Chapter: Methods and Data

Methods and Data

WHAT'S THE ANSWER?

Why is it not enough for an advertiser to say, "Brushing with Ripsnort yields 16% fewer cavities"?

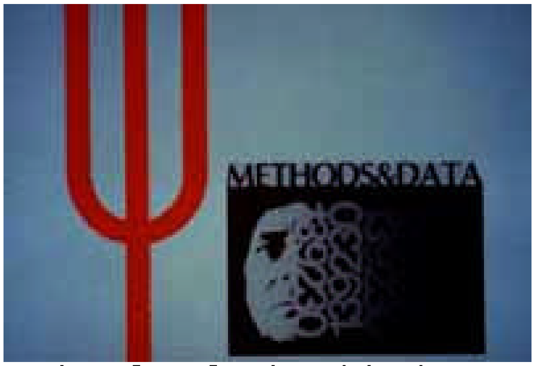
"Sheila, I don't think I'll ever forget an argument I heard when I visited the Canadian Parliament during the Spring of 1998. Two Members of Parliament (MPs) were arguing on the floor about the average annual income for Canadian households."

"What's so memorable about that, Bart?" asks Sheila.

"A Liberal MP said something like, 'The average annual income is almost $63,000 (CD) per household throughout Ontario!' Another MP who was a Conservative got up -- in fact, he interrupted -- to point out, 'Nonsense! The average Ontario household earns less than $55,000 (CD) per year!' What made it so interesting is that both MPs were right."

"That can't be," replies Sheila. "One of them has to be wrong!"

Who's right here, Sheila or Bart? Why?



As an empirical science, psychological studies require the creation of experiments, with proper control conditions and regard for the ethics of conducting experiments. Ideas for experiments come from curiosity, past research, and theory. To establish a functional relationship between a cause and an effect, psychologists use control groups. They compare experimental- and control-group behavior between groups or within groups. In doing research psychologists must be on guard against various errors, including demand characteristics, and experimenter biases, caused by the process of measuring behavior. All psychological research is done following a strict set of ethical guidelines for treatment of participants.

Statistics are important in analyzing the data generated in many psychological experiments. The simplest type of descriptive statistic is a frequency distribution of the raw data. Averages (mode, median, or mean) can also be calculated, as well as measures of variability and of the skewness of data. Inferential statistics are used to help a scientist decide whether experimental-control differences could occur by chance or are more likely due to the effects of the independent variable. Correlations summarize the size and nature (positive, zero, or negative) of a relation between two sets of data. A high correlation does not mean one variable has caused the other.

Once the results are collected and the results analyzed, a psychologist must then write a report regarding his or her findings. Such reports typically have four main sections (Introduction, Method, Results, and Discussion), but may also have two or more ancillary sections including an Abstract and References.

Experiments Controls and Ethics

As demonstrated throughout this book, psychology is an empirical science. This means that psychological studies focus on things which can be measured. Empirical observations yield objective data -- numbers or frequencies of events. And thus psychologists ultimately face two major challenges -- how are our empirical studies to be designed, and how are the data generated from those studies to be analyzed?

The conduct of psychological research continues as we try to answer questions about human (and animal) behavior. Whereas astrologers appeal to ancient oft-unconfirmed truths, psychologists -- and scientists generally -- are distinguished because of their use of control groups. The classic forms for such studies involve the use of between-groups comparisons in which control group performance is compared with that of one (or more) experimental groups. Or in some situations participants serve as their own control and psychologists are making within group comparisons of the same humans' or animals' performance under two or more conditions. Unlike physical scientists, psychologists face the unusual situation of studying interacting, thinking, (oft-)human organisms who are usually trying to determine exactly what the researcher him- or her-self is trying to study. All such human-to-human interactions between researchers and participants are subject to a variety of sources of error generated out of those interactions. The human-human or human-animal interaction is a central part of modern-day psychological research. Because of this, the American Psychological Association -- a national association of over 100,000 psychologists -- endorses a series of ethical principles encouraging courteous, safe, proper treatment of all who participate in psychological research studies.

Psychologists' research generates numbers in studying everything from testing of intelligence to hunger in animals. Such numbers lead to the need for statistics which are important in summarizing and interpreting data. To help you understand statistics, it is important to talk about how psychologists conduct research. From that we can show you how statistics are used.

What is an Experiment?

As we discuss in the Psychology: Its Nature and Nurture chapter, do you know what an experiment is? It's an attempt to establish a functional relationship between independent and dependent variables. In short, it's an effort to find out what stimulus causes what response or what response is related to what response. We suggested that to do that you could gather data using experimental (laboratory) methods, naturalistic (field-research) observations, or statistical methods. You could also conduct interviews or surveys. A number of examples of such procedures for gathering data are given throughout this book.

From where do ideas for experiments come? That's a problem you may have already faced when your instructor asked you to "do an experiment studying behavior"! As you might suspect, most experiments don't occur "out of the blue." Thomas Edison once said, "Genius is 5 percent inspiration and 95 percent perspiration!" He's not far from right, but the ideas for experiments come from a number of sources. One is the "I wonder what would happen if process. Have you ever been curious as to why people almost always walk in the right door when a building has two doors, even without signs telling them to do so? Observing our environment, noting something that usually happens, and then trying to develop an explanation for it is one ready source of ideas.

Another source is the research that some other person has done in the past. Many psychologists engage in what is called "programmatic" research meaning that each study may be but a small part of an ongoing research program. One psychologist published almost 30 studies back in the 1950s and 1960s. They were attempts to explain why it is that distributed practice is better than massed practice, as we discuss in the Learning Chapter!

A third source is to test hypotheses derived from a theory. If a theory is correct, it ought to be possible to extend it and make predictions about what should happen in situations never tested before. A fourth related source of ideas is simply to review the research literature. Checking recent publications -- journals, magazines, books, texts or even e-mail -- is a good way to discover the latest findings on current topics.

Regardless of the source of the ideas, once the question is posed, the difficulties of finding the right answer have only begun. Research studies in psychology can often be classified as one of three designs. Between-groups studies involve comparing the performance of separate groups of randomly sampled participants. Want to know the optimum temperature for sleeping? You can quickly narrow the range of temperatures to be studied by assessing the sleep behavior of separate groups of randomly selected participants at 10o increments from 30o to 110 o.

Alternatively a within-group comparisons can be used if the experimental manipulation does not permanently alter the subject in any way. If we want to determine the most effective diet for weight loss, we could try each of several strategies on the same group of people and make within-group comparisons to identify the best. The reason? In this case we don't care what the starting weight of our participants is, we are mainly interested in which diet strategy causes the greatest weight-loss from the starting weight.

A third design involves statistical analyses which are attempting to establish the extent to which a participant's responses in one situation can be predicted from his or her performance in another situation. Assessing your life-time driving skills with a 10-minute driving test is an example of such an assessment. In these studies the statistic of interest is usually a correlation between the two sets of responses. All three types of studies must be designed so as to reduce the likelihood that unanticipated sources of error will adversely impact the results. Likewise, a strict code of ethics should also be followed in conducting any psychological study.

Between-Groups Comparison

We stress throughout the book how important it is to have control groups. Do you remember why? In any experiment to establish a functional relationship between an independent and a dependent variable, we must isolate the effect of the independent variable. So we establish a situation in which participants experience every important variable. That's our control group. In addition, then, we have a second group or condition that includes all of the normal variables plus the independent variable in which we're interested. That's our "experimental" group. If differences occur between what our experimental group does and what our control group does, what can we conclude? It must be due to the effects of our independent variable. Inferential statistics help us decide whether such differences are important.

One important alternative must be considered when we set up our control. Should we use two separate groups -- one experimental, one control? Or, should we use one group for both experimental and control conditions and let each person experience both conditions which leads to comparisons within participants? There are a number of guidelines to help us make that decision. Regardless of the design choice we make, both approaches face the necessity of keeping errors in data collection to a minimum.

Suppose we're interested in learning how the lighting in one of your classrooms or lecture halls influences your ability to learn material by reading. We might set the lights at some level, give your class a piece of material to learn, and then test to see how much you've learned. But now we can't use your class again, because you've already learned the material! You are no longer experimentally naïve. If we set the lights at another level and gave your class the same material, you'd instantly know it as well as the first time you'd learned it! Instead, a between-groups comparison -- using similar, but different students -- is needed.

In the Social Behavior of Groups Chapter we talk about leadership and about what makes a person attractive to us. In order for any group to operate effectively, it must include people with good ideas -- new views, new ways to achieve a group's goals, and so forth. Often the leader is such a person, yet it has been observed that frequently the most intelligent person in a group is neither the most popular nor the leader. Might leadership be better if an intelligent person could be made more attractive, and thus better liked?

Feature 1 describes a piece of research intended to measure exactly this effect. It was suggested that smart people might (sometimes) simply seem less attractive. How could you make such a person appear more approachable? Why not have him or her make a mistake -- something that could happen to any of us? Read the Feature to see how this was done.

FEATURE 1

Falling on Face to Gain Face

Forty eight 19-year-old college sophomore males were recruited to listen to one of four tape recordings. One recording contained a taped interview consisting of 50 difficult questions and the replies to them—with 46 questions answered correctly, indicating a very bright person; another recording featured the same person answering 15 of the same questions correctly—quite average. A separate tape was created which included the noises of a cup of coffee being spilled. There was much noise and chatter, and the person being interviewed was heard to say, "0h, I've spilled coffee all over my new suit."

Each group included 12 males listened to one of four tapes, created by inserting the coffee-spilling incident near the end of each of the two interviews to create four groups: Group I heard the person being interviewed answer 15 of the questions correctly; no coffee was spilled. Group II heard the same tape; coffee was spilled. Group lII listened to the tape of the bright person being interviewed; no coffee was spilled; Group IV heard the bright-person interview and coffee was spilled. During the taped interview, the person being interviewed also made statements about his high school activities. These reinforced the impression of his being quite average in high school on the average-performance tape or the aside comments suggested he was of superior ability (if he'd answered 46 of 50 questions).

After listening to the interview, each participant was individually interviewed by one of two interviewers, who did not know which tape the participant had heard. The participants were questioned about their feelings concerning the ability and attractiveness of the person interviewed on the tape. The person who gave the superior performances (Groups III and IV) was rated significantly more intelligent by his listeners than was the person giving the average performance rated by Groups I and II. But what is more interesting, the person of superior ability was rated to be more attractive if he had made the social blunder of spilling his coffee, while the person of average ability was rated significantly less attractive by virtue of spilling his coffee!

Thus, in this particular experiment the hypothesis was confirmed. A person thought to be of high ability was seen as more attractive, and thus more approachable, when his behavior made him more human.

Have you read the Feature? All right, now think about it for a moment. Why couldn't we just let one group listen to the smart person, rate him, and then collect the ratings again after the smart person had committed the blunder of spilling the coffee? Because the group has been permanently altered by listening to the first tape. They've already formed an impression and then committed themselves to paper. They are no longer naive -meaning that they now "know" something about the person. Thus they do not have the necessary qualifications any more to serve as the experimental group in our experiment. This important change is the main reason why the experiment described in the Feature was conducted using four separate groups. It's called a between-groups design. Experimental and control group ratings are being compared here between groups.

Within Groups Comparison

There are situations in designing an experiment when a between-groups comparison will not work. Have you ever said a word or a phrase over and over again out loud? Remember as a child how the phrase "toy boat" became very hard to pronounce if you repeated it out loud as rapidly as possible? Soon you couldn't pronounce it correctly. One thing that might have happened to you when you said the word over and over again is that the word began to lose some of its meaningfulness for you.

When what is being measured is the loss in meaningfulness of individual words, the same measures can be made over and over again within the same person -- because your repetition affects the meaningfulness of only one word at a time. You could rate the meaningfulness of a series of words or phrases such as "toy boat" or "putrid." Then say "toy boat" many times and rate it again. You could then say "putrid" many times and rate it meaningfulness, and so forth. The loss of meaningfulness is not influenced by the fact that the same person is saying and rating many words. Here, then, a within-groups experimental procedure is used with the same participants saying and rating each of the words.

Throughout this book there are examples of both types of research, some within-groups and some between-groups designs. The effects of the experiment itself on the participant are what force us to use one or the other procedure. The Think About It suggests the importance of control groups in enriching our ability to interpret data from experiments. Regardless of the procedure selected, an experimenter must conduct his or her research consistent with the guidelines for ethical research.

Think About It

The question: Why is it not enough for an advertiser to say "Brushing with Ripsnort yields 15 percent fewer cavities"?

The answer: Fewer than what? Because no control-group data has been reported, it's possible that brushing with something else would yield 28 percent fewer cavities. To be expressed properly, any statistic should include enough information to allow it to be properly analyzed. Aren't you better informed if you're told, "Brushing with Ripsnort yields 16 percent fewer cavities than no brushing at all," or "Brushing with Ripsnort with Cramdec yields 13 percent fewer cavities than brushing with Ripsnort alone"?

Sources of Error

We all make errors. Sometimes they're deliberate -- which raises ethical concerns -- and sometimes they're totally unintentional. There are three types of error that psychologists must always be on guard against in running experiments.

Psychologists sometimes forget that humans are human. Humans are smart; they're (almost) always thinking; and they're often trying to please each other. So when an experiment is conducted, the people who serve as participants are going to be guessing about everything that happens to them. That includes the instructions, the room, the task, and even the smile on the researcher's face. They may develop hypotheses (guesses) about what's being studied that have absolutely nothing to do with the real purpose of the study. Yet those guesses may cause the participant to behave other than naturally. Such changes in behavior may influence the conclusions that the researcher is trying to reach.

Demand characteristics are those aspects of an experiment that influence a participant's responses unintentionally. Every time you get an injection of medication to cure an illness, the drug itself is probably causing changes in your body. But what else is involved? Calling for the appointment, driving to the doctor's office, waiting, being examined, rolling up your sleeve, and feeling the pin-prick of pain are also part of the total "treatment." They're irrelevant to the effects of the drug; they're demand characteristics. They must be controlled in an experiment, to be sure that they impact both the experimental and the control group similarly.

To control demand characteristics, researchers use a single-blind technique. Here the experimenter knows what group a participant is in, but the participant himself or herself does not.

Experimenter bias is another problem. We humans smile when we're pleased. And it pleases researchers to have their theories confirmed in the laboratory. So it would be natural for a researcher to smile when participants perform "correctly" and to frown when they don't. Since participants notice how the experimenter reacts, they may try to guide their responses in order to receive that smile. In short, unless care is taken to make sure every participant is treated exactly the same in an experiment, it's possible for experimenter biases to influence the results of an experiment unintentionally.

Experimenters need to be alert and honest in the collection and analysis of data. A scientist might communicate his or her biases to research assistants, so care must be taken to prevent this. For instance, in the Language and Communication Chapter you learn that as little as seven percent of an emotional message may be communicated by the verbal portion of that message. The other 93 percent may be communicated nonverbally by a tone of voice, a frown or smile, or even a delay in responding.

Such communication errors are almost always unintentional in an experimental; one estimate is that about one percent of data may be incorrect. Thus, one of the best controls is simply to make everyone aware of the errors. In addition, in some experiments it's possible to use a double-blind technique. Here neither the experimenter nor the participant knows the experimental conditions to which a participant has been assigned. The experiment we describe in Feature 2 involves a double-blind procedure. Each participant knew only about the one tape to which he had listened. Each interviewer did not know which tape the participant had heard. Since no one participating in the interview knew enough about the experiment to bias the results, it was a double-blind.

FEATURE 2

(Ab)Using Statistics

Here are some examples of statistics being used to confuse, not to communicate. Can you find the flaws?

First, "With proper treatment, a cold can be cured in seven days, but left to itself a cold will hang on for a week!" All that's happened here is a change of terms, not a change of message!

Second, "80 percent of all fatal accidents occur within five miles of home." But, then, how many trips have you ever taken where the first and last five miles weren't within five miles of home? Here's another view of the same situation: How much of your driving is done within five miles of home?

Third, and this one is a bit tricky, "If I buy an article every morning for $.99 and sell it for $1.00 every afternoon, then I make only one percent on total daily sales." However, we should point out that the same merchant is making 365 percent annually on the money he or she has invested in the business (by repeating the same purchase-and-resell cycle each day)!

Fourth, read carefully: "For every VW sold in Italy, three Fiats are sold in Germany!" Does that mean Fiat is three times as popular? Where? In Germany? In Italy? The important missing fact here is the absolute number of cars that Germans and Italians buy. If Germans buy more than three times as many cars as Italians, then the Fiat is actually a less popular car in Germany than the Volkswagen is in Italy! Without all the data, we can't make a decision.

Finally, "The more races Mario Andretti runs on Pennsyl Motor Oil, the more he wins!" Does that mean we should buy Pennsyl Motor Oil? Another equally true statement would be, "The fewer races Andretti runs on any motor oil, the fewer he wins!" Sometimes statisticians simply seem to be assuring that we have a firm grasp of the obvious!

Any time a psychologist wants to measure someone's behavior, he or she must do so without interfering with the behavior being measured. In the Learning Chapter we note that one of the measures of the strength of a learned response is how long it takes to extinguish the response. The problem, of course, is obvious: By the time we know how well learned the response is, we've destroyed it! Naturalistic observation offers possibilities for watching and recording natural behavior without influencing it at all.

Psychologists must always be on guard to assure that their methods don't influence the behavior being measured, and to assure this they've developed a number of protective strategies. We've already reviewed two of them -- single- and double-blind techniques of data collection. Deception is a third way; under certain conditions participants may not be informed until after the experiment about the true purposes of the experiment. A fourth way involves assessing a participant's awareness. Asking the participant a series of questions after the experiment in order to learn whether he or she has detected what was being done does this. In this way adjustments can be made if they are needed. In addition, participants who knew (or guessed correctly) what was being measured can be compared with those who remained naive. This will reveal whether or not knowledge of being measured influenced anyone's behavior.

Ethics in Research

What happens to a participant? There are some very strict ethical guidelines that are endorsed by psychologists through the American Psychological Association. Table 1 lists the "rights" of anyone who is asked to participate in an experiment. One of the most important rights is that of informed consent. Few experiments would be destroyed by telling the participants beforehand what is going to be done to them.

But what if we're studying honesty -- for instance, as it is demonstrated (or not!) by a participant's response to finding money on a sidewalk? Suppose we're interested in what you'd do when (1) alone, (2) with a friend, or (3) as part of a large crowd of people. Clearly, we couldn't tell you beforehand that we were studying your honesty! If we'd told you, then of course you'd try to return the money. But what if you didn't know you were being watched?

So there are situations where the experiment could be conducted before your permission to use the results would be asked. This, however, is done very rarely! All of us have a right to know the benefits and dangers that might result from any experiment before deciding whether or not to take part in it.

Table 1

The rights of research subjects

SUBJECT HAS A RIGHT TO:	**WHICH MEANS:**
Give informed consent	The decision to participate should be based on all relevant information that can be released before the experiment.
No pressure to participate	Psychologists must be sure their subjects are not forced to participate.
An honorable "contract"	Both psychologist and subject should perform as if morally and ethically bound to a clearly stated series of mutual obligations upon agreeing to participate.
Freedom from physical and mental stress	Any possible psychological or physical harm should be fully explained before the start of an experiment to assure the subject understands any risks.
Complete debriefing and follow-up	Once the experiment is completed, the subject should be fully informed of all relevant aspects. The psychologist assumes a responsibility to correct any possible physical or mental damage created by the experiment.
Anonymity and confidentiality	The subject's individual data will be held in secret and never reported in a form allowing identification of a specific subject's performance without prior permission from the subject.

(Adapted from APA, 1973)

In addition, some experiments -- several are discussed in the Psychology: Its Nature and Nurture Chapter involve animals. Again, there are very strict ethical rules describing the care with which research animals must be treated. These rules provide specific guidance as to what can and cannot be done.

How are these human and animal-care decisions reached? This question must be asked: Are the benefits to society as a whole greater than any potential (or actual) harm that might be done? That is, do the benefits of the knowledge gained outweigh the costs of gaining the knowledge sought? If the answer is yes, then the research should be done. However, reaching that decision is not always easy. And the responsibility for making it is an important factor in designing any experiment.

The Importance of Statistics

The "What's the Answer?" section shows you two ways in which statistics are used in our everyday world. We're surrounded by statistics. There seems to be no way to avoid them. We're tempted to say that past studies have shown average North Americans to be 21 percent confident of their knowledge of statistics, 38 percent worried about it, 16 percent uncertain, and 25 percent eager to forget the whole thing! Do statistics scare you? They shouldn't. The fear of statistics that many people have is based on a lack of understanding. Statistics themselves are not difficult to understand. The only thing you ought to be worried about is how people use statistics to try to convince you of something, because statisticians can "lie."

There are basic principles by which psychologists design experiments, specify their independent and dependent variables, and calculate statistics. The processes by which they create their control and experimental groups and how they collect their data, or results usually lead to numbers. Numbers lead to confusion. Confusion leads to mistakes. To prevent mistakes we've got to reduce confusion. To reduce confusion we've got to reduce the number of numbers. To reduce the number of numbers we've got to have a system. That system is statistics! Sometimes they can be misused, but most often that happens only because they are misunderstood. Statistics should be considered an aid to, not a substitute for, common sense. Read Feature 17.2 for some examples of where common sense should suggest the next question that needs to be asked. A statistic is a numerical fact or datum. Statistics has to do with the collection, arrangement, and use of those numerical facts or data.

In doing an experiment we are usually trying to reach conclusions about a much larger group of people than the ones we are actually studying. In the Psychology: Its Nature and Nurture Chapter we try to decide whether it was true that blondes have more fun. Obviously we can't study all blondes. Instead we draw a sample. Doing so immediately involves us in some very basic descriptive statistics.

The population in psychological research is any group of people, animals, concepts, or events all of which are alike in (at least) one respect. In our study described in the Psychology: Its Nature and Nurture Chapter the population is all blondes. But we might study all automobile accidents, or all salespeople, or all males weighing less than 130 pounds. All of these are populations. We almost never know something for sure about a population, because we can never study all members of a population -- it's usually too large a group.

Because of this problem we study a sample, which is any subgroup drawn by a nonbiased method from a population. A sample is always smaller than the population from which it is drawn -- else we'd have no need for our sample. Obviously, the method by which we draw our sample from the population is very important. The usual method is to draw a random sample. That's a sample for which every member of a population has an equal chance of being chosen.

An estimate is any characteristic of a sample. In selecting your sample of blondes you might be interested in the age of your sample. That average age would be an estimate. Clearly, the "fun" reported by members of our sample would be another estimate.

A parameter is any characteristic of a population. What is the average age of all blondes in North America? We have no way of knowing for sure, but we can estimate it by calculating that age for our sample of blondes. Thus, a sample average is an estimate of the corresponding population parameter. We hope our sample has been drawn at random -- that is, without bias--from the population. If it has, then our estimates should also be true for the parameters of the population from which that sample was drawn. Using sample statistics to draw conclusions about populations frequently involves the use of inferential statistics, though this process may also be aided through the use of correlations.

Descriptive Statistics

To summarize and analyze any data, three types of statistics can be used. The simplest are the descriptive statistics. Inferential statistics are used for making decisions, and they can be much more complex than descriptive statistics. Correlational statistics are descriptive, but they can also be used to make certain kinds of decisions. The primary use for descriptive statistics is to collapse large amounts of data into a few numbers. These numbers will convey an impression about -- a description of -- the whole group of numbers.

For example, if you scored 45 on a test in your psychology class, would you be happy? If you've been paying attention so far, then you should have mumbled to yourself that we haven't told you enough yet. A frequency distribution will list all the possible scores (the distribution), and show how many people gained each possible score (the frequency of that score). Developing a frequency distribution for a set of data is the simplest statistical analysis we can perform.

One of the most common forms of such distributions is called a normal distribution, as seen in the Figure. However, in addition to how you placed in the distribution, you'd also want to know some other information. The first piece of information most people would want to know is the "average" score -- also called the central tendency -- for the class. That's not an easy question to answer -- there are several "averages." The average may be a point around which scores are grouped, or the most typical score, or sometimes the most frequent one. Different distributions yield different averages.

The mode is the simplest measure of central tendency. It's the score that occurs most often. To calculate it, develop a frequency distribution, and then report the most common score. When would you use it? When the data are grouped mainly at one end of the distribution of scores. For example, giving the mode would be the most revealing way to show the salary structure of a company. This would show the salary that most people earn.

The median is a slightly more complex measure of central tendency. It is the score that is the midpoint of the distribution; half the scores are above it and half below. How do you find the median? First, count all the scores and divide in half to identify the middle score (or scores). Second, you would rank-order your data in ascending or descending order. Then start at the top or bottom and count halfway. With an odd number of scores, the middle value is your median; with an even number, the median is halfway between the two middle scores.

The mean is the "average" you're probably most used to reporting. To find the mean you just add up all the scores and divide by the number of scores. The formula for this calculation is (SX)/N. The capital Greek letter S , or sigma, is a statistical operator. It tells the person manipulating data to Sum or add all of the X values. S(X+M) would indicate to add up all of the quantities (X+M). The mean is the most sensitive measure of "average," because every score affects the mean.

When should we use the mean, the median, or the mode? The Think About It suggests one basis for deciding which to report.

Think About It

The question: We reported to you a conversation in which two members of the Canadian Parliament were said to be arguing about the average annual income for households in Ontario Province. The Liberal MP reported the average was less than $55,000 (CD); the Conservative MP indicated it was almost $63,000 (CD). Who was right, and why?

The answer: Both were. Remember, the Liberal Party was not in power at the time. Thus a Liberal MP would wish to select "averages" that would not make Canadians seem well off. The Conservative Party was in power, so its MPs would select figures suggesting great progress by Canadian citizens. Since they were both working with the same set of data, the Liberal MP reported median income, while the Conservative MP reported mean income.

Any distribution of a nation's annual income by household is not normally distributed. Most citizens make adequate salaries, but a few make very high salaries—meaning the mean is most inflated. Thus, the median will be somewhat lower, but the mean will be higher because of the few extremely positive earnings figures. Without identifying their "averages" each was correct—but only partly so.

If your distribution is normal, then the mode, median, and mean are identical -- right at the mid-point of your data. As the scores shift so that more and more scores fall at one end or the other, the mean is most affected. Under conditions of severe skew, the mode is least affected while the median typically falls between the mode and mean, usually closer to the mode

Inferential Statistics

You've just gotten a new car during Christmas vacation. Just as you drive up in front of your best friend's house to show it off, your friend backs out of the driveway in her family's brand new car. The two of you are planning to go somewhere together, but whose car are you going to use? Naturally, an argument follows. The two of you finally decide to flip a coin. It's an arbitrary but fair decision, since it leaves the "decision" up to the laws of chance. Your friend loses, so she suggests, "Two out of three?" She wins the next flip, but you win the third one. As she asks, "Three out of five?" we hope the impossibility of it is beginning to occur to you. How many times will you get a "head" if you flip a coin 100 times? On average, fifty -- we hope. So we always resolve arguments by a previously negotiated odd number of random events.

What if you flipped coins with her 100 times and you won 60? Would you worry about the coin? What if you won 70 times -- always guessing "heads"? How about 80? When would you start to worry that your coin was biased, or that something other than chance was causing the results? In resolving that question you will make an inferential decision using an elementary form of statistics. If extreme things happen too often, then we assume it was not chance that was operating. Something -- probably your independent variable in an experiment--was causing the data to behave other than randomly. And your coin? Table 2 shows how likely it is that you would flip 10 coins and by chance end up with various ratios of heads to tails.

Table 2

The Laws of Chance in Tossing 100 Coins

If you tossed 1 coin in the air 100 times, then you would find at least:

______ heads and	_____ tails about	______ times
10	0	0.1
9	1	1.0
8	2	4.4
7	3	11.7
6	4	20.5
5	5	24.6
4	6	20.5
3	7	11.7
2	8	4.4
1	9	1.0
0	10	0.1

Descriptive statistics are used to describe data or collapse numerous data to a represent few. Correlations are used to describe the degree of relationship existing between two sets of data, though they can be used to aid decision-making in designing an experiment. However, inferential statistics are specifically designed and used to help scientists make decisions. Could the differences that occur between an experimental group's performance and that of the properly matched control group occur by chance? The less the likelihood that a chance difference could have occurred, the more the likelihood that your independent variable is having an effect on your dependent variable. A significant difference is simply a statement that the difference between two (or more) groups' performances is too great to be due only to chance factors -- that is, your independent variable had a (statistically) significant effect.

When measured, many human skills, abilities, and features turn out to be normally distributed. Plotted as seen in the Figure, few people have a marked deficit of an individual skill -- such as high jumping -- but equally few have marked skills in the sport. The vast majority of us fall somewhere near the middle. The normal distribution occurs so often in psychological data that statisticians have studied the distribution itself. Doing so, a number of major consistencies in such distributions have been identified.

As seen in the figure, slightly less than 100% of the data in a normal distribution falls within three standard deviations (SD) of the mean. We can use this fact to isolate the effects of an independent variable. We've invented a new toothpaste ingredient; we want to measure its effectiveness as a deterrent to decay. We set up two groups. Our experimental group of 100 children brushes using Ripsnort with Cramdec for three years. Our control group of 100 children selected by the same method as our experimental group brushes for three years with Ripsnort (without Cramdec). At the conclusion of the experimental trials we count all the cavities in every child's mouth. For each group we calculate the mean number of cavities and the standard deviation of that data. We conduct that entire experiment in order to obtain one number: The mean difference in the average number of cavities in our experimental and control groups. The final question is whether the obtained difference is significant -- does the Cramdec actively retard cavity development more than just brushing with Ripsnort.

So far, everything we've calculated is a descriptive statistic, but these are used to make inferences (reach conclusions) about our findings. Imagine that we conducted this same 200-child experiment a million times. Each experiment yields one mean difference between our two groups' cavities -- a single number. If we plot that number on a distribution of differences between the means, we would achieve data something like that pictured in the Figure. If adding Cramdec has had no effect on cavity-reduction, on what number would our distribution of the differences between the means center? Zero. In the long run across all of our experiments, we would be as likely to have greater cavities in one group as the other, so our distribution of differences would center on zero.

If we find a difference between the average cavities in each group, we can use inferential statistics to estimate the likelihood that the difference we found is due purely to chance factors. We can use the variability of our data in the two groups to estimate the likelihood that we have achieved a significant difference between our two groups. Let's take an extreme example. Suppose we found that every single child in our control group had 4.0 cavities, while every child in our experimental group had 3 cavities, as seen in the upper part of the Figure. At this point if I tell you my own child had four cavities, you know that if he was in the study, he was in the Control group -- he brushed without Cramdec. The lack of variability in our two groups' data gives us great confidence in our findings.

By contrast, suppose we found 4 cavities in each Control Group child, but the range was from 0 to 20 cavities in that group, and the Experimental Group averaged 3 cavities, but also had a range from 0 to 20 among its participants as seen in the lower part of the Figure. Now if I said my child had 4 cavities could you confidently predict the condition in which he participated? No. There are two factors which affect our confidence in an experimental finding: the variability of the data within each group and the size of any difference obtained between groups -- that is, the size of the difference in the means.

Inferential statistics are usually a ratio in which the between-groups differences in the means is the numerator, and the within-groups variability becomes the denominator. We could (theoretically) create a distribution of that ratio just as we could with our difference between the means. Such distributions (which can be calculated theoretically -- a process about which you do not need to worry -- are often normally distributed. Thus, if we find a large difference between the means (Experimental minus Control) and/or very low variability in the data within our groups, our confidence that we have identified an important independent variable increases. When a researcher says he or she has obtained a significant difference (between the means) they are simply saying that the likelihood is very low that a difference as big as they obtained would have occurred simply by change. Most likely, the obtained effects can be attributed to their independent variable. At that point the researcher is ready to start writing up his or her report.

Correlation

A correlation is a very useful statistic -- especially in the early stages of research. It indicates the amount of variation in one variable (such as your weight) that can be predicted from knowledge about another variable (such as your height). A correlation is always reported as a number that can vary from -1.00 to +1.00. The size of the number tells you how strong a correlation exists between two sets of data. A correlation of 0 means that there is no relation. And remember the laws of chance. Just by chance there might appear to be a slight correlation between two sets of data even when no actual relationship exists. Thus, a correlation smaller than + or -.40 is seldom of interest to a psychologist. A correlation of + or -1.00 means there is a perfect correlation -- every change in one variable can be predicted exactly by a change in the other variable. Obviously, we don't find correlations that big very often -- and when we do, we usually call them laws, not correlations!

A correlation also tells us something about the nature of the relationship -- whether positive, negative, or zero (meaning nonsignificant). A positive correlation usually exists between your salary and the amount of income tax you have to pay. As your salary goes up, your taxes also tend to go up. Positive correlations can range from + .01 to +1.00.

A zero, or nonsignificant correlation means simply that it is of no help to you to know anything about one variable if you must predict the other. An example of a zero correlation would be an attempt to use the speed of a truck to try to predict the height of its load. They simply are not significantly related.

A negative correlation means that as one variable increases, the other is actually decreasing. A fall in one is associated with a rise in the other.

By now you've probably been able to guess for yourself: Correlation research is usually passive. This means that the researcher simply watches and records height, speed through a stop sign, or whatever. Such data simply report the extent to which two sets of data are related to each other. Doing experiments, on the other hand, involves actively changing the independent variable to note changes in the dependent variable. Experiments can establish cause-and-effect relationships, especially with the aid of inferential statistics. Correlational studies simply demonstrate or describe degrees of relationship although they can aid decision-making.

USING PSYCHOLOGY: Writing a Research Report

Once you've finished your research you've got this huge stack of data. What next? In the Psychology: Its Nature and Nurture Chapter we say that one part of the total activities of a scientist is communication. In some ways, it's the most vital activity of all. No research can have an impact until it's been published or word of it has spread around.

The key to good research is keeping accurate, detailed, readable records. Without them, a researcher is lost. The body of a report describing completed research contains an introduction, and a methods-, results-, and discussion-sections. However, there are several possible ancillary sections which are just as crucial to accurate communication as those primary components. For instance, a single issue of a research journal may have from three to fifty or more research reports in it! No scientist is interested in all of them. The Title and Abstract of a report attract attention and provide cursory details of the report.

There is general agreement on the form in which reports should be written. Moreover, a lot of help is available to anyone wishing to write a research report. The proper form for writing a report is detailed in the *Publication Manual* of the American Psychological Association. This is the sourcebook for the details of APA-style reports. The APA style is markedly different from that of the Modern Language Association's MLA style -- especially a different style of referencing the work of others, and the APA style places much less emphasis on footnotes and ancillary materials. All psychology papers -- as well as many papers in the physical sciences -- are written APA style.

USING PSYCHOLOGY: Main Sections of a Written Report

Paul Chu, a research physicist at the University of Houston, published his first major paper announcing his work on superconductivity -- involving efficient conduction of electricity -- in the *Physical Review Letters* in 1987. The report unleashed a furious array of scientific research in physics labs all around the world. This was possible because Chu reported all of the pertinent details of his research. Such reports are crucial to fostering scientific research. In addition to stimulating additional research, such reports also make science self-correcting. Attempts to replicate what turn out to be erroneous reports typically lead to published corrections of such errors.

A standard report includes four major sections: An introduction, a methods section, results, and discussion. In the Introduction you state why you conducted the research on which you're reporting. In this section, you summarize (1) what has already been found out about what is being studied, (2) any conflicts in previous research (which may be noted or commented upon), (3) hypotheses (guesses) about what may be occurring, and (4) the specific factors to be studied and how they will be investigated. The goal here is simply to tell the reader what you're studying and why.

In the Method section you may have as many as three or four separate sections. (1) Participants will describe the population you were studying and the sample you drew from that population. (2) Design will describe how many groups were involved. You will identify the independent and dependent variables, as well as the relationship of the control group to the experimental group(s). (3) Apparatus will include a description of any unusual equipment used for the experiment. Everyone knows what a slide projector is, so naming it is enough. However, if you're using a "modified one-way platform avoidance chamber," you'll want to describe it in more detail. Finally, (4), procedure will explain what the participants were asked to do, and often will include the various activities of the experimenter. The instructions you gave to your participants may be included; if not, the reader should at least know what each participant was told.

One effective way to check on how good your method section is, is to give it to a friend. Let him or her read it, and then ask your friend to explain it to you -- with no hints from you! If he or she can't accurately describe your experiment, then you should add to or change the information you've included (or find a smarter friend)!

The third is the Results section in which you report what you actually found. Don't interpret it in this section, just report your findings. Here you would include tables of numbers such as averages and measures of variation. In addition, any graphs of the participants' behavior would also be included. In this section you might also add any other descriptive, inferential, or correlational statistics you have calculated. Also report the values showing the probability your differences (between experimental and control groups) could have occurred just by chance. You do not usually show the calculations themselves, just the results.

The last major section is your Discussion. Bad experiments are not usually reported -- more often they're just called "pilot studies!" That's an indirect way of saying the discussion section should not simply be a list of apologies.

Here you analyze what you've found in your experiment. You talk about how your results confirm the hypotheses you developed in the introduction, or how they fail to do so. Aside from these major sections of a standard report, it is also important to include any necessary ancillary sections of support materials such as a required Abstract and References--necessary if you cite the work of others.

USING PSYCHOLOGY: Ancillaries of a Research Report

Because of the hundreds of research articles published on many widely studied topics, a psychologist is hard-pressed to read everything in his or her field. In addition to the main elements of a report -- including a Methods and a Results section -- there are several ancillary sections that serve a vital purpose. The title is crucial for attracting a reader's attention. The abstract of a report is the first component where you lay out the skeleton of your research -- your hypothesis, the design (method) of your experiment, the very basics of what you found (or, alas, failed to find), and what you conclude. Each fact is given in no more than one sentence. The goal is not to relay all the information, but to mention the essential parts.

After the major sections of properly written research reports there is a section that is totally boring to read -- most people don't -- but crucial to putting the report is its proper context: References. Harry Harlow -- a very prominent research psychologist -- once commented that "plagiarism is stealing from one person without credit; scholarship is stealing from many without credit"! There's a nugget of truth to the statement, one not easily learned by young scientists. Very few original pieces of research are done. There's only been one Freud in the world of psychoanalysis. Others have built on his work. It's the same in most fields of research.

Always include in your references the work of others from which your research is being drawn. It is not uncommon even for short research reports to contain as many as 100 references. You'll notice that hundreds of references are cited at the end of this book. Research studies build on the work of others and may stimulate future work. Thus, references are the mortar holding the brickwork of science together. Particularly the introduction and discussion of your work should typically cite a number of other studies. This identifies the intellectual arena and foundation of your work and allows those who read your work to reach the papers on which your work is based.

A third component among the support materials are the figures and graphs which accompany your work. These should summarize the report's findings -- not simply present them in another form. Occasionally, reports may also include a fourth support component: footnotes or appendices -- which may detail a unique questionnaire used in the study, or a description of some procedure within the experimental protocol.

Review

EXPERIMENTS, CONTROLS, AND ETHICS

1. From where do scientists get their ideas for experiments?
2. Why are control groups necessary in experiments?
3. Compare and contrast "between-group" experiments and "within-group" experiments.
4. What are "demand characteristics," and how do we keep them from influencing the results of experiments?
5. What is "experimenter bias," and how can it be prevented?
6. What kinds of errors can be caused by the processes of measuring behavior?
7. What ethical restrictions limit experimentation?

THE IMPORTANCE OF STATISTICS

1. Define and give an example of each of the following: population, sample, estimate, parameter.
2. Name and distinguish among the three most common types of statistics.
3. What is a frequency distribution?
4. How are the mode, the median, and the mean of a set of data similar and how are they different? When will their numerical value be closest to equal?
5. How do inferential statistics differ from descriptive statistics?
6. Name and describe two types of information given by a correlation.
7. What does it mean when two variables in an experiment are found to be highly correlated?

ACTIVITIES

1. Career Search. Write to the Publications Office, American Psychological Association, 1200 Seventeenth Street, N.W., Washington, D.C. 20036, for a copy of their ethical standards that apply to research conducted with animals. Before

the publication arrives, imagine that your class is a group of research psychologists. From what you've learned in this book, develop your own set of standards. When the APA standards arrive, compare them with yours.

2. To watch the laws of chance and the principles of sampling in operation, find a supply of marbles-red ones, white ones, and blue ones. Place sixty marbles of one color, thirty of another, and ten of the third into a box. Allow a number of your friends to draw one marble at a time without looking into the box. After you have recorded which color marble each one drew, have them replace the marble in the box. Shake the box and allow them to draw again. Do this ten times and then ask each of your friends to guess how many marbles of what color there are in the box. Keep track of all the data you collect, and after ten or more friends have drawn their marbles (and replaced them), see what your totals are for each color of all the sets of ten marbles. How accurate would your own guess be, based on the repeated sets of ten that were drawn? Identify the population, the sample, the estimate, and the parameter in this experiment. And ... don't lose your marbles!

3. Run an experiment to develop a frequency distribution. Does the main entrance to your school have at least four doors? If so, get some friends to help and then observe for fifteen minutes one morning. Record how many people enter each door and (on a separate distribution) how many leave. Or you could ask as many people as possible to choose their favorite number between zero and nine, inclusive. Is there a favorite?

4. If you ran the experiment in Activity 3 in the morning, as suggested, try repeating it for the same length of time in the afternoon. Compare the two frequency distributions. Is there a difference in the number of people entering as opposed to leaving? Now, look only at those who enter in the morning, and those who leave in the afternoon: Is there any difference in the frequency with which they choose to use each door? What can you conclude from this?

INTERESTED IN MORE?

AGNEW, N. M. & PYKE, S. (1976). *The Science Game: An Introduction to Research in the Behavioral Sciences, 2nd ed*. Prentice-Hall. Discusses how to design, conduct, and write a research report on an experiment in psychology.

BRADLEY, J. I. & McCLELLAND, J. N. (1978). *Basic Statistical Concepts: A Self-Instructional Text, 2nd ed*. Scott, Foresman. A programmed-learning text introducing descriptive, inferential, and correlational statistics.

CAMPBELL, S. K. (1974). *Flaws and Fallacies in Statistical Thinking*. Prentice-Hall. A painless and funny look at how statistics can be used to misrepresent the data they are summarizing. Includes examples from research and advertising.

DOHERTY, M. E. & SREMBERG, X. M. (1978). *Asking Questions About Behavior: An Introduction to What Psychologists Do, 2nd ed*. Scott, Foresman. Shows how to develop a question and design a simple experiment to answer it. Emphasizes the study of stress.

DOWNIE, N. M. & HEATH, R. W. (1983). *Basic Statistical Methods, 5th ed*. Harper & Row. A good means of mastering introductory statistics independently: includes 3 to 16 problems per chapter, with answers listed in the appendix.

HUFF, D. & GElS, I. (1954). *How to Lie With Statistics*. W.W. Norton & Co., Inc. An old book but still an excellent source of funny examples of the use and misuse of statistics. Nine out of ten people like it.

LOVEJOY, E. P. (1975). *Statistics for Math Haters*. Harper & Row. Approaches the introduction of statistics as a logical exercise, rather than a series of mathematical techniques. Not light reading, but interesting.

MOSTELLAR, F., et al. (1973). Statistics by Example. Addison-Wesley. Designed for high-school students, this four-volume series covers exploring data, designing patterns, finding models, and weighing chances. Uses real data to show the value of statistics in clarification.

REED, J. G. & BAXTER, P M. (1983). *Library Use: A Handbook for Psychology*. American Psychological Association. An excellent guide to the use of library resources to find published information on psychology. Written by a psychologist and a librarian, it conducts the reader through all major reference resources for psychological literature.

SARBIN, T. & COE, W. (1969). *The Student Psychologists' Handbook: A Guide to Sources*. Schenkman. Shows how to locate

sources, use a library, and organize, write, and reference a research paper.

SPENCE, J. T., et al. (1983). *Elementary Statistics, 4th ed*. Prentice-Hall. A well-written introduction to standard statistical techniques. This classic text is accompanied by a workbook that helps students determine the type of statistical test most appropriate.

WOOD, C. (1977). *Fundamentals of Psychological Research, 2nd ed*. Little, Brown. Analyzes the major research strategies, with heavy emphasis on statistical analysis.