

Statistics and Data Analysis for






SOCIAL WORKERS



JOHN L. CRAFT

Statistics and Data Analysis for





SOCIAL WORKERS

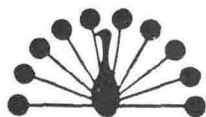


JOHN L. CRAFT

University of Iowa

with the assistance of

LAWRENCE R. ASKLING



F.E. PEACOCK PUBLISHERS, INC.
ITASCA, ILLINOIS 60143

Copyright © 1985
F.E. Peacock Publishers, Inc.
All rights reserved
Printed in the U.S.A.
Library of Congress
Catalog Card No. 84-061422
ISBN 0-87581-305-4

Preface

This book is about the use of numbers in social work and other human service professions. It is designed to present basic statistical concepts directly applicable to the types of analytical problems encountered by social workers and others.

This book is intended for undergraduate and graduate students in social work and related disciplines. For several years students have requested a book written "just like you present it and explain it in class." This book is an attempt to honor those requests. The authors' classroom experience and work with human-service agencies over the years have clearly indicated that workers and administrators need a text that will help them understand and interpret the data generated by themselves and others in the delivery of human services.

We are deeply indebted to the scores of social-work students at the University of Iowa who have provided encouragement for and helpful criticism of the content of the book. They are truly "co-authors." A special debt of gratitude is due Dixie Kramer, who spent many hours skillfully translating the authors' handwritten narrative into the final product. We are grateful to John F. Else for his many helpful criticisms and suggestions.

We are especially grateful to Joyce Usher and Linda Pierce of F. E. Peacock Publishers, Inc., for their thoroughness, diligence and general positive demeanor throughout the preparation of this book.

Special thanks go to Tom Walz, who has contributed to the authors' professional and personal development in so many ways, many of which are not measurable by the statistical techniques presented in this book.

This book is dedicated to the memory of our friend and buddy, Bill Sackter. In his unique way Bill taught us and countless others the true meaning of love, compassion, and friendship. We are truly richer for having known him and shared his life.

I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S. to Dr. Frank Yates, F.R.S. and to Longman Group Ltd, London for permission to reprint Tables 8.1, 8.2 and 8.4 from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th Edition, 1974).

We are especially grateful to our wives, Marty and Jan, for their help, understanding, and sharing their lives with us. Their tolerance and encouragement have been very important to us.

JOHN L. CRAFT
LAWRENCE R. ASKLING

Contents

CHAPTER

1. Introduction	1
2. The Nature of Data	6
3. Organization and Presentation of Data	12
4. Data Reduction Techniques	26
5. Measuring Variability	40
6. Crosstabulation Analysis	56
7. Correlational Methods	75
8. Tests of Statistical Significance	94
9. Data Processing, <i>Lawrence R. Aspling</i>	108
References and Suggested Readings	123
Appendix A. Police Perceptions Study	125
Appendix B. ADC Study	141
Appendix C. Coding Sheet and Organization Form	157
Index	159

Chapter 1

Introduction

The focus of this book is on the analysis and interpretation of data generated by social workers in practice. Social-work practice involves a variety of decision-making levels and activities, from policy making, planning, and administration (macro-practice) to direct-service intervention (micro-practice) with individual clients. Social work practice involves continual decision making and research skills are an important and integral part of sound decision making.

Quality decision making requires that judgments be formed on the basis of good information. Social work research is concerned with the systematic collection, analysis, and interpretation of information for input into the decision-making process. Social work research can be considered a tool to aid the practitioner and policy maker in making the best informed decisions possible. The methods of social work research guide the production of practice-relevant information; the methods of social work practice determine how the information is utilized.

The research conducted in social work can be characterized by its diversity in both methods used and the investigators who use these methods for carrying out their inquiries. Social work research is noted for its eclectic nature in choosing a method or methods to investigate a particular problem.

Much of what we consider as social work research is carried out in the practice setting by practitioners themselves. These information gathering and analyzing activities are often applied in nature and limited in scope. Other social work research involves more extensive investigations, either in terms of time and effort invested in the study or in terms of the particular research design and data-analysis techniques employed.

University-based research conducted by social work faculty and students (latent practitioners) has grown markedly in the last decade. Some of the research done in the academic setting is directly related to practice concerns and some to more theoretical issues. University researchers (including students) often work hand in hand with caseworkers, supervisors, planners,

and administrators within agency settings in investigating and obtaining solutions to both short-range and long-range problems faced by these practicing professionals. Rather than maintain an artificial dichotomy between social work research and social work practice, it is becoming more and more recognized that social workers today can comfortably deal with both; thus, the term practitioner/researcher is now widely used to denote the integration of the two aspects of the social-worker role.

The social work research process is a multi-stage process that involves problem formulation, choosing an appropriate research design, devising data collection instruments, data collection, and data analysis and interpretation. The content of this book is concerned primarily with the data analysis and interpretation stages of the research process. This is not to say that the earlier stages are less important. Excellent readings are available on the logic of research, research designs, instrument construction, ethical considerations in conducting research, and other important research issues. But, unless social workers have a good understanding of what they can and should do with data once they have them, knowledge of methods and techniques for producing data lead them only part way to the end of the research maze. Social workers can and should use an in-depth understanding of data analysis as a guide in the selection and application of various research methods and techniques. So in our discussions here we will assume that the data under perusal were gathered while taking into account these other important research considerations. We will then begin from that point to see how we can better analyze and understand the data generated from our data-collection efforts.

Before we begin our journey through the labyrinth of data processing and analysis, it will be helpful to set the context or framework within which social work research is carried out.

Our primary interest in this book is to provide aid to the social worker in practice. We shall concentrate our topical presentations and discussions on those concepts and techniques that we feel are most important and useful in analyzing the kinds of information (data) that social workers deal with in action. Our goal is to help social workers learn how the appropriate use of statistical analysis will in turn help them make better sense out of the myriads of information surrounding them on a regular basis.

In doing social work research (broadly defined), social workers sometimes collect information on what are termed *samples*, and at other times on what are called *populations*. Sometimes these terms become confused. Both terms refer not only to people or events *per se*, but also rather to a collection of measurements made on these people or events. Thus, a population is a set of measurement scores on a defined number of people or events under study. If we are interested in studying all of the clients that receive services during a one-year period at a particular agency, then by definition the measurements on the clients over this one-year period are our population of scores. A *sample* is merely a subset of scores from the larger population of scores. If we picked one half of the clients from the one-year period, then this would be a *sample* of scores derived from the population of scores from the total one-year period.

Whether we consider a set of scores to be a sample or population of scores depends in part on what the investigator intends to do with the outcome from an analysis of the data. If our interest is only in describing and interpreting a set of scores (data) that we have *in hand*, then we use *descriptive statistics*. The data can be considered as coming from either a sample or a population. The distinction is not useful in this case.

However, if our intent is to use this set of data in hand to try to understand an even larger *body of data*, then we apply *inferential statistics* on the sample data and make estimates about the characteristics of a larger body of data (population data). We will return to the concepts of samples and populations of scores at several points in this book.

As the preceding discussion indicates, social-work “researchers” today carry out their activities in a variety of settings, picking from a fairly extensive selection of research methods and techniques. Social workers use a variety of techniques to collect information (data) about clients, families, other workers, agencies, and communities. These include data derived from client-case research, agency records, questionnaires, interview schedules, management-information systems (MISs), and census data. Once a set of data is in hand, the social worker needs a large array of statistical tools available to aid in organizing, summarizing, describing, interpreting, and presenting the data. The objective is to use these tools to help expand the knowledge base and aid in improving delivery of service.

The following chapters detail what these tools are and when and how they can be used. The outcome will be the ability to do a complete *descriptive* analysis of a set of data, including frequency distributions; measures of central tendency, dispersion, and relative standing; measures of association and correlation; and communication of data by tables and graphs. Further, you will learn how to use a few of the more common tests of statistical significance. Real data from a variety of social work research efforts are used throughout the book to present the concepts and procedures.

Quite often we have a computer available to help us with our data analyses. If we are really fortunate, we have an experienced computer programmer attached to the computer. However, many times we have only a hand calculator at our disposal to aid with our computations. And, sometimes, we have just a pencil and paper. In any case, it is useful and even intrinsically pleasing to do some of the calculations while learning to use the variety of statistical procedures and techniques presented in this book. By doing some of the calculations, we get a “feel” for the data, as well as an understanding of what the procedures do and when they are appropriate to use. And, of course, after we have done several calculations, we will come to appreciate more fully just how valuable a computer is as an analytic aid.

Many students start a course in social work research and data analysis with either an abhorrence or trepidation of numbers (or both) and anything or anyone associated with them. Quite often these negative feelings die and are replaced by positive feelings of accomplishment and appreciation. The following simple numerical exercises are introduced to help rehearse some

fundamental mathematical manipulations. To many students these exercises will appear to be banal or just plain simple. Of course, you are entitled to your opinion, and we respect your right to object. But after you are through sneering, *work the exercises anyway, for there is method in our madness.* Everything we do has a purpose.

PROBLEMS

1. List the numbers 21 through 30, inclusive, in a single column on a piece of paper. Write in a second column the *number of times* each of the following numbers (21 through 30) occurs:

29	25	21	26	30
24	29	23	27	26
24	30	29	27	28
21	28	27	26	27

2. Which number in problem 1 appears most often? least often?
3. In the following list there is one number that is greater than three of the others but also less than three of the others. What is the number?
313; 98; 101; 7; 3613; 69; 25
4. There is a number, let us call it M, which is equal to your age plus the ages of two of your friends, divided by 3. Calculate the value of M.
5. Subtract 9 from each of the following numbers, being careful to watch the sign of the remainder.

13; 14; 4; 8; 5; 10

Now square each of the remainders and add the squares; that is
 $(13-9)^2 + (14-9)^2 + (4-9)^2 + (8-9)^2 + (5-9)^2 + (10-9)^2 = ?$

6. Now divide the answer to problem 5 by 6 and take the square root of the quotient. Use a calculator or table of square roots.
7. Add the following numbers and square their sum: 2; 4; 6; 8; 10. This may be written as $(2+4+6+8+10)^2$.
8. Square each of the following numbers and add their squares: 2; 4; 6; 8; 10. Once you have done this, compare your answer to the answer to problem 7, and you will see that the *square of the sum* (problem 7) does not equal the *sum of the squares* (problem 8).
9. Arrange the following numbers in order from greatest to least:
35; 17; 2; 14; 21; 7; 11

Rank these seven numbers by assigning a 1 to the greatest number, a 2 to the next greater number, and so on until you assign a 7 to the least number.

10. Add the following set of numbers: 22; 24; 26; 28. Divide the sum by 4. Call this quotient A. Then add 10 to each of the four numbers, and again divide by 4. Call this quotient B. How does B compare with A?
11. Now divide 22, 24, 26 and 28 each by 2, then add the quotients, and then divide by 4. Call this result C. How does C compare with A in problem 10?

12. We assume you have worked each of the eleven problems and found them enjoyable. Those of you who meet our assumption have performed every single mathematical operation and applied many of the formulas used in this text. You are to be commended. You are now ready to enter the remaining chapters of the book. Before you start, a word of warning! You must remember not to lug dirt to a hilltop.

Chapter 2

The Nature of Data

When we do a descriptive analysis of a set of data, we say we are *doing* descriptive statistics on the data, or we are *using* descriptive statistics on the data.

But what are these data to which we are applying all of these statistical procedures? We will define data as a set of numbers that represent categories or instances of variables. How we get and assign these numbers is a basic issue we will address shortly. But, first we will present two examples of the types of data with which social workers might be involved.

Table 2.1 presents a portion of a data set generated from a social work research project. The project is concerned with collecting information about persons who use the services of domestic-violence shelters in a midwestern state; it is part of an ongoing agency information system used to analyze and evaluate the agency's clientele and services.

What do we see when we look at Table 2.1? At first glance, all we see is a large set of numbers that appear to be meaningless. What do these numbers represent? Where did they come from? What do they tell us about shelter clients? We will have to do something to this set of numbers that will help us impart meaning to the numbers.

The numbers in Table 2.1 represent responses to a large number of questions asked of a sample of fifteen shelter clients. All responses were coded to numerical form for statistical analyses. One set of responses from one client is referred to as one *case*. Table 2.1 contains numerically coded responses from 15 cases. The first two lines (rows) contain responses from case 1. The next two lines are data from case 2, and so on through case 15. Each question in this context is called a *variable*. This study involved obtaining data related to 117 different variables. Thus, the data set in Table 2.1 contains numerically coded information on 117 variables for 15 cases.

Table 2.2 contains a small sample of data obtained from a central child-abuse registry in a midwestern state. The original data are generated by child-protective service workers who fill out a reporting form for each of their cases.

TABLE 2.1. Selected data from a study of domestic-violence shelters

13252111	2212421652	86	8	66	1115	164121	3	1111121212
11222222221122122					32222222122222221412221	5	222212	
132951	121	22111263121			1115511			21222221
1222221121122222	122212223212222222221221				5	211123		
13252111	12211115511566				3111155	163122	1	21111122212
11112211221112221	112222121121222222221222262221				22			
1317211131121121631346416111111	111336163132				1121121222221			
12122221222222121211211222122222212222122	192222112				2212221			
12282121	1222	21	112	1	241114	153	31	1111121221211
11112211221122121	12221222212212122				2242122112			3212221
122121112	22233222112	1			3311134	144132		1121111222222
12122222221212222	2				312222222			124222113223212221
122421222112211	12112	666			2	1213	161172	2
12222221212	1112222212222222211221122				3154			211222211221
122421212111221112112	46	24			1	14	1431221	221212212222
1212222212121222112122222221222222					23523121			1221211222
122822	2114122211341136	666666			211124	12512111211111221222		
12222221122	142222221	121211212122122232221211121211221						
0420212121222	216612	66	6	66	521114	1	222	11211112221221
1222222221	1512221222122212222				1202222	22		1222212
0439212121222111621171					3211177	162272		1221111111212
1222222122	122221222122222212212222				149212	15111223221		
1249212121122122256111444					2	1116	152142	1221211221221
2222221211	111222222	12222122					1502121	22
123221	2	122215213511362	2		3111155	12317112		11121221222
12222221222	1	1221122221222222				1262222	52	2212221
12221121	21211113111362	8	28		3211111	141112	1	21111121222
111222111211222141122212211211222222122122112221111211221								
12322121	221212111511566				21216	162		2111122211221222
11212211121211221311112222211222222						13122		211112211222

These are sent to a central office where the numerically coded information is entered into a computer for analysis. The analyses of these data are used by department staff for planning and allocation of resources.

What do the numbers in Table 2.2 tell us about child abuse in this midwestern state? Without further information, very little or nothing. What we must do is apply certain statistical procedures to the numbers to see what information they contain about child-abuse cases.

Once again, the information obtained about one client is called one case. Table 2.2 contains numerically-coded information on 34 variables for a sample of 40 child-abuse cases. Note in Table 2.2 that it only requires one line

TABLE 2.2. Selected data from a central child-abuse registry

040	518792	1	1	7211	21431003108	534416403931000	11
723	52179016	1	9613	20732003	86	325 3 26	33000150
790	52979011	199922	31331003178	764416343743000	16		
240	6 579014	48999131617312011801	12221645553	28			
070	61279010	1	4813	20531030	0	334426303500000	3
900	61379010	1	2625	21031000130	41	16435100000	3
140	615790 5	1	6723	20831000170	424	2 40	00000 5
773	61879010	3	4813	20031 00170	513312212101000	12	
780	6 179014	2	9613	20732000170	314	6 24	10100 10
570	6 779016	22	9411	21331003178	45	5	03000 9
773	61979016	7299933	20332000200	523	1 29	00000	6
770	62279012	4	6821100512000162	414	4 32		6
083	62279016999	731	20232000211	512222265242122	7		
083	62279016999	741	20032000211	512222265242100	7		
780	52979015	2	4511	21431000120	444	6 40	00000 3
780	6 879010	3	9633120832002137	91	222273442105	13	
771	62179015	5699933	20942000170	413	6 40	00000	7
901	62279016	1	9111	10831000218	543316253400000	10	
770	62279224	60	8721	21031000190	634426353500000	9	
551	6 779011	7	2943	20532003201	713316333443003	32	
551	6 779011	7	2953	20431003201	713316333443003	32	
821	61179011	1	2421	203320 170	45	00000	5
425	521790 9	6	9633	20031002171	444416232433000	5	
170	614790 9	1	823	20532003106	3	30	12000 5
220	52979016	1	7211	11632062104	422266505130010	24	
780	61879010	1	2623	20831000170	767814353500000	5	
780	620790 9	3	4833	20531000160	434	5 32	00000 4
780	62879014	1	9613	20931000200	534416282900000	17	
821	7 279015	1	2333	20232000190	5	00000	1
820	7 3790 9	3	4823	21131000170	8	00000	2
290	7 679016	1	2453	101320 210	73336634350	3	
820	62779220	1	2411	20632000170	8	00000	7
090	62979015	1	713	20631100106	414466212430100	5	
090	7 079011	1	2911	21632000218	474466404043000	10	
090	7 279011	1	2921	21531000218	474466404043000	10	
780	61879011	3	9623	20832000200	335456303100000	12	
400	62379219	1	7214	51531003178	654314363631000	7	
771	7 179222999	9111	10031000198	3133111734	7		
07	7 579011	1	412	11632000101	314414424342100	6	
775	62979012	52	7423	0631000208	623316322710100	11	

to represent the information obtained from each of the 40 cases. In Table 2.1 two data lines were necessary for each of the 15 cases.

What do the numbers in Table 2.2 stand for? The first two numbers in each line indicate the county in which the child-abuse report was filed. The 26th and 27th numbers depict the perpetrators of the abuse. The 23d and 24th numbers (columns 23 and 24) show the age of the child. In a similar fashion, the remaining numbers in each line represent the categories of the remaining 31 variables, such as year of the report, type of service offered, sex of the child, and so on. It should be pointed out here that the 40 cases in Table 2.2 are only a small sample from a population of approximately 70,000 cases reported over a four-year period.

Now that we have seen some examples of numbers that make up a data set, it is necessary to take a closer look at how the numbers were obtained and what numerical meaning can be attached to them.

MEASUREMENT AND LEVELS OF MEASUREMENT

In the process of collecting data relevant to social work, we typically measure more than one variable at a time. A *variable* is some observable entity, object, or event that can take on more than one characteristic or value. These different characteristics or values that variables take on are often referred to as the categories of variables. The biological sex of people is a variable. Male and female are categories or instances of this variable. Age of people is a variable. The various years are categories of this variable. Type of counselling—individual, family, or group—is a variable. The three types of counselling listed are categories of this variable.

As is true with other professions and disciplines, social work practice is replete with variables. In fact, it is the delineation and study of sets and classes of variables that serve to help provide identity for a particular profession or discipline. It is the study of these variables and the linkages among them that foster the advancement of knowledge. Measurement plays a key role in this enterprise.

Recording an observation is known as *measurement*. Measuring a variable involves assigning values or symbols to observed instances or categories of a variable, according to specified rules. The rules that define the assignment of an appropriate value determine the *level of measurement* for a particular variable. The different levels are defined on the basis of the ordering and distance properties inherent in the measurement rules. These rules are very important because which particular statistical technique one should use on data depends, in large part, on at which level the data is measured.

There are four types or levels of measurement to be considered. For analytical purposes these four levels will usually involve the assignment of numerals to observed variables. Some variables take on fairly prescribed values, such as the variable age and number of service hours. Other variables take on less obvious values and often require an arbitrary coding scheme to

identify various categories adequately. Variables of this type include marital status and ethnicity.

Nominal Measurement

The nominal level of measurement is sometimes called the qualitative level and is the simplest form of measurement. The word *nominal* is derived from the Latin word for name, *nomine*. Hence, nominal measurement involves the assignment of names to the various categories of variables. We are primarily interested in the assignment of numeral names to variable categories. For example, instead of naming people as married, single, or divorced, we can name these categories of the variable marital status 1, 2, or 3. Or we can label them 5, 6, or 7. It is important here that we not consider these numeral names as numbers. We can perform neither arithmetic nor logical operations on data measured at the nominal level. Nothing is implied about the relationships that might exist among different categories. We must remember that at the nominal level of measurement, 1, 2, 3, and 4 are not numbers, they are numerals. But because nominal measurement is simple, does not mean it is of little use. In many situations, nominal-level statistics are the only appropriate measures to use. We will see evidence of this in the following chapters.

Ordinal Measurement

Ordinal measurement allows us to rank objects, events, and people according to some quality. Classifying people as low, moderate, or high in self-esteem implies that there is an ordering among the three categories. Assume we assign the numerals 1, 2, and 3 to these categories. Then people who fall in category 3 are judged to possess more self-esteem than those in category 2. Similarly, those in category 2 possess a higher level of self-esteem than those in category 1. What we have done is simply ranked people into these categories. Therefore, we know that there is a “distance” or interval between categories, but we do not know what that distance is. We can neither add these categories to one another nor perform other mathematical operations on them. But having our variable-categories rank ordered does give us a higher level of measurement than simply naming or labeling the categories.

Interval and Ratio Measurement

It is at the interval level of measurement that we can begin to perform arithmetic operations on data. Interval numbers have magnitude because they are based on a common unit of measurement. This means that the distance from one number to its adjacent number is equal to the distance from any other number on the continuum to its adjacent numbers. We say that such numbers are along an *equal-interval scale*.

Consider three people with self-esteem scores of 25, 35, and 45. First of all, these people are classified into one of three self-esteem categories (nominal). Second, the scores follow in order along the dimension (ordinal). If we consider the difference between 25 and 35 units of self-esteem to be equal to the self-esteem difference between 35 and 45, then we are placing the variable self-esteem on a scale or dimension with equal intervals. We have thus measured the variable self-esteem along an interval scale. Measuring variables on an interval scale allows one to perform the operations of addition and subtraction on the numbers on the scale.

The *ratio level* is distinguished from the interval level in that ratio scales of measurement have an absolute zero. An absolute zero point for a measurement scale means that there is a point on the scale where none of the variable being measured exists. Commonly occurring ratio scales are the several physical measurements: length, weight, and time. Other variables of interest that can be considered to be ratio level are income and family size.

In social work research the distinction between the interval and ratio levels is usually not important; the same statistic often is used for both levels.

Many social work research projects include the analysis of data measured at several levels of measurement. The data in Tables 2.1 and 2.2 include several variables at each of the levels of measurement. One of our goals in research is to obtain such precision of measurement that we can say we are studying variables that lie along at least an interval scale. This then enables us to use more sophisticated statistical techniques in our analysis of data. Some social workers (and others) would argue that measurement in social work research at best has reached the ordinal level. There are, however, those who contend that variables of a psychological or social nature most likely approximate interval measurement. Hence, these variables can be treated as if they were, in fact, interval-level scales. It may be a long and arduous task to "prove" that we have measured such concepts as self-esteem and group cohesiveness on an interval scale. In most cases, the game of proof may not be worth the candle. The position taken here is that it is often necessary and appropriate to consider some ordinal-level measurement as interval-level measurement. But it is important that we and others recognize the possibility of error in analysis and interpretation that may arise from doing so.

Chapter 3

Organization and Presentation of Data

Descriptive statistics are used to untangle a seemingly unorganized set of data, such as those in Tables 2.1 and 2.2, and to provide meaning to the data. The purpose of this chapter is to examine some of the basic techniques for organizing and presenting data. Throughout this book we stress the utility of data. For data and data analyses to be useful to the social work practitioner and administrator, the data must first be meaningfully organized and presented.

FREQUENCY DISTRIBUTIONS

One of the first things we can do with a set of data to facilitate comprehension is to organize the data into a set of frequency distributions, one distribution for each variable in the data set. If we have ten variables in a data set, we can

TABLE 3.1. Ages of children from small sample of child-abuse registry

14	07	09	16	06
07	13	08	08	16
13	03	10	05	15
17	05	05	09	08
05	02	04	02	15
10	00	03	11	00
08	14	00	01	16
00	08	05	06	06