Sampling and Sampling Distributions

We want to know about our population, but only have a sample.

We need to know how GOOD information from a sample is…how close to the population mean and standard deviation are the sample mean and

TYPES OF SAMPLES:

Non-Probability Sampling

Don't know the probability of choosing each member of the population.

Quota Sampling--Interviewer stands on a corner and interviews until a certain number are conducted

Volunteer Sampling--Newspaper asks readers to complete questionnaire and send it in; an internet survey to which people have to go to respond

Snowball Sampling--interviewer talks to one person and asks if they know anyone else. The interview then goes and talks to those people (a way of sampling hard-to-reach populations).

Probability Sampling

Know the probability that each member of the population is sampled. Probability sampling is essential to obtaining a representative sample.

Simple Random Sample (SRS)-- Each member of the population has an equal probability of being in the sample.

Stratified Random Sample-- SRS from within defined strata (groups) within the population. Yields a more efficient estimate.

Cluster Sample -- Random selection of grouped elements (like schools, households, buildings, neighborhoods) and then random selection of members within group.

Sampling Distribution--

Distribution of possible values of a <u>sample statistic</u> assuming it is from a random sample.

Sample Proportion sampling distribution–

$E(\hat{P}) = P$ $$Var(\hat{P}) = \frac{P(1-P)}{n}$$

Sample Mean sampling distribution –

$E(\overline{X}) = \mu$ $$Var(\overline{X}) = \frac{\sigma^2}{n}$$

→The sample proportion and the sample mean statistics are random variables.

→Sampling distributions reflect randomness due to who or what is randomly chosen

Standard Error -- Standard deviation of the distribution of <u>sample mean</u> or proportion

How far a typical sample mean is from the population mean (How far a typical $\overline{X}$ is from μ.)

For a sample mean:

$$SE \equiv SD(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

For a sample proportion:

$$SE \equiv SD(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}$$

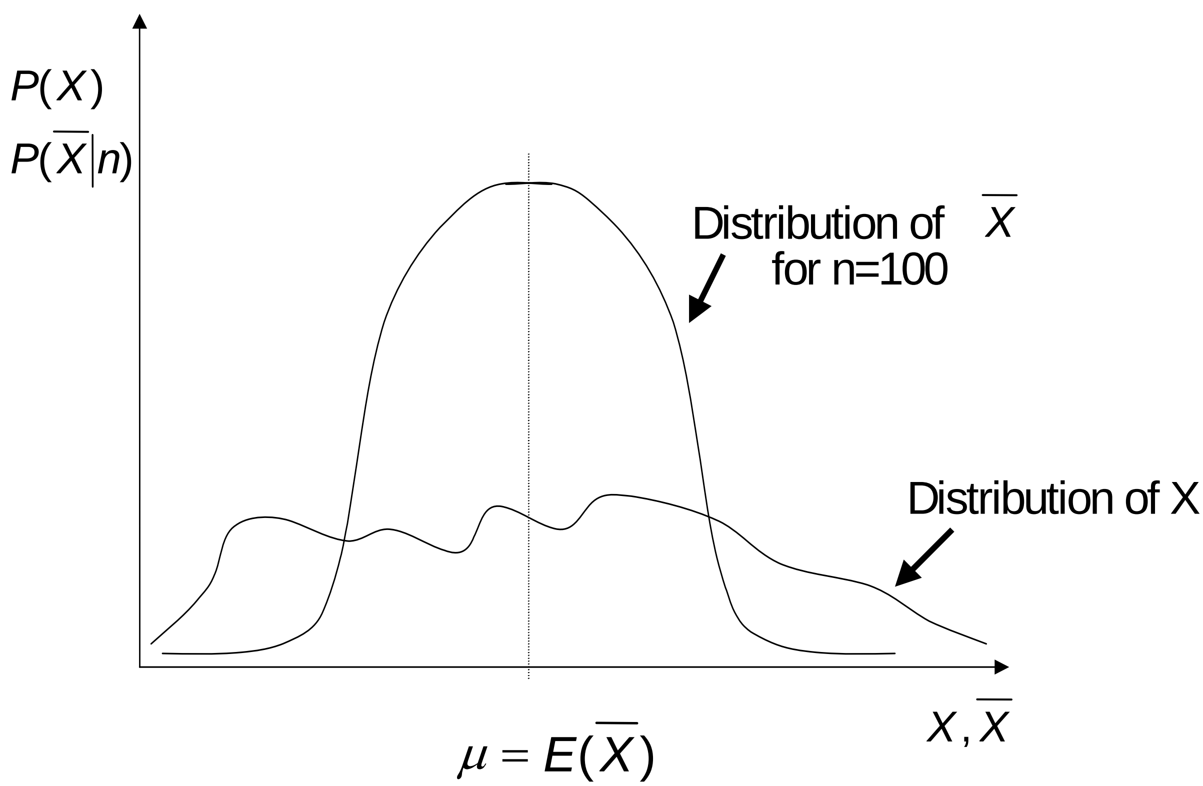
- Larger with more variation in variable values for population (σ)
- Smaller for larger sample sizes
- Doesn't depend on population size unless sample is large % of population.

Central Limit Theorem

- Sampling distribution of the sample mean is approximately normal *regardless* of the original variable distribution for large samples.

- Sampling distribution gets "more normal" as sample size increases.

The $\overline{X}$ distribution with a large sample size.



What percent of possible $\overline{X}$s are within 1 SE of μ?

What percent of possible $\overline{X}$s are within 2 SE of μ?

Properties of the Sampling Distribution for $\overline{x}$

Unbiased--The mean of the sampling distribution equals the mean of the population.

$$\mu_{\overline{x}} = E\left(\overline{x}\right) = \mu$$

Consistent—the estimates get better as the sample size increases because the variance and bias decrease.

$$SE \equiv SD(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

Efficient—the estimate is the "best" we can get.

- Might have to trade off bias for lower variance
- Efficiency may differ by sample size and population distribution
- This is determined theoretically or with Monte Carlo methods

The 1999 household income in the US had a mean of $44,000 and a standard deviation of $62,000. If I were to pick a random sample of 100 U.S. households, how likely am I to get a sample mean of household income more than $1000 above the population mean?

1. Write down the target and draw picture.

2. Find the standard error:

3. Find the standard score for our sample mean target value:

4. Look up score in Z table.

In 1999, 12 % of people in the U.S. lived in households with income under the poverty level. What's the probability that the poverty rate in a sample of 1000 people will be within 1 percentage point of the "true" proportion?

To find the probability for a sample proportion:

1. **Find target and draw picture**.

2. **Find SE** (Need pop SD first):

3. **Find Z scores:**