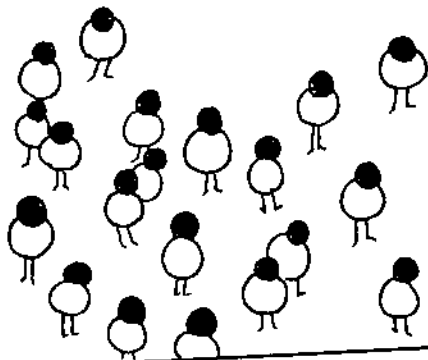
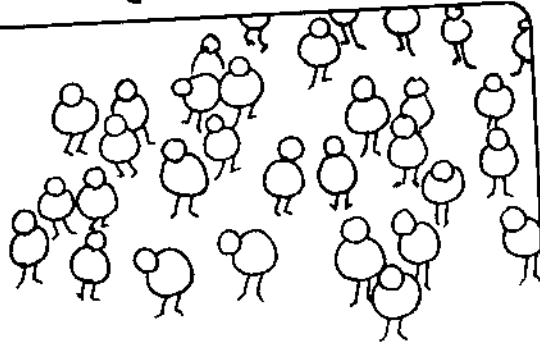


SAMPLING DESIGN



BEFORE DOING THE NUMBERS, WE SHOULD POINT OUT THAT THE QUALITY OF THE SAMPLE IS AS IMPORTANT AS ITS SIZE. HOW DO WE ASSURE OURSELVES THAT WE'RE CHOOSING A REPRESENTATIVE SAMPLE?



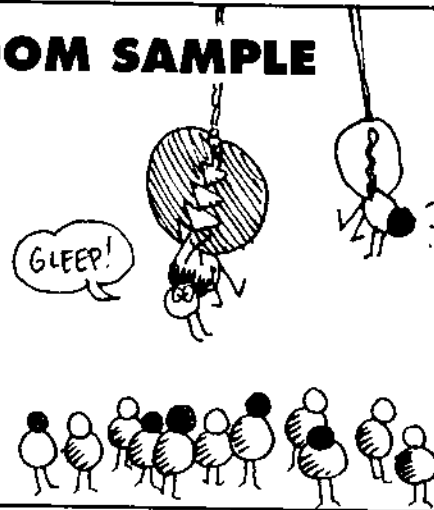
THE SELECTION PROCESS ITSELF IS CRITICAL. FOR EXAMPLE, A VOTER SURVEY THAT SYSTEMATICALLY EXCLUDED BLACK PEOPLE WOULD BE WORTHLESS, AND THERE ARE A HOST OF OTHER WAYS TO RUIN, OR BIAS, A SAMPLE.

NOT TO PROLONG THE MYSTERY, THE WAY TO GET STATISTICALLY DEPENDABLE RESULTS IS TO CHOOSE THE SAMPLE AT **random**.

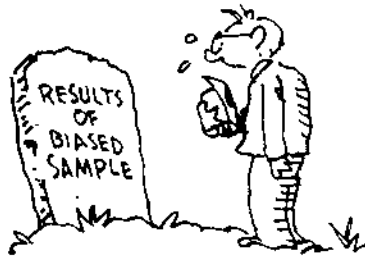


THE SIMPLE RANDOM SAMPLE

SUPPOSE WE HAVE A LARGE POPULATION OF OBJECTS AND A PROCEDURE FOR SELECTING n OF THEM. IF THE PROCEDURE ENSURES THAT ALL POSSIBLE SAMPLES OF n OBJECTS ARE EQUALLY LIKELY, THEN WE CALL THE PROCEDURE A **simple random sample**.



THE SIMPLE RANDOM SAMPLE HAS TWO PROPERTIES THAT MAKE IT THE STANDARD AGAINST WHICH WE MEASURE ALL OTHER METHODS:

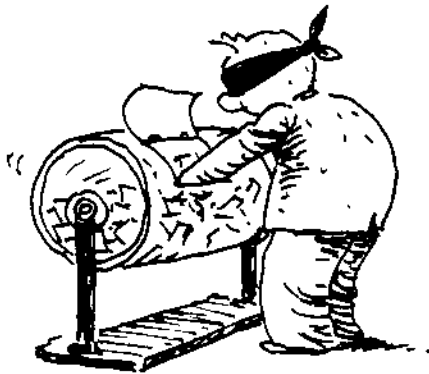


- 1) UNBIASED: EACH UNIT HAS THE SAME CHANCE OF BEING CHOSEN.
- 2) INDEPENDENCE: SELECTION OF ONE UNIT HAS NO INFLUENCE ON THE SELECTION OF OTHER UNITS.

UNFORTUNATELY, IN THE REAL WORLD, COMPLETELY UNBIASED, INDEPENDENT SAMPLES ARE HARD TO FIND. FOR INSTANCE, SURVEYING VOTERS BY RANDOMLY DIALING TELEPHONE NUMBERS IS BIASED: IT IGNORES VOTERS WITHOUT A TELEPHONE AND OVERSAMPLES PEOPLE WITH MORE THAN ONE NUMBER.



IT'S THEORETICALLY POSSIBLE TO GET A RANDOM SAMPLE BY BUILDING A **SAMPLING FRAME**: A LIST OF EVERY UNIT IN THE POPULATION. BY USING A RANDOM NUMBER GENERATOR, WE CAN PICK n OBJECTS AT RANDOM.



EQUIVALENTLY, WE CAN PUT ALL THE NAMES ON CARDS AND PULL n OF THEM OUT OF A DRUM.

BUT THIS IS NOT ALWAYS EASY. MAKING THE FRAME MAY BE PROHIBITIVELY COSTLY, CONTROVERSIAL, OR EVEN IMPOSSIBLE. FOR EXAMPLE, AN E.P.A. WATER QUALITY STUDY NEEDED A SAMPLING FRAME OF LAKES IN THE U.S., SO THEN SOMEBODY HAS TO DECIDE:

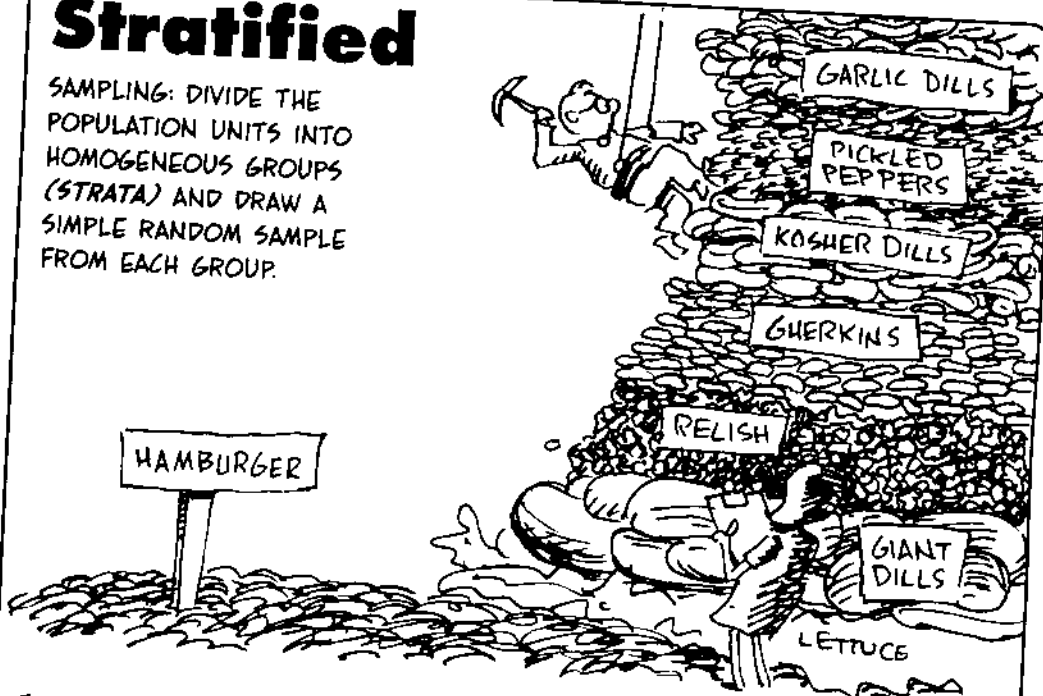
WHAT WET SPOT IS A LAKE?



ARE THERE OTHER WAYS TO SAMPLE THAT ARE MORE EFFICIENT AND COST-EFFECTIVE THAN A SIMPLE RANDOM SAMPLE? YES—IF YOU ALREADY KNOW SOMETHING ABOUT THE POPULATION. FOR INSTANCE...

Stratified

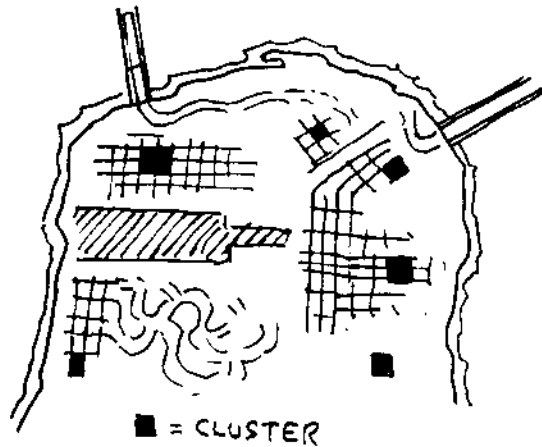
SAMPLING: DIVIDE THE POPULATION UNITS INTO HOMOGENEOUS GROUPS (STRATA) AND DRAW A SIMPLE RANDOM SAMPLE FROM EACH GROUP.



FOR EXAMPLE, THE POPULATION OF ALL PICKLES CAN BE STRATIFIED BY TYPE OF PICKLE. WITHIN EACH TYPE OR STRATUM, THE SIZE SHOULD BE LESS VARIABLE.

Cluster

SAMPLING GROUPS THE POPULATION INTO SMALL CLUSTERS, DRAWS A SIMPLE RANDOM SAMPLE OF CLUSTERS, AND OBSERVES EVERYTHING IN THE SAMPLED CLUSTERS. THIS CAN BE COST-EFFECTIVE IF TRAVEL COSTS BETWEEN RANDOMLY SAMPLED UNITS IS HIGH.



AN EXAMPLE IS A CITY HOUSING SURVEY WHICH DIVIDES A CITY INTO BLOCKS, RANDOMLY SAMPLES THE BLOCKS, AND LOOKS AT EVERY HOUSING UNIT IN EACH SAMPLED BLOCK.