Sampling and Sampling Distributions

Population

- Researchers in the social science almost never have enough time or money to collect information about the entire group that interests them.
- The entire group is known as the population, all the cases in which the researcher is interested.

Sample

- Fortunately for social scientists we can learn a lot about a population if we carefully select a subset of it. The subset is called a sample.
- By selecting a sample we attempt to generalize to the population, that is why it is so important to select a good sample. This is the basis for inferential statistics.

Parameter

- The term parameter is associated with the population, it refers to measures used to describe the distribution of the population in which we are interested.
- For example, if I asked this class if you wanted quizzes, that would be a population, so the mean, proportions, and standard deviations would be population parameters.

Sample Statistic

- We use the term sample statistic when referring to a corresponding characteristic calculated for the sample.
- For example, if I asked you as a class if you wanted to go to a 16 week semester, and then took your answer and applied it to the whole school, the description of the distribution would be a sample proportion and sample standard deviation.

Sample and Population Notations

Measure	Sample Notation	Population Notation
Mean	$\overline{Y}$	μ_y
Proportion	p	π
Standard Deviation	S_y	σ
Variance	s^2_y	σ^2_y

Probability Sampling

- Social researchers are very careful in their effort to obtain representative samples of the population.
- Probability sampling is a method of sampling that enables the researcher to specify for each case in the population the probability of its inclusion in the sample.

Non-probability Samples

- Most social science researchers use non-probability samples because they are more convenient and cheaper to collect.
- When you use a non-probability sample, you cannot generalize to the population.

Types of Probability Sampling

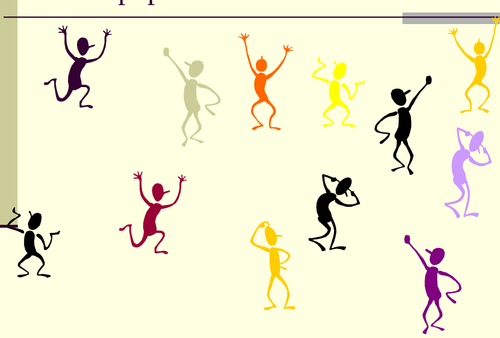
- Simple Random Sample, A sample designed to ensure that every member of the population has an equal chance of being chosen.
- Systematic Random Sample, A method of sampling in which every Kth member in the total population is chosen for the sample.
- Stratified Random Sample, A method of sampling obtained by dividing the population into subgroups based on one or more variables central to the analysis then drawing a simple random sample from each of the subgroups.

Simple Random Sampling

- Simple random sampling is the most basic probability sampling design.
- In a simple random sample, every member of the population has an equal chance of being chosen for the sample, and every combination of members has an equal chance of being chosen.

Population inferences can be made...

...by selecting a representative sample from the population



Example of simple random sampling:

- Let’s say I wanted to study this class with a sample of ten. To do a simple random sample, I could put everyone's name on a piece of paper and draw ten names out of a hat.
- The sample would be random because it was pure chance that determined who was chosen, every student had the same chance, and every combination of students was possible.

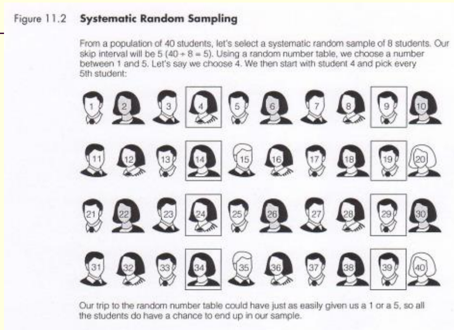
Systematic Random Sample

- A systematic random sample is easier to implement than a simple random sample, but it is not a true probability sample.
- The results are similar to those obtained with a simple random sample.
- It uses a ratio (K). obtained by dividing the population size by the desired sample size.
  - K = Population Size/Sample Size

Systematic Random Sample

- In systematic random sampling we chose every Kth member of the total population.
- So if the total population of this class is 48, and I wanted a sample of 12, what would K be?
- Then you choose any one student at random from among the first four and then select every forth person after than until you have 12 students.

Systematic Random Sampling



Stratified Random Sample

- To obtain a stratified random sample you first must divide the population into subgroups based on one or more variables central to our analysis.
- Then draw a simple random sample form each of the subgroups.

Example of stratified random sample:

- Let's say I want to study this class's views on women working outside the house.
- I want to see if there is a difference based on sex. So I would divide the population based on sex.

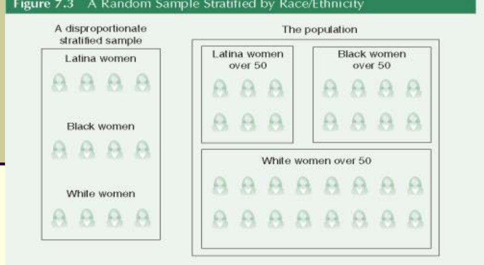
Example of stratified random sample:

- The population of interest consists of 48 students, with 36 (0r 75%) females, and 12 (0r 25%) males.
- Because we know the proportion of each subgroup we could draw a stratified sample that would reflect these exact proportions.
- If I wanted a sample of twelve, I would include nine females and 3 males.

Stratified Random Sampling

- Disproportionate Stratified Sample

Figure 7.3 A Random Sample Stratified by Race/Ethnicity



The Concept of the Sampling Distribution

- The sampling distribution helps us estimate the likelihood of our sample statistic and therefore, enables us to generalize from the sample to the population.

Population: Personal Income

Individual	Income (Y)
Case 1	11,350 (Y_1)
Case 2	7,859 (Y_2)
Case 3	41,654 (Y_3)
Case 4	13,445 (Y_4)
Case 5	17,458 (Y_5)
Case 6	8,451 (Y_6)
Case 7	15,436 (Y_7)

μ_Y = \$16,521 δ_Y = \$11,619

The Dilemma

- Few if any sample estimates correspond exactly to the actual population parameter.
- If sample estimates vary and if most estimates result in some sort of sampling error, how much confidence can we place in the estimate?

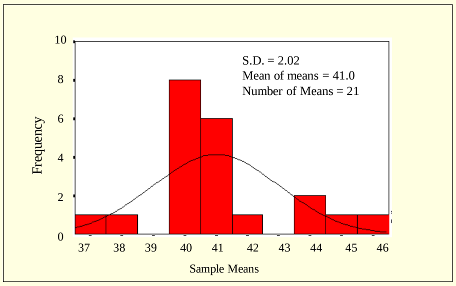
Sampling Distributions

- **Sampling distribution of the mean** – A theoretical probability distribution of sample means that would be obtained by drawing from the population all possible samples of the same size.

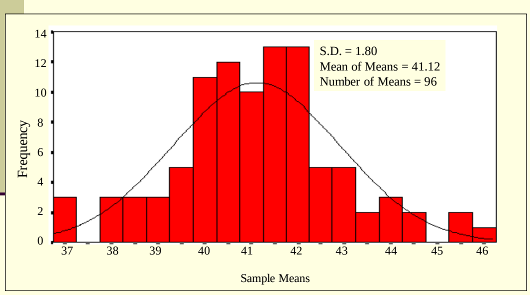
If we repeatedly drew samples from a population and calculated the sample means, those sample means would be normally distributed (as the number of samples drawn increases). The next several slides demonstrate this.

- **Standard error of the mean** – The standard deviation of the sampling distribution of the mean. It describes how much dispersion there is in the sampling distribution of the mean.

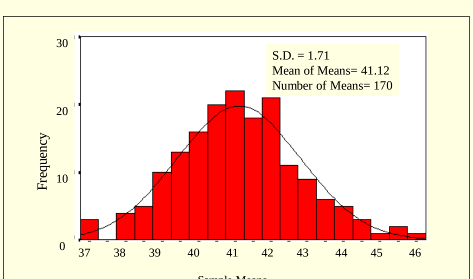
Distribution of Sample Means with 21 Samples



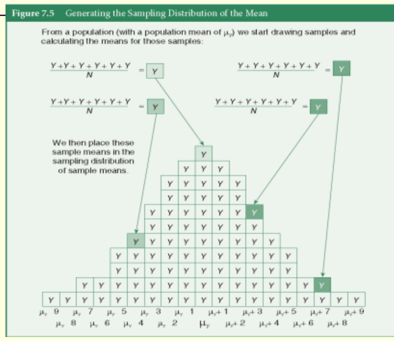
Distribution of Sample Means with 96 Samples



Distribution of Sample Means with 170 Samples



Sampling Distributions



The Central Limit Theorem

- If all possible random samples of size N are drawn from a population with mean μ_y and a standard deviation σ_y, then as N becomes larger, the sampling distribution of sample means becomes approximately normal, with mean $\mu_{\bar{Y}}$ equal to the population mean and standard deviation equal to $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{N}}$

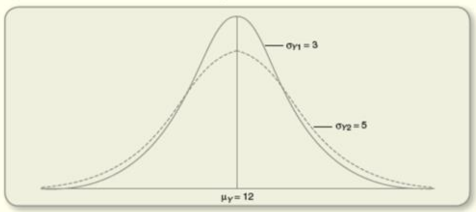
The Central Limit Theorem Cont.

1. As sample becomes larger, the mean of the sampling distribution becomes equal
2. As the sample gets larger, the standard error of the mean (the standard deviation of the sampling distribution) decreases in size.
3. When N is 50 or more, the sampling distribution of the mean will be approximately normal regardless of the shape of the distribution.

Normal Distribution

- Notice that the standard deviation changes the relative width of the distribution
  - The larger the standard deviation, the wider the curve

Figure 6.2 Two Normal Distributions With Equal Means but Different Standard Deviations



Computing a z-score

- For a sample (we did this in chapter six)

$$\frac{(Y - \bar{Y})}{S_y}$$

Computing a Z-score

- For a population

$$\frac{\bar{Y} - \mu}{\sigma_y \div \sqrt{n}}$$

Let Us Try! ☺

- My statistics class has a population mean of 3.5 hours of studying for exam three and a standard deviation of 3.27.
- Say I took a sample of 20 students and wanted to find the probability that they would have a mean between 3 and 4 hours of studding.

$$\frac{\bar{Y} - \mu}{\sigma_y \div \sqrt{n}}$$