Flaws and Fallacies in Statistical Thinking

Stephen K. Campbell

FLAWS AND FALLACIES IN STATISTICAL THINKING

STEPHEN K. CAMPBELL
UNIVERSITY OF DENVER

DOVER PUBLICATIONS, INC.
Mineola, New York

Bibliographical Note

This Dover edition, first published in 2004, is an unabridged republication of the work originally published by Prentice-Hall, Inc., Englewood Cliffs, New Jersey, in 1974.

Library of Congress Cataloging-in-Publication Data

Campbell, Stephen K. (Stephen Kent)
Flaws and fallacies in statistical thinking / Stephen K. Campbell.
p. cm.
Originally published: Englewood Cliffs, N.J. : Prentice-Hall, 1974.
Includes bibliographical references and index.
ISBN 0-486-43598-9 (pbk.)
1. Statistics. I. Title.

QA276.C272 2004
519.5—dc22

2003065966

Manufactured in the United States of America
Dover Publications, Inc., 31 East 2nd Street, Mineola, N.Y. 11501

Contents

Preface

When a person loves something he hates to see it abused. Such is my feeling toward statistics. Hardly a day goes by that I fail to see the subject I love put to use in faulty and misleading ways in newspapers, magazines, books, speeches, and, especially, advertising.

Curiously, little of the statistical literature is aimed at helping the nonstatistician recognize such abuses. Statisticians, understandably, tend to write for other statisticians. Still, one of life's rare truisms is that most people in the world are not statisticians and don't aspire to become statisticians. Nevertheless, it is an ususual person who isn't frequently confronted with the need to evaluate, if only very informally, statistical information when making decisions about a variety of subjects ranging from which political candidate to vote for to which brand of detergent to buy. Professional statisticians don't need help along such lines; keen critical judgment about statistical information is, for most, a natural by-product of their professional training. But who is to help the statistical layman if it is not the statistical professional?

This book was written with two purposes in mind—the first, less important, purpose being that of getting something off my chest. For many years I have been distressed by the frequency with which (1) relatively simple statistical tools such as percents, graphs, and averages are misused and (2) faulty conclusions are drawn from, perhaps, flawless data in our news media simply because the purveyors of the information don't know any better. Moreover, I have been annoyed—indeed, made damned mad—by the frequency with which bogus statistical evidence is used intentionally by some unconscionable people to sell their products or pet ideas to others. Writing this book has been good therapy for me.

The second, more important, purpose of this book is loosely related to the first. I have long felt that the university student who is likely to take only one or two courses in statistics and the ordinary citizen who maybe lacks any classroom exposure whatever to the subject could benefit from a nontechnical book written with a view to helping him increase his ability to judge the quality of statistical evidence and, in turn, to make better-informed decisions about many facets of everyday life. This book, therefore, has been written both as a supplemental reading text for the student taking his first course in statistics and as a self-help guide for the nonstudent who feels the need to evaluate statistical evidence more judiciously than he is presently capable.

The sequence of topics covered is in rough accord with that of many beginning textbooks in statistics. The terminology used and the manner in which the subjects are treated are based on the assumption that the reader has had little or no prior exposure to the subject of statistics and has studied precious little formal mathematics.

Many people have helped me with the preparation of this book—more than I can possibly thank individually. I must, however, single out three statistics professors for special thanks. Professors Richard B. Ellis of Northern Essex Community College and Richard E. Lund of Montana State University both reviewed early drafts of the first five chapters and offered many incisive criticisms and imaginative suggestions for improving upon my proposed project. My colleague and friend, Professor Paul R. Merry of the University of Denver, went over the final manuscript with a thoroughness that is so characteristic of everything he does and offered numerous constructive suggestions.

I would also like to thank my statistics students at the University of Denver who gathered hundreds of examples of statistical fallacies from which many of those appearing in the following pages were selected. Also deserving of thanks are several authors, editors, and advertisers who granted me permission to quote from copyrighted material even though they must have suspected that I intended to be more critical than complimentary.

Last, but certainly not least, I wish to thank my wife Judy who not only relinquished without complaint many hours of time with me which were rightfully hers but who also did more than her fair share to create an atmosphere within which the work could proceed with a bare minimum of discord or distraction.

Needless to say, any blame for errors, omissions, or bad manners should be directed at me.

STEPHEN K. CAMPBELL

1

Dangers of Statistical Ignorance

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

—H. G. WELLS

This book is unusual. Textbooks show you facts and the right methods. This book shows you fallacies and the wrong methods. It will serve as a companion volume to any textbook on statistics. It will also serve as a self-help guide to distinguish between valid and faulty statistical reasoning.

Furthermore, it deals with a very important subject because statistics influence our daily lives in a great many ways. By enlisting the aid of statistics, we measure economic activity; record social progress; elect Presidents and keep abreast of their current popularity (or, more often, unpopularity); measure intelligence, interests, and aptitudes; compare his sexual habits with various norms; determine which television shows will survive and which will not; compare the profit potential of several alternative business strategies; decide whether to invest in bonds or stocks and, if the latter, whether now is a good time to get into the market; keep track of batting averages; assess the likelihood of rain tomorrow; and, in general, keep informed about what is going on in the world with the aid of statistical data gathered, presented, and interpreted by others. Even if you and I have nothing to do with the actual calculations implied by the items in this list (or the many other items that might easily have been included), as socially conscious citizens we should be able to interpret the results of such calculations with some sophistication, for these are the figures that serve as a basis for so many vital newspaper and magazine articles, books, and speeches.

Although something of an exaggeration, the quote from H. G. Wells that introduced this chapter is basically sound. I would amend it only in one important respect and have it read "Statistical [straight] thinking will one day be as necessary for efficient citizenship as the ability to read and write." I prefer "statistical straight thinking" to "statistical thinking" because it seems unlikely that fuzzy or erroneous thinking could contribute much to efficient citizenship—or efficient anything else, for that matter. Unfortunately, there is enough fuzzy and erroneous statistical thinking around these days to justify my focusing on it in this book.

This book deals with erroneous and sometimes deliberately misleading statistical arguments. It deals with fallacious statistical thinking—how to avoid doing it yourself and how to recognize when others do it. This point requires elaboration. First, however, let us touch on some essential background topics, not the least important of which is what is meant by *statistics*.

The Two Meanings of "Statistics"

What is or are statistics? The word has two widely used meanings. The most generally familiar—and for many people the least interesting—can probably be introduced most painlessly by the following excerpt from O. Henry's *Handbook of Hymen*:

> "Let us sit on this log at the roadside," says I, "and forget the inhumanity and ribaldry of the poets. It is in the columns of ascertained facts and legalized measures that beauty is to be found. In this very log we sit upon, Mrs. Sampson," says I, "is statistics more wonderful than any poem. The rings show it is sixty years old. At the depth of two thousand feet it would become coal in three thousand years. The deepest coal mine in the world is at Killingworth near New Castle. A box four feet long, three feet wide, and two feet eight inches deep will hold one ton of coal. If an artery is cut compress it above the wound. A man's leg contains thirty bones. The Tower of London was burned in 1841."
>
> "Go on, Mr. Pratt," says Mrs. Sampson, "Them ideas is so original and soothing. I think statistics are just as lovely as they can be."

Although not all of Mr. Pratt's original and soothing ideas are really statistics, enough of them are to convey the idea that a statistic is a fact. More precisely, it is a fact expressed as a number and can be a measurement, a count, or a rank. A statistic in this first sense can even be a summary measure such as a total, an average, or a percentage of several such measurements, counts, or ranks.

In addition to referring to numerical facts, the term "statistics" also applies to the broad discipline of statistical manipulation in much the same way that "accounting" applies to the entering and balancing of accounts. "Statistics" in this broader sense is a set of methods for obtaining, organizing, summarizing, presenting, and analyzing numerical facts. Usually these numerical facts represent partial rather than complete knowledge about a situation, as is the case when a sample is used in lieu of a complete census. Generally speaking, numerical facts are subjected to formal statistical analysis in order to help someone make wise decisions in the face of uncertainty or to help researchers arrive at scientifically-sound generalizations or principles.

The word "statistics" will be used in both senses throughout this book. The context within which the term is used should make the intended meaning clear.

The Statistical Fallacy

No one knows just when the first statistical lie was foisted upon a trusting listener. For that matter, no one knows for certain when or where statistics first appeared. We do know that the earliest written records contain numbers, a fact suggesting that the ability to count goes way back. The Bible tells us that statistics in the purely descriptive sense were used to provide information about taxes, wars, agriculture, and even athletic events. Nevertheless, there probably was a time when counting, and therefore statistics, was unknown; a time when a shepherd, for example, did not describe his flock as consisting of twenty, fifty, or one hundred sheep but instead kept track of his woolly charges by assigning each a name. If two sheep turned up missing, the shepherd searched not for two anonymous animals but for, say, Peter and Paul.

Although the first uses of statistics are lost in antiquity, I would wager that misuses of statistics—intentional and unintentional—first appeared at about the same time as valid statistics. We all know people whose honesty we have good reason to doubt as well as people who are chronically careless or conspicuously stupid. Certainly there must have been counterparts of such people from the very beginning just as there must have been counterparts of you and me—honest, meticulous, and highly intelligent. Not much imagination is required to envision our shepherd, recounting the harrowing challenges he faced while retrieving his two wayward sheep, Peter and Paul, and, rather than spoiling a good story by undertelling it, claiming that Mary and Ruth had also gone astray (had gotten lost, that is). The advent of counting and statistics certainly didn't create the all-too-human tendencies to lie, exaggerate, or make honest mistakes, but it did introduce a whole new, very colorful means of giving vent to such tendencies.

Today, statistical fallacies abound in our newspapers, magazines, advertisements, and conversations. I am not suggesting that all statistical evidence is faulty. Indeed, the proliferation of statistical fallacies in recent decades has been, to a considerable extent, the natural result of burgeoning statistical data and formal techniques for analyzing such data. But even so, the mere existence of statistical fallacies imposes a responsibility upon the citizen who would call himself well informed to learn to distinguish between erroneous and valid statistics or statistical arguments.

How dangerous are statistical fallacies? No general answer is possible. Some statistical fallacies are undoubtedly perfectly harmless even when widely believed. But some are much more potentially dangerous than you might suppose. Let us consider a few examples.

If an award is ever granted the fabricator of the world's phoniest but possibly least harmful statistic of a descriptive nature, I suggest that the honor be bestowed posthumously upon a German named Weirus. Weirus, who served as physician to the Duke of Cleaves during the latter part of the sixteenth century, a time when most of Europe was gripped by the fear of demons and witches, had some definite opinions about the number of demons in existence. Whereas most of his contemporaries lazily assumed that demons were too plentiful for their numbers to be determined, Weirus, using methods beyond anyone else's comprehension, calculated that exactly 7,405,926 demons inhabited the earth; these, he claimed, were divided into seventy-two battalions, each under a prince or captain.[1]

How serious Weirus was when he revealed his remarkable findings is anybody's guess. According to Sir Walter Scott, "Weirus was one of the first

[1] Joseph Jastrow, *Error and Eccentricity in Human Belief* (New York: Dover Publications, Inc., 1962), p. 86. [This title currently out of print.]

who attacked the vulgar belief and boldly assailed, both by serious argument and by ridicule, the vulgar credulity on the subject of wizards and witches."[2] Quite possibly the good doctor had his tongue planted firmly in his cheek when he revealed his spurious figures. If so, many historians have not been in on the joke for they have written of the incident as if they believed Weirus to be dead serious. Moreover, you can bet that many people hearing of the doctor's calculations accepted the bogus results without question, thinking that 7,405,926 sounded "about right." And why not? To them demons were a reality and Weirus was a learned man. In the final analysis, however, it is hard to imagine how Weirus' figures could have done anyone harm, except perhaps Weirus himself if he really was poking fun at the prevailing beliefs of his day.

Here is a more modern-day example. A 1968 advertisement for Volvo automobiles is a treasure trove of statistical fallacies, but that unfortunate fact didn't keep it from appearing in many of the country's top magazines. The advertisement states that, according to statistics, the average American drives 50 years in his lifetime and the average car is traded in on a new one every three years and three months. The logical conclusion, we are informed, is that if one drives an average number of years in average cars, he will own

[2] *Letters On Demonology and Witchcraft* (London: W. Tegg, n.d.), p. 186.

15.1 cars in his lifetime. But not so if he owns a Volvo. The lucky Volvo owner, the ad asserts, can get by with only 4.5 Volvos because in Sweden Volvos last an average of 11 years.

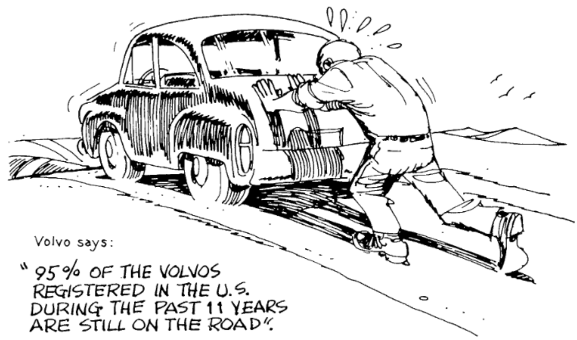
The advertisement goes on to say, "We don't *guarantee* they'll last that long here where being a car is relatively easy. But we do know that over 95% of all Volvos registered here in the last 11 years are still on the road."

The Volvo may indeed be a durable, well-constructed automobile—I have no convictions one way or the other—but, despite the collection of presumably authentic figures, this advertisement conveys no genuine information about the Volvo's durability relative to that of other cars.

Aside from the question of whether the number of years a car is driven is the most meaningful possible measure of durability (why not number of miles driven instead?), two clear-cut fallacies stand out. First, an improper comparison is made. The almost adjacent statements "The average car is traded in on a new one every three years and three months" and "Volvos last 11 years . . ." is an apples and oranges comparison if there ever was one. Obviously, the number of years an automobile can be driven before it becomes incapable of providing transportation and the number of years an automobile is driven before the owner tires of it and voluntarily trades the still usable vehicle in on a newer model are two very different matters. No meaningful comparison whatever can be made between the two figures.

Second, the last part of this ad—the part that states ". . . over 95% of all Volvos registered here in the last 11 years are still on the road"—is no more informative than the first part and is just as potentially misleading. Suppose, for example, that all Volvo sales made during the 11 year period referred to (apparently 1957 through 1967 inclusive) had been made during the most recent year and none whatever during the preceding ten years. In such a case, the figure of 95 percent would be indicative of poor rather than outstanding durability because it would mean that five percent of the Volvos sold had to be scrapped during their first year of use. Actual sales admittedly were not bunched so dramatically, but, according to a 1968 issue of *Ward's Automotive Reports*, approximately 45 to 50 percent of the Volvo sales in this country were made during the most recent four years of the 11-year period in question. The 95 percent figure, therefore, says little or nothing about the Volvo's durability.

What can we conclude about this advertisement other than it is much more misleading than informative? Is it dangerous? That is quite possibly a different matter altogether. If the Volvo is in fact an unusually durable car, then the advertisement presumably did no real harm. (Of course, one can hardly help wondering why, if the car really is all that durable, the company has to resort to half-truths to sell it.) But if the Volvo really is no more durable than other makes, and certainly if it is less durable, then many automobile buyers might have been led astray by the ad, and, according to any criteria I can think of, the statistical fallacies that helped to sell the car would have to be viewed as dangerous.



Now let us consider two heavyweights. These are examples of statistical fallacies whose propensities for causing trouble are incalculably great. The first example comes from a *Playboy* Magazine interview with the late George Lincoln Rockwell, commander of the American Nazi Party:[3]

> Rockwell: A psychologist named G.O. Ferguson made a definitive study of the connection between the amount of white blood and intelligence in niggers. He tested all the nigger school children in Virginia and proved that the pure-black niggers did only about 70 percent as well as the white children. Niggers with one white grandparent did about 75 percent as well as the white children. Niggers with two white grandparents did still better and niggers with *three* white grandparents did almost as well as the white kids. Since all these nigger children shared exactly the same environment as niggers, it's impossible to claim that environment produced these tremendous changes in performance.
>
> Playboy: In his book, *A Profile of the Negro American,* the world-famed sociologist, T. F. Pettigrew states flatly that the degree of white ancestry does not relate in any way to Negro I. Q. scores. According to Pettigrew, the brightest Negro yet reported—with a tested I. Q. of 200—had no traceable Caucasian heritage whatever.
>
> Rockwell: The fact that you can show me one very black individual who is superior to me doesn't convince me that the average nigger is superior. The startling fact I see is that the lighter they are, the smarter they are, and the blacker they are, the dumber they are.

Rockwell's faith in the Ferguson study might be rather touching if the moral stakes weren't so high and if the study enjoyed any scientific repute,

[3] "Playboy Interview: George Lincoln Rockwell," *Playboy,* April, 1966, pp. 71ff.

a consideration that Rockwell didn't bother to worry about. But as the magazine's editors point out:

> Ferguson's study, conducted in 1916, we later learned, has since been discredited by every major authority on genetics and anthropology; they call it a pseudoscientific rationale for racism, based on inadequate and unrepresentative sampling, predicated on erroneous assumptions, and statistically loaded to prove its point.

Credit for the second heavyweight statistical fallacy goes to Joseph Stalin and concerns statements he made about the success of his first Five Year Plan. The story is related most colorfully in Eugene Lyons' *Workers' Paradise Lost*:[4]

> No other economic enterprise in history has been so vastly publicized, so glamorized and misjudged, as Stalin's first Five Year Plan. As originally charted, the Plan covered every department of the nation's life, promising great advances in consumer industries, food production, housing. Meticulously the planning agency, Gosplan, detailed higher living standards. The purchasing power of the Soviet currency would rise by 20 percent, real wages by 66 percent, the cost of living would be lowered by 14 percent.

Lyons continues by describing a speech Stalin himself made only eighteen months prior to the end of the five-year period, a speech in which he came very close to admitting that the Plan had proved a dismal failure. Nevertheless, eighteen months later, in January of 1933, Stalin announced the quantitative fulfillment of 93.7 percent of the entire Plan! What kind of statistical trickery is reflected in this figure? Lyons explains as follows:

> . . . The Kremlin simply compared total result with the total planned instead of weighing the actual increase against the *planned* increase. For example, steel output in 1928 was 4.2 million tons. The Plan foresaw an increase to 10.3 million tons. Actual production in the final year was 5.9 million tons—up 1.7 million instead of 6.1 million, or 28 percent of the planned expansion.
>
> The Kremlin, however, said in effect: "We aimed at 10.3 and got 5.9, therefore, out Plan was fulfilled by 57 percent." On this basis, if production had not increased by a single ton, the Plan would have been carried out by over 40 percent—progress while standing still!
>
> When such sleight of hand is revealed, the official claims collapse. New housing, credited with 84 percent fulfillment, in fact increased only 44 percent. . . . The actual increase in cement was 37 percent, in brick 28 percent, in

[4] (New York: Funk & Wagnall, 1967.) Quotations presented here are from the *Reader's Digest* condensed version of the book (November 1967, pp. 233ff.). The wording but not the essence of the message differs slightly from the original.

automobiles 13 percent. Meanwhile, living costs zoomed, wages declined, hunger spread, consumer goods were tragically short.

Lyons' summary of the situation just described is a ringing testimonial to the potential treachery of a statistical lie when it is told by a strategic political figure at a strategic point in world history. Lyons concludes:

> But amazingly, the Plan has gone down in history as a fabulous success. Indeed, the belief that Communism is a virtual guarantee of rapid economic progress for underdeveloped nations stems primarily from this stubborn delusion which began when Stalin's boasts were accepted across a large part of of the world.

Clearly, faulty statistical reasoning can be relatively harmless or extremely dangerous. Much depends upon what the figures measure, how they are interpreted, and how the conclusions are acted upon.

What Is a Statistical Fallacy?

"The question is," said Humpty Dumpty to Alice, "which is to be master—that's all . . . . When I use a word it means exactly what I choose it to mean —neither more nor less." This, I must confess, is very nearly the same dictatorial attitude I have assumed while culling through many hundreds of statistical arguments to determine the fallacious ones. If an example impressed *me* as a fallacy I labeled it such. Period. In this respect, I have been a little like Lewis Carroll's Humpty Dumpty and the baseball umpire who insisted that a pitch is neither a ball nor a strike until he calls it one or the other.

The task of identifying statistical fallacies is fun but frustrating because the debatable ones and the borderline cases probably outnumber the clear-cut ones. For this reason, I have not attempted to define "statistical fallacy" in a rigorous way and then religiously stick with that definition throughout this book. Maybe a look at the dictionary definitions of *fallacy* will convince you of the impossibility of such as approach:

> 1 (a): Guile, trickery (b): deceptive appearance: deception. 2 (a): A false idea (b): erroneous or fallacious character: erroneousness. 3: An argument failing to satisfy the conditions of valid and correct inference.

Quite a few things, therefore, can be placed under this definitional umbrella. But who is to say what is deceptive and what isn't? In order for something to be deceptive, someone must presumably be deceived. Or is it enough that someone might possibly be deceived? Graphs with broken vertical axes, for example, are often labeled fallacious in articles about misuses of statistics on the grounds that they can give the viewer a false impression of both the level and the rate of change in the data charted. But if such a graph were prepared by a reputable economist for presentation before a group of fellow economists, especially if it were constructed with a view to highlighting important short-term fluctuations, it certainly couldn't be very deceptive.

Similarly, who is to say what is a false idea and what isn't? Sometimes a false idea is very easy to recognize; but when doubt exists, by what infallable criteria is such doubt erased? Consider, for example, this excerpt from a widely read magazine article:

> The production of pornography is a $19-million annual business in my state [California]. Nationwide, the production and sale of pornography is perhaps a $500-million industry. I estimate that more than 50,000 Californians participate in some way in the filth racket.[5]

These assertions can be faulted on the grounds that nowhere in the article is pornography defined. In the absence of either a generally accepted or a Friendly Definition[6] of this colorful word, the specific figures cited are meaningless. Nor are they made more meaningful by their proximity to such vague phrases as ". . . perhaps a $500-million industry" or ". . . participate in some way. . . ." But are the figures false? That, I suppose, depends on what the author is calling pornography—and he is keeping that a carefully guarded secret. I feel that less secrecy would have been desirable and am most willing to attach the label "fallacy" to this example. Someone else, on the other hand, might take the position that "Any darn fool knows what pornography is."

In the final analysis, determining what is and what is not a statistical fallacy will often involve differences of opinion. But perhaps that is all to the good from your standpoint. Since the principal goal of this book is to help you sharpen your critical judgment concerning statistical evidence, you should be prepared to question my position on any or all examples presented in the following chapters.

A Note About the Examples

Not only are statistical fallacies difficult to distinguish from the borderline cases, they are hard to categorize as well. Because the categories often used are not mutually exclusive, a single example can illustrate measurement

[5] Max Rafferty, "Crack Down on the Smut Kings!" *Reader's Digest*, November 1968, p. 98.

[6] By *Friendly Definition* I mean a definition selected from among several contending possibilities. The user of the data is in effect asked to accept that specific definition when interpreting the data. In return, the supplier of the data promises to adhere rigorously to that definition. Friendly Definitions are discussed more fully in Chapter 2.

problems, spurious accuracy, faulty comparisons, and maybe several other kinds of fallacies as well.

The difficulties of categorizing statistical fallacies are compounded when the would-be categorizer relies heavily (as I have done) upon examples from real life rather than upon hypothetical examples tailor-made to illustrate a single point in an exaggerated way. Admittedly, I have not eschewed the use of hypothetical examples altogether any more than I have held myself morally aloft from stealing examples from other authors. For the most part, however, the examples presented have been selected from among many hundreds of similar real-life examples collected by myself and several of my former statistics students (although I have admittedly taken the liberty of disguising some for reasons that should be obvious). The examples are drawn from real situations and relate to such diverse subjects as business, economics, psychology, biology, education, sports, entertainment, law, politics, and a great many others. My goal in using this approach has been not only to acquaint you with common misuses of statistics but also to convince you—subliminally if all else fails—that statistics does indeed play a vital role in today's world.

To some extent, the virtues of the examples selected are also their weaknesses. Here are a few caveats:

First, for reasons already given, the categories shouldn't be taken too seriously.

Second, when discussing a particular example, I have usually limited my comments to points bearing directly upon the subject of the chapter or section in which the example appears. Because the same example could perhaps be used to illustrate other kinds of fallacies as well, you may find some of my discussions lacking in thoroughness. Rather than being annoyed by this failing, however, why don't you take it upon yourself to supply the desired thoroughness? It should be good practice.

Third, my approach has necessitated much quoting or paraphrasing out of context. In no case, however, have I intentionally misrepresented another's argument.

In Defense of Being Negative

I promised earlier to elaborate on my statement that this book is about fallacious statistical thinking. The time has come to keep that promise. Such elaboration seems advisable because I am keenly aware that not everyone will approve of my emphasis upon faulty rather than valid statistical reasoning.

Prior to acquiring real statistical sophistication, most of us pass through two distinct stages in our attitudes toward statistical evidence. Early in life we tend to accept statistical conclusions uncritically on the assumption that figures don't lie. Often we even wilt upon mere exposure to statements beginning with "according to statistics. . ." or "statistics prove. . . ."

As we grow older, however, we tend to swing over to the opposite extreme. We have been too often deceived by advertisers, politicians, prominent citizens with various kinds of axes to grind, journalists seeking to increase the dramatic impact of a point, and on and on—in general, people who have developed considerable skill in drawing faulty conclusions from perhaps flawless data. We find ourselves believing that statistics can prove anything, and therefore can really prove nothing at all. Whereas we once believed that figures can't lie, we now conclude that figures can do nothing but lie. We come to share with Stephen Leacock the cynical notion that "In earlier times they had no statistics, and so they had to fall back on lies. Hence the huge exaggerations of primitive literature—giants, or miracles, or wonders! They did it with lies and we do it with statistics; but it is all the same." Some critics will undoubtedly accuse me of trying to swing you completely toward this attitude of bald skepticism. Such is definitely not my intention as is attested to by the fact that I am an enthusiastic user and teacher of statistics and have been so for quite a few years now.

Libraries are awash with excellent textbooks showing you how to use statistics the right way. I urge you to expose yourself to several of these. Unfortunately, relatively little has been written on the misuses of statistics, and as a result, a potentially worthwhile approach to teaching this broad and fascinating subject has been all but overlooked.

In this book I simply offer a different slant on the subject of statistics in the hope that by studying examples of how things shouldn't have been done you will not only be entertained but also find your judgment sharpened and your ability to appreciate good statistics enhanced. Also, I hope to lead you, as painlessly as possible, toward an understanding of why statistical tools must be used in conjunction with a near-fanatic love for truth.

My own views are quite accurately summarized by Ernst Wagemann:

> We share with Socrates the pious hope that men avoid mistakes once they are aware of them. But we are frivolous enough to suppose that men do this out of a spirit of pure contrariness, and hence are more affected by the sight of a horrible example than a good precept."[7]

[7] *A Fool's View of Statistics: The Outline of a Statistical View of the World* (Bern: A. Francke Ag. Verlag, 1950.) This translation from the German appears in W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach* (Glencoe, Illinois: The Free Press, 1956), p. 65.

2

Some Basic Measurement and Definition Problems

When you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.

—LORD KELVIN

In either of its two meanings, statistics is intimately tied in with the problem of measurement—*the use of numbers to represent properties.* Before one can study something scientifically, he must be able to express it in numbers, for only then can he distinguish easily and minutely between different but similar properties.

Unfortunately, the ideal way of expressing a property as a number may not be self-evident. Or if it is self-evident, the physical procedures required may be prohibitively expensive or in some other way impracticable. Researchers and professional data gatherers, therefore, must often resort to second- or third-best ways of measuring whatever interests them. The procedures adopted determine to a considerable degree the validity of the figures and the precise manner in which we, as consumers and critics of statistics, interpret them.

In this chapter we consider several kinds of problems related to the task of measuring things. Because it is impossible to measure something meaningfully without knowing what that something is, we must begin by concentrating on the crucial subject of definitions.

The All-Important Definition

Sometimes the task of measuring a property is quite simple. Determining the weight of a sack of potatoes, for example, is for most of us a task of less than staggering proportions. If someone—let us say a philosophy student—were to ask between long, thoughtful draws on his favorite briar, "Exactly how are you defining 'sack of potatoes'?" or "Are you planning to use avoirdupois weight or troy weight?" or "Can you prove the scales are accurate?" we might be moved to respond with something abrasive to a philosophy student's finer sensibilities.

On the other hand, when the thing being measured is "unemployment," "poverty," "marital compatibility," "mental health," "political popularity," or some other concept lending itself to many interpretations, the kind of cautious inquisitiveness displayed by our hypothetical philosophy student becomes absolutely essential. What could be less informative than a collection of figures purporting to measure, say, unemployment, when we are not even certain what kinds of people are counted as unemployed? Are nonworking children included in the count? Are housewives? Professionals? Part-time employees? People on temporary layoff? In most cases, it probably matters less how such problems are handled—provided, of course, that they are handled wisely—than whether we are told or left to guess about which categories of people are counted among the unemployed and which are not.

Whenever a term can be defined in a variety of ways, the data gatherer must decide which of the possible definitions seems most sensible, and, often just as important, which definition lends itself best to efficient, relatively inexpensive data collection. As a result, the definitions used are usually what I shall call *Friendly Definitions*, a term meaning that one of several contending definitions has been settled upon and the data user is asked to accept that specific definition when interpreting the figures. In return for this acceptance, the supplier promises to adhere rigorously to that definition.

Unfortunately, some purveyors of statistical information do not define the concepts they are allegedly measuring, a fact suggesting that our task is primarily one of learning to separate the statistical offerings of suppliers who play by the rules from those who do not. Granted, we must also pass on the adequacy of the definition offered, when one is offered; but when asked to go along with a clearly stated Friendly Definition, chances are we will usually oblige.

Here is a good example of the great importance a clear definition can have in giving meaning to a measurement, in this case a count:

> Whether the city with the world's greatest population is New York or London depends on what areas are referred to by "New York" and "London." The city of London proper had a population in 1955 of only about 5,200, and New York County, or Manhattan, one of the five boroughs of New York City, had 1,910,000. The analogous political units, however, are the City of New York, with a population of 8,050,000 in 1955, and the county of London, 3,325,000 in 1955. Each of these is a municipality made up of boroughs, 29 in London and 5 in New York. A comparison often made (though inaccurately) is that between greater London and the City of New York—probably because of the coincidence that the City of New York, when it was formed by the consolidation of New York, Brooklyn, and other areas in 1898, was referred to as "Greater New York." "Greater London," with a 1955 population of 8,315,000, is defined as the area within 15 miles of the center of the City of London.
>
> It has been estimated that the area within 15 miles of the center of New York has a population of 10,350,000. The "New York Standard Metropolitan Area," however, had a 1955 population of 13,630,000. (A Standard Metropolitan Area is defined by the U.S. Bureau of the Census as a county or group of counties containing at least one city of 50,000 or more, plus such contiguous counties as are metropolitan in character and integrated with the central city by certain specified criteria.) A metroplitan area defined for London on a basis similar to that used for New York have a population of approximately 10,000,000.[1]

The figures are admittedly dated, but the moral isn't. Imagine how meaningless a comparison between the populations of London and New York City would be if the geographic boundaries used were not clearly spelled out. The

[1] W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach* (Glencoe, Illinois: The Free Press, 1956), p. 68.

necessity of using Friendly Definitions is also demonstrated in this example, for whatever definitions were decided upon, they would necessarily be Friendly according to my Friendly Definition of this term.

Additional examples of the strategic role of definitions are easy to find. In cost studies, for instance, confusion sometimes occurs between the economist's and the accountant's definition of "overhead cost." In economic analysis, overhead costs do not change with changes in the volume of production, whereas accountants sometimes allocate these costs among different years or different products in proportion to the volume of production. Either practice can be justified, but the definition used in a specific instance should be explicitly stated.

The familar and innocuous-sounding word "industry" can plague business economists whether they are charged with measuring the degree of monopoly in an "industry" or studying the extent to which a specific "industry" utilizes the output of other "industries." The measuring devices used are constructed so that the statistical results are totally determined by the manner in which one "industry" is distinguished from others. Any way such a distinction is made is bound to prove imperfect. Is there such a thing as an "automobile-tire industry," for example, or are automobile tires manufactured by the "rubber industry" or some other industry?

Examples of the Statistical Leverage of a Definition

If terms like "overhead cost" and "industry" are trouble makers, "poverty" is a hardened criminal. Efforts to measure the extent of poverty in the United States have been both numerous and well-publicized in recent years. The usual procedure is to (1) designate some level of family income that will serve to distinguish poverty families from the others, and (2) calculate the number of families whose income falls below the specified cutoff point. Despite some shortcomings, which I'll touch on momentarily, this basic approach is probably as good as any other reasonably simple one. At least the figures are quite easy to interpret, a condition which might not exist if the concept were too greatly refined. It should be remembered, however, that the level of family income that spells poverty completely determines the number of families receiving that designation and is necessarily always picked somewhat arbitrarily.

While not wishing to make light of the serious human and social problems implied by the poverty statistics, I must admit to having been grimly amused by the poverty numbers game that some politically oriented economists have been playing, particularly since 1964. To play the game, all one really needs, it seems, is a plausible Friendly Definition of "poverty" that differs somewhat from the definitions used by the other contestants.

The game was being played prior to 1964, but its visibility was increased greatly in that year with the release of a report prepared by the President's Council of Economic Advisors and entitled "The Problem of Poverty in America."[2] The CEA decided that households with incomes of less than $3,000 per annum lived in poverty. Using this income criterion, they calculated that 20 percent of all U.S. households containing some 30 million persons fell into this poverty class. How was the $3,000 figure arrived at? In brief summary, it was obtained by starting with a concept of minimum nutritional adequacy, consisting of specified amounts of calories, proteins, vitamins, etc., and translating these requirements into food items that would do the job. It was determined that, with 1964 prices, these food needs could be supplied for about $5.00 per week per person. For an urban family of four, the diet comes to about $1,000 a year. On the assumption that low-income families should spend about one-third of their incomes on food, a poverty standard of $3,000 was arrived at.

As previously mentioned, any income criterion of poverty is necessarily arbitrary and woefully inadequate in a number of ways, a fact that the CEA was more keenly aware of than anybody. Here are a few of the more apparent inadequacies of the one used: First, it does not take assets into consideration or count as income such things as money from sale of property, borrowed funds, gifts, lump-sum inheritences or insurance payments. Hence, a retired couple who own their own home and enjoy a modest income from stocks and bonds and who, of course, are free to cash in some of their assets at any time they choose, might be classified as a poverty family even though their subsistence needs are well taken care of.

Second, the $3,000 figure is itself, to quite an extent, a reflection of American affluence rather than an indication of the bare minimum income requirement for keeping body and soul together. About one-third of the poverty families had homes and one-half had cars. Most had telephone service and various household durables that might have seemed like extravagant luxuries to our great grandfathers had they been available at all.

Third, transitory poverty is not distinguished from permanent poverty. Temporary unemployment of the main breadwinner due to a sluggish national economy could make the family eligible for inclusion in the poverty count one year but not the next. Maybe that's the way it should be, but we must keep this transitory component in mind when interpreting the figures. Most of us, I believe, think of poverty as a more or less chronic condition.

These disadvantages of the $3,000 poverty criterion, as well as others that might be mentioned, are certainly not news to the Council of Economic Advisors. Nor are they news to the Social Security Administration. (The SSA has recently refined the concept of poverty by getting inflation into the

[2] Presented in *Economic Report of the President* (Washington, D.C.: United States Government Printing Office, 1964), pp. 55–84.

calculations and by taking account of family size and composition, the age of members, and whether the residence is farm or nonfarm.) Despite its several inadequacies, the $3,000 figure for 1964 was probably as good as any reasonable alternative figure that might have been arrived at.

The official count of 30 million persons living in poverty in 1964, however, rather than settling the matter, served to fire the imaginations of other poverty researchers who proceeded to see who could give the count the biggest boost. The poverty numbers game grew quite exciting for a time with several contestants placing the figure in the 40 to 50 million range. But the winner, the one whose definition of poverty exerted the most *Statistical Leverage* (a term I like to use to indicate the extent to which the count changes, given a small change in the definition), seems to be the zealous fellow who counted a startling 80 million poverty-stricken people, a figure representing more than a third of all the people living in the United States.

The game seems to have settled down a little in recent years. Still, each time a new report is issued, a greater commotion follows than is justified by the quality of the figures. According to the SSA report for 1970, for example, the number of people living below the poverty line was 25.5 million, some 1.2 million (5.1 percent) above 1969. This news was widely interpreted as an indication that as a nation we are losing the war against poverty. Is this conclusion justified? Maybe. But let us not forget that 1970 was a recession year and the transitory component, mentioned above, was undoubtedly much inflated.

How many people would you guess were unemployed during the Great Depression of the 1930's? Well, picking out a specific month, we find that in November of 1935 the figure was around nine million. Or 11 million. Or 14 million. Or 17 million. It all depends on whose figures you like. The following bewildering and contradictory list of unemployment estimates all pertain to this one month and all were prepared by reputable agencies:

Table 1. Estimates of Unemployment for the Month of November 1935 According to Five Reporting Agencies

Agency Preparing Estimate	*Estimate of Number Unemployed*
The National Industrial Conference Board	9,177,000
Government Committee on Economic Security	10,913,000
The American Federation of Labor	10,077,000
National Research League	14,173,000
Labor Research Association	17,029,000

Source: Jerome B. Cohen, "The Misuse of Statistics," *Journal of the American Statistical Association*. XXXIII, No. 204, (1938), 657.

To make matters worse, just six months later, as the United States Chamber of Commerce issued a report estimating the number unemployed at 4 million, the *New York Sun* announced that on the basis of a survey of 30 million workers, unemployment amounted to between 3 and 3.5 million. The Labor Research Association insisted that all estimates lower than its own were erroneous. The Chamber of Commerce held that all estimates higher than its own were inaccurate.[3]

These estimates differ, of course, primarily because of differences in the definitions of unemployment used by the various sources. Some estimates took into account unemployment among farm labor and some did not; some included estimates of people leaving school and seeking employment for the first time and some did not; some considered unemployment among professionals and some did not. And so it goes—the considerable variation among the estimates testifying to the sometimes substantial Statistical Leverage exerted by a difference in definition.

As far as this country is concerned, such differences of opinion have been eliminated—officially, at least. The Bureau of Labor Statistics releases unemployment statistics each month that are the most all-embracing in the world. The figures include people who have (1) voluntarily quit their jobs, (2) been discharged for misconduct or poor work, (3) recently found jobs but have not yet reported for work, (4) are available only for part-time or temporary work, (5) simply haven't bothered to look for work because they don't believe that any is available, or (6) are on temporary layoff due to a strike affecting a major customer or supplier of their employer. So just about any age-eligible person who can possibly be construed as unemployed is included in the count. As long as we are aware that the Friendly Definition of unemployment in this country is so all-embracing, we can use the figures in many meaningful ways. Problems still arise, however, when making comparisons between countries. Most foreign unemployment estimates are based on registrations at employment exchanges, a method that results in a relatively lower count than does this country's approach.

Recently, members of a Senate Small Business subcommittee had difficulty determining just how many franchise operations exist in the United States. Robert M. Dias, president of the National Association of Franchised Businessmen, told the Senators there are 1,200 franchisers with 670,000 franchisees doing a total business of $100 billion. John V. Buffington, general counsel of the Federal Trade Commission, citing latest Commerce Department figures, put the franchisers at 1,100 and the franchisees at 400,000. Thomas H. Murphy, publisher of the *Continental Franchise Review*, said there are "con-

[3] Cited in Jerome B. Cohen, "The Misuse of Statistics," *Journal of The American Statistical Association*. XXXIII, No. 204, (1938), 657.

servatively" 500,000 franchisees doing a total business of $90 billion.[4] The source of the trouble was, again, lack of agreement on what should be included in the count. Some witnesses construed "franchise business" to be limited to small hamburger and fried chicken, fast-service operations, while others viewed it as including automobile dealers and service stations.

Important Omissions in the Data

Needless to say, once a definition has been decided upon all relevant subcategories should be fairly represented in the reported data. Omission or under-representation of a strategic component can lead to annoying errors in interpretation.

Between 1950 and 1956, for example, some 1.5 to 2 million houses, worth maybe $15 billion, were lost in this country. They weren't lost through fires, floods, or other natural disasters; they were simply misplaced, so to speak, by government statisticians. Instead of 8 million houses going up in those years, as was originally reported, the correct figure was probably almost 2 million (25 percent) higher than that. Recognition of this underestimate led to a program of aerial data gathering over specific counties where reported data on housing starts and value of construction put in place were considered by government statisticians to be especially inadequate.[5]

A more recent example of the potential treachery of an excluded or under-represented component was a $7-billion error made by the Federal Reserve Board when reporting money supply figures for October 1970. These figures are watched closely by analysts of business conditions because changes in the rate of change in the money supply carry implications about future economic growth, inflation, and, in turn, forthcoming Federal Reserve policy decisions. The error resulted from a failure to include a large volume of international dealings in *Euro-dollars*—U.S. dollars held by foreigners. When the figure of $206 billion was boosted to $213 billion, much colorful debate arose around the country regarding the likelihood of an imminent reversal in Federal Reserve policy.[6]

Unfortunately, there is little that even a conscientious user of statistical data can do about recognizing such omissions. And there is absolutely nothing he can do about correcting for them himself. If he is wise, however, he will develop the habit of viewing most highly aggregated national economic data as a tentative mix of fact, estimate, and judgment. A variety of problems related to definitions and methods of data collection still remain and few such problems will be resolved in the very near-term future.

[4] *Business Week*, January 31, 1970, p. 30.

[5] *The Wall Street Journal*, November 2, 1959.

[6] *The Wall Street Journal*, November 30, 1970.

Spurious Accuracy

The story is told about a man who, when asked the age of a certain river, replied that it was 3,000,004 years old. When asked how he could give such accurate information, his answer was that four years ago the river's age was given as three million years.

Clearly, the man in this story was unaware that the three-million figure was a crude estimate rather than a precisely known fact. His tacking on the four years was not only unnecessary but potentially misleading as well, for it gave the impression that a degree of accuracy had been achieved that was really unattainable. This is an example of *spurious accuracy.* Many things simply cannot be measured with as much accuracy as some purveyors of statistical information like to pretend.

An automobile advertisement caught the reader's attention with the assertion that on a certain day an estimated 262,825,033.74 tons of snow fell upon the United States.

The official publication of the Austrian Finance Administration stated that the population of Salzburg Province in 1951 was 327,232 people—4.719303 percent of the entire population of Austria.

A large distillery declared that over the years the company had squeezed 191,752 oranges, 580,582 lemons, and 453,015 limes to make its whisky sours, daiquiris, and margaritas.

The New York Times reported that the St. Patrick's Day parade cost the city $85,559.61 whereas the Puerto Rican Day parade cost only $74,169.44.

The Automobile Manufacturers Association reported in *1971 Automobile Facts and Figures* that employment in automotive parts production by companies outside the automobile industry proper was: for narrow fabrics, 1,127 people; for apparel findings and related products, 20,196 people; for points and allied products, 4,959 people; for hardware, 34,303 people; and the list goes on. All told, 37 categories of automotive parts are listed with the related number of employees shown with seeming accuracy right down to the last man or woman. To the AMA's credit was the explanation in an attached footnote revealing that the figures were estimates subject to error.

Examples could be multiplied indefinitely. Seldom in statistical work of any kind can precise figures like those in the above examples be obtained. But the appearance of accuracy suggests to many trusting readers that the source "really knows what he's talking about."

This fallacy is found in all branches of statistical investigation. Whatever the context, one is always wise to be skeptical of figures pretending great accuracy. Usually the simple test of asking yourself "Judging from what I know about the thing being measured, can I really believe that such accuracy is possible?" will be sufficient to keep you from blindly accepting data pretending unattainably high orders of accuracy.

Valid Measures Used Inappropriately

Sometimes a perfectly good measure is used as an imperfect proxy for something else. The following advertisement, for example, is hypothetical but based loosely upon an actual one that appeared in many magazines around the country:

> Smythe's Elixer corrects a variety of scalp diseases and stops the hair loss they cause. Smythe's has been used by over half-a-million people on our famous Double-Your-Money-Back-Guarantee. Only 1% of those men and women were not helped by Smythe's and asked for a refund. This is truly an amazing performance.

The most obviously misleading part of this advertisement is the use of requests for refunds as a measure of the number of customers not helped by the product. What do you suppose is the ratio of dissatisfied customers to the number of customers who actually take advantage of a money-back guarantee for a relatively inexpensive product? One-to-one as implied by the ad? Two-to-one? Ten-to-one? One hundred-to-one? The true ratio, of course, is unknown and quite likely unknowable. Almost certainly, however, using the number of customers requesting a refund to measure the total number of customers who weren't helped by the product leads to a too-low estimate of the truth.

Who has not heard the oft-repeated advertising claim that nine out of ten doctors recommend the ingredients of a certain patent medicine? Let us accept for now the part about the nine out of ten (presumably, though not necessarily, every ten) doctors recommending the ingredients in this product; it does not necessarily follow that they recommend the product by brand name. I am informed by a doctor friend that when the company researchers conduct their marvelous surveys they indicate only ingredients, not brand name, and these ingredients are common to many commercial medicines. This advertising claim appears to be an earnest attempt to mislead consumers into thinking that the product is recommended more often by name than is actually the case.

Automobile registration figures are frequently used to measure the number of automobiles in the hands of the public. However, such figures are not entirely satisfactory for this purpose for at least three reasons: (1) Some states issue a new registration upon sale of a car while some transfer the old registration to the seller's new car; (2) Station wagons, taxis, and some other types of automobiles are classified as passenger cars in some states and not in others; and (3) Some cars are registered to dealers before they are sold to customers.

The use of proxy measures is often justifiable because the thing of interest doesn't lend itself to accurate and/or relatively inexpensive measurement. However, one should at least be able to recognize when a proxy measure is being used and be prepared to pass judgment on its quality. Some proxies are better than others and many are decidedly bad.

In the next chapter we focus more closely on problems of measurement and definition when we consider the rather colorful, though totally unenlightening, subject of *Meaningless Statistics.*

3

Meaningless Statistics

. . . a tale told by an idiot, full of sound and fury, signifying nothing.

—SHAKESPEARE

My favorite Meaningless Statistic is this one attributed to humorist Robert Benchley: "It is not generally known, I believe, that one comic editor dies every 18 minutes, or, at any rate, feels simply awful." It seems to me that if someone is going to waste my time and insult my intelligence by foisting Meaningless Statistics upon me, he should at least have the common decency to see that his offerings are funny. So far, Benchley is the only one to meet my rigorous requirements. Most Meaningless Statistics aren't even very funny. In this chapter we deal briefly with these little time wasters.

Some Typical Examples of Meaningless Statistics

The executive of a certain company claimed that about 75 percent of the entire organization had been with the company for many, many years.[1] Is that not a truly impressive record? On second thought, just how impressive

[1] C. I. Daugherty, "How Dedicated People Build Sales," *Specialty Salesman*, August 1968, p. 10.

is it? The answer, of course, depends entirely upon what is meant by "many, many years," a detail this executive saw fit to spare us.

The organization referred to is a manufacturer and distributor of stainless steel cookware, one of several brands sold exclusively by direct salesmen. I admittedly have no idea what the salesman-turnover rate is for this particular company, but I do know that in most direct selling it is pretty high—so high, in fact, that "many, many years" could conceivably mean four or three or even fewer. (Not that it necessarily does, mind you. We simply have no way of telling from the information given.) The 75-percent figure sounds precise enough, but the entire claim fails to deliver the precision promised because the rest of it is so vague.

This is a fairly typical example of a Meaningless Statistic. A *Meaningless Statistic* is a precise figure used in conjunction with a term sufficiently vague that a Friendly Definition is sorely needed to endow the figure with meaning. But such a definition is not provided, or, if one is provided, it is itself so vague that it doesn't really help.

The term "Meaningless Statistic" was coined by Daniel Seligman in a delightful article called "We're Drowning in Phony Statistics."[2] Seligman cites, among many other examples, the following assertion made by a former U.S. Attorney General: "Ninety percent of the major racketeers would be out of business before the end of the year if the ordinary citizen, the businessman, the union official, and the public authority stood up to be counted and refused to be corrupted." That the underlying thought is plausible as well as praiseworthy is beyond dispute. Nevertheless, in the absence of either generally accepted or Friendly Definitions of terms like "major racketeer" and "stood up to be counted," the 90-percent figure used to dress up the argument is meaningless.

An author asserted that he had studied the food intake of more than 50,000 men and women and found, to his own astonishment, that in one group of 4,500 cases, 83 percent were found to be overweight while undereating. Moreover, only 17 percent were found to be overweight because they overeat. These facts might have seemed astonishing to the reader as well as to the writer if the latter had only been more clear about what he was calling "overweight," "undereating," and "overeating." If the actual criteria used were no more precise than the description given in this article, then the 17- and 83-percent figures are worthless. Either way, the reader of this particular article is left poorly informed.

An advertisement claimed that 95 percent of key government officials read a certain newspaper but failed to let the reader in on how one goes about distinguishing between a "key official" and one who is less "key."

[2] *Fortune*, November 1961, pp. 146 ff.

More Subtle Examples

Now and then, we run across statistical information where definitional details are super-abundantly present, but we still find ourselves poorly informed. The following, a statistical tautology of sorts, is one such case:

Toward the end of 1967 considerable publicity was given a figure released by the U.S. Bureau of Labor Statistics to the effect that it costs $9,191 for a family to buy "a moderate living standard." The spurious accuracy of the figure (the amount being shown right down to the last dollar) gives the impression that if anyone *really* knows about "moderate living standards," it is the Bureau of Labor Statistics.

Nevertheless, the Bureau's attempted explanation of the term, presented in a 40-page publication entitled "City Workers' Family Budget," provides a paradigm of circular thinking. First come the qualifications. We are told that the figure (1) pertains only to families in urban areas, (2) is not exactly current, based, as it is, upon 1966 data, and (3) represents a national average. So far, so good. We could probably live with these qualifications without undue strain.

We next must know what, according to the Bureau of Labor Statistics, is a "family." A "family," we are told, is a group consisting of a man who is 38, a wife (age unspecified), a boy of 13, and a girl of eight. To deal with families not conforming to these specifications, the Bureau provided a so-called "equivalence scale," which describes other kinds of households and tells us how their costs compare with the standard family's. As the editors of *Fortune* remarked in a scorching editorial:

> For example, an adult under thirty-five living with three children requires 88 percent as much as the standard family in order to live moderately. These relationships were established by studying in staggering detail, the spending patterns of different kinds of consumers. Data in "City Worker's Family Budget" show, for example, that husbands in metropolitan areas buy straw hats, on the average, once in twenty years; in nonmetropolitan areas, husbands seem to need a straw hat every six years. Boys in metropolitan areas get 12.24 pairs of socks in a year; in nonmetropolitan areas they get only 10.31 pairs. Perhaps that is all their fathers can afford after buying all those straw hats.[3]

Splendid. But what is a "moderate living standard?" That, after all, is the key to understanding the $9,191 figure. Well, here are some of the things it isn't: It is not a minimum or "subsistence" standard. It is not an average living standard. Nor is it a standard required for a sense of well-being. BLS states: ". . . many families can and do spend less than the total amount specified in

[3] "Shadowy Statistics (Contd.)" (Editorial), *Fortune*, December 1967, p. 98.

ficients."[5] From this table we learn that physicists have a status coefficient of 7.64 whereas pilots only achieve 7.62. Radio mechanics are also 7.62, ahead of mathematicians who only merit 7.34 and well ahead of geologists with their mere 7.22.

As readers, we are assured that these status coefficients permit government heads ready information for determining how easy or difficult it will be to provide personnel in a given occupation. Unfortunately, we are told nothing about procedures used in computing these coefficients. Granted, understanding them might require such a high order of sophistication that our inferior minds might not be up to the challenge. But if that is so, why are they shown at all? The simple ranks of 1, 2, 3, . . . would be every bit as informative as status coefficients devoid of clarifying details.

In the next chapter we consider statistics whose meanings are clear but which were arrived at by highly questionable procedures. These I call *Far-Fetched Estimates.*

[5] Vladimir Shubkin, "The Occupational Pyramid: Low and High Status Jobs," *Soviet Life*, September 1971, p. 21.

tion being out of the question), but the people questioned either do not know the answers or are reluctant to give accurate answers. Perhaps the questions require the respondent to remember things that he is not in the habit of keeping track of—as, for example, "How many cups of coffee have you consumed during the past thirty days?" Or perhaps the questions deal with illegal or potentially embarrassing activities.

I recently heard a speaker declare: "There are at least five to seven million couples who swap mates in this country in one way or another, sometimes at parties especially for the purpose, and at other times in small groups or sometimes just a pair of couples." The difficulty of obtaining accurate information on mate swapping, especially since only a pair of couples is often involved, is too obvious to merit comment. When the speaker was asked where he got his statistics, he answered that it was "common knowledge among sociologists and a conservative estimate." Such "common knowledge" is at worst misleading and at best pointless and silly.

One of the most unsettling Unknowable Statistics I have run across illustrating the point that sometimes the only people who might know don't know was turned up in a poll, conducted by a university student newspaper, of marital status among freshmen. The results: Single– 1,568; Married– 16; Undecided– 11!

Example 4: A Young Man Rejects All Girls Under Five Feet Tall as Possible Dating Partners Because Three Such Short Girls Have Jilted Him

In this case, although the sample is easy to identify, being made up of the three girls under five feet in height who have jilted this embittered young man, the population is much more vague than in the previous examples. Presumably, the population consists of all eligible girls under five feet tall whom this young man might conceivably date if he were aware of their existence and if external conditions permitted.

You would have good reason to doubt whether the three girls in the sample were representative of all the girls in this nebulous population. It goes without saying that people, including short girl friends, differ a great deal from one another. That is, they are highly variable with respect to many important characteristics. A sample of only three short girls would certainly be much too small to reflect accurately the substantial variation in personality traits in the population being judged.

Some Fundamentals of Sampling Theory and Practice

Throughout this book I have assumed that your goal is to learn to evaluate with some sophistication statistical evidence presented by others. I have not assumed that you will be doing any elaborate statistical analysis yourself (although you may, and I hope you do). But if you are to evaluate the worth of information obtained from samples and the inductive conclusions derived therefrom, you must be familiar with at least a handful of basic sampling concepts, terms, and procedures. In this section I shall touch briefly on the bare minimum collection of sampling topics with which you should be quite familiar.

An Important Basic Assumption

The single most important sampling concept is this one: If sample items are chosen at random from the total population, the sample will tend to have the same characteristics, in approximately the same proportion, as the entire population. Notice the emphasis on random selection. If the sample really is selected in a random manner, we can place great confidence in this basic assumption provided we show proper respect for the word "tend." Sample characteristics only *tend* toward or approximate the corresponding population characteristics.

Probability Samples, Judgment Samples, and Convenience Samples

Three fundamentally different approaches to sampling can be distinguished. These are probability sampling, judgment sampling, and convenience sampling. Statisticians, generally speaking, have relatively little to say about the last two approaches save for warning of their inherent limitations. Probability sampling, on the other hand, is the cornerstone of much formal statistical analysis wherein information about a sample is utilized to make precise guesses about the corresponding population.

The unique characteristic of all probability sampling procedures is that the selection of items from the population for inclusion in the sample is made according to known probabilities. This characteristic of probability sampling implies three other features: (1) A specific statistical design is followed, (2) The selection of items from the population is determined solely according to known probabilities by means of a random mechanism, usually a table of random digits (to be discussed shortly), and (3) The sampling error—that is, the difference between the result obtained from a sample survey and that which would have been obtained from a census of the entire population conducted using the same procedures as in the sample survey—can be estimated and, as a result, the precision of the sample result can be evaluated. Notice particularly that with probability sampling, personal judgment about which population items should be included in the sample is ruled out. Moreover, once a sample item has been selected using the random mechanism, it must be included in the sample and not arbitrarily discarded.

So what are judgment and convenience samples? These terms must be touched on even though the main focus of the remainder of this section will be probability sampling. Sometimes a judgment or a convenience sample will be presented as if it were a probability sample. For that reason alone, the statistical critic should know the difference.

In a judgment sample, personal judgment plays the key role in determining which population items are selected for inclusion in the sample. The selection of "representative" sample items is a matter of personal conviction rather than the outcome of an impersonal random mechanism.

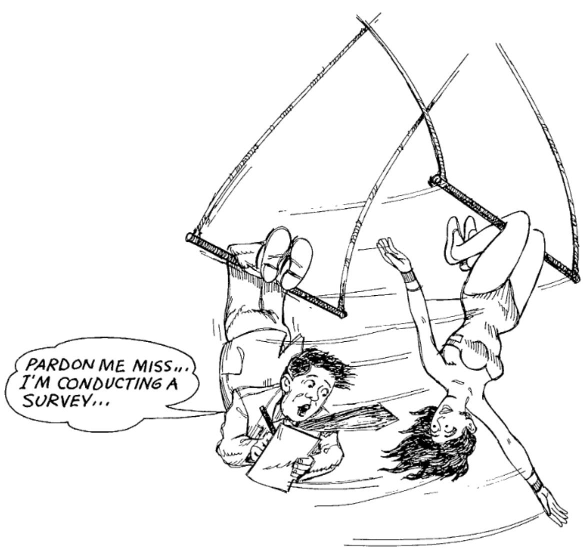
A convenience sample is merely a part of the population that happens to be conveniently at hand. A former marketing professor of mine, for example, one day brought some disposable paper ashtrays to class for the purpose of soliciting the opinions of the class members regarding the probable usefulness and popularity of this idea. He was obviously utilizing a convenience sample.

Judgment and convenience samples have legitimate places in research work. But they do lack certain advantages of probability sampling, namely lack of bias in selection and amenability to measurement of sampling error, and should not be passed off as true probability samples.

Before leaving this last point, I must emphasize that a convenience sample bearing only the most superficial possible resemblance to a probability sample is sometimes passed off as the latter. Reference to a fairly recent television commercial should help to clarify this point. In this commercial, a man runs out into the street and asks each of three people in cars—apparently waiting for the light to change—whether he or she uses a certain product. Two answer "yes," and one answers "no." The viewer, unless he is on his guard, not only tucks away the impression that two out of three people use this product but also that the sample is random. The interviewer picks out three cars quite arbitrarily, hence, the seeming similarity to random sampling. But this is still a convenience sample and not a random one. A truly random sample would give the stay-at-homes and people in other parts of the city a chance to be included as well as drivers in that particular neighborhood.

In an article called "What Makes A Perfect Husband?" author Sam Blum states:

> I interviewed, at random and with relish, secretaries, college girls, schoolteachers, a young woman standing in line to see a movie, two dancers, a salesgirl, several women temporarily between marriages, the wives of friends —a whole array of assorted husband watchers.[3]



[3] "What Makes the Perfect Husband?" *McCall's*, August 1967, p. 61.

Index