A statistical generalization has the form:

STATISTICAL GENERALIZATION

X percent of observed As are f. Therefore, probably X percent of all As are f.

In such an argument, the premise or premises contain numerical data that characterize an observed sample of the population. The conclusion is an assertion of the probability that the same characterization will be true of the population.

How do we assess the reliability of an inductive generalization? How do we tell whether it is strong? Considering that an inductive generalization is an inference from some to all, common sense would suggest that the larger the sample, the stronger the generalization. After all, the more "As" found to be "f," the more likely all "As" are "f." There is truth to common sense here, but the truth is below the surface. Consider this example:

Example 9

Of a sample of 60 percent of all Boston teachers, all of whom are primary grade teachers, all are underpaid. Therefore, it is likely that all Boston teachers are underpaid.

The sample, 60 percent of all Boston teachers, is remarkably large for a survey. If size were the definitive factor, then we should consider this a strong inductive generalization. But notice that the sample consists of primary grade teachers only. Despite the fact that the sample is well over half of all teachers, can we be confident that what we learned about the salaries of primary grade teachers will be reflected in the salaries of all teachers? Our suspicion is that the salaries of primary grade teachers may not be representative of all teachers' salaries. This suggests that the size of the sample is less important than how representative it is. As a matter of fact, public opinion polls, such as Gallup polls and Nielsen TV ratings, are typically based on surveys of a very small number of people carefully selected to accurately represent the larger group. So the truth behind the commonsense idea that size determines reliability is really a function of the fact that size *may* increase representativeness. In the assessment of an inductive generalization, the degree to which a sample is representative is what matters most.

The more representative the sample, the stronger the inductive generalization. We have seen that it is much more important that a sample be representative than that it be large relative to the population. But what makes a sample *representative?*

representative *A sample is* representative *of a population to the degree that the target characteristics found in the sample occur with the same frequency or in the same proportion as they occur in the population.*

That defines representativeness, but how do we know when we have it? As you can see from the definition, one way to tell that a sample is representative is to examine the entire population! But if we could do that, we wouldn't bother with inductive generalization. So, in light of the fact that we want to infer rather than examine what may be the case about all members, how can we tell whether we have a sample whose characteristics are typical?

To tell whether a sample is representative, we must already know something about the population and about how the target characteristic is related to other characteristics of members. In Example 9, for instance, we assess the representativeness of that sample by using certain background information we already have. For instance, we know that there are different levels of teaching and that salaries vary accordingly. We also know that salaries are influenced by such factors as seniority and school district. Thus, to be representative in this case, a sample should be composed with those differences taken into account. The fault with Example 9 is that the sample lacks the variety that we know to be present in the larger population.

To consider how a representative sample is selected, let's design our own study. Suppose we want to determine what percentage of students at the university owns automobiles. Suppose further that we believe that a sample consisting of 100 members will be sufficient. The method by which we compose that sample will bear on the reliability of our generalization. For instance, if we interview 100 students at the entrance to the parking lot, we are almost certain to have skewed results. On the other hand, if we interview 100 students near the campus bus stop, again our sample is not likely to be representative. What we need to consider is whether other characteristics of the members might bear on the occurrence of the target characteristic. In those two instances, they do. Clearly, students found near a parking lot are more likely to be car owners. Students near a bus stop are not. Neither group gives us a representative sample. Thus, since where we select members bears on the results, it is best to avoid selecting members who have the characteristic of location in common. Our interviewers should take their data from various locations.

Simple Random Sampling The science of statistics provides carefully devised procedures for designing samples. One of these, *simple random sampling,* is a commonly used way to achieve variety in our sample. If we can reasonably assume that car ownership is randomly distributed throughout the student population, then we may use a simple random sample. One way to do this would be to assign a number to each university student—something the registrar's office is likely to have done already—and to randomly select students with the use of a random number generator. Even simpler, we could pick 100 names out of a hat.

What makes either of these methods a *random* sample? The selection is random in the mathematical sense that each name has an even chance of being selected. The theory of a simple random sampling as a method for designing a representative sample assumes that (1) each possible member has an equal chance of appearing in the sample and (2) the characteristic sought is evenly distributed throughout the population. In our example, we can use simple random sampling because we have assumed that the characteristic in question, car ownership, is likely to be distributed evenly throughout the population. If that is an acceptable assumption, then the generalization we infer about all the students has a high likelihood of being accurate.

Stratified Random Sampling A more sophisticated method of sampling is appropriate when we know that the population consists of different groups and the groupings bear differently on the presence of the characteristic we are interested in. The generalization in Example 7 is more appropriately handled using this method, *stratified random sampling*. We know that teachers in Boston are not a homogenous group but comprise instead several overlapping subgroups, or *strata*. Therefore, a more representative sample will draw from those groups in proportion as they occur in the population. Suppose that 38 percent are primary grade teachers, 27 percent are secondary grade teachers, 19 percent are two-year college teachers, and so on. Suppose further that we factor in representative proportions of teachers at different levels of seniority for each stratum. The final sample will consist of randomly selected members from each substratum and, thus, each stratum. This sample, now quite complex, is more likely to reflect the characteristics we can expect to find throughout the population.

> random sample *The random sample is a sample selected by a method that gives each element in the population an equal chance of being selected. The idea is that if each element has an equal chance to appear in the sample, then whatever characteristics are typically distributed throughout the group have the same probability of occurring in the sample.*

As the two sampling methods above show, assessing representativeness involves bringing background information to the argument. The more we already know about the population, the better we can assess the sample. If we know the population is diverse and we know that those differences bear on the presence of the target characteristic, then we know the sample must reflect that relevant diversity to be representative. Suppose, for example, we are interested in estimating the percentage of good eggs produced at an egg ranch. Suppose also that egg size is related to quality. Then, the sample should contain a variety of sizes. On the other hand, those differences that make no difference can be ignored. For example, if the size of a marble makes no difference in its color and we are estimating the percentage of red marbles, a representative sample need not include variety of sizes.

To summarize: To assess the representativeness of a sample, we need to know (1) what different characteristics occur in the population and (2) whether those characteristics are relevant to the occurrence of the target characteristic. In general we can say that the more relevant diversity in the sample, the more representative it is. Interestingly enough, it also follows that the more relevant diversity in the population, the larger the sample will need to be to reflect that diversity. Thus, size is a function of the requirement for representativeness.

If strong inductive generalizations are ones that are based on representative samples, then weak ones are those that are not. A sample that is unrepresentative is called a *biased sample*. In statistics the concept of sample bias is mathematically defined. For our purposes a biased sample occurs when the sample fails to reflect relevant differences in the population either because it is too small or because it is not proportionately composed. One of the virtues of random sampling is that it avoids the likelihood of a biased sample. Nevertheless, sometimes an argument purportedly based on a random sample has members too much alike to be randomly selected. When a sample is biased, the resulting generalization is unreliable, and the argument is said to commit the *fallacy of hasty generalization* (see Section 9.11).

Other telling flaws in an inductive generalization have to do with the reliability of the premises themselves. Consider asking of the premises: How is the target characteristic defined? How is the information obtained? In Example 7, the premise reports that so many teachers are underpaid. What does "being underpaid" mean here? Is it the subject's perception of his or her salary? Is it a measurement of the subject's salary relative to the cost of living, to other comparable jobs, or to teachers elsewhere? If teachers are interviewed, as the example suggests, what questions are they asked? Consider the difference between being asked "How does your salary compare with that of other comparable jobs?" and "Are you underpaid?" The way in which a question is framed can affect the responses; thus, the way the information is obtained can introduce what is called *interviewer bias* into the results.

Three reasons why size is not what matters most!

1. If sample size is the measure of strength of an inductive generalization, then equally sized surveys would have the same degree of strength. However, this is not true, as the following arguments show.

   Example A

   Mr. Smith, a mathematician, is cold and unemotional.

   Therefore, all mathematicians are cold and unemotional.

   Example B

   This raccoon is rabid.

   Therefore, all raccoons are rabid.

   Example C

   Mr. Spock, a Vulcan, has a two-chambered heart.

   Therefore, all Vulcans have a two-chambered heart.

   All three arguments are based on samples with only one member. Are they equal in strength? No. Argument C is stronger than A and B. Why? The answer has to do with the nature of the target characteristics and what we know about how they occur. For example, emotional qualities are highly variable within a population. Some people are unemotional; some are not. Similarly, rabies is highly variable. It is not a defining feature of mammals but tends to occur in special circumstances. If we know that a feature is variable within a population, then a sample size of one case is not good enough evidence. On the other hand, if a feature is invariable, such as anatomical structure, then a sample of one may be good evidence. One Vulcan with a two-chambered heart is a good indicator of what other Vulcans are like because heart structures do not vary in species. Therefore, *equally sized surveys are not equal in strength. Size is not what matters.*

2. A large sample, constructed in a biased manner, is not good evidence. Imagine a study to determine what percentage of students out of a student body of 2,500

is married. In sample S1, 1,200 students were interviewed in the college dormitories. In sample S2, 400 students, randomly selected by ID number, were surveyed by phone. Sample S2 is smaller but more reliable because Sample S1 is biased. S1 is not representative of the student body, since not every student has an equal chance of appearing in the sample. Therefore, it is not true that of two generalizations, the one with the larger sample is the stronger. Again, *size is less important than how the sample is constructed.*

3. Providing survey information is big business. Large samples are not only expensive to conduct but unnecessary. Very small surveys may provide very good evidence if they are constructed carefully.

Summary: Inductive Generalization

1. Inductive generalizations are arguments concluding that something is the case about all or many things on the basis of what is observed about some of them. Presupposed in every inductive generalization is the idea that what we observe in the sample is likely to be true of all members of the group.
2. An inductive generalization consists of: (1) premises describing a *sample* (2) as having a *target characteristic* as reason for (3) a conclusion that all or some percentage of the *population* has that target characteristic.
3. The strength of an inductive generalization is a function of the representativeness of the sample. The more representative the sample, the stronger the argument.
4. A sample is *representative* of a population to the degree that the target characteristics found in the sample occur with the same frequency or in the same proportion as they occur in the population.
5. To assess the representativeness of a sample, we need to know (1) what different characteristics occur in the population and (2) whether those characteristics are relevant to the occurrence of the target characteristic. In general we can say that the more relevant diversity in the sample, the more representative it is.
6. In general, a random sample is a sample selected by a method that gives each member in the population an equal chance of being selected.

   A simple random sample is a sample selected by a method that gives each member in the population an equal chance of being selected without regard to differences. The assumption is that if each member has an equal chance to appear in the sample, then whatever characteristics are typically distributed throughout the population have the same probability of occurring in the sample.

   A stratified random sample is a sample consisting of subgroups, or strata, in proportion as they occur in the population, with each member randomly selected. The assumption is that the occurrence of the target characteristic in the population is related to the occurrence of other characteristics. Strata are differentiated by relevant differences within the population.
7. A sample that is unrepresentative is called a biased sample. An inductive generalization based on a biased sample commits the fallacy of hasty generalization.
8. The premises describing what is learned about a sample may be unreliable if the target characteristic is not clearly defined or if the method of obtaining information adversely influences the results. Either is a case of interviewer bias.

that: The men who passed the disease to the women often do not get treated. One study that looked at 60,000 cases of P.I.D. found that relatively few of the women's partners were treated. Because some of the women had more than one sexual partner at risk, the number of men treated should have been more than 60,000. In fact, only 29,000 were treated.

Answer

1. A study of 60,000 cases of P.I.D. in women showed that only 29,000 men were treated.

2. P.I.D. is a sexually transmitted disease.

---

3. Therefore, men who pass the disease to women often do not get treatment.

a. *Sample:* 60,000 cases of P.I.D.

b. *Population:* all cases of men and women with P.I.D.

c. *Target characteristic:* evidence of treatment for P.I.D.

d. It is not clear whether the sample is representative. Although the sample is large, the report does not tell us how the sample was constructed. Presumably, the data reflect reports of the disease made to the ACOG.

e. If we assume that most reports of the disease are reported to the ACOG, then we may conclude that the sample is representative. Therefore, the argument is a strong inductive generalization.

1. In 1991 a study of a nursery ward for newborns in Chicago showed that nurses followed appropriate hand-washing guidelines about half the time; doctors in the ward followed the guidelines half as frequently as that. Apparently the effectiveness of antibiotics in U.S. hospitals has caused generations of hospital staff to rely on antibiotics rather than on hygiene. (Paul E. Ewald, *Plague Time,* 2000, p. 19)

2. In trying to decide how to spend an additional $8,000 in entertainment funds, the UConn Student Council took a poll of students entering the campus center one Thursday morning when, by coincidence, the UConn Birders Society meets. Sixty-eight percent of those polled favored donating the money to the American Birding Association Wildlife Refuge. The rest were evenly divided among "more movies," "hiring a magician," and "the Bread and Puppet Theatre." On the basis of the sample the council concluded that the majority of UConn students would want to donate the money to the American Birding Association.

3. Don has skied the same terrain weekly for the past five ski seasons. He always sees at least one person skiing with reckless abandon. "You can expect at least one reckless person every time you come here," he says.

4. A survey by the American Academy of Actuaries reports that 72 percent of pension fund actuaries polled predict that half the baby boomers won't have the wherewithal to retire at age 65. The number of actuaries polled was 326; the number of registered actuaries in the nation is 7,854. The reasonable conclusion to draw is that well over half of all actuaries would agree that half of the nation's baby boomers will not be able to retire at age 65.

5. A recent study at a large teaching hospital involving 82 physicians and 75 patients found that there were 154 cases of resuscitation. Although 86 percent of

the patients who received resuscitation were considered competent to make medical decisions, only 19 percent were asked for their consent prior to resuscitation being administered. From this study it is reasonable to conclude that the practice at most hospitals is to administer resuscitation without first discussing it with patients.

6. According to an essay by Mary McGrory, the American Bar Association aims to have its 129,000 law students throughout the country contribute 50 hours of public service before graduation. The ABA bases this recommendation on a report that 54 percent of law students at 100 out of 175 law schools voted in favor of mandatory public service.

7. Tim has done a nonscientific survey of a two-square-mile tract of woodlands near his home in order to estimate this coming spring's population of gypsy moth caterpillars. Following the same path through the woods at approximately the same time of year, Tim counted the number of egg cases on tree trunks visible from the path. The first count totaled 1,090 egg cases; the second year's count totaled 650. Tim concluded that there will be 40 percent fewer gypsy moth caterpillars feeding on the trees this coming spring because he observed 40 percent fewer egg cases.

8. The National Medical Care Expenditure Survey of 1977 conducted a survey, consisting of six household interviews, of over 40,000 individuals over an 18-month period during 1977 and 1978. Among their findings were that approximately 18 million Americans are without health insurance the entire year, and as many as 34 million may be uninsured for some period of time during the year.

9. Macro Market Research of Burlington, Vermont, conducted a phone survey of 508 Vermonters representing 1 percent of the total state population of approximately 500,000 people. Callers were selected through random digit dialing. The number of calls within each of the state's fourteen counties was proportional to the counties' population and distributed geographically according to population. Callers were given the names of candidates for election and asked which they would be inclined to vote for if the election were held today. For governor, the survey showed that 49 percent were likely to vote for Richard Snelling, 29 percent for Peter Welch, and 21 percent were undecided. The margin of error is plus or minus 4.5 percent with a confidence level of 95 percent. That means that 95 times out of 100 this survey would produce results within 4.5 percent of these findings. Based on these statistics it is nearly certain that Snelling will be the state's next governor. (*Rutland Herald*)

10. To determine the percentage of Alameda County drivers wearing seat belts, tollbooth operators were instructed to make observations of all drivers going through the toll at the Hill Bridge. A survey of approximately 8,000 drivers over a 24-hour period showed that 37 percent were wearing seat belts. County officials concluded that approximately 37 percent of all local drivers wear seat belts.

11. By randomly selecting names from the phone book, surveyors for Ace Phone Company asked people whether they owned an Ace phone. Out of 450 calls, 14 percent owned an Ace phone, 36 percent did not know, 2 percent hung up, and 48 percent owned another brand. Marketing researcher Victor Kay concluded that 14 percent of all area phone owners have an Ace and that Ace should print its name boldly across the front of each phone it produces. He is convinced, he argued, that a large percentage of those who "did not know" were Ace owners who couldn't find the label.