Sample Designs

Gaston Sanchez

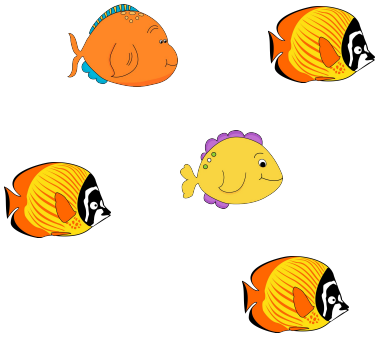
Recap:
Sampling Idea

Population

Avg Weight

parameter

Sample



Avg Weight

statistic

To help you remember

Population

***P**arameter*

describes an aspect of the population

Sample

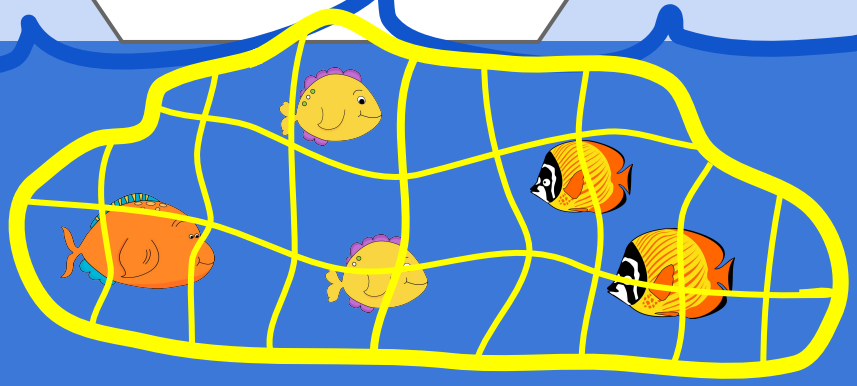
***S**tatistic*

describes an aspect of the sample


population difficult to observe

But...

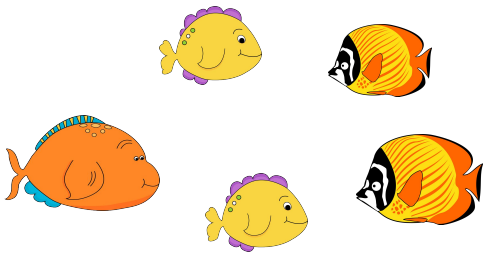
We can get **samples** ...



to make confident statements about populations

Toward statistical inferences

What can these fish...



sample

tell us about these fish?

population

Sample Designs

How to make a sample?

Samples with a probability element

Samples without random component

Samples with a probability element

Does every unit has the same chance of being selected?

Do some units have a different chance of being selected?

Sampling Types

Samples

Units individually chosen

Cluster: groups of units (students in 4th row)

Strata: divide population in non-overlapping groups (freshman, sophomore, junior, senior)

Multi-stage: involve taking different stages to partition the population

Multi-stage Sampling

Make a list of states

Select two states; then list counties in those states

Select two counties from each of the states, and list the blocks in the counties

Select three blocks from each county, and list the housing units in the blocks

Select five housing units from each block, and list the residents of the housing units

Select one person from each of the selected housing units

Convenient Sampling

What fraction of the Berkeley faculty are registered republicans?

I could take the phone directory and take the 1st 100 names

I could go the Faculty Club launch and interview 1st 100 persons

I could interview all faculty in the Stat Department

Quota Sampling

Deliberately match the demographics in the sample to those of the population

These fraction of students are female

These fraction of students have financial aid

These fraction of students are state residents

Have a sample matching those proportions (does not guarantee representative sample)

Simple Random Sample

This is the golden standard for sampling

Random Sample without replacement

Assign a random number to every element in the population

Systematic Random Sample

Start with a random position

And then take every unit *k-th* steps away

The 10 equally likely samples for systematic random sampling of 10 elements from 100

Sample 1	1, 11, 21, 31, 41, 51, 61, 71, 81, 91
Sample 2	2, 12, 22, 32, 42, 52, 62, 72, 82, 92
Sample 3	3, 13, 23, 33, 43, 53, 63, 73, 83, 93
Sample 4	4, 14, 24, 34, 44, 54, 64, 74, 84, ,94
Sample 5	5, 15, 25, 35, 45, 55, 65, 75, 85, 95
Sample 6	6, 16, 26, 36, 46, 56, 66, 76, 86, 96
Sample 7	7, 17, 27, 37, 47, 57, 67, 77, 87, 97
Sample 8	8, 18, 28, 38, 48, 58, 68, 78, 88, 98
Sample 9	9, 19, 29, 39, 49, 59, 69, 79, 89, 99
Sample 10	10, 20, 30, 40, 50, 60, 70, 80, 90, 100

Another Example

Estimate the average size of undergraduate classes

Take a random sample of 100 courses from the current course schedule and average their sizes. Simple random sample

Take a random sample of 50 students, list the courses each student is taking, and average the list. Clustered sample

Take a random sample of 50 instructors, list the courses they are teaching, and average the combined list. Clustered sample

Take separate random samples of 5 courses from each department, and average the sizes in all the samples.

Stratified sample

Estimate the average size of undergraduate classes

Take a random sample of 5 science/engineering departments and a random sample of 5 humanities/professional deparments. For each department in the sample, list the sizes of all the courses, and average the sizes in the list.

Stratified clustered sample

Take a random sample of 5 science/engineering departments and a random sample of 5 humanities/professional deparments. For each department in the sample, take a random sample of 5 instructors. Average the sizes of the courses those instructors are teaching.

Stratified multi-stage clustered sample

Estimate the average size of undergraduate classes

Pick a random number *K* between 1 and 10. Starting with the *K*-th course in the course schedule, list the size of every 10-th class. Average the list of sizes.

Systematic Random Sample

Hypothetical Populations

Sampling from Hypothetical Populations

Drawing samples from population that don't actually exist

What is the average test score from some group of students?

Effect of a drug: how many people would have gotten better if everybody have gotten the drug?

Chance Error and Bias

Chance Error

Observed Value = Expected Value + Chance Error

Chance Error with SRS

$$\text{Statistic} = \text{Parameter} + \text{Chance Error}$$

Chance error = sampling error

Error because the sample is only part of population

In practical situations

$$\text{Statistic} = \text{Parameter} + \underset{\text{Error}}{\text{Chance}} + \text{Bias}$$

- Response and Non-response Bias
- Question bias
- Selection bias

What's more accurate?

Statistic = Parameter + Chance Error + Bias

Sampling with replacement?

Sampling without replacement?

Small chance error