C—WHY SAMPLE?

Sampling was used in the Gallup poll—even though it involved some uncertainty—because polling the whole population is much too large a task. There are a variety of reasons why sampling is done in general, including:

1. *Limited resources.* Not only in political polls, but in many other cases such as market surveys, neither funds nor time are available to observe the whole population.

   There are many other examples in business. An allegedly more efficient machine for producing valves may be tested before purchasing, for instance. The manager of quality control simply cannot wait around to observe the entire population of valves that this machine will produce. Instead, a sample run is observed, and the decision on the machine is based on an inference from this sample.

2. *Scarcity.* Sometimes only a small sample is available. For example, in heredity-versus-environment controversies, identical twins provide ideal data because they have identical heredity. Yet very few such twins are available.

3. *Destructive testing.* For example, suppose that we wish to know the average life of all the switches produced by a factory. What use would it be to test the whole population of switches until they burn out? Or, for a hemoglobin count, would you go to a doctor who drew *all* your blood?

4. *Sampling may be more accurate.* How can a mere sample be more accurate than looking at the whole population? If the sample is carefully done, its error can be pretty well confined to the confidence allowance given in (1-2). On the other hand, a survey of the whole population might be such a gargantuan task that much worse errors could occur: A large force of inadequately trained personnel, for instance, may make measurement errors that a smaller and better trained force would avoid.

   For example, the U.S. Census Bureau conducts a monthly sample of about 100,000 Americans called the *Current Population Survey* (to determine unemployment, among other things). This is a model of careful work, and in some ways is more accurate than the complete census of all 250,000,000 Americans taken every 10 years.

PROBLEMS

1-1 Project yourself back in time to six recent U.S. presidential elections. In parentheses we give the results of the Gallup pre-election poll of 1500 voters. (As we mentioned already, each sample has about the same accuracy as a simple random sample. And we continue to ignore third parties.)

Year	*Democrat*	*Republican*
1968	Humphrey (50%)	Nixon (50%)
1972	McGovern (38%)	Nixon (62%)
1976	Carter (51%)	Ford (49%)
1980	Carter (48%)	Reagan (52%)
1984	Mondale (41%)	Reagan (59%)
1988	Dukakis (44%)	Bush (56%)

a. In each case, construct a 95% confidence interval for the proportion of Democratic supporters in the whole population.

b. Mark each case where the interval is wrong—that is, fails to include the true proportion π given in the following list of *actual* voting results:

1968	Humphrey	49.7%
1972	McGovern	38.2%
1976	Carter	51.1%
1980	Carter	44.7%
1984	Mondale	40.8%
1988	Dukakis	46.1%

1-2 In order to serve its advertisers better, a radio station specializing in FM music commissioned a market research survey of 500 listeners to determine their preference for classical or popular music. The survey results broken down by age and sex were as follows:

Numbers Who Prefer Classical

	Sex	
Age	*Male*	*Female*
under 25	19	26
25–50	38	34
over 50	48	60

Numbers Who Prefer Popular

	Sex	
Age	*Male*	*Female*
under 25	63	45
25–50	38	33
over 50	44	52

a. For each of the following, calculate the appropriate estimate and then a 95% confidence interval around it:

i. The percentage of young males (under 25) who prefer popular.

ii. The percentage of the young (under 25) who prefer popular.

iii. The percentage of males who prefer popular.

iv. The percentage of females who prefer popular.

v. The percentage of people who prefer popular. How is this answer related to the previous two?

b. What assumption did you make in part **a**?

1-3 Criticize each of the following sampling plans, pointing out some possible biases and suggesting how to reduce them.

a. In order to estimate how many of her constituents support a gun-control bill, a Senator found from her mail that 132 supported it while 429 opposed it.

b. In order to predict the vote on a municipal subsidy to child day-care centers, a survey selected every corner house and asked whoever answered the door which way they intended to vote. Out of 2180 corner homes canvassed between 9 a.m. and 5 p.m., 960 replies were obtained.

c. To estimate the average income of its MBA graduates 10 years later, a university questioned all those who returned to their tenth reunion. Of the 281 graduates, 56 returned to their reunion and 14 were willing to provide information on their income.

1-2 RANDOMIZED EXPERIMENTS: TESTING A HOSPITAL ROUTINE

Thus far we have seen how randomizing ensures a sample against bias. In this section we will see how randomization similarly frees an experiment of bias.

Is the emotional bond of a mother to her infant weakened by the traditional hospital routine (allowing the woman only a glimpse of her newborn and then keeping them apart for about eight hours)? To test this, a group of mothers were provided with a different treatment—*extended contact* with their infants. They were allowed a full hour with their baby just after birth, plus extra contact for the first three afternoons (Klaus and others, 1972).

A—TREATMENT VS. CONTROL GROUPS

Rather than giving the extended contact treatment to all 28 women selected for this study, half were kept in the traditional routine as a *control group* for comparison. The question was: Who should be given the new treatment, and who should be kept in the control group? The new treatment should not be given just to the women who requested it, since these women might inherently be the ones most interested in their children. Then, if the mother-child relationship thrived, how could one tell if it was because of the treatment, or because of this confounding factor—that is, because the women who got the treatment were initially better mothers?

In fact, there are probably many other confounding factors as well. For example, the woman's education, age, or marital status might also influence the bond with the infant. The most effective way to neutralize *all* these confounding factors at once is to randomize.

PROBLEMS

1-4 Underline the correct choice in each pair of brackets.

a. A "randomized controlled experiment" is an experiment where some of the subjects [have the new treatment deliberately withheld, are given random and unknown doses of the new treatment]—the so-called control group. Whether any given subject gets assigned to the treatment or control group is then determined by [chance, the expert judgment of the experimenter, the wishes of the subject].

b. The purpose of random assignment to treatment and control groups is to [keep, initially start] both groups equal, whereas the purpose of blind care throughout the experiment is to [keep, not keep, initially start] both groups equal.

At the end, the evaluator generally [should, should not] know whether the subject she is evaluating is in the treatment or control group. This is called being ["blind," "randomized"] and it prevents the evaluator from being [blind, biased]. This is especially important when the evaluation is [merely technical, a difficult judgment].

1-5 Give an example of some "treatment" whose full effects are still unknown (for example, alcohol, compulsory education, certain forms of psychotherapy). Suggest how you might conduct an experiment to evaluate its effects. Include in your discussion:

a. Who would be the subjects?

b. Who would get the treatment, and who would not?

c. Who would be kept blind?

d. How would the effects be evaluated?

1-6 To determine how well vitamin C prevents colds, a study divided volunteers randomly into two groups of about 400 each—a treatment group given vitamin C, and a control group given a placebo (a "dummy" pill that had no drug, but looked indistinguishable). The proportions who were free of colds during the winter were 26% and 18%, respectively. (Consumer Reports, 1976.)

a. What was the purpose of the placebo?

b. It is customary to imagine that the volunteers represent a random sample from a large hypothetical population, for which we can construct a 95% confidence interval. For the proportion of people treated with vitamin C who would be free of colds, construct this interval.

c. Repeat part **b** for the people treated with a placebo.

d. Write a brief verbal conclusion about the effect of vitamin C.

PROBLEMS

1-7 "To find out what happens when you change something, it is necessary to change it." (Box, Hunter, and Hunter, 1978, p. 495). What does this mean, in terms of randomized controlled experiments and observational studies?

1-8 Underline the correct choice in brackets:

a. In an observational study, it is [difficult, easy] to determine the effect of the treatment, because its effect is [confounded with, clarified by] the effects of all the other variables that happen to be changing at the same time.

b. Confounding is particularly serious if the effect we're looking for is relatively [large, small] and consequently easily [masked, double-blinded] by the confounding variables—as, for example, in a study that tried to measure the effect of [seat belts, capital punishment].

c. In an observational study, multiple regression can [reduce, completely eliminate] confounding if the confounding variables that can be measured are [included in the regression, held constant by the experimenter].

d. To be ethical, a human experiment to evaluate a potentially hazardous treatment ought to be [never done, done only if the potential good seems likely to outweigh the potential harm], and should be carried out first in [volunteers, people randomly selected from the population].

e. One difficulty with limiting an experiment to volunteers is that [it is unethical, the effect the treatment has on volunteers may not hold for the whole population]. Nevertheless, if the treatment turns out to be beneficial to the volunteers, it is likely that in later experiments [fewer, more] people would volunteer, so that the "volunteer bias" would be of [growing, decreasing] concern.

1-9 As we have seen, the outcome of an observational study may be influenced more by confounding variables than by the actual treatment given. Let us now look at another possible source of confusion: Cause-and-effect may run in the *opposite* direction to what is first supposed. For the following examples, underline the correct choice in brackets:

a. Death rates from respiratory disease are unusually high in Arizona. A foreigner might guess that this is caused by Arizona's terrible air. But any American knows that, on the contrary, Arizona's air is so dry and clean that it attracts people who suffer from respiratory disease. We therefore conclude that:

Arizona has a lot of people dying from respiratory disease [because Arizona causes the disease, because the disease causes these people to move to Arizona.]

The high death rate from respiratory disease in Arizona occurs because the effect of people with respiratory disease entering Arizona [overwhelms, is overwhelmed by] the good effects of the climate.

b. In Sri Lanka, a negative relation has been observed between population density and malaria; that is, low density is associated with a high malaria rate (Gourou, 1966). Two possible reasons for this are:

1. High malaria rates [drive out, attract] people, thus producing a relatively [low, high] population density in a district, or
2. A high population density might help [maintain, destroy] drainage and consequently [reduce, increase] the breeding grounds of malaria-carrying mosquitoes.

From the negative relation alone, it is [impossible, possible but difficult, quite easy] to tell which of these explanations are true.

c. Some studies have shown that states with capital punishment have higher murder rates.[4] Again, two possible reasons for this are:

1. Capital punishment [works, is counterproductive], causing the murder rate to be [lower, higher], or
2. States with [low, high] murder rates tend to be the most anxious to reduce them, by resorting to drastic measures such as capital punishment.

From the positive relation alone, it is [impossible, possible but difficult, easy] to tell which explanation is true.

1-10 To generalize 1-9, suppose an observational study shows Y related to X. Consider three possible ways this could occur:

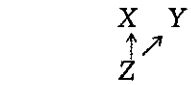
i. X might cause Y.

ii. Some confounding factors might cause both X and Y.

iii. Y might cause X.

a. Match up these three phenomena with the following "path diagrams" (where the arrow indicates the direction of causation):

[4] Other studies have shown *lower* murder rates. So these studies prove nothing, because of the inherent limitations discussed in this question, and elaborated on by Fisher, 1980.

$X \leftarrow Y$ $\qquad$  $\qquad$ $X \rightarrow Y$

b. Is it possible that the relation of Y to X could be due to two or three of these phenomena in combination?

c. Is it possible for X to have a positive *effect* on Y, yet the *relation* of X and Y be negative? If not, why not? If so, how?

1-11 Naive studies earlier in this century showed a positive relation between population density and crime. (Higher density areas of cities had higher crime rates.) However, more careful studies using regression (Choldin, 1978 or Simon, 1981) have shown that for confounding factors kept constant, the relation often disappears or even becomes slightly negative.

a. Suggest some confounding factors that often accompany high density, and might be responsible for some of the crime.

b. Underline the correct choice:

i. This illustrates how [double-blindedness, observational studies, randomized experiments] can often be deceptive when analyzed naively.

ii. Specifically, the [confounding factors, confidence intervals, sociological theories] may be what produce some—or all—of the effects.

iii. We therefore say the effect of population density is [biased by, interchangeable with, multiplied by] the effect of the confounding factors.

1-12 An experiment was conducted to evaluate the relative benefits of a new welfare policy (the "negative income tax") in several American cities in the 1970s (Ferber and Hirsch, 1982). To determine which families would continue with the old program, and which would get this new program, random assignment was proposed. Underline the correct choice:

a. Random assignment permits a [less biased, perfectly accurate] evaluation of the new program.

b. Instead of being randomly assigned, suppose the group assigned to the new program included just those families who had been on welfare the longest, and therefore had greatest need of a new approach. In the final evaluation the new program would consequently [be more fairly compared, have its effect confounded with the effect of duration of welfare]. On the whole, this form of selection would therefore [slightly improve, greatly improve, invalidate] the comparison of the new and the old welfare programs.

REVIEW PROBLEMS

These review problems included at the end of each chapter give an overview of all the material. Because they do not fit neatly into a pigeonhole, they require more thought. Consequently, they provide the best preparation for meeting real life—and exams.

1-13 "The possession of such a degree as the MBA will double lifetime earnings as compared to those of a high school graduate." (Bostwick, 1977)

To back this up, suppose the author found out that MBAs' annual incomes are twice as high on average as high school graduates'. Is the author's conclusion accurate? Explain.

1-14 Do seat belts reduce injury and death? They definitely do, according to experiments that have been done, smashing up cars and measuring the damage to dummy passengers. It is also valuable to look for confirming evidence from real people, in observational studies. For example, in a study of accidents, we could compare injury rates of those who were wearing seat belts to those who were not. In order to keep confounding factors to a minimum, which of the two alternatives in each of the cases below would be better, and why?

a. Using data from all accidents, or from just those involving cars equipped with seat belts?

b. Using data on all the occupants, or on just the drivers?

c. Using data that lumps together all injuries, or that categorizes them into several levels of severity?

d. Having the doctor who evaluates the injury be informed, or not informed, of whether the patient had been wearing a seat belt?

1-15 The following data on seat belt usage and injury rates was collected in the 1970s from selected states that kept good records:

Belt Usage and Injury Rates for Accident-Involved Occupants With Safety Belts Available

		Injury Rate	
Seat Belt Usage	*Percent of All Occupants*	*Moderate Injury* $(2 \leq AIS < 3)^a$	*Serious or Greater Injury* $(AIS \geq 3)$
Unbelted	85.9	.023	.013
Lap Belt	3.9	.011	.009
Lap-and-Shoulder Belt	7.8	.005	.004
Unknown	2.4	—	—

U.S. Dept. of Transportation, 1981

[a] AIS: Abbreviated Injury Scale 1980 Revision, Amer. Assoc. for Automotive Medicine.

a. Regard the data as a random sample of 100,000 accident-relate occupants over the whole decade from the whole country. Cor struct a confidence interval for the rate of moderate injur among those wearing lap-and-shoulder belts. Then do the sam for those wearing no seat belts.

b. Repeat part **a** for injury that is serious or worse.

c. What can you conclude about the value of lap-and-shoulde belts?

1-16 **a.** Of the 18,000 deaths in Arizona in 1977, 1440 were from respi ratory disease. Assuming these deaths can be regarded as random sample from a hypothetical population of million who might have lived and died in Arizona, calculate a 95% confidence interval for the population proportion of deaths tha are due to respiratory disease.

b. For the whole United States in 1977 calculate the correspond ing proportion, given the data that out of 1,900,000 deaths 110,000 were from respiratory disease.

c. To what extent does the data show that:

i. Arizona has a higher proportion of deaths from respirator disease?

ii. Arizona's climate worsens respiratory disease?

1-17 U.S. unemployment statistics are obtained through the Curren Population Survey, a sample of about 100,000 adults conducte monthly by the U.S. Census Bureau. Like the Gallup poll, it is combination of multistage and other kinds of random sampling tha altogether provide about the same accuracy as simple random sam pling, so that equation (1-2) gives a good approximation. In July 1987, the sample gave roughly the following figures for the nonin stitutional population aged 16 and over:

Employed	62,690
Unemployed	4,080
Outside the Labor Force	33,230
Total	100,000

(*U.S. Survey of Current Business*, November 1987.)

a. The "participation rate" is the proportion who are in the labor force, that is, who are employed or looking for employment (unemployed). Construct a 95% confidence interval for the participation rate for the entire U.S. population.

b. Construct a 95% confidence interval for the unemployment rate.

c. When an unemployment rate is quoted by the media, they don't bother with the confidence limits, of course. Nevertheless this uncertainty must be appreciated by anyone who tries to interpret unemployment figures carefully.

For example, if the quoted rate of unemployment went down from 5.8% to 5.7%, does this mean the actual *population* rate dropped 0.1%?

1-18 *A Final Challenge: How Hard Was It In 1973 For Women To Get Into Berkeley?*

In that year, their Graduate School admitted 3700 out of 8300 men applicants, and 1500 out of 4300 women applicants (Bickel and O'Connell, 1975).

a. What is the difference in admission rates, between men and women? How good is this evidence of sex discrimination?

b. To find out where this difference arises, the data was broken down by faculty, as follows (although this breakdown is hypothetical, it preserves the spirit of the problem while simplifying the computations):

	Men		Women	
Faculty	*Number of Applicants*	*Number Admitted*	*Number of Applicants*	*Number Admitted*
Arts	2300	700	3200	900
Science	6000	3000	1100	600
Totals	8300	3700	4300	1500

Now what is the difference in admission rates between men and women, for Arts? And for Science?

c. Explain why your answers to parts **a** and **b** seem to be in conflict. (This is an example of *Simpson's paradox*: What is true for the parts may not be true for the whole.)

d. Underline the correct choice:

i. If faculty is kept constant, men and women are admitted about equally. However, there is a tendency for women to apply to the [easier, tougher] faculty, which may explain why their overall admission rate is considerably [lower, higher].

ii. In part **a,** the effect of sex on admissions could not be properly understood because there was a confounding factor (faculty) that wasn't being controlled. When appropriate control was introduced in part **b,** however, a more [incriminating, accurate] picture emerged.