CHAPTER 7

DATA PREPARATION AND DATA PROCESSING:

A PRELUDE TO DATA ANALYSIS

Once data have been collected it is incumbent upon the researcher to prepare this information for systematic analysis, whether or not the computer will provide mathematical assistance. Data preparation and data processing are the operations that bridge the data collection and data analysis phases of the research act. These research processes entail the conversion of data into a category system and the transformation of these categories into codes that facilitate analysis.[1] In this chapter we will focus upon some of the critical aspects of coding, data processing, and measurement.

An Example. To illustrate several major ingredients of data processing and preparation consider the research instrument used by Leonard and Schmitt to study sport.[2] Figure 7.1 reproduces the questionnaire administered to several samples of subjects. The researchers' purposes were to assess the meanings of sport and the socio-demographic correlates of these meanings.

Constructing Coding Schemes. Coding is the name applied to the process of preparing raw data for quantitative or qualitative analysis by constructing meaningful categories for data classification. The function of coding is to simplify and facilitate dealing with large amounts of individual data. This is accomplished by classifying subjects' responses into a finite number of categories, so that the responses in each category share similarity in content.[3]

As you might guess, coding the samples' responses to the research instrument in Figure 7.1 varies according to the nature of the item. Coding is greatly facilitated when fixed-response alternative items are used. Consider first the question (F) on gender. Since we know that sex varies on two dimensions only--male and female--we can precode the item, that is, we can construct the categories of response in advance of the questionnaire administration. Numerals (e.g., 1=female and 2=male) can be assigned to

Figure 7.1

BIOGRAPHICAL/OPINIONNAIRE DATA USED IN SOCIOLOGY OF SPORT COURSE

This information will be used to illustrate certain principles in the sociology of sport. No individual data will be identified and only aggregate findings will be reported in class.

A. Please list the activities you think of when you hear the word "sport"?

1.	11.
2.	12.
3.	13.
4.	14.
5.	15.
6.	16.
7.	17.
8.	18.
9.	19.
10.	20.

B. List your favorite <u>spectator</u> sports:

1.	6.
2.	7.
3.	8.
4.	9.
5.	10.

C. List your favorite <u>participant</u> sports:

1.	6.
2.	7.
3.	8.
4.	9.
5.	10.

D. How much do you talk about sports? (Circle correct number)
1 = rarely 2 = occasionally 3 = frequently 4 = very frequently

E. In what situations or contexts do you discuss sports? e.g., bars, dorms, etc.

1. 4.

2. 5.

3. 6.

F. Gender (circle appropriate number): 1=female; 2=male

G. Year in School (circle appropriate number): 1=frosh; 2=sophomore; 3=junior; 4=senior; 5=grad student; 6=other

H. Age (circle appropriate number): 1=17, 18; 2=19, 20; 3=21, 22; 4=23, 24; 5=25+

I. Gradepoint (overall) average (circle): 1=1.499 or below; 2=1.5-1.999; 3=2.0-2.499 4=2.5-2.999; 5=3.0-3.499; 6=3.5 or above

J. Number of brothers and sisters (circle): 1=none; 2=one; 3=two; 4=three; 5=four; 6=five; 7=six or more

K. Major: ______________________________

L. Race (circle): 1=black; 2=white; 3=other than black or white

M. Hometown Size (circle): 1=under 1000; 2=1,000-4999; 3=5000-15000; 4=15001-50000; 5=50001 and over

N. Religion (circle): 1=Protestant; 2=Catholic; 3=Jewish; 4=other; 5=none

O. Marital Status (circle): 1=single; 2=married; 3=divorced/separated/widowed

P. Family's Social Class (circle): 1=upper; 2=upper middle; 3=middle; 4=working, 5=lower

Q. Family Income (circle): 1=$0-4999; 2=$5000-9999; 3=$10000-14999; 4=15000-19999; 5=$20000 and over

R. Father's education (circle): 1=some elementary (1-6); 2=some secondary (7-12); 3=some college; 4=college graduate; 5=post graduate work

S. Mother's education (circle): 1=some elementary (1-6); 2=some secondary (7-12) 3=some college; 4=college graduate; 5= post graduate work

T. List sports in which you participate and indicate whether intercollegiate or intramural:

each alternative response before the questionnaire is administered and when the instruments are returned the preassigned numerals (e.g., 1 or 2) rather than the written responses (e.g., male or female) can be recorded. The procedure of requiring data to be recorded according to some preconceived coding scheme is technically known as <u>deductive coding</u>.

Coding responses is more problematic when <u>open-ended</u> questions are employed. For example, the question (A), "Please list all those activities you think of when you hear the word 'sport'"?, is of the open-ended variety. With this item, the researcher is not cognizant of the range or variability of subjects' responses. Hence precoding cannot be meaningfully employed. In this case, coding is best completed <u>after</u> the data are collected. Figure 7.2 contains a sampling of the different responses to these items. Note that many responses are "conventional," that is, refer to popular and well known sports. However, notice that many responses are "unconventional," (i.e., rather unique and idiosyncratic). It is easy enough to count the number of mentions of football, baseball, basketball, etc. But what about the "atypical" responses? Here the researcher must use seasoned judgment. Leonard and Schmitt, after reviewing the entire list of responses, arrived at several broad categories under which the "novel" responses could be subsumed. For example, nine categories were devised to include the more "elaborate" responses. These were: 1) doing mentions (e.g., moving, crying, panting), 2) group mentions (e.g., games, teams), 3) individual or group expenditures (e.g., effort, intensity), 4) performance mentions (e.g., competition, aggression), 6) organizational mentions (e.g., NFL, NCAA), 7) spectator mentions (e.g., fans), 8) sport related objects (e.g., referee, fouls, superstars, boundary lines) and 9) effect mentions (e.g., brotherhood, injuries). Notice that the researchers' intent was to categorize similar specific mentions under a broader meaningful heading. In exploratory

Figure 7.2

SOME RESPONSES MADE BY SUBJECTS TO THE QUESTION:

"Please list the activities you think of when you hear the word 'sport'?"

basketball	football	hockey
moving	crying	panting
softball	running	skiing
baseball	games	teams
diving	swimming	effort
intensity	tennis	dance
bowling	competition	aggression
NFL	NCAA	billiards
bicycling	parachuting	fans
referee	official	fouls
superstars	rugby	boundary lines
archery	auto racing	brotherhood
injuries	checkers	basketball
hockey	football	baseball
swimming	softball	archery
lacrosse	surfing	racketball
table tennis	sailing	fishing
hunting	track	weight lifting
snow skiing	gymnastics	wrestling
volleyball		

studies and studies containing "open-ended" responses the coding scheme is typically inductive. An inductive coding scheme does not use pre-designated response categories; instead, the responses are recorded after the subjects make them and categorization follows suit.

Each of these coding schemes--deductive and inductive--has merits and liabilities. The deductive system is the simpler of the two but has been criticized on the grounds that it fails to fully deal with the complexity of human behavior as well as behavioral continuity.[4] Additionally, the a priori coding scheme is somewhat restrictive and may bar new insights on the part of the researcher.

The inductive system is characterized but its flexibility and "insight-inducing" potential. It permits a variety of coding schemes to be applied to the same data.[5] Its shortcoming revolves around the possibility that the researcher may get lost in a maze of detail and fail to see the forest for the trees.

Criteria for Constructing Useful Coding Schemes. Regardless of which type of coding system is employed, there are certain requirements for constructing a good coding scheme.[6] First, the coding scheme should be closely aligned with the theory guiding the research endeavor and the problem under scrutiny. Second, the coding scheme must be exhaustive. The response categories must provide an unambiguous category system to which all responses can be allocated. Third, the response categories must not overlap, that is, they must be mutually exclusive. Fourth, when a researcher must decide between a large number of fairly detailed response categories and a relatively small number of gross categories, the decision should lean in the direction of the former. The reason for this is that detailed categories can always be subsequently combined whereas gross categories generally preclude this logistic manipulation.

The Computer.

The major characteristics of the modern day computer are its tremendous speed of operation and its ability to perform thousands of repetitive operations with a high degree of accuracy.[13] How fast do computers operate? They approximate the speed of light and you will recall that light travels about 186,000 miles per second. See Figure 7.5. Researchers are a major beneficiary of the computer. Statistical/mathematical operations that formerly took days, weeks, or even months can now be performed in seconds or minutes. This means that research projects formerly neglected because of the excessive amounts of time required for data analysis can be easily undertaken. Hence, whole new research vistas have been opened up.

On the negative side, the computer is a "stupid monster." Like a robot, it will perform the programmer's instructions brilliantly. But if the programmer provides incorrect or nonsensical instructions the computer will faithfully and obediently carry out these "errors." Computer scientists have an acronym for this: "GIGO" means "Garbage In, Garbage Out." This characteristic of the computer is called ductility, that is, stupidity.[14]

Despite the revolutionizing features of computer processing several caveats are in order. All too often computer consumers utilize the technology without understanding the statistical/mathematical principles behind the programs. This then leads to an uncritical acceptance of the computer-generated results. The user must be wary of

Figure 7.5

In the computer, the basic operations can be done within the order of a

NANOSECOND

One thousandth of a millionth of a second.

Within the half second it takes this spilled coffee to reach the floor, a fairly large computer could—

(given the information in magnetic form)

Debit 2000 checks to 300 different bank accounts,

and ***examine the electro-cardiograms of 100 patients and alert a physician to possible trouble,***

and ***score 150,000 answers on 3000 examinations and evaluate the effectiveness of the questions,***

and ***figure the payroll for a company with a thousand employees.***

and a few other chores.

the undiscriminating acceptance of the output. The computer can and has been a boon to researchers but users must remind themselves that it is a stupid electronic gadget that faithfully and obediently carries out the instructions however asinine the instructions might be.

Canned Programs and the Computer. Only a small proportion of computer users actually know the intricacies of how the machine works. Fortunately, this lack of knowledge does not, generally, pose as a major stumbling block. The field of computer science is a technical specialty requiring concentrated courses of study and a great deal of "on-the-job" experience. People with the ability to tell the computer what to do are called programmers. Researchers do not have to become computer scientists to use and exploit the abilities of the computer.

Most computer centers maintain a library of computer programs and the user need only know of their availability and be able to follow instructions (described in manuals) for executing the programs. Among the most popular canned programs--'canned' in the sense that the user need only follow a few relatively simple instructions to use--are SPSS, BMD, SAS, Minitab, and Psystat.

Statistical Package for the Social Science (SPSS). According to Nie, Hull, Jenkins, Steinbrenner, and Bent, SPSS

> . . .is an integrated system of computer programs designed for the analysis of social science data. The system provides a unified and comprehensive package that enables the user to perform many different types of data analysis in a simple and convenient manner. SPSS allows a great deal of flexibility in the format of data. It provides the user with a comprehensive set of procedures for data transformation and file manipulation and it offers the researcher a large number of statistical routines commonly used in the social sciences.[15]

Many different kinds of statistical analyses can be routinely handled with SPSS. For univariate (single variable) distributions the program can construct frequency distributions, compute various measures of central tendency (e.g., mean, median, and mode) and dispersion(e.g., range, variance, and standard deviation). For bivariate (two variable) distributions various measures of association can be calculated (e.g., Pearson's r, Spearman's r_s, Kendall's tau, gamma, and Pearson's C). For multivariate (more than two variables) distributions the programs enable one to produce partial correlations, multiple correlations, multiple regressions, path analyses, etc. In addition, several options exist for special outputs (e. g., generating random samples, Guttman scaling, etc). In short, SPSS is capable of providing practically any statistical output normally required in the course of data analysis.

Other canned programs like BMD (originally designed for medical science research) and SAS essentially perform the same functions. Rather than describe them at this juncture, it is important to realize that most computer centers have these and other "packaged" programs accessible to facilitate data processing and analysis.

While the mechanized procedures described so far may seem simple and straight-forward, they do require certain statistical understanding in order to legitimately use. In other words, the computer will produce virtually any statistic you request even when that statistic may be nonsensical, meaningless, or inappropriate for the kind of data or information with which you are working. The appropriate use of computer technology is dependent upon your understanding the meaning of measurement and the kinds of variables with which researchers typically deal. We will now turn to these very significant and fundamental considerations.

MEASUREMENT

We saw in Chapter 2 that the processes involved in moving from abstract theory to empirical hypotheses are indirect. Similarly, the selection of an appropriate mathematical or statistical model to be used in research involves a number of decisions. This process is so important that two social scientists have written: "The central problem of the application of statistics is in the relation of model to practical circumstance."[16]

A germane question at this juncture is this: "What do we mean by 'measurement'?" In part, the conceptual morass surrounding the idea of measurement stems from our implicit tendency to think in terms of quantitative scales. Does the number "2" always imply a quantity larger than "1"? The answer is no if we are dealing with telephone numbers, street addresses, or football jersey numbers. The use of numbers in these cases is merely to unambiguously identify or categorize the different phenomena.

Consider this illustration. If we rank three individuals, assigning "1" to the tallest, "2" to the middle person, and "3" to the shortest, does the height of the first person plus the height of the second person equal the height of the third person? Of course, the answer is no! Statisticians make a distinction between numerals and numbers to avoid this confusion. Numerals refer to the use of symbols to classify or order events--no arithmetical operations are implied. Numbers, in contrast, may involve arithmetical manipulations. This distinction is more than academic since many of the numbers we confront daily (e. g., zip codes, uniform numbers, address numbers), do not possess the arithmetical properties we ordinarily ascribe to them. Since confusion surrounds the use of numerals and since they can be used to name (nominal numerals), to represent position in a series (ordinal numerals), and to represent quantity (interval-ratio numerals), a precise definition of measurement is necessary.[17]

Probably the broadest and most widely-used definition of measurement comes from S. S. Stevens.[18] He notes that "measurement is the assignment of numerals to objects and events according to rules." This definition captures the basic nature of measurement.[19] Let's take an example from the annual Combine in Indianapolis. As we know, the judges rate the contestants on numerous dimensions. Does this constitute measurement?

To answer this query let us refer back to Stevens' definition. The judges assign values or numerals (perhaps 1 to 10) to objects (in this case females) according to the instructions (in this case whatever rules they are expected to abide by). Since the operations involved conform to Stevens' conceptualization we have an illustration of measurement.

Levels of Measurement

Statistics texts traditionally discuss four levels of measurement (nominal, ordinal, interval, and ratio) to emphasize that data contain different information depending upon how scores/values are assigned. These different strata provide the basic conceptual foundation for choosing the most appropriate statistical techniques for discovering the meaning of collected empirical observations. It should be borne in mind that other scholars consider "in-between" measurement levels such as a partially ordered scale or an ordered metric scale. However, in basic statistics these latter concerns are not ordinarily important.[20]

The notion of "levels of measurement" refers to the relationship(s) between categories of a variable.[21] For a nominal variable like gender, there are two categories (male and female) that are simply different. For an ordinal variable such as social class, there are several possible categories--upper, middle, lower; upper, middle, working, lower, et cetera--but in all instances the categories can be located on a continuum or hierarchy from high to low or more to less on such dimensions as power, privilege, and prestige. For interval and ratio variables like temperature and years of formal schooling, respectively, we can talk in terms of clearly defined units of measurement. With standardized measurement units we can deal with magnitudes of difference (such as a high school graduate has twice as much formal schooling as a grammar school graduate).

There are several distinct levels of measurement, each with its own rules and properties. These strata are cumulative, meaning that it is legitimate to drop back one or more levels but not move up in analyzing our data. These levels of measurement refer to the manner in which we've operationalized and assigned numbers to the concepts in question. The four conventional levels will now be discussed.

Measurement and Quantification in Social Research. When researchers quantify or measure a phenomenon they assign to it a numerical value through a process called mapping.[25] However, the numbers assigned to individuals, objects and events are not all of the same type. Furthermore, depending upon the process by which the numbers were assigned, investigators are more or less restricted in the kinds of statistical analysis they can legitimately make. In order to highlight the different kinds of data with which researchers deal, take a moment and answer the following questions. Place a check mark or fill in your response at the appropriate point.

1. Gender: male___ female ___
2. Religion: Protestant ___ Catholic ___ Jew ___ Other ___ None___
3. How happy are you? ecstatic ___ very ___ so-so ___ unhappy ___ depressed ___
4. What do you think the maximum temperature will be today?______
5. How much do you weigh?_____

Levels of Measurement in Social Research. With your answers to these items one of the basic building blocks for statistical analysis--level of measurement--will be easier to appreciate and comprehend. To facilitate your grasping the nature of measurement, imagine that we must prepare these data for computer processing.

Nominal Level Measurement. Suppose you checked female to item one. The researcher must attach a unique code (usually numerals or numbers for data processing) to each response category. There are two categories to gender, male and female. Because there are only two subdivisions the variable is called a (natural) dichotomy. The key to understanding the level of measurement of a variable resides in the relationship that exists between/among the categories of the variable. Take gender with its two categories of male and female. The two categories are simply different, one isn't better, bigger, higher, etc., than the other. When the relationship between two categories is simply one of difference or nonequivalence, the level of measurement is nominal. Returning momentarily to the coding procedure in preparation for the computer run, the number 1 may be assigned to all males (and therefore a 1 is punched in the appropriate column) and the number 2 to all females (and a 2 is punched in the appropriate column). What is most critical is this: whether the variable is hand calculated or computer processed, not all statistical techniques would be meaningful. The measurement process for gender results in a variable that is nominal in nature and also a dichotomy (two categories).

In question number two you were asked to check one of the five religious preference categories. In preparing this item we would again have to use numerals to code the response alternatives. Anytime more than two choices of selection exist, the variable is said to be a polychotomy (or polytomy), literally meaning it contains many ("poly") cuts ("chotomy") or subdivisions. To determine the level of measurement of the variable we look at the relationship between the categories of the variable. The categories are merely different from one another and therefore are nominal in character. It would be permissible to arbitrarily assign Protestants a "1" code, Catholics a "2" code, Jews a "3" code, others a "4" code, and none a "5" code. The important thrust to the measurement level of a variable surrounds the fact that some analytical techniques are applicable whereas others may not be. In short, as we've measured the variable of religious preference we have a nominal variable that is also a polychotomy (or manifold classification).

Ordinal Level Measurement. The third question asked, How happy are you? Notice the choices. The response alternatives are different but they also have an added property, namely they stand in some kind of order with respect to each other. In other words, we can imagine the categories of the variable being arranged on a continuum or hierarchy from most (e. g., ecstatic) to least (e. g., depressed). The relationship among the categories displays some kind of rank ordering in addition to being different. Whenever the sub-classifications of a variable have an underlying order to them, the criteria for ordinal level measurement have been achieved. There are five response categories, making it a polychotomy, to which we might arbitrarily assign a 1 to the category representing the "most" of the characteristic (happiness) being measured, a 2 to the category representing the next most, . . . and 5 to the category displaying the least of the characteristic being measured.[26] Suppose we used the following codes to identify the response alternatives:

1 = ecstatic

2 = very happy

3 = so-so

4 = unhappy

5 = depressed

While it is true that anyone with a smaller number is more subjectively happy than anyone with a larger number (e.g., a person with a 2 is happier than a person with a 4), it is not possible to say that a person with a 2 is "twice" as happy as a person with a 4. This is because the interval distance between each and every category is not necessarily the same or equal. Without an equal interval scale researchers must be content with having ordinal level measurement. The third variable conforms to the ordinal level of measurement and is also a polychotomy.

Interval Level Measurement. The fourth query asked you to predict today's maximum temperature. No response categories are provided, making this item an open-ended one unlike the fixed alternative items (1 through 3, inclusive). Suppose you believe the highest temperature will be 70 degrees and a friend of yours indicates that the highest temperature will be 35. It is tempting, albeit incorrect, to assume that you believe it will be twice as warm today as your friend. This is entirely false because the zero point on conventional temperature scales like Fahrenheit and Celsius is arbitrary. Zero on these scales does not represent the absence of heat; instead it is a convenient starting point from which to assess temperature and temperature changes. The interval between points on such a scale, say 70 - 75 degrees and 30 - 35 degrees, is the same making it what is called an equal (appearing) interval scale, but it is not permissible to talk of ratios between scale points because there is an absence of an absolute zero. Physicists know that absolute zero is a - 273 degrees measured on the Kelvin thermoscale. The variable in question conforms to the interval measurement level and is also a polychotomy.

Ratio Level Measurement. The final query produces a variable different from the previous four. How much do you weigh? Suppose you weigh 160 pounds and your kid brother weighs 80. It is appropriate and correct to say that you weigh twice as much as your brother. Why? In addition to having a standardized measuring unit (e.g., pound, ounce, or gram) which we also had in measuring temperature, an absolute zero representing the complete absence of the phenomena being measured exists. It is possible to have an object that weighs virtually nothing or zero. Because the zero point is theoretically meaningful, the use of ratios such as one person weighing half of what another weighs or three times what another weighs is legitimate. The variable "weight" conforms to ratio level measurement and is also a polychotomy.

Figure 7.7highlights the properties of the four levels of measurement.

Absolute Scale Measurement. Data at the conventional four levels can be transformed into another level of measurement called an absolute scale.[27] Heretofore we have either labeled the categories of a variable (e.g., nominal level measurement) or assigned a value to the categories of the variable (e.g., ordinal and interval/ratio level measurement). If, instead of proceeding in this fashion, we count the frequency with which a particular category occurs, we have performed absolute scale measurement. Returning to item one, suppose we count the number of males and females in the class and, in item 2, the number of cases in each of the religious preference categories. We would be remiss if we computed the arithmetic mean for the variable of gender or religion but, if we wished, we could compute the arithmetic average of the

Figure 7.7

PROPERTIES OF FOUR LEVELS OF MEASUREMENT

Properties	Levels of Measurement: Nominal	Ordinal	Interval	Ratio
mutually exclusive categories	X	X	X	X
order among categories		X	X	X
equal intervals between categories			X	X
absolute zero point				X

Figure 7.8

CROSS - CLASSIFICATION OF LEVEL OF MEASUREMENT AND SCALE CONTINUITY

Scale Continuity	Level of Measurement: Nominal	Ordinal	Interval/Ratio
Discrete	gender, religion		family size
Continuous	logically impossible		weight

number of students for each variable. If in a class of 25 the number of students in the ecstatic, very, so-so, unhappy, and depressed categories were 3, 8, 11, 3, and 1, respectively, it would not be reasonable to compute the average happiness. However, it would be legitimate to compute the average number in each of the categories.

The distinction advanced here is very important since data at any level of measurement can be converted into an absolute scale permitting the utilization of virtually any statistical techniques. Just the same, it is necessary for you to understand the type of question that can be asked and answered from the data. The nature of the question changes (i.e., it is not possible to determine the average happiness score from ordinal level data, but it is possible to determine the average number of students in each category).

Statistical Importance of the Theory of Measurement. Some methodologists whom we might christen "purist statisticians" would argue that statistical techniques appropriate for higher levels of measurement never be used for data quantified at lower levels.[28] Statistical techniques, to them, are more-or-less measurement-level specific. For example, the arithmetic mean and standard deviation are "geared" to deal with interval/ratio level variables. The statistical purist would believe it anathema if an interval level statistic like the mean were used with nominal level data (assuming the data were not transformed into dummy variables).[29] Moreover, some statistical techniques like multiple regression and path analysis assume the variables are measured at a particular level and therefore make certain qualifying assumptions regarding the measurement level.

Other methodologists whom we might call "statistical pragmatists" acknowledge the importance of measurement levels and their implications, but advocate the utilization of virtually any technique if that approach sheds additional light on the nature of the phenomenon being investigated.[30] The pragmatists, for the most part, caution such use but, nevertheless, are a bit more lenient than their conservative counterparts. Another salient and related point concerns the robustness of many techniques. A robust statistical technique is one that allows certain violations of assumptions with little or no damage to the resulting computed value or its interpretation.

The concept of level of measurement also provides the background for a couple of other important statistical considerations. Sometimes a distinction is made between quantitative and qualitative (or attribute) variables. When these rubrics are used, the reference for quantitative variables are those phenomena measured at either the interval or ratio level. Qualitative variables are ordinarily thought to subscribe to nominal and ordinal measurement level assumptions. A second related distinction is between parametric and non-parametric statistics. Regarding the level of measurement the former generally refer to interval and ratio measured variables and the latter to those at the nominal and ordinal levels. Finally, variables may be classified as discrete and continuous. A discrete variable is one that comes in integer or whole number form. For example, gender or religion, when quantified, can only be assigned a whole number. A continuous variable is one that can take on practically any intermediate value along a continuum. For example, temperature and weight could be fractionated almost indefinitely although there are practical and measurement instrument restrictions. The level of measurement and scale continuity (i.e., discrete vs. continuous variables), dimensions can be visualized in Figure 7.8. Notice that it is logically impossible to have a continuous nominal level variable although a discrete nominal level variable (e.g., gender and religion) is perfectly permissible. Also it is legitimate to have a discrete interval/ratio level variable (e.g., family size) as well as a continuous (e.g., weight) interval/ratio level variable.[31]

Summary

Once data have been collected the time for data processing has arrived. This phase prepares the data for analysis. Since the computer vis-a-vis hand processing is typically used, we discussed the manner in which data are prepared for computer processing. This phase necessitates understanding the function of coding, the types of coding schemes employed (e.g., inductive and deductive) along with the manner in which the information is transformed.

scanned. Constructing a codebook--identifying the nature and location of one's variables on the input medium (tapes, or discs)--is an important step at this juncture.

To fully comprehend the role of the computer it is necessary to discuss computer hardware, the physical components that make up the computing system

and computer software, programs that help the programmer create and execute other programs, such as SPSS, SAS, and BMD.

The appropriate use of computer technology is dependent upon the researcher's understanding of the meaning of measurement and the kinds of variables with which researchers typically work. The general meaning of measurement was discussed along with the defining characteristics of the traditional four levels of measurement: 1) nominal, 2) ordinal, 3) interval, and 4) ratio. In addition, the nature of absolute scale measurement was considered along with the statistical importance of the theory of measurement. Finally, distinctions were made between quantitative and qualitative variables and parametric and nonparametric statistics.

Important Concepts Discussed in This Chapter

Data Preparation

Data Processing

Coding

Deductive Coding

Inductive Coding

Canned Programs, e.g., SPSS, BMD, SAS

Nominal, Ordinal, Interval, Ratio Scales

Absolute Scale

Polychotomy

Quantitative Variables

Nonparametric Statistics

Codebook

Computer

Measurement

Ordinal

Dichotomy

Parametric Statistics

Qualitative Variables