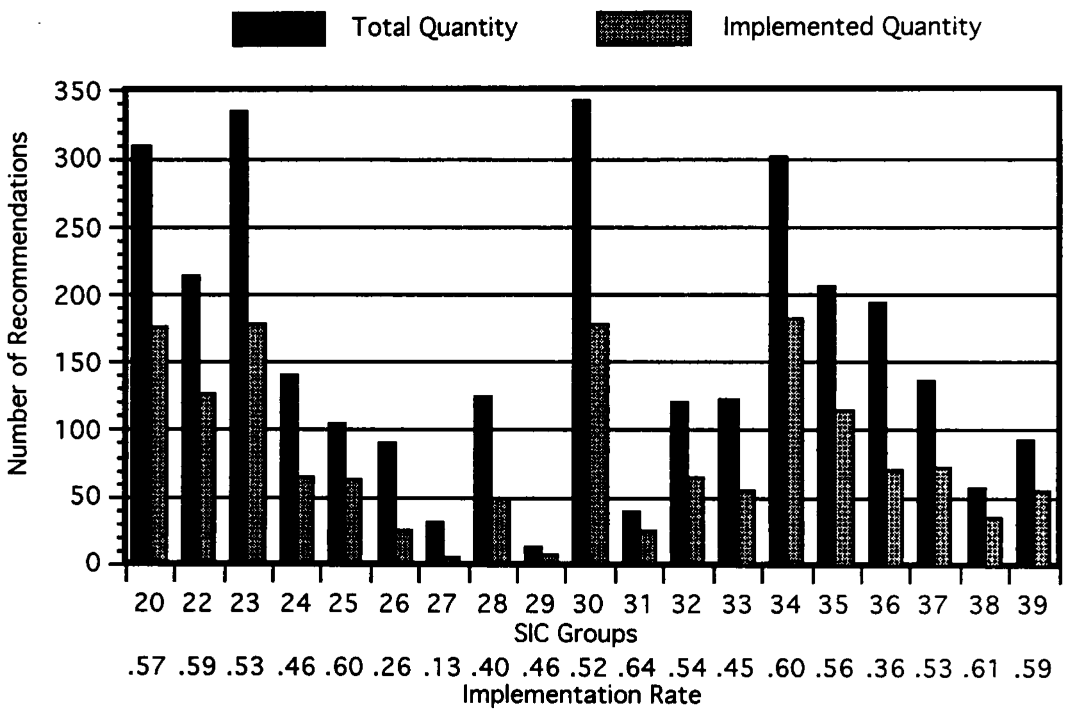


Figure IV-16. TOTAL AND IMPLEMENTED NUMBER OF RECOMMENDATIONS BY SIC GROUPS specific SIC groups reaches 200 or more. However, the groups containing the highest implementation rates (60% and up) are found throughout the range of number of recommendations made (e.g., SIC group 31 @ 64% and 34 @ 60%), and this exposes the fact that implementation rates do not follow a specific trend for under 200 recommendations.

Figures IV-17 and IV-18 again displays the distribution of the number of recommendations made, number of recommendations implemented and implementation rates, but are divided into ARC categories instead of SIC groups. Figure IV-17 depicts information about energy recommendations alone, and Figure IV-18 illustrates corresponding information about waste recommendations. Initially, it can be seen that the number of energy recommendations made far outweighs the number of waste recommendations, which, as mentioned previously, is due largely to the contract time periods associated with the EADC program (energy audits only from 1976-1993) and the broadening IAC program (energy and waste assessments 1994-present, with productivity included for first time in 1996-97 contract year).

Within Figure IV-17, it can be seen that the implementation rates within ARC categories *do* appear to be number-of-recommendations specific, with the five categories with the total number of recommendations above 100 having higher magnitude implementation rates, and the remaining three ARC categories having lower magnitude implementation rates, ARC category 91 excluded (reference Table A-9 for category descriptions.) Owing to the scale used in Figure IV-17 it is not apparent that the ARC energy category 91

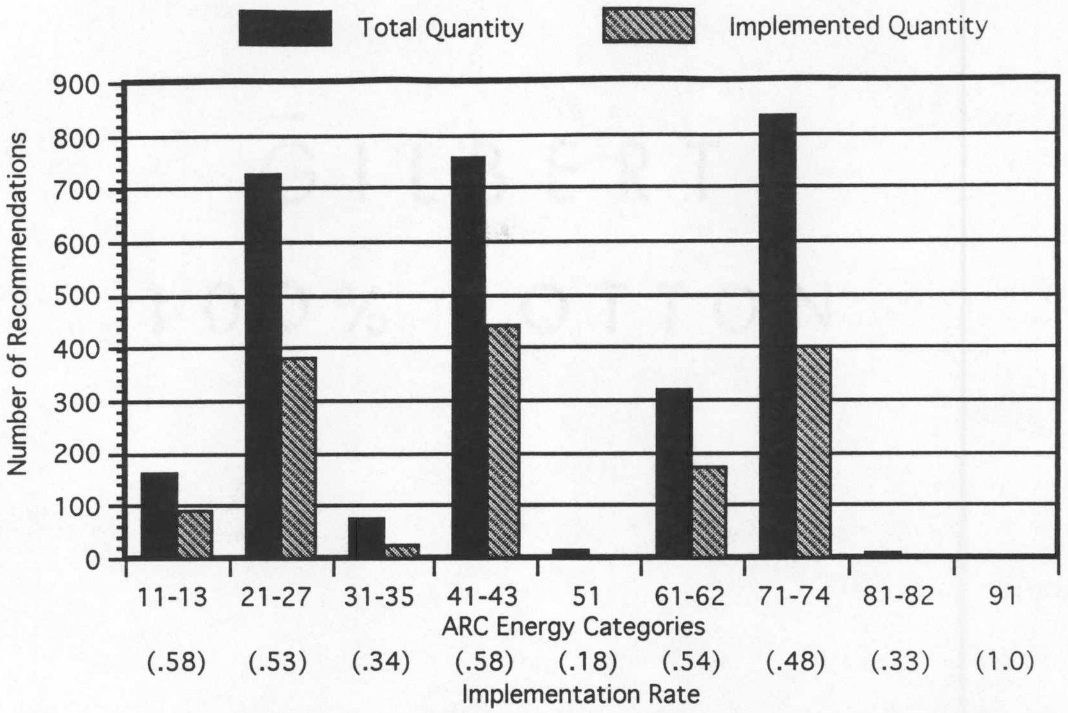


Figure IV-17. TOTAL AND IMPLEMENTED NUMBER OF ENERGY RECOMMENDATIONS BY ARC ENERGY CATEGORIES

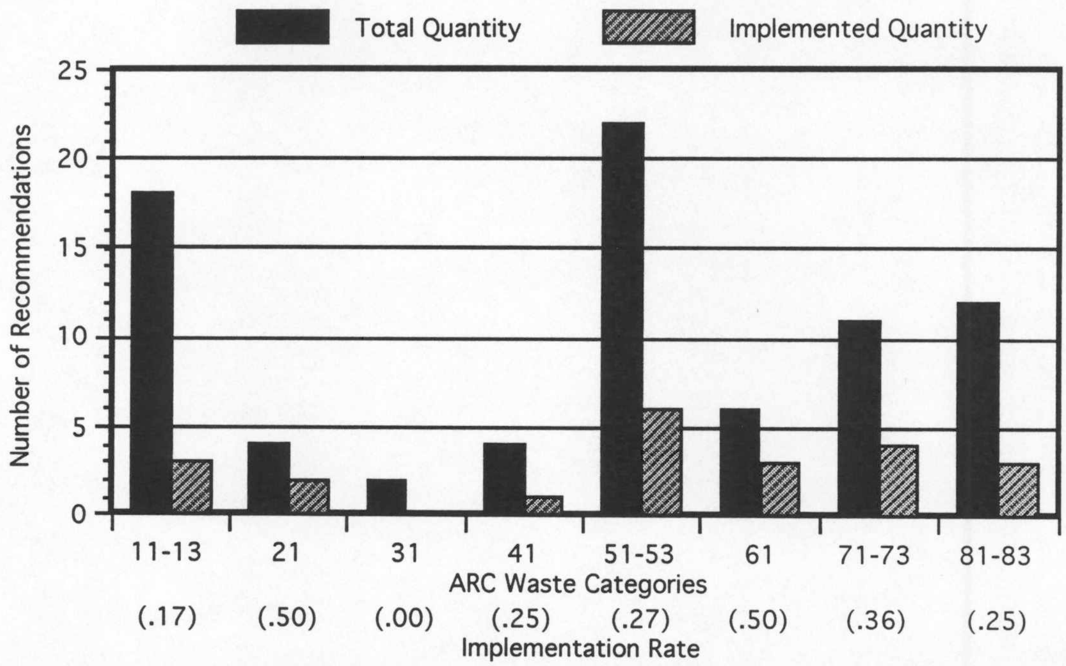


Figure IV-18. TOTAL AND IMPLEMENTED NUMBER OF WASTE RECOMMENDATIONS BY ARC WASTE CATEGORIES (alternate energy use) had only one recommendation made and that that one recommendation was reported as implemented. The fact that the several ARC categories in which there have been a greater number of recommendations made have higher implementation rates (thermal systems, motor systems and building and grounds) may have a plausible explanation. Those categories contain the systems that are more common to the majority of manufacturing environments, therefore, this trend may simply reveal that the assessed facilities are most interested in assuring that these systems are operating as efficiently as possible.

Within Figure IV-18, the implementation rate for waste recommendations does not appear to be number-of-recommendations specific, with a wide range of implementation rates being found throughout the ARC waste categories.

Regression Analyses

The developed UTIAC database contains both quantitative and qualitative information of obvious importance in specifying salient manufacturing plant characteristics. For example, for any single manufacturing plant, business activity reflected in annual sales is important to specifying the plant's economic impact within its specific type of industry. Similarly, simple computed quantities such as average sales for several plants within the same SIC category may reflect the regional impact of a particular industry class. These, and other indices can yield useful insights into relative impacts of single plants and groupings of plants. In addition, based on data derived from the database data, it is possible to characterize energy use and waste generation associated with particular industries. However, beyond such simple characterizations, it is also of interest to determine how sets of variables are related to each other. For example, if it was discovered that a given industrial class' annual sales was directly proportional to that class' plant floorspace, sales for a future plant to be constructed might be predicted on the basis of size (for constant values of other important variables). For this reason, it is important to test for selected relationships, which if existent, would facilitate optimizing certain future actions.

For regression analysis, one dependent variable can be regressed against either one independent variable (simple regression) or two or more independent variables (multiple regression). For one independent variable, the available models include a straight line fit (SLR) or a first-order model, or a curvilinear fit (SCR) or a second-order or higher model. Which of these models is most appropriate depends on the number and quality of the raw data available and the intended use of the developed correlation. In many simple first-order analyses where scatterplots of the data reveal no apparent relationships, it is sufficient and expedient to study simple linear fit models before utilizing more complex models.

For the developed database components containing 147 numerical and categorical variables total, it would be possible to perform 10,731 different simple linear regression analyses alone! Even though only a twentieth of this quantity might be of interest, that is still over 500 possible analyses. Therefore, among this very large potential set, a small subset was selected for intensive regression analysis in this work based on a subjective selection of relationships of particular interest. Broadly, the issues of major concern were identified as the following:

- Which plant characteristics can be used to predict the magnitude of energy use or waste generation in manufacturing operations?
- Which factors most influence the decisions to implement measures recommended in assessment reports?

From these two areas of interest, the following five questions were posed to be answered through regression analysis of the data compiled in the new UTIAC database:

1) To what extent are plant total annual electricity and natural gas consumption related to facility size?
2) To what degree are the UTIAC recommended plant total energy and waste cost savings related to facility size?
3) To what degree are plant total annual energy consumption and waste generation related to the annual production hours of the facility?
4) To what extent are the UTIAC recommended plant total energy conservation and waste reduction related to the number of recommendations made to each facility?
5) To what degree is recommendation implementation rate related to the capital cost of implementation and the projected simple payback period?

All of these questions were posed, where feasible, utilizing data from all types of industries (non-SIC-specific) and from specific types of industries (within specific SIC codes), to look for relationships inherent to all manufacturing types and to specific manufacturing types. To best illustrate the forms of statistical analysis utilized, the first question's answer and interpretation are expounded upon where necessary to better describe the analysis method used and subsequent results.

QUESTION #1:

To what extent are plant total annual electricity and natural gas consumption related to facility size?

To answer this question, it was important to first determine which variable among those in the developed database would best describe facility size for all industrial manufacturing types, selecting between plant square footage, number of employees and annual sales. This was necessary because, as explained previously in Chapters II and IV, facility square footage data was not recorded during an initial program period limiting the quantity of plant floorspace data analyzable. From the preliminary analysis performed, annual sales was found to have the highest correlation with facility square footage (see Chapter III, pp. 27-32 for the statistical output and accompanying discussion). Therefore, to answer this question, annual electricity consumption and annual natural gas consumption were regressed as independent variables against annual sales as the dependent variable. The JMP output tables and graphs are shown in Figure IV-19. (It should be noted here that the statistical software output shown in the following figures may vary in size somewhat due to page formatting requirements.)

Initially, the R^2 value of about 12.4% reveals some utility in the model for associating annual sales with the two annual energy consumption values, and the ANOVA table p value (Prob>F) reveals that the model as a whole (with all regressors included) is significant at lower than a 0.0001 level of

Response: Sales

Summary of Fit

RSquare	0.123774
RSquare Adj	0.120014
Root Mean Square Error	25.47031
Mean of Response	24.98699
Observations (or Sum Wgts)	469

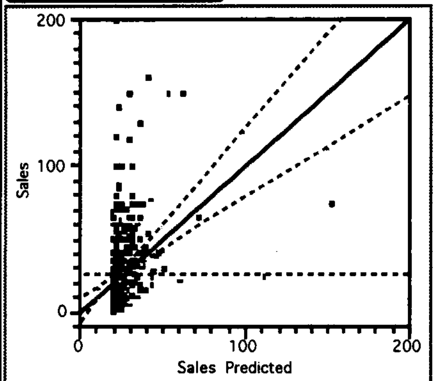
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.822454	1.341265	14.78	<.0001
Ann. E Cons.	0.0002325	0.000034	6.92	<.0001
Ann. NG Cons.	0.0000059	0.000015	0.39	0.6947

Effect Test

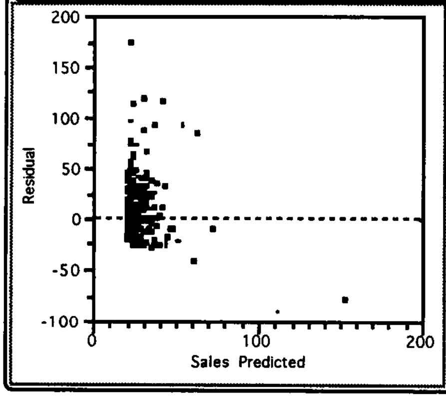
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Ann. E Cons.	1	1	31105.502	47.9478	<.0001
Ann. NG Cons.	1	1	100.071	0.1543	0.6947

Whole-Model Test



Analysis of Variance

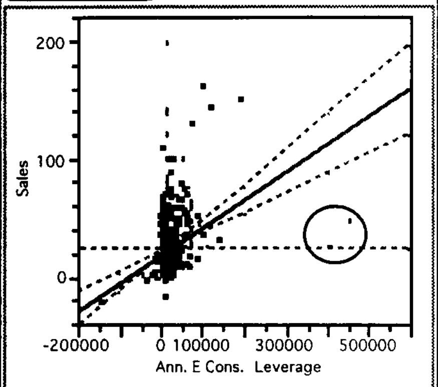
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	42704.11	21352.1	32.9133
Error	466	302311.34	648.7	Prob>F
C Total	468	345015.45		<.0001



Correlations

Variable	Ann. E Cons.	Ann. NG Cons.
Ann. E Cons.	1.0000	0.4792
Ann. NG Cons.	0.4792	1.0000

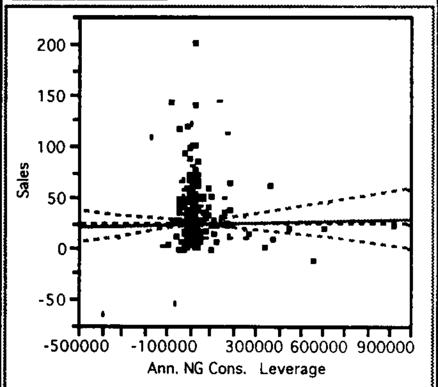
Ann. E Cons.



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
31105.502	47.9478	1	<.0001

Ann. NG Cons.



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
100.07138	0.1543	1	0.6947

Figure IV-19. QUESTION #1 - JMP OUTPUT TABLES AND GRAPHS significance. The equation that this output generates to describe the data is (values in the Parameter Estimates table):

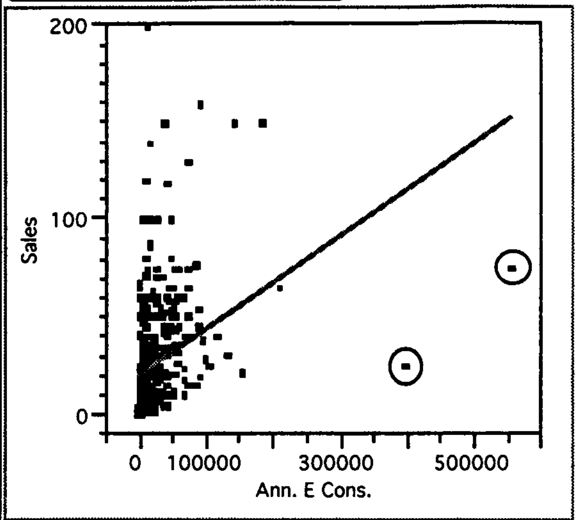
$$\hat{y} = 19.82 + (2.33E\text{-}4)x_1 + (5.9E\text{-}6)x_2 \qquad \text{(Equation \#14)}$$

There are several additional points to note about this model. First, in the Parameter Estimates table, the Prob>|t| value for the annual natural gas consumption is 0.6947 (statistically insignificant for the overall alpha value of 0.05 chosen for these analyses). The Prob>|t| values reveal each regressor's contribution to the model, *while all currently included regressors are in the model*. Stated differently, this value would not have the same magnitude if annual natural gas consumption alone was regressed with annual sales.

A second notable point pertaining to Figure IV-19 is the linear correlation between the two independent variables. The value in the Correlations table indicates that there is a 0.4792 level of correlation between the annual electricity and natural gas consumption parameters. If there is significant correlation between variables then the correlations value tends toward 1; almost no correlation reveals a correlations value tending toward 0. Thus, it can be concluded that there is a moderate level of correlation between these two independent variables.

Noting the two points made about the overall MLR fit, another test was run excluding the natural gas consumption values, reducing the analysis to SLR of annual sales against annual electricity consumption. Figure IV-20 presents the JMP output tables and graphs from this analysis.

Sales By Ann. E Cons.



—— Linear Fit

Linear Fit

Sales = 19.8749 + 0.00024 Ann. E Cons.

Summary of Fit

RSquare	0.123484
RSquare Adj	0.121608
Root Mean Square Error	25.44724
Mean of Response	24.98699
Observations (or Sum Wgts)	469

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	42604.04	42604.0	65.7915
Error	467	302411.41	647.6	Prob>F
C Total	468	345015.45		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.874876	1.333398	14.91	<.0001
Ann. E Cons.	0.0002389	0.000029	8.11	<.0001

Figure IV-20. QUESTION #1a - JMP OUTPUT TABLES AND GRAPH

The first point to note is that the R^2 value was reduced only minimally from about 12.38% to 12.35%. Second, the ANOVA table p value is still significant below the alpha 0.0001 level. Third, the Parameter Estimates tables t Ratio for the independent variable increased from 6.92 to 8.11, suggesting improved effect on the dependent variable (notice the Prob>|t| value). All factors considered, this information reveals that for all types of industrial manufacturers, there is a statistically significant relationship between annual electrical energy consumption and annual sales.

However, by closer observation of the scatterplot in Figure IV-20, one can see that there appears to be some outliers (or influential observations; these can be either one or both depending upon their effect on the model). An outlier can be described graphically (or spatially) as a point which is an extreme value in relation to the remaining plotted data in a scatterplot. An outlier can also be described statistically as a data point that has a large residual value (where a data point's residual value is the difference between its actual and predicted values). An influential observation is one that has a large impact on the regression analysis. The values encircled on the scatterplot in Figure IV-20 are examples of points that seem to be outliers (because of their distance from the remaining data) and might be influential observations (because of their effect on the regression analysis). As can be seen from the X-axis of the scatterplot, these values are extremely high in the annual electrical energy consumption category (approximately 400,000 and 550,000 million BTU/yr).

When outliers are detected in regression analysis, whether or not to omit them is dependent on several factors, including the variables being analyzed and the results sought from the analysis. Before deciding on the fate of the outliers in this case, the outliers were set aside and the analysis performed again to see if any improvement in model fit was revealed. The results are presented in Figure IV-21 wherein a significant improvement in regression fit is achieved. The portion of variation of annual sales accounted for by the variation in annual electrical energy consumption almost doubled, with the R^2 value increasing from 12.35% to 21.28%. The equation developed by the model fit is shown underneath the Linear Fit box, and can be interpreted in this manner:

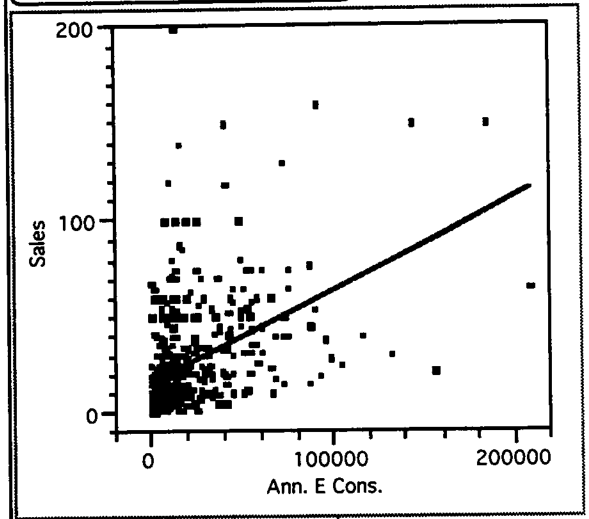
> *With every unit increase in annual energy consumption of 10,000 million BTUs, this model reveals that the annual sales should increase by $4.8 million. More practically, if a company's annual energy consumption were 75,000 million BTUs, the 95% confidence limits suggest that the expected annual sales range would be between about $49 million and $54 million.*

The confidence interval range was calculated utilizing the following formula

$$= \hat{y} \pm t_{\alpha/2}^{n-2}\, s\sqrt{\left(\frac{1}{n} + \frac{(x_o - x_{avg})^2}{SS_{xx}}\right)} \qquad \text{(Formula \#15)}$$

where $\hat{y}$ is the estimated value of y for a given value of x (x_o in the formula), t is a value from a Student's t distribution critical values table (n = 467, alpha = 0.05), s is the standard deviation, n is the number of observations, x_{avg} is the average value of all the independent variable observations and SS_{xx} is the Sum of Squares value for the all the independent variable observations. With the correct values inputted and calculated, the formula becomes that shown next, and can be solved for the values given in the statement.

Sales By Ann. E Cons.



— Linear Fit

Linear Fit

Sales = 15.5206 + 0.00048 Ann. E Cons.

Summary of Fit

RSquare	0.212757
RSquare Adj	0.211064
Root Mean Square Error	24.0804
Mean of Response	24.87987
Observations (or Sum Wgts)	467

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	72871.28	72871.3	125.6693
Error	465	269637.51	579.9	Prob>F
C Total	466	342508.79		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	15.520647	1.392376	11.15	<.0001
Ann. E Cons.	0.0004813	0.000043	11.21	<.0001

Figure IV-21. QUESTION #1b - JMP OUTPUT TABLES AND GRAPH

$$\$51.52 \text{ million} \pm (1.960)(1.10441) \qquad \text{(Formula \#16)}$$

Thus, from the previously presented results, the model improves significantly with the removal of the two outliers. Therefore, if they are chosen for removal, the interpretation of the model fit changes. The range limitation on annual electrical energy consumption *for this analysis to be applicable* is limited to approximately 250,000 million BTU/yr on the higher consumption end. The previous statement of the results would have to include this limitation in the interpretation of the outliers to be omitted.

The model discussed appears to be relatively significant for the use of observational data. For completeness, curvilinear (second degree polynomial) analysis was performed on the last data set. However, the increase in model fit was insignificant, therefore the results from the linear analysis were considered the best "fit" for this data. (Curvilinear analysis will be discussed further with regard to the additional questions posed for analysis.)

The next step undertaken to investigate the relationship between annual sales and annual electricity and natural gas consumption was to examine relationships within data pertaining to specific SIC codes. Table IV-2 reveals the results from an SIC-specific breakdown and statistical analysis. As can be seen from the table data, only one SIC group (29) had fewer than 3 data points, and since connecting two dots with a line results in a perfect fit (thus the R^2 = 1.0), this result has minimal significance. There were 14 SIC groups with 30 or fewer observations, leaving five groups (shown in italicized and bold

Table IV-2. TABULAR RESULTS FROM SIC-SPECIFIC JMP ANALYSIS OF QUESTION #1

SIC Group	n	R^2	Prob>F value	SIC Group	n	R^2	Prob>F value
20	***53***	***0.2521***	***0.0007***	31	7	0.1158	0.7818
22	30	0.0207	0.7544	32	24	0.4719	0.0012
23	***53***	***0.4341***	***<0.0001***	33	18	0.5618	0.0021
24	23	0.5539	0.0003	***34***	***44***	***0.1699***	***0.0220***
25	20	0.1421	0.2718	***35***	***32***	***0.3220***	***0.0036***
26	16	0.2249	0.1909	36	30	0.1112	0.2036
27	10	0.5418	0.0651	37	27	0.5727	<0.0001
28	21	0.2992	0.0408	38	11	0.8134	0.0012
29	2	1.0000	———	39	14	0.0955	0.5757
30	***34***	***0.8440***	***<0.0001***	——	—	——	———

typeface) with a sufficient number of observations to justify further analysis and possible inference.

Initially, it is clear that all five of these SIC groups have statistically significant fit (from p value; at alpha 0.03 level or better). Moreover, the results for SIC group 30 reveals an R^2 value greater than 0.75. Finally, the results for SIC groups 20, 23 and 35 reveal a similar relationship with R^2 values better than 0.25.

From the MLR analyses' results for these five SIC groups, the outputs were studied more closely. For SIC groups 20, 23, 34 and 35, one of the two independent variables contribution to the model was found to be statistically insignificant at the alpha level of 0.05, with natural gas use being the insignificant factor in each analysis except for SIC group 23 where the association with electricity use was found to be insignificant. The statistically insignificant variable was removed in each case and the R^2 values dropped (leaving SLR), but not significantly for any SIC group. Thus, the addition of the second order term did not add notable utility to any of the models investigated.

For SIC group 30 (rubber and miscellaneous plastics manufacturing), both independent variables' contribution to the model were significant to an alpha level of 0.02. Additionally, the correlation between the two regressors is minimal at only 0.0859. Figure IV-22 depicts the results from this SIC-specific analysis. Upon removal of an apparent outlier, the model utility decreased to approximately 0.68. However, there appears to be a significant, positive relationship between annual electricity and natural gas consumption and annual sales for this particular SIC group. As another example, the interpretation of the equation generated from the model fit reveals that for every 10,000 million BTU increase in both electricity and natural gas annual consumption, an increase in annual sales of about $6.5 million can be expected.

Thus, the results from the first question's analyses reveal some significant relationships both within specific SIC groups and across all manufacturing types between the main forms of energy utilized in industrial manufacturing facilities today and the annual sales realized from use of that energy.

Response: Sales

Summary of Fit

RSquare	0.844044
RSquare Adj	0.833982
Root Mean Square Error	11.59659
Mean of Response	23.67059
Observations (or Sum Wgts)	34

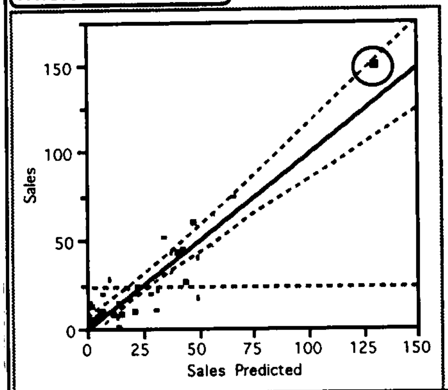
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.509712	2.818892	-0.54	0.5961
Ann. E Cons.	0.0007164	0.000058	12.44	<.0001
Ann. NG Cons.	0.0001024	0.000041	2.51	0.0173

Effect Test

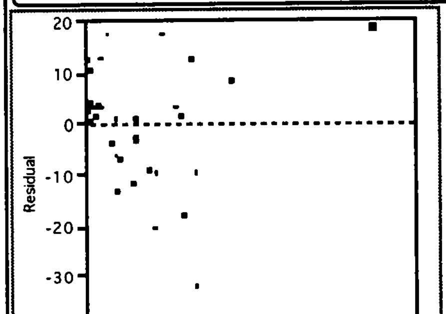
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Ann. E Cons.	1	1	20822.890	154.8390	<.0001
Ann. NG Cons.	1	1	849.970	6.3204	0.0173

Whole-Model Test



Analysis of Variance

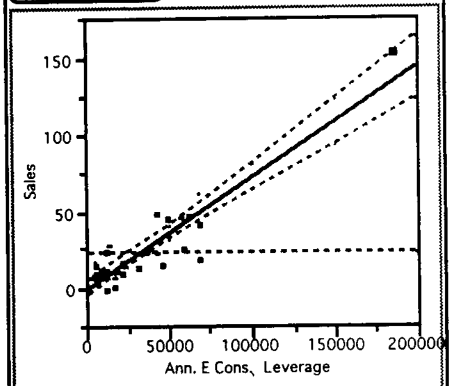
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	22562.423	11281.2	83.8871
Error	31	4168.908	134.5	Prob>F
C Total	33	26731.331		<.0001



Correlations

Variable	Ann. E Cons.	Ann. NG Cons.
Ann. E Cons.	1.0000	0.0859
Ann. NG Cons.	0.0859	1.0000

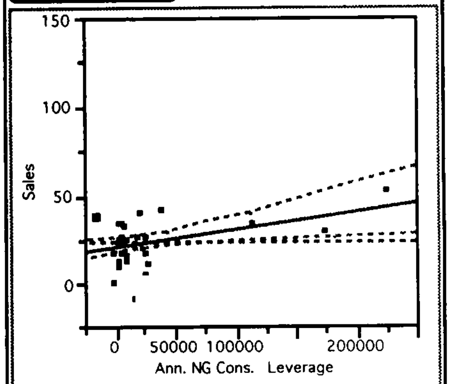
Ann. E Cons.



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
20822.890	154.8390	1	<.0001

Ann. NG Cons.



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
849.97049	6.3204	1	0.0173

Figure IV-22. QUESTION #1c - JMP OUTPUT TABLES AND GRAPHS

QUESTION #2:

To what degree are the UTIAC recommended plant total energy and waste cost savings related to facility size?

For this analysis, facility annual sales was again utilized as the dependent variable for facility size. Table IV-3 illustrates the results from the non-SIC-specific JMP analysis in tabular format. As can be seen from the Regression Type column, these analyses included multiple linear, simple linear and curvilinear regression of the chosen variables.

A very important first point to make is the reasoning for the small number of data points analyzable (n) in the first, fourth and fifth analyses (reference numbers). Included in this question as a variable is the total UTIAC recommended annual waste cost savings, and since the UTIAC has been performing waste analysis since only the 1994 fiscal year, there were a limited number of assessments from which to draw data.

Now reviewing the more important results from these analyses, it is obvious that the only variables that appear to have any relationship between them are facility gross annual sales and the UTIAC recommended annual energy cost savings (as is evidenced by the Prob>F and Prob>|t| values). The Prob>|t| values indicates statistical significance at an alpha level greater than 0.0004; however, the R^2 reaches only 0.079 at best. The set of two numbers shown within each cell in the columns for R^2, Prob>F and Prob>|t| are indicative of the first model analyzed and then the second model analyzed after one or two extreme values were removed. The text at the end of each analytical set

Table IV-3. TABULAR RESULTS FROM NON-SIC-SPECIFIC JMP ANALYSIS OF QUESTION #2

Ref. #	Variables Analyzed	Regression Type	n	R^2	Prob>F	Prob>|t|	Associated Term	Outlier/Influential Observation Information
		(see Key)	*From Summary of Fit Table*		*From ANOVA Table*	*From Parameter Estimates Table*		
1	① X ②, ③	MLR	47/ 46	0.009/ 0.022	0.8112/ 0.6132	0.5269/ 0.4063	②	One outlier was removed, reducing applicable facility total annual waste cost savings range to under approximately $1,000,000.
						0.9473/ 0.4554	③	
2	① X ②	SLR	475/ 474	0.045/ 0.043	<0.0001/ <0.0001	<0.0001/ <0.0001	②	One outlier was removed, reducing applicable facility total annual energy cost savings range to under approximately $450,000.
3		SCR	475/ 474	0.062/ 0.079	<0.0001/ <0.0001	<0.0001/ <0.0001	②	
						0.0033/ <0.0001	②^2	
4	① X ③	SLR	50/ 48	0.039/ 0.005	0.1694/ 0.6191	0.5386/ 0.8065	③	Two outliers were removed, reducing applicable facility total annual waste cost savings range to under approximately $1,500,000.
5		SCR	50/ 48	0.056/ 0.010	0.2551/ 0.8052	0.3550/ 0.6656	③	
							③^2	

Key:

	①	=	Facility Gross Annual Sales
	②	=	UTIAC Recommended Annual Energy Cost Savings
	③	=	UTIAC Recommended Annual Waste Cost Savings
	MLR	=	Multiple Linear Regression
	SLR	=	Simple Linear Regression
	SCR	=	Simple Curvilinear Regression

describes how the removal of one or more outliers effects the interpretation of the JMP output.

Now to analyze these variables on an industry-specific level, only annual sales and UTIAC recommended energy cost savings can be utilized due to the previously stated lack of waste-related data. Therefore, the data was broken down into SIC groups and reanalyzed in both SLR and SCR modes. Table IV-4 reveals the SIC groups that had enough datasets for regression-related statistical significance. The results from each of the SIC groups chosen is discussed next.

For SIC group 20 (food and kindred products), the curvilinear fit is an improvement over the linear fit as is evidenced by the improved R^2 and Prob>F value of less than 0.05 (after outlier removal). However, that the value of the second term of the polynomial (the squared term - 0.2204) is greater than the chosen alpha value of 0.05 indicates that the addition of that term does not significantly improve model fit. The results reveal that approximately 14% of the variation within the UTIAC recommended annual energy cost savings can be accounted for by the differences in the annual sales data.

For SIC group 22 (textile mill products), there appears to be no statistically significant relationship. Neither the individual variable significance nor the whole model significance can break the alpha level of 0.05 needed for significance in this work. There were four influential observations; two sets of two working in tandem against each other to reduce model fit. However,

Table IV-4. TABULAR RESULTS FROM SIC-SPECIFIC JMP ANALYSIS OF QUESTION #2

Ref. #	SIC Group	Regression Type	n	R^2	Prob>F	Prob>|t|	Term Order	Outlier/Influential Observation Information
		(see Key)	*From Summary of Fit Table*		*From ANOVA Table*	*From Parameter Estimates Table*		
1	20	SLR	52/ 50	0.067/ 0.143	0.0643/ 0.0067	—	—	Two outliers were removed, reducing applicable facility total annual energy cost savings range to under about $100,000 and applicable annual sales range to under $125 million.
2		SCR	52/ 50	0.072/ 0.170	0.1619/ 0.0124	0.2633/ 0.0418	1	
						0.6154/ 0.2204	2	
3	22	SLR	29	0.117	0.0689	—	—	No outliers removed.
4		SCR	29	0.154	0.1130	0.1049	1	
						0.2954	2	
5	23	SLR	53/ 51	0.077/ 0.028	0.0446/ 0.2365	—	—	Two outliers were removed, reducing applicable facility total annual energy cost savings range to under $80,000 and applicable annual sales range to under about $100 million.
6		SCR	53/ 51	0.132/ 0.042	0.0291/ 0.3590	0.0175/ 0.2488	1	
						0.0806/ 0.4183	2	
7	30	SLR	34/ 33	0.718/ 0.371	<0.0001/ 0.0002	—	—	One outlier was removed, reducing applicable facility total annual energy cost savings range to under $200,000 and annual sales to under about $100 million.
8		SCR	34/ 33	0.745/ 0.422	<0.0001/ 0.0003	0.0946/ 0.0120	1	
						0.0817/ 0.1140	2	

Key: SLR = Simple Linear Regression
SCR = Simple Curvilinear Regression

Table IV-4. (Continued)

Ref. #	SIC Group	Regression Type	n	R^2	Prob>F	Prob>|t|	Term Order	Outlier/Influential Observation Information
		(see Key)	*From Summary of Fit Table*		*From ANOVA Table*	*From Parameter Estimates Table*		
9	34	SLR	44/ 43	0.034/ 0.146	0.2291/ 0.0114	—	—	One outlier was removed, reducing applicable facility total annual energy cost savings range to under $125,000.
10		SCR	44/ 43	0.166/ 0.165	0.0244/ 0.0270	0.0069/ 0.0917	1	
						0.0149/ 0.3448	2	
11	35	SLR	33/ 31	0.333/ 0.345	0.0004/ 0.0005	—	—	Two outliers were removed, reducing applicable facility total annual energy cost savings range to under about $100,000.
12		SCR	33/ 31	0.337/ 0.353	0.0021/ 0.0022	0.4751/ 0.1070	1	
						0.6649/ 0.5626	2	
13	36	SLR	32/ 30	0.265/ 0.003	0.0026/ 0.7696	—	—	Two outliers were removed, reducing applicable annual sales range to under $175 million and energy cost savings range to under about $100,000.
14		SCR	32/ 30	0.308/ 0.005	0.0048/ 0.9347	0.9982/ 0.9083	1	
						0.1874/ 0.8234	2	
15	37	SLR	27	0.168	0.0340	—	—	No outliers removed.
16		SCR	27	0.199	0.0702	0.1293	1	
						0.3450	2	

Key: SLR = Simple Linear Regression
SCR = Simple Curvilinear Regression

they were not outliers, and thus there is now reasonable justification for removing them. Therefore, no statistically significant relationship appears to exist within these variables and within this industry type.

SIC group 23's (apparel and other textile products) results bring to light yet another interesting facet of statistical modeling. The relationship improves with the transition from linear to curvilinear model fit. However, the relationship almost disappears with the removal of the outliers (note the R^2, Prob>F and Prob>|t| values). What this reveals about this SIC group's data is that the two outliers that were removed were not only outliers, but also influential observations. The initial relationship was defined to a large extent by those particular observations that were removed and well outside the range of the remaining 51 observations. This results in the fact that there is no statistically significant relationship between these two variables within this particular SIC group.

Within SIC group 30 (rubber and miscellaneous plastic products) is the most significant relationship between annual sales and UTIAC recommended annual energy cost savings for all the manufacturing SIC groups analyzed. Again, the move from linear to curvilinear model fit improves the R^2 value, but the individual contributors' p values (Prob>|t|) signifies that the second order term does not significantly improve the overall model fit. However, the outlier removal definitely improves the relationship. Therefore, the relationship between these two variables seems to be rather significant for observational data and attains about 37% data variation explanation through the linear model fit.

SIC group 34's (fabricated metal products) results reveal improvement in model fit with the change from linear to curvilinear fit and with the outlier removal. However, the p values for the individual contributors in the curvilinear model increase above the acceptable alpha value of 0.05 (0.0917 and 0.3448), justifying the choice in this case of the linear model with outliers removed as the best choice for model fit. With an overall model fit significance value of 0.0114, this relationship is definitely statistically significant.

For SIC group 35 (industrial machinery and equipment), the curvilinear model shows an insignificant improvement in variation explanation (see R^2 values) and the p values for the individual contributors reveal insignificance at the alpha 0.05 level. The removal of the outliers improves the fit slightly, thus the most applicable fit for this data is the linear model. The relatively good R^2 value of about 35% reveals that a substantial amount of the variation within the UTIAC recommended annual energy cost savings can be accounted for by differences in the annual sales data.

For SIC group 36 (electronic and other electric equipment), the same result is obtained as that for SIC group 23. The outliers that were removed were defining the fit of the model almost entirely, therefore, there is no real statistical relationship between these two variables for this SIC group.

Finally, for SIC group 37 (transportation equipment), there were no outliers in the scatterplot. Utilizing SCR, the prob>|t| values are too high for statistical significance in this analysis, thus the SLR model is the chosen one.

At almost 17% data variation explanation, this model reveals medium range significance between these two variables. It should be noted however, that the number of data points available for analysis in this SIC group is below the lower limit for statistical inference, therefore, these results should be restricted from use for statistical inference.

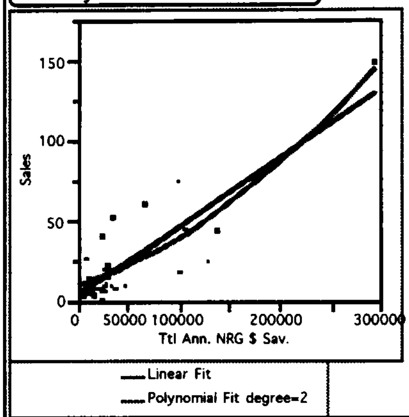
In summation, there appears to be significant SIC-specific and non-SIC-specific relationships between two of the three variables considered for analysis in this question: facility gross annual sales and the total UTIAC recommended annual energy cost savings. SIC groups 30 and 35 appear to have an approximate 35% model utility, and SIC groups 20 and 34 appear to have on the order of 15% model utility. No statistically significant relationship exists between annual sales and UTIAC recommended annual waste cost savings. For illustrative purposes, the JMP results from SIC group 30's analysis are shown in Figure IV-23.

QUESTION #3:

<u>To what degree are plant total annual energy consumption and waste generation related to the annual production hours of the facility</u>?

For this analysis, the units utilized were millions of BTUs for energy consumption, and total pounds of waste generated (including all reported waste streams) for waste generation. Table IV-5 reveals that no significant relationship was found between these variables. One outlier (but not an influential observation) was found and removed and the fit worsened, disclosing that the outlier removed was actually improving the fit to that found in the initial analysis.

Sales By Ttl Ann. NRG $ Sav.



Linear Fit

Sales = 3.34627 + 0.00043 Ttl Ann. NRG $ Sav.

Summary of Fit

RSquare	0.718348
RSquare Adj	0.709547
Root Mean Square Error	15.3388
Mean of Response	23.67059
Observations (or Sum Wgts)	34

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	19202.405	19202.4	81.6155
Error	32	7528.926	235.3	Prob>F
C Total	33	26731.331		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.3462673	3.461393	0.97	0.3409
Ttl Ann. NRG $ Sav.	0.0004307	0.000048	9.03	<.0001

Polynomial Fit degree=2

Sales = 8.91041 + 0.00022 Ttl Ann. NRG $ Sav. + 8.4e-10 Ttl Ann. NRG $ Sav.^2

Summary of Fit

RSquare	0.744977
RSquare Adj	0.728524
Root Mean Square Error	14.82925
Mean of Response	23.67059
Observations (or Sum Wgts)	34

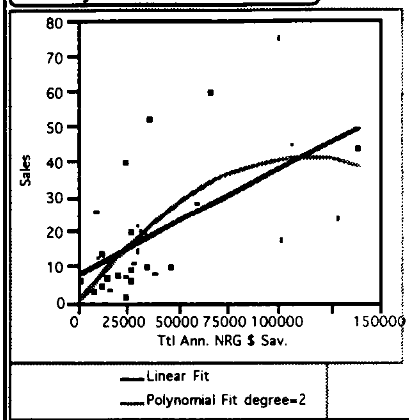
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	19914.229	9957.11	45.2789
Error	31	6817.102	219.91	Prob>F
C Total	33	26731.331		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	8.9104135	4.556636	1.96	0.0596
Ttl Ann. NRG $ Sav.	0.0002185	0.000127	1.72	0.0946
Ttl Ann. NRG $ Sav.^2	8.449e-10	4.7e-10	1.80	0.0817

Sales By Ttl Ann. NRG $ Sav.



Linear Fit

Sales = 7.70483 + 0.00031 Ttl Ann. NRG $ Sav.

Summary of Fit

RSquare	0.370875
RSquare Adj	0.35058
Root Mean Square Error	14.44993
Mean of Response	19.84242
Observations (or Sum Wgts)	33

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3815.782	3815.78	18.2748
Error	31	6472.819	208.80	Prob>F
C Total	32	10288.601		0.0002

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.704833	3.793249	2.03	0.0509
Ttl Ann. NRG $ Sav.	0.0003056	0.000071	4.27	0.0002

Polynomial Fit degree=2

Sales = -0.4142 + 0.00074 Ttl Ann. NRG $ Sav. - 3.31e-9 Ttl Ann. NRG $ Sav.^2

Summary of Fit

RSquare	0.421926
RSquare Adj	0.383388
Root Mean Square Error	14.08022
Mean of Response	19.84242
Observations (or Sum Wgts)	33

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	4341.026	2170.51	10.9482
Error	30	5947.575	198.25	Prob>F
C Total	32	10288.601		0.0003

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.41416	6.208262	-0.07	0.9473
Ttl Ann. NRG $ Sav.	0.0007446	0.000279	2.67	0.0120
Ttl Ann. NRG $ Sav.^2	-3.309e-9	2.033e-9	-1.63	0.1140

Figure IV-23. QUESTION #2 - JMP OUTPUT TABLES AND GRAPHS

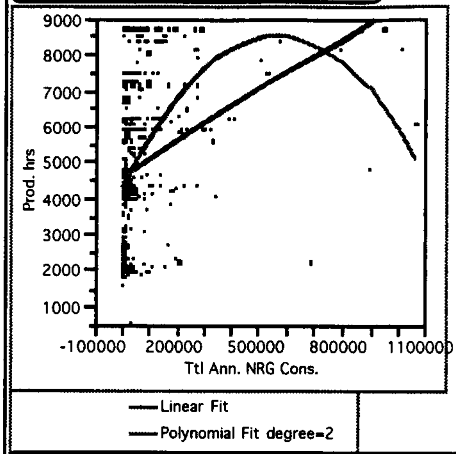
Table IV-5. TABULAR RESULTS FROM NON-SIC-SPECIFIC JMP ANALYSIS OF QUESTION #3

n	R^2	Prob>F	Energy Consumption Prob>|t|	Waste Generation Prob>|t|
From Summary of Fit Table		*From ANOVA Table*	*From Parameter Estimates Table*	
Initial analysis:				
33	0.0204	0.7336	0.4482	0.8469
Analysis after the removal of one outlier:				
32	0.0108	0.8537	0.6059	0.8606

To continue with the analysis of this question, annual energy consumption alone was regressed against production hours. The initial results yielded minimal statistical correlation, however, two very significant outliers were exposed (in total annual energy consumption) and they were removed. Retesting the correlation revealed a significant increase in model fit, not only from the outlier removal, but also between linear fit and curvilinear fit. The scatterplot of the data in Figure IV-24 graphically reveals the linear and curvilinear fits applied to the data. For both models, all p values (Prob>F and Prob>|t|) were extremely significant, yet the R^2 value increases from about 0.09 to approximately 0.17.

The next phase of the question #5 analysis was to explore the relationships between production hours and annual energy consumption within specific SIC groups. Again, annual waste generation was removed for lack of a sufficient number of data values to infer statistical significance. To begin, the

Prod. hrs By Ttl Ann. NRG Cons.



Linear Fit

Prod. hrs = 4656.63 + 0.00474 Ttl Ann. NRG Cons.

Summary of Fit

RSquare	0.091935
RSquare Adj	0.090078
Root Mean Square Error	2050.188
Mean of Response	4974.277
Observations (or Sum Wgts)	491

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	208095364	2.081e+8	49.5080
Error	489	2055398668	4203269	Prob>F
C Total	490	2263494032		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4656.6322	102.9497	45.23	<.0001
Ttl Ann. NRG Cons.	0.0047389	0.000673	7.04	<.0001

Polynomial Fit degree=2

Prod. hrs = 4267.35 + 0.01526 Ttl Ann. NRG Cons. - 1.35e-8 Ttl Ann. NRG Cons.^2

Summary of Fit

RSquare	0.167508
RSquare Adj	0.164096
Root Mean Square Error	1965.033
Mean of Response	4974.277
Observations (or Sum Wgts)	491

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	379153004	1.8958e8	49.0959
Error	488	1884341028	3861355	Prob>F
C Total	490	2263494032		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4267.3501	114.7052	37.20	<.0001
Ttl Ann. NRG Cons.	0.0152552	0.001707	8.94	<.0001
Ttl Ann. NRG Cons.^2	-1.351e-8	2.03e-9	-6.66	<.0001

Figure IV-24. QUESTION #3 - JMP OUTPUT TABLES AND GRAPH entire set of data was broken down by SIC groups to see how many data points would be available for regression analysis. The results revealed that only seven of the 19 SIC groups had 30 or more data points, therefore, those seven were selected for further analysis.

Table IV-6 highlights the SIC groups having enough data points and the number within each group (n). Again, similar to the analysis of question #2, Table IV-6 lists the separate values for the pre and post-outlier removal, where applicable. Three of the four SIC groups did not have significant outliers and thus, no data points were removed. (Before removal, all outliers were checked against actual report information to ensure human error was not the reason for the extreme value.) Of the seven SIC groups considered, there are two sets of groups with similar regression results. First, models for SIC groups 20, 22 and 35 all improved with the removal of the outlier(s) and the transition from linear to curvilinear model fit. The individual parameter p values all remained below 0.04 and the whole-model p values below 0.003. The R^2 values reveal data variation explanations of about 24% (SIC group 20), 31% (SIC group 35) to 40% (SIC group 22). Only SIC group 22 becomes questionable with the outlier removal due to the reduced number of data points (29) available for analysis.

The second similarity between SIC group results is relative to the 23, 30 and 34 SIC groups. In each of these cases, there were no outliers and thus none were removed. The resulting data revealed in Table IV-6 discloses that for each SIC-specific analysis, the curvilinear fit was not a significant enough improvement to justify usefulness. The R^2 values increased for each SIC

Table IV-6. TABULAR RESULTS FROM SIC-SPECIFIC JMP ANALYSIS OF QUESTION #3

SIC Group	Regression Type	n	R^2	Prob>F	Prob>|t|	Associated Term	Energy Consumption Range Reduced to:
	(see Key)	*From Summary of Fit Table*		*From ANOVA Table*	*From Parameter Estimates Table*		
20	SLR	54/53	0.021/0.093	0.2954/0.0262	0.2954/0.0262	Cons.	Approximately 300,000 million BTUs
	SCR		0.116/0.235	0.0437/0.0012	0.0136/0.0005	Cons.	
					0.0235/0.0037	Cons.^2	
22	SLR	32/29	0.078/0.169	0.1221/0.0265	0.1221/0.0265	Cons.	Approximately 250,000 million BTUs
	SCR		0.160/0.396	0.0797/0.0014	0.0608/0.0008	Cons.	
					0.1027/0.0043	Cons.^2	
23	SLR	53	0.200	0.0008	0.0008	Cons.	No outliers
	SCR		0.233	0.0013	0.0136	Cons.	
					0.1515	Cons.^2	
30	SLR	34	0.312	0.0006	0.0006	Cons.	No outliers
	SCR		0.333	0.0019	0.0322	Cons.	
					0.3389	Cons.^2	
34	SLR	47	0.121	0.0168	0.0168	Cons.	No outliers
	SCR		0.139	0.0371	0.0978	Cons.	
					0.3367	Cons.^2	
35	SLR	35/34	0.059/0.210	0.1578/0.0064	0.1578/0.0064	Cons.	Approximately 125,000 million BTUs
	SCR		0.322/0.314	0.0020/0.0029	0.0005/0.0058	Cons.	
					0.0013/0.0381	Cons.^2	
36	SLR	36/33	0.117/0.024	0.0415/0.3911	0.0415/0.3911	Cons.	Approximately 150,000 million BTUs
	SCR		0.133/0.079	0.0950/0.2903	0.1924/0.1312	Cons.	
					0.4366/0.1895	Cons.^2	

Key: SLR = Simple Linear Regression
SCR = Simple Curvilinear Regression

group, but for only a minimal amount (maximum increase of 0.033 - SIC group 23), and range from about 0.12 (SIC group 34) to about 0.31 (SIC group 30) in the linear models. Moreover, the Prob>|t| values for the second degree term are all above 0.15, with two being above 0.30 (significantly breaking the alpha value borderline limit of 0.05).

The final SIC group, 36, had a barely significant relationship (Prob>F value - 0.0415) at an R^2 of almost 0.12 before outlier removal, however, the relationship deteriorated with each change after that. The removal of the three outliers and subsequent reanalysis disclosed that those three outliers were an *influential set*: working together to establish a fit for all the data in that SIC group. This SIC group was the only one of the seven analyzed with post-outlier removal Prob>F values of greater than 0.05.

QUESTION #4:

To what extent are the UTIAC recommended plant total energy conservation and waste reduction related to the number of recommendations made to each facility?

It was chosen to analyze this question in several different ways, of which the first was to utilize MLR to check the relationship between the total number of recommendations made to a facility (including energy and waste measures) as the dependent variable, to the UTIAC recommended plant total energy conservation and waste reduction as the independent variables. (Due to the limited number of waste-related assessments that have been performed, this analysis was not possible to perform on a SIC-specific level.) Upon obtaining the results, it was revealed that a statistical relationship existed with an R^2

value of about 0.26 and an ANOVA table p value of 0.0150. However, the individual parameter p values revealed that only the waste savings was significantly contributing to the model. Upon retesting with UTIAC recommended waste savings alone regressed against the total number of recommendations made to a facility, the R^2 was reduced to 0.11, which infers only borderline variation explanation through model fit. Moreover, the Prob>F value increased above the 0.05 alpha limit and several outliers were found. With removal of the outliers the model had only 27 variables, and this is below the quantity determined significant to infer statistical significance in this work. Thus, no statistically significant relationship was found between the chosen variables in this first analysis of this question.

For the second analysis, each type of UTIAC recommended savings (i.e., energy and waste) was regressed against the number of recommendations made to the facility specifically relative to that type of savings. The waste portion of the analysis was performed first, due to fact that the energy analysis performed would be followed directly by an SIC-specific analysis of those variables. The waste analysis results yielded a statistically insignificant relationship, with a low R^2 value (0.0612) and an insignificant overall model p value (0.1582). Similar to the first analysis of question #4, outliers were also discovered and removal of them would have yielded too few data points to infer statistical significance.

For the third analysis of question #4, the number of energy recommendations made to a facility, the independent variable, was regressed against the UTIAC recommended energy savings, the dependent variable. The non-SIC-specific analysis was performed first, in both SLR and SCR formats, and the output yielded statistically significant results in both cases with respect to the Prob>F values (both less than 0.0100), however, the R^2 values were 0.0144 and 0.0160, respectively (revealing almost no variation in the UTIAC recommended energy savings explained by variation in the number of energy recommendations made). An outlier was found and checked for incorrect data entry. After the number was verified, it was removed and the variables retested. The relationship retained its statistical significance (Prob>F value less that 0.05), however the R^2 values improved to no greater than 0.02. Thus, no statistically significant relationship exists between these variables at the non-SIC-specific level.

Delving within specific SIC groups, the fourth analysis within question #4 yielded the information shown in Table IV-7. The SIC groups that did not have enough data points for statistical inference were excluded.

SIC group 20 was the only SIC group that did not have any significant outliers, however, it was not the only SIC group that did not have a statistically significant relationship between the variables. SIC groups 23, 30 and 35 also showed no signs of a statistically significant relationship between the number of energy recommendations made and the total energy savings recommended by the UTIAC. The low R^2 values and Prob>F values of at least 0.09 relay this information.

Table IV-7. TABULAR RESULTS FROM SIC-SPECIFIC JMP ANALYSIS OF QUESTION #4

SIC Group	Regression Type	n	R^2	Prob>F	Prob>|t|	Associated Term	Energy Savings Range Reduced to:
	(see Key)	*From Summary of Fit Table*		*From ANOVA Table*	*From Parameter Estimates Table*		
20	SLR	53	0.057	0.0853	0.0853	Savings	No outliers
	SCR		0.085	0.1101	0.1442	Savings	
					0.2257	Savings^2	
22	SLR	31/30	0.029/0.353	0.3618/0.0005	0.3618/0.0005	Savings	Approximately 75,000 million BTUs
	SCR		0.037/0.389	0.5893/0.0013	0.4863/0.8270	Savings	
					0.6267/0.2239	Savings^2	
23	SLR	53/51	0.013/0.056	0.4227/0.0937	0.4227/0.0937	Savings	Approximately 6,000 million BTUs
	SCR		0.023/0.066	0.5523/0.1928	0.3769/0.2992	Savings	
					0.4602/0.4764	Savings^2	
30	SLR	34/33	0.007/0.022	0.6358/0.4152	0.6358/0.4152	Savings	Approximately 15,000 million BTUs
	SCR		0.031/0.034	0.6131/0.5956	0.3488/0.4494	Savings	
					0.3878/0.5390	Savings^2	
34	SLR	48/46	0.110/0.154	0.0212/0.0071	0.0212/0.0071	Savings	Approximately 10,000 million BTUs
	SCR		0.132/0.174	0.0410/0.0163	0.6099/0.7205	Savings	
					0.2896/0.3060	Savings^2	
35	SLR	35/33	0.023/0.052	0.3847/0.2009	0.3847/0.2009	Savings	Approximately 10,000 million BTUs
	SCR		0.063/0.053	0.3528/0.4401	0.1960/0.6747	Savings	
					0.2507/0.8577	Savings^2	
36	SLR	35/34	0.008/0.128	0.6095/0.0380	0.6095/0.0380	Savings	Approximately 30,000 million BTUs
	SCR		0.010/0.138	0.8480/0.0992	0.7079/0.2821	Savings	
					0.7887/0.5372	Savings^2	

<u>Key:</u> SLR = Simple Linear Regression
SCR = Simple Curvilinear Regression

SIC groups 22, 34 and 36 all disclosed some statistically significant relationships between the variables, with reported R^2 values in the linear models of about 0.35, 0.15 and 0.13, respectively. The whole-model p values were all below the chosen alpha value of 0.05 for the linear models. The most interesting point to note though about these SIC groups' results is that in each case, the inclusion of the squared (second degree) term in the regression model resulted in statistical insignificance for the contribution of each term, as can be seen in the Prob>|t| values. As has been described previously in this chapter, the Prob>|t| values are a measurement of that particular variable's contribution to the model fit while the other currently included terms are in the model.

Thus, we can conclude for these three SIC groups that the inclusion of the second order term actually reduces the overall model fit.

For all the results discussed pertaining to question #4, it appears that the relationship between these variables exists only at the industry-specific level and only within the energy portion of the data.

QUESTION #5:

To what degree is the recommendation implementation rate related to the capital cost of implementation and the projected simple payback period?

Before proceeding with the analytical breakdown and results discussion for this question, some background information must be provided. So far in this work all of the statistical regression investigations have included variables that are categorized as *continuous* in regression analysis. That labeling is for assigning a modeling type to the variable, and the two main categories for modeling types include continuous and nominal (another category, ordinal, is rarely utilized). Continuous variables are defined as being numerical values that are of a continuous nature, and are treated as such in regression analysis platforms. Nominal is described as either character or numerical values that are treated as discrete values in regression analysis platforms. These modeling definitions relate to the quantitative and qualitative variable definitions in the following way: quantitative variables are almost always modeled as continuous variables, yet, qualitative variables can be modeled as either continuous or nominal variables.

For the particular question analyzed, the response, or the dependent variable (the implementation rate), is actually the 'Yes' or 'No' response provided by the industrial clients served. Therefore, the data for this response was input into the database as a 'y' or an 'n.' Because of the character state of this regressor, it must be labeled nominal. When regressing nominal variables in the dependent role, JMP utilizes a different type of testing termed nominal logistic modeling. Nominal logistic modeling outputs similar tables as standard least squares modeling, however, the values are calculated in a different manner (utilizing a Chi squared statistic for comparison in place of a F or t statistic). Even though this difference exists, the output is read in a very similar manner. The similarity between the output of the two modeling methods is stated most succinctly by JMP in describing the Whole-Model Test in a help panel:

The Whole-Model Test table shows tests that compares the whole-model fit to the model that omits all the regressor effects except the intercept parameters. The test is analogous to the Analysis of Variance table for continuous responses. The negative log likelihood corresponds to the sum of squares, and the chi-square test corresponds to the F test.

Additionally, the Whole-Model Test table presents the R^2 value. Even though this value is calculated differently than in the least squares model, it still reveals the percent of total variation within the data that is described by the model fit.

Figure IV-25 presents the results from the nominal logistic regression analysis. The Whole-Model Test table reveals that the fit that exists is significant, but the fitted model explains only approximately 1% of the variation within the data. The Parameter Estimates table values show that

Whole-Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	22.3388	2	44.67763	<.0001
Full	2029.9566			
Reduced	2052.2954			

RSquare (U)	0.0109
Observations (or Sum Wgts)	2964

Lack of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack of Fit	2205	1515.2291	3030.458
Pure Error	756	514.7275	Prob>ChiSq
Total Error	2961	2029.9566	<.0001

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-0.2225064	0.0443259	25.20	<.0001
Cap. Cost	0.00000182	0.0000011	2.79	0.0950
PB	0.14697756	0.0288518	25.95	<.0001

Figure IV-25. QUESTION #5 - JMP OUTPUT TABLES the simple payback period is a statistically significant contributor to the fit, and that the implementation capital cost is not (at the alpha 0.05 level). The table in between those two tables is the Lack of Fit table, and this table is displayed when there are observations that are exact replicates of each other in terms of the X variables when there are exact replicas in terms of the x variables; with these exact replicas is associated an error entitled "pure error" that cannot be explained or predicted. [JMP Guide, p. 84, 1994] The Lack of Fit table relays a statistic that reveals the proportion of the total error that is residual error from the model versus pure error. In this case, the data in this table indicates that there is significant lack of fit in the model. The capital cost term was removed and the model rerun, however, no improvement in model utility was realized. Basically, there appears to be little to no association between the capital cost of implementation and/or simple payback period for recommendations and the implementation rate.

This test was run again within specific SIC groups to see if a relationship exists between these parameters within a certain type of industry. Table IV-8 highlights the results from that analysis. For the majority of SIC groups, there was almost no relationship whatsoever, considering the R^2 value and Prob>ChiSq values from each of three listed tables. However, several groups seemed to have enough fit to justify further analysis, and the following SIC groups were tested: 31, 32, 33 and 36.

With respect to the Table IV-8 output data, a common problem in statistical analysis can be seen within some of the observations, and is entitled *collinearity*. "Collinearity exists when two or more of the independent

Table IV-8. TABULAR RESULTS FROM SIC-SPECIFIC JMP ANALYSIS OF QUESTION #5

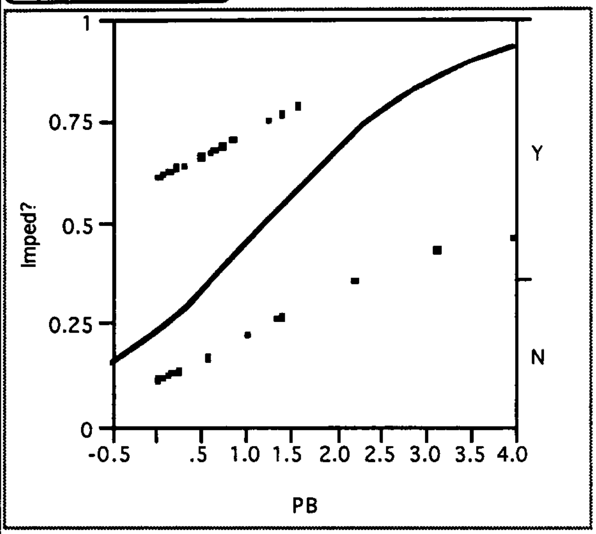
SIC group	n	R^2	Prob> ChiSq	Prob> ChiSq	Prob> ChiSq (Capital Cost term)	Prob> ChiSq (PB term)
	From Whole-Model Test Table			*From Lack of Fit Table*	*From Parameter Estimates Table*	
20	308	0.0368	0.0004	0.0007	0.0026	0.5030
22	214	0.0226	0.0383	0.0026	0.2574	0.0993
23	334	0.0130	0.0503	0.0001	0.1227	0.2433
24	150	0.0161	0.1904	0.0060	0.2569	0.7715
25	118	0.0056	0.6365	0.0059	0.8570	0.3734
26	91	0.0118	0.5382	0.0978	0.6999	0.3969
27	54	0.1016	0.0546	0.5227	0.1482	0.3033
28	123	0.0566	0.0093	0.0290	0.3544	0.0132
29	13	——	*Not*	*enough*	*data*	——
30	238	0.0272	0.0120	0.0012	0.4683	0.0082
31	***39***	***0.1070***	***0.0656***	***0.1442***	***0.6590***	***0.0492***
32	***129***	***0.0306***	***0.0655***	***0.0127***	***0.8836***	***0.0547***
33	***131***	***0.0871***	***0.0004***	***0.0512***	***0.2676***	***0.0210***
34	316	0.0415	0.0002	0.0016	0.4285	0.0020
35	205	0.0197	0.0628	0.0024	0.8503	0.0861
36	***194***	***0.0341***	***0.0132***	***0.0242***	***0.0218***	***0.1145***
37	159	0.0025	0.7606	0.0030	0.8619	0.5635
38	56	0.0229	0.4230	0.0473	0.2980	0.4735
39	92	0.0467	0.0543	0.0440	0.1157	0.2794

variables are correlated." [Mendenhall and Sincich, 1996, p. 355] An example observation would be SIC 23, where the overall model p value suggests borderline significance, however the two independent variable p values imply that neither contributes to the overall model fit. This is due to the overlapping contribution of each regressor. Collinearity can have several different effects on the results, and results must be interpreted very carefully when collinearity is detected.

To begin the SIC-specific analysis for this question, SIC group 31 (leather and leather products) was studied. The major results yielded an R^2 of just over 0.10, but with a Prob>ChiSq from the Whole-Model table of just over 0.06. Additionally, the Lack of Fit table suggests that there *is not* sufficient lack of fit, and the Prob>ChiSq from the Parameter Estimates table reveals that the contribution of the payback term is significant while that of the capital cost term is not. To further analyze this SIC group, it was chosen to drop the capital cost term and see if model fit improved in anyway. Figure IV-26 presents the results obtained.

When the dependent variable is nominal and there is only one regressor (SLR), the output includes a graph. For this particular SIC (31), the output information does infer that there is a relationship between the payback period and the implementation rate. The R^2 value held steady at about 10%, the Prob>ChiSq for the Whole-Model decreased to about 2.3% and the p value for the payback period was *almost* 5%. In a strict interpretation of these results, the fit would not be worthy of the chosen alpha value of 0.05. However, this

Imped? By PB



Converged by Gradient

Whole-Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	2.591519	1	5.183039	0.0228
Full	22.868687			
Reduced	25.460206			

RSquare (U)	0.1018
Observations (or Sum Wgts)	39

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-1.1843233	0.4602395	6.62	0.0101
PB	0.96939831	0.5018835	3.73	0.0534

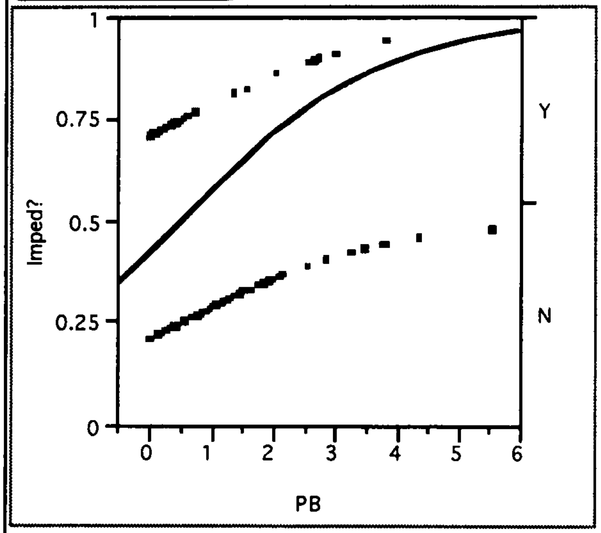
Figure IV-26. QUESTION #5a - JMP OUTPUT TABLES AND GRAPH output as a whole suggests some relationship between these two variables for clients in SIC group 31.

For SIC group 32 (stone, clay and glass products), the initial results were very similar to those obtained for SIC group 31, except for a reduced R^2 value (0.0306) and significant lack of fit. However, the capital cost term was again removed to determine if the overall fit improved. Even though the overall model fit was significant, the R^2 value remained at approximately 0.03. Therefore, this SIC group has very little relationship between measure implementation rate and the payback period.

For the pool of SIC group 33 firms (primary metal industries), initial results were notable with over 0.085 error explanation with the model and whole-model significance to 0.0004. However, the lack of fit was borderline and the contribution of capital cost of implementation was insignificant. The capital cost term was removed and the model was rerun. The JMP results are shown in Figure IV-27, and are similar to those obtained for SIC group 32. There is a relationship between the implementation rate and the payback period, however, it is not particularly strong. The R^2 value of just over 0.07 suggests an association, but a fairly weak one.

The last SIC-specific analysis was for SIC group 36 (electronic and electric equipment), for which the initial analysis suggested a weak relationship, but with the payback period insignificantly contributing to the model. With the payback period term removed, the test was run again. The new results revealed no improvement in the relationship, thus, what little correlation

Imped? By PB



Converged by Gradient

Whole-Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	6.321450	1	12.6429	0.0004
Full	83.834730			
Reduced	90.156180			

RSquare (U)	0.0701
Observations (or Sum Wgts)	131

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-0.3020068	0.2290819	1.74	0.1874
PB	0.6224578	0.1990647	9.78	0.0018

Figure IV-27. QUESTION #5b - JMP OUTPUT TABLES AND GRAPH existed originally was attributed to both implementation capital cost and payback period.

From the analysis of this question, the are two main results. First, from a non-industry specific stance, there does appear to be some relationship between the implementation rate and the recommendation capital cost of implementation and payback period. However, the relationship is primarily dependent upon payback period (and that relationship is weak at best), and is definitely not sufficient for making future inferences about the implementation rate. Second, from an industry-specific standpoint, several SIC groups reveal no relationship between the variables, however for plants in two SIC groups (31 - leather and leather products, and 33 - primary metal industries) there is a moderate relationship (at best) with recommendation payback period alone.

Chapter V
Summary Findings and Recommendations

Summary Findings

The objective of this work was two-fold, with the first goal being to develop a comprehensive, yet easily usable database of industrial manufacturing environment data from the assessments performed by the UTIAC during the last 20 years. The second goal was the subsequent analysis of the data in the newly-developed database in two ways: characterizations and statistical regression analyses. The purpose of the data analysis was to specifically investigate trends, linkages or relationships within the data, for use as a guide for future work and in summarizing program accomplishments.

In brief, the newly developed database is very easy to update with current information. Additionally, the creation of the database will greatly increase the availability of large quantities of UTIAC data for future statistical or other analysis.

From the four different types of characterizations performed, the following information and/or trends were discovered:

Background Characterizations

- Approximately two-thirds of the facilities visited by the UTIAC fall into the smaller size range category, being less than 200,000 ft^2 in size and having fewer than 200 employees and less than $20 million in annual sales.
- Almost 80% of all the assessments performed by the UTIAC have been performed in the state of Tennessee.

SIC Characterizations

- The SIC groups that have been served most frequently by the UTIAC are 20, 22, 23, 30, 34 and 36 (see Table I-1, p. 4, for industry descriptions).
- SIC group 21 is the only one that has never been visited by the UTIAC.
- Considering all the different types of industries served by the UTIAC, there is not a clearly distinct pattern in the relationship between facility square footage, number of employees and annual sales. However, there did appear to be some trends within particular SIC groups.
- The variation in the number of recommendations made per assessment for the different SIC groups served when considering energy recommendations is moderate (varying between 4 and 6). However, the number of recommendations made per assessment when considering waste recommendation varies more significantly (between 1 and 5).

ARC Characterizations

- Considering the energy-related ARC categories, over 80% of the recommendations made to facilities fall into one of the following categories: thermal systems, motor systems and buildings and grounds (see Table IV-1 for a listing of all the categories).
- Considering the waste-related ARC categories, over 55% of the recommendations made to facilities fall into one of the following categories: operating procedures, recycling of miscellaneous materials and waste disposal (see Table IV-1 for a listing of all measure categories).

Recommendation Implementation Characterizations

- The implementation rate for recommended measures has been approximately 10% higher in Kentucky, North Carolina and Virginia (about 60%) than in Tennessee and South Carolina (about 49%).
- With respect to different levels of clients' annual sales, the implementation rate varies little, with a respective minimum and maximum of 51% and 56%.
- The only significantly different implementation rate discovered when implementation rate is normalized with facility production hours is the range of 1,000-1,999 annual working hours, with a 77% success rate.

- There is a distinct, increasing trend in successful implementation of the UTIAC recommended measures as the number of recommendations made per report increased from 1 to 9. However, between 9 and 14 recommendations per report, this trend disappears and the implementation rate levels off, 12 recommendations per report notwithstanding where the implementation rate reaches almost 90%.
- With the implementation rate broken down by the SIC groups served, large variation was discovered. Two SIC groups had lower than a 40% implementation rate, with SIC group 27 at 13% and SIC group 36 at 36%, and two SIC groups had a higher than 60% implementation rate (31 at 64% and 38 at 61%).
- Distributed by ARC categories, the implementation rate for energy-related recommendations varies between 34% and 58% for ARC groups having more than 50 total recommendations, and between 25% and 50% for waste-related recommendations and ARC groups having more than 5 total recommendations.

From the five specific questions posed and answered with regression analyses, several statistically significant relationships were uncovered. To initiate study of each question, a preliminary analysis had to be performed to determine which facility size-related variable could be utilized in lieu of square footage (due to a lack of data). After several analyses, it was discovered that annual sales is most statistically similar to facility square footage under roof. Upon successful completion of the preliminary analysis, the results discovered were as follows:

- Through utilizing MLR to investigate the association between plant size and annual electricity and natural gas consumption (question #1), a significant, positive relationship was found across all SIC groups between plant size and annual electricity consumption alone. As the plant size increases, the annual electricity consumption also increases.

Additionally, when focusing the analysis within specific SIC groups for the same relationship, four SIC groups (20, 23, 34 and 35) revealed positive, modest to better-than-average statistical significance between plant size and annual electricity consumption, and 1 SIC group (30) revealed a *highly* significant relationship between all three variables.

- For question #2, the relationship between facility size and UTIAC recommended plant total energy and waste cost savings was explored. There was no statistically significant relationship found between all three variables in the MLR analysis, however, a modest curvilinear relationship was found between plant size and the recommended annual energy cost savings. Again, a positive association was found with the UTIAC recommended energy cost savings increasing as facility size increases. Furthermore, when focusing within specific SIC groups, moderate, positive linear and curvilinear relationships were discovered within SIC groups 20, 30, 34 and 35.
- Question #3 was to analyze the relationship between the annual production hours of the facility and the total annual energy consumption and waste generation. No relationship was found between all three variables. However a mediocre, positive relationship was found between annual production hours and annual energy consumption. For the industry specific analysis, fair to moderate linear and curvilinear relationships were exposed within SIC groups 20, 22, 23, 30, 34 and 35.
- To completely answer question #4, investigating the relationship between the number of recommendations made per report and the total UTIAC recommended energy conservation and waste generation quantities, several all manufacturing type tests were performed. No statistically significant relationships were revealed from these analyses. However, several SIC groups were found to have a moderate level of statistically significant, positive association between the number of recommendations made per report and the total UTIAC recommended energy conservation quantity, and include SIC groups 22, 34 and 36.
- For question #5, it was determined that there is almost no association between the implementation rate and the capital cost of implementation and/or payback period for all industrial manufacturing types. For the

SIC-specific investigation relative to this question, SIC groups 31 and 33 revealed a meager, positive relationship between implementation rate and payback period alone.

Pertaining to recommendations for future changes by the UTIAC related to a shift in analytical focus or techniques utilized, or to improve the overall performance of the UTIAC, several points can be made. From the characterizations performed, two possible avenues for improving overall implementation rates would be to simply perform more assessments in Kentucky, North Carolina and Virginia, and to attempt to include 9 or more worthwhile recommendations for each client served. Additionally, focusing auditing efforts on those specific SIC groups with higher-end implementation rates (or reducing the frequency of serving those with lower-end implementation rates) would most likely produce an increase in the overall measure implementation rate.

From the regression analyses performed, some data could be carefully utilized as a gauge of energy consumption magnitude, and they appear to be annual sales and annual production hours (with range limitations on those categories specified in Chapter IV). For the purposes of illuminating which factors influence facility personnel decisions to implement measures, it was found that neither the payback period nor the capital cost of implementation have an association with the implementation rate.

In summary, several positive, non-SIC-specific relationships were revealed relative to each of the categories of questions posed, and in almost all cases there existed at least one or two SIC groups with statistically significant and positive relationships.

Recommendations

Now that the onsite UTIAC database has been created, the most important recommendation is that it be maintained and used periodically to characterize changes in program performance.

Pertaining to recommendations for improving or continuing the analysis of this type of data, one recommendation is simply to collect more waste related information. As the limiting factor in this study, the lack of waste related information eliminated the opportunity to search for relationships within different SIC groups. It will likely require several years' of additional assessments to gain the significant quantity of information needed to perform industry-specific waste issue-related analyses.

Another recommendation for improvement would be to perform the regression analyses utilizing an analytical method other than the least squares method. One example might be the weighted least squares method, which allows for the weighting of some observations more heavily than others. This method of analysis allows for greater control over individual observations such as reducing the effect on the model of observations that have greater error variances.

Additionally, future work with this type of data could include analyzing different independent variables for significance and extent of relationship with other dependent variables of interest. Since only five questions were explored in this work, a vast quantity of other possible, interesting relationships could be analyzed.

Follow up analysis could also include analyzing other dependent variables as indicators of influence on recommendations' implementation. It is often the case in statistical analysis that variables are found linked in some way to what was previously thought unrelated information. Therefore, performing a MLR with many independent variables (perhaps 10 or more) might yield some previously disguised relationships as relates to energy consumption, waste generation or recommendation implementation in the industrial manufacturing environment.

Bibliography

NOTE: The bibliographies marked by an asterisk (*) at the end of the reference are noted as being generated by field management agencies.

[1] Walker, Esteban E. Personal communication. February 24, 1997.

[2] SAS Institute, The. "JMP Statistical Discovery Software." <www.sas.com/jmp>. February 1997.

[3] Wandell, Dell. "News About the New SIC Manual." <www.ntis.gov/business/sic.html>. December 1996.

[4] Walker, Esteban. Applied Regression Analysis, Course Notes for Statistics 461. Knoxville, TN: Graphic Creations - Academic Services Division, Fall 1996.

[5] Woodruff, Madeline G., et al. Analysis of Energy-Efficiency Investment Decisions by Small and Medium-Sized Manufacturers. DOE/PO-0043, Washington, DC, PO-64, March 1996.*

[6] Mendenhall, William, Terry Sincich. A Second Course In Statistics: Regression Analysis, Fifth Edition. Upper Saddle River, NJ: Prentice-Hall, Inc., 1996.

[7] Muller, Michael R., et al. IAC Database Manual, Version 5.0. Piscataway, NJ: OIPEA, October 1995.

[8] Muller, Michael R., and P. P. Polomski. "Industrial Assessment Database for Energy Efficiency and P2." Proceedings of the 21st Annual RREL Research Symposium. pp. 29-35, 1995.*

[9] Muller, Michael R.,T. Barnish, and P. P. Polomski. "On Decision Making Following an Industrial Energy Audit." Proceedings of the 17st National Industrial Energy Technology Conference. pp. 150-155, 1995.*

[10] Thomas, Todd M. "Analysis of the Characteristics and Interrelationships of Industrial Energy Consumption and Waste Generation for Selected Tennessee Industries." Thesis University of Tennessee, December 1994.

[11] Muller, Michael R. "Industrial Support Through the US DOE's EADC Program." Proceedings of the International Energy Agency Workshop on Industrial Energy Efficiency: Policies and Programmes. pp. 107-124, 1994.*

[12] Groliers Multimedia Encyclopedia CD-ROM, 1994.

[13] JMP Statistics and Graphics Guide. Cary, NC: SAS Institute, Inc., 1994.

[14] Kirsch, F. William, William F. Kershner and William J. Clark, III. Energy Analysis & Diagnostic Centers: Five Years of Assisting Small and Medium-Size Manufacturers to Conserve Energy and Reduce Costs. Philadelphia, PA: University City Science Center, December 1993.*

[15] Muller, Michael R., and P. P. Polomski. "Insights Into Industrial Energy Auditing From the EADC Program: An Academic Perspective." Proceedings of the 15th National Industrial Energy Technology Conference. pp. 278-285, 1993.*

[16] Montgomery, Douglas C. Designing and Analysis of Experiments, Third Edition. New York, NY: J. Wiley & Sons, 1991.

[17] Webster's New Collegiate Dictionary. 9th ed., 1989, s.v. "database."

[18] Standard Industrial Classification Manual. Office of Management and Budget, 1987.

Appendix

Table A-1. SIC CATEGORIES AND THEIR ASSOCIATED ALPHABETICAL AND NUMERICAL GROUPINGS

Industry	SIC Industry Division	SIC Groups
Agriculture, forestry and fishing	A	01 - 09
Mining	B	10 - 14
Construction	C	15 - 17
Manufacturing	D	20 - 39
Transportation, communications, electric, gas and sanitary services	E	40 - 49
Wholesale trade	F	50 - 51
Retail trade	G	52 - 59
Finance, insurance and real estate	H	60 - 67
Services	I	70 - 89
Public administration	J	91 - 97
Nonclassifiable establishments	K	99

Table A-2. RUTGERS UNIVERSITY DATABASE COMPONENT #1: ASSESSMENT-SPECIFIC INFORMATION

Field	Field Name	Description
1	id	Unique identifying number given to all records based on IAC Name and Report Number. This number is used when linking the two databases.
2	iac	The identifier assigned to each IAC.
3	repnum	The number assigned by the IAC to their visit and subsequent report.
4	visitdate	The date the assessment was performed.
5	sic	The Standard Industrial Code that represents the principle product manufactured by the plant.
6	sales	The annual sales in dollars for the site reported by the client.
7	employees	The total number of employees on the site as reported by the assessment client.
8	plant_area	The total amount of in square feet of area used for production and office support purposes.
9	products	Principle products of the plant (in words).
10	resources	Total number of resources tracked at the plant.
11	produnits	The units of production for the principle product.
12	prodlevel	The total number of units produced annually as reported by the assessment client.
13	prodhours	Client reported annual production hours.
14	numars	The total number of ARs recommended in this report.

Table A-2. (Continued)

Field	Field Name	Description
15-41	————	The annual usage and cost of electrical consumption, electrical demand, electrical fees, natural gas, L.P.G., fuel oil #2, fuel oil #4, fuel oil #6, coal, wood, paper, other gases and others - taken from actual bills provided by the client prior to the assessment.
42	nrgcosttot	Total energy cost for this client. Figure is produced by summing energy costs reported in columns 15-38.
43-54	————	The annual production and cost of waste: water disposal, non-hazardous liquid, hazardous liquid, non-hazardous solid, hazardous solid, and gaseous waste in dollars and waste stream units.
55	wstcosttot	Total waste cost for client. Figure is produced by summing the waste costs reported in columns 40-51.
56	comments	General comments about the assessment.
57	fy	The fiscal year in which the assessment was performed.

Table A-3. RUTGERS UNIVERSITY DATABASE COMPONENT #2: RECOMMENDATION-SPECIFIC INFORMATION

Field	Field Name	Description
1	superid	The unique identifying number given to all records based on IAC Name, Report Number and the Assessment Recommendation Number. This number is used when linking to the assessment database(s).
2	ar_number	The recommendation number as it appears in the report.
3	appcode	The application code for the recommendation.
4	arctype	The recommendation type.
5	arc	The code representing the specific recommendation made.
6	impdate	Client reported date of implementation of this Assessment Recommendation. This date may be estimated if the Assessment Recommendation is not implemented by the time of client contact.
7	impstatus	Client reported implementations status of this Assessment Recommendation.
8	impcost	Client reported implementation cost. This cost may be estimated if the Assessment Recommendation is not yet implemented by the time of client contact.
10	psourccode	The Primary resource coded per "Resource Identification Code". This resource may not necessarily be the most important resource involved in the Assessment Recommendation, but it is usually chosen based on greatest usage before conservation measures are suggested.
11	pconserved	The amount of primary resource conserved.
12	psaved	The primary resource's dollar savings for this Assessment Recommendation.
13	ssourccode	The Secondary Resource involved in this Assessment Recommendation. This resource is usually chosen based on second highest amount of usage before conservation measures are suggested.

Table A-3. (Continued)

Field	Field Name	Description
14	sconserved	The amount of secondary resource conserved.
15	ssaved	The secondary resource's dollar savings for this Assessment Recommendation.
16	tsourccode	The Tertiary Resource involved in this Assessment Recommendation. This resource is usually chosen based on second lowest amount of usage before conservation measures are suggested.
17	tconserved	The amount of tertiary resource conserved.
18	tsaved	The tertiary resource's dollar savings for this Assessment Recommendation.
19	qsourccode	The Quaternary Resource involved in this Assessment Recommendation. This resource is usually chosen based on the least amount of usage before conservation measures are suggested.
20	qconserved	The amount of quaternary resource conserved.
21	qsaved	The quaternary resource's dollar savings for this Assessment Recommendation.
22	rebate	Indicative whether the Assessment Recommendation included a rebate for implementation.
23	incremental	Indicates if the Assessment Recommendations to be implemented on an incremental basis. Incremental data is included in the database for the first two years only.
24	descript	Description in words of the individual Assessment Recommendation.
26	imp_comm	Description in words of any variation between the suggested recommendation and the actual implementation.

Table A-4. LISTING OF UTIAC PLANT DATABASE DATA

Sectional Information Category	Catalogued Data Description	
BASIC PLANT & ASSESSMENT INFORMATION	Assessment number	
	Assessment date	
	Assessment year	
	Assessment season	
	Assessment state	
	Principal assessor	
	Number of student assistants	
	SIC code	
	SIC group	
	Annual sales	
	Number of employees	
	Plant square footage	
	Annual production hours	
	Number of energy-related recommendations	
	Number of waste-related recommendations	
WASTE INFORMATION	Number of reported	Solid
	waste streams in each	Water
	of the following	Air
	categories:	Special
		Hazardous
		Unspecified
		Total
	Estimated total annual waste costs for assessed facility	

Table A-4. (Continued)

Sectional Information Category	Catalogued Data Description	
WASTE INFORMATION (Continued)	Recommended annual waste cost savings through waste-related recommendations	
	Recommended percent (%) waste cost savings	
ENERGY INFORMATION	Electricity	Annual consumption
		Average monthly demand
		Total annual cost
		Average unit cost
	Natural gas	Annual consumption
		Total annual cost
		Average unit cost
	Propane	Annual consumption
		Total annual cost
		Average unit cost
	Fuel oil #2	Annual consumption
		Total annual cost
		Average unit cost
	Fuel oil #6	Annual consumption
		Total annual cost
		Average unit cost
	Coal	Annual consumption
		Total annual cost
		Average unit cost
	Waste wood	Annual consumption
		Total annual cost
		Average unit cost
	Other fuel	Annual consumption

Table A-4. (Continued)

Sectional Information Category	Catalogued Data Description	
ENERGY INFORMATION (Continued)	Other fuel (continued)	Total annual cost
		Average unit cost
	Total energy	Annual consumption
		Annual cost
	Recommended total annual energy savings	
	Recommended total annual energy cost savings	
	Recommended percent (%) energy savings	
RECOMMENDATION TOTALS	Electricity	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Natural gas	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Propane	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Fuel oil #2	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Fuel oil #6	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Coal	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Waste wood	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Other fuel	Estimated annual savings (MMBTU)
		Estimated annual savings ($)
	Total energy	Estimated annual savings (MMBTU)
		Estimated annual savings ($)

Table A-5. LISTING OF UTIAC RECOMMENDATION DATABASE DATA

Catalogued Data Description	
Assessment number	
Assessment recommendation number	
ARC type	
ARC code	
ARC application code	
Implementation cost	
Implementation date	
Implementation status	
Implemented quantity conserved	
Implemented cost savings	
Implemented (Yes or No)	
Primary Resource	Code
	Quantity conserved
	Cost saved
Secondary Resource	Code
	Quantity conserved
	Cost saved
Tertiary Resource	Code
	Quantity conserved
	Cost saved
Quaternary Resource	Code
	Quantity conserved
	Cost saved
Rebate utilized (Yes or No)	
Incremental implementation (Yes or No)	
Assessment recommendation description	

Table A-6. EXPLANATION OF ABBREVIATIONS USED IN UTIAC DATABASES

Column Heading	Abbreviation	Explanation
General Abbreviations (both databases)		
Non-column specific	MM$	million dollars
	k	one thousand
	Imp.	implemented
Plant Database Abbreviations		
Assessment season	g	spring
	s	summer
	f	fall
	w	winter
Assessment state	K	Kentucky
	N	North Carolina
	S	South Carolina
	T	Tennessee
	V	Virginia
	A	Alabama
	G	Georgia
	M	Maryland
Principal assessor	S	William R. Snyder
	J	Richard J. Jendrucko
	F	Frederick W. Symonds
	G	Bernie McGraw
	B	Robert Baugh
	M	Phylissa S. Miller
	W	William S. Johnson
Non-column specific	Ass.	assessment
	Prod.	production
	NRG	energy

Table A-6. (Continued)

Column Heading	Abbreviation	Explanation
Plant Database Abbreviations (continued)		
Non-column specific (continued)	Rec.	recommendations
	WST	waste
	TTL	total
	Wtr	water
	Spcl	special
	Hzds	hazardous
	Est.	Estimated
	MMBTU	one million British thermal units
	Ann.	annual
	Sav.	savings
	Recc.	recommended
Recommendation Database Abbreviations		
Source code	E1	electrical consumption (older version)
	EC	electrical consumption (newer version)
	ED	electrical demand
	EF	other electrical fees
	E2	natural gas consumption
	E3	L.P.G. (propane) consumption
	E4	fuel oil #1 consumption
	E5	fuel oil #2 consumption
	E6	fuel oil #4 consumption
	E7	fuel oil #6 consumption
	E8	coal consumption
	E9	wood consumption

Table A-6. (Continued)

Column Heading	Abbreviation	Explanation
Recommendation Database Abbreviations		
Source code (continued)	E10	paper consumption
	E11	other gas consumption
	E12	other energy consumption
	W1	water disposal
	W2	other liquid nonhazardous
	W3	other liquid hazardous
	W4	solid waste nonhazardous
	W5	solid waste hazardous
	W6	gaseous waste
	R1	personnel changes
	R2	administrative costs
	R3	primary raw material
	R4	ancillary material cost
	P1	primary product
	P2	by-product product
Non-column specific	App.	application

Table A-7. INFORMATION ON SELECTED AVAILABLE STATISTICAL SOFTWARE PACKAGES

Statistical Software Package	**Vendor**	**Contact Information**	**Current Version**
BMDP	BMDP Statistical Software, Inc.	1440 Sepulveda Boulevard Los Angeles, CA 90025 (310) 479-7799	PC90
GLIM	Numerical Algorithms Group (NAg Ltd.)	Wilkinson House, Jordan Hill OX2, 8DR, Oxford, UK (44) 1865-511245	4.8
JMP	SAS Institute, Inc.	SAS Circle, Box 8000, Cary, NC 27512-8000 (919) 677-8000	3.2
MICRO-TSP	Quantitative Micro Software	4521 Campus Drive #336 Irvine, CA 92715 (714) 856-3368	7.0
MINITAB	Minitab, Inc.	3081 Enterprise Drive State College, PA 16801-3008 (800) 448-3555	11.0
SAS	SAS Institute	SAS Circle, Box 8000, Cary, NC 27512-8000 (919) 677-8000	6.07
SPSS	SPSS, Inc.	444 North Michigan Avenue Chicago, IL 60611 (800) 543-5831	4.1
STATPAC GOLD	Walonick Associates, Inc.	3814 Lyndale Avenue South Minneapolis, MN 55409 (612) 822-8252	4.2
SYSTAT	SPSS, Inc.	1800 Sherman Avenue Evanston, IL 60201-3793 (708) 864-5670	6.0

Table A-8. CALCULATED AVERAGE NUMBER OF RECOMMENDATIONS MADE PER ASSESSMENT WITHIN SIC GROUPS

SIC Group	Total Number of Assessments	Total Number of Recommendations	Average Number of Recommendations per Assessment
20	54	341	6.31
22	32	212	6.63
23	53	346	6.53
24	24	241	10.04
25	20	112	5.60
26	18	87	4.83
27	10	48	4.80
28	24	141	5.88
29	3	13	4.33
30	34	354	10.41
31	7	39	5.57
32	24	119	4.96
33	22	101	4.59
34	48	306	6.38
35	35	211	6.03
36	35	212	6.06
37	27	144	5.33
38	11	62	5.64
39	14	92	6.57
Totals	495	3181	116.49
Averages	26.05	167.42	6.13

Table A-9. ARC APPLICATION CODES

Code	Application	Examples
1	Manufacturing Process	Process heat recovery, variable speed drives on process equipment, solvent recovery
2	Process Support	Air compressors, steam, nitrogen, cogeneration
3	Buildings and Grounds	Lights, HVAC, burn waste for heat
4	Administrative	Taxes, inventory control, sale of wastes

Vita

Jonathan Gregory Overly was born in Newport, Rhode Island on September 22, 1968. After his father was relocated several times in several years through the Naval service, the Overly family landed in Knoxville, Tennessee in approximately 1971. Jonathan attended Bearden Elementary, Middle and High School in Knoxville, and graduated in May of 1986. He entered the University of Tennessee at Knoxville in the fall of 1986, and ten years later graduated with his first degree, a Bachelors of Science in Engineering Science. Along the way he experienced acute kidney failure and was on dialysis for approximately six months before receiving a kidney transplant, allowing him to pursue his degree more vigorously.

Approximately 2 years before finishing his Bachelors degree, he began working for the University of Tennessee Industrial Assessment Center, performing no-cost industrial assessments to regional manufaturing facilities. Upon completing his Bachelors degree Jonathan began pursuit of a Masters Degree at UT in Engineering Science with a concentration in Environmental Engineering, continuing his work with the UTIAC. Jonathan expects to graduate in May 1997 with his Masters degree, and has already accepted a position with the Energy, Environment and Resources Center at UT as a Research Associate, working in the pollution prevention arena.