Introduction to Social Statistics

Day 11

Instructor Rob Kemp

Monday February 18, 2012

Announcements

- Tests Graded and Grades posted.
- HW 1: If you did it, you got the full points.
- Tests Back at the end.

Agenda:
Sampling and Sampling Distributions

- Aims of Sampling
- Basic Principles of Probability
- Types of Random Samples
- Sampling Distributions
- Sampling Distribution of the Mean
- Standard Error of the Mean
- The Central Limit Theorem

Sampling

- **Population** – A group that includes all the cases (individuals, objects, or groups) in which the researcher is interested.
- **Sample** – A relatively small subset from a population.

Notation

Table 7.1 Sample and Population Notations

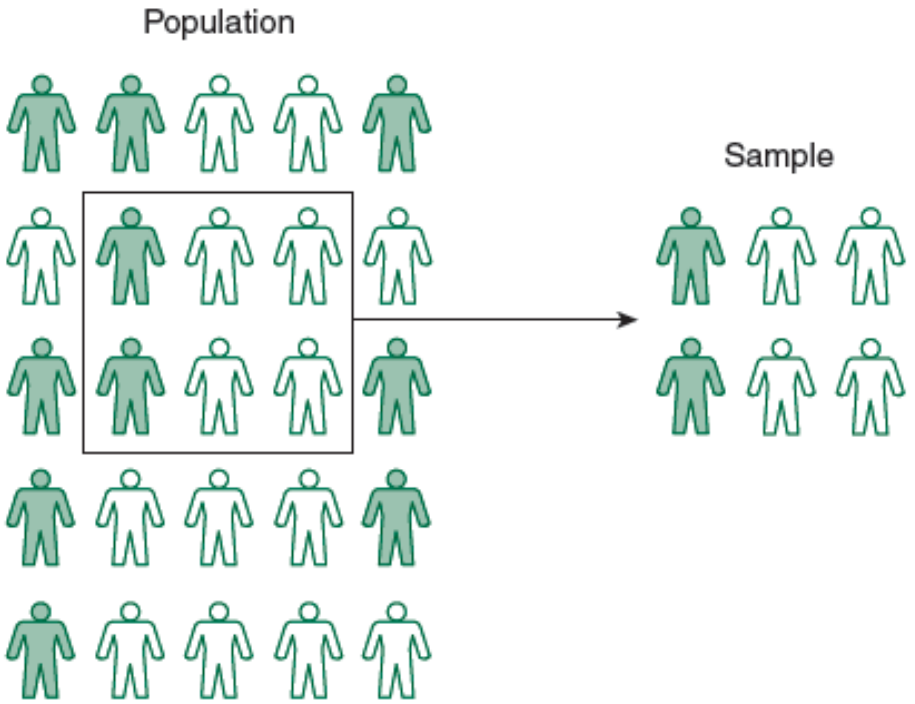
Measure Notation	*Sample Notation*	*Population*
Mean	$\overline{Y}$	μ_Y
Proportion	p	π
Standard deviation	S_Y	σ_Y
Variance	S_Y^2	σ_Y^2

Sampling

- **Parameter** – A measure (for example, mean or standard deviation) used to describe a population distribution.
- **Statistic** – A measure (for example, mean or standard deviation) used to describe a sample distribution.

Sampling: Parameter & Statistic

Figure 7.1 The Proportion of White Respondents in a Population and in a Sample



Parameter

Proportion of white respondents in the population

$$\pi = \frac{15}{25} = .60$$

Statistic

Proportion of white respondents in the sample

$$p = \frac{4}{6} = .67$$

Probability Sampling

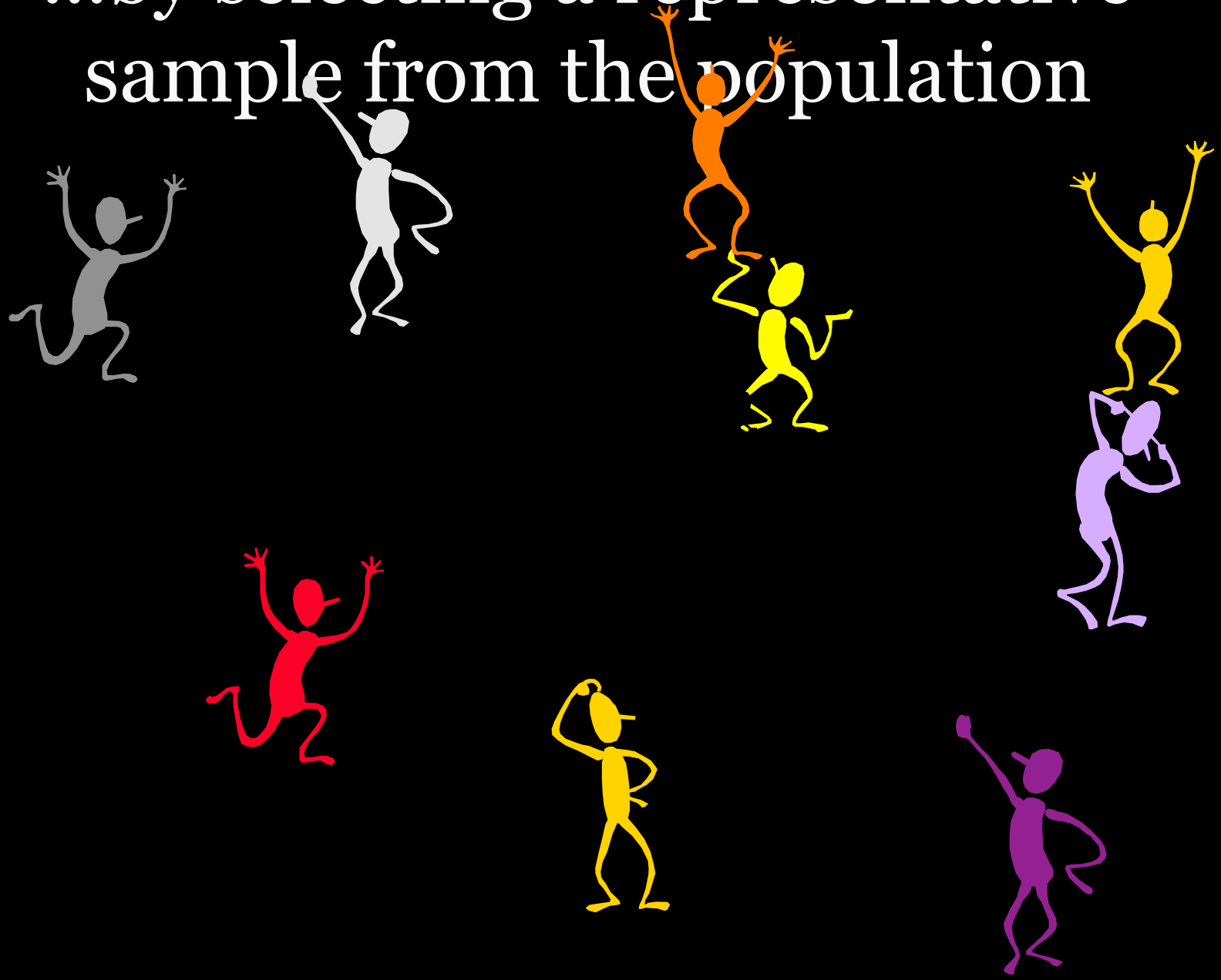
- **Probability sampling** – A method of sampling that enables the researcher to specify for each case in the population the probability of its inclusion in the sample.
- Goal is to select a sample that is as representative as possible of the population.
- Nonprobability sampling techniques are sometimes used (e.g., snowball sampling), but they do not allow us to make inferences about the population

Random Sampling

- **Simple Random Sample** – A sample designed in such a way as to ensure that (1) every member of the population has an equal chance of being chosen and (2) every combination of N members has an equal chance of being chosen.
- This can be done using a computer, calculator, or a table of random numbers (or a hat!)

Population inferences can be made...

...by selecting a representative sample from the population



Random Sampling

- **Systematic random sampling** – A method of sampling in which every *K*th member (*K* is a ratio obtained by dividing the population size by the desired sample size) in the total population is chosen for inclusion in the sample after the first member of the sample is selected at random from among the first *K* members of the population.

Systematic Random Sampling

Figure 7.2 Systematic Random Sampling

From a population of 40 students, let's select a systematic random sample of 8 students. Our skip interval will be 5 (40 ÷ 8 = 5). Using a random number table, we choose a number between 1 and 5. Let's say we choose 4. We then start with student 4 and pick every 5th student:

Our trip to the random number table could have just as easily given us a 1 or a 5, so all the students do have a chance to end up in our sample.

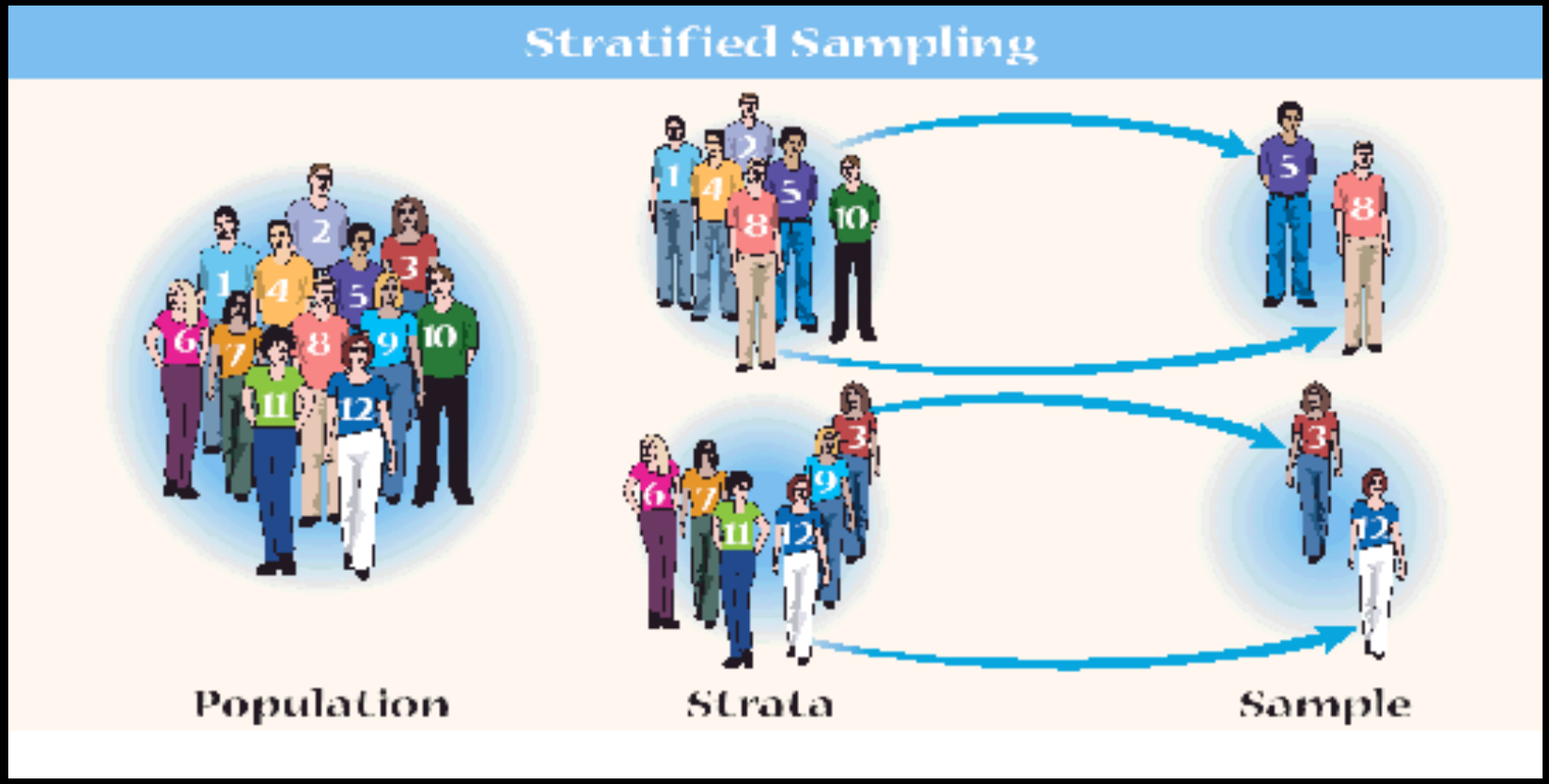
Stratified Random Sampling

- **Stratified random sample** – A method of sampling obtained by (1) dividing the population into subgroups based on one or more variables central to our analysis and (2) then drawing a simple random sample from each of the subgroups

Stratified Random Sampling

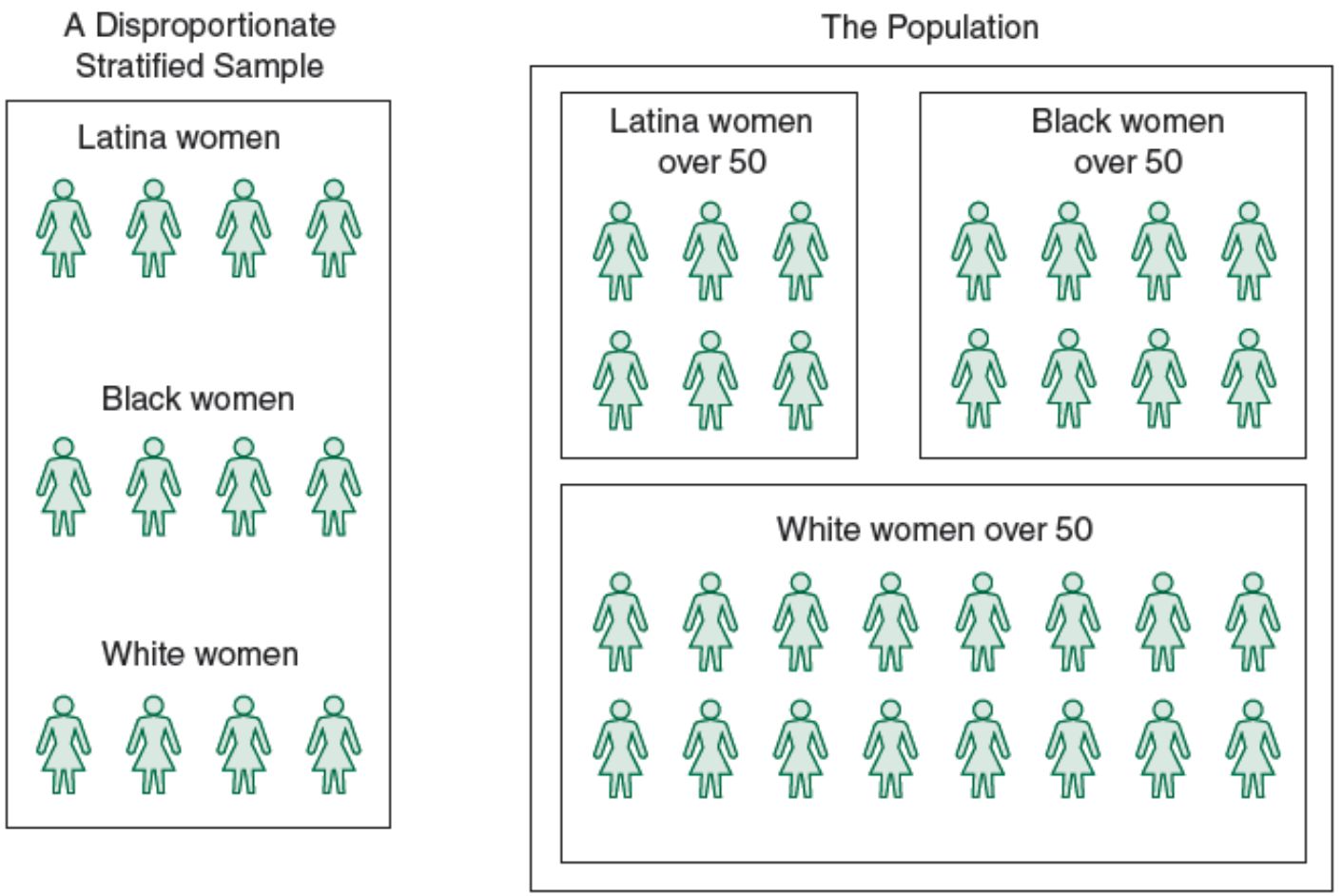
- **Proportionate stratified sample** – The size of the sample selected from each subgroup is proportional to the size of that subgroup in the entire population.
- **Disproportionate stratified sample** – The size of the sample selected from each subgroup is disproportional to the size of that subgroup in the population.

Proportionate Stratified Sample



Disproportionate Stratified Sample

Figure 7.3 A Random Sample Stratified by Race/Ethnicity



How do we draw inferences from our sample to the population?

- Need to have a random sample
- The larger the sample, the better our ability to draw inferences
- We can use the sampling distribution (a theoretical tool)

Sampling Distributions

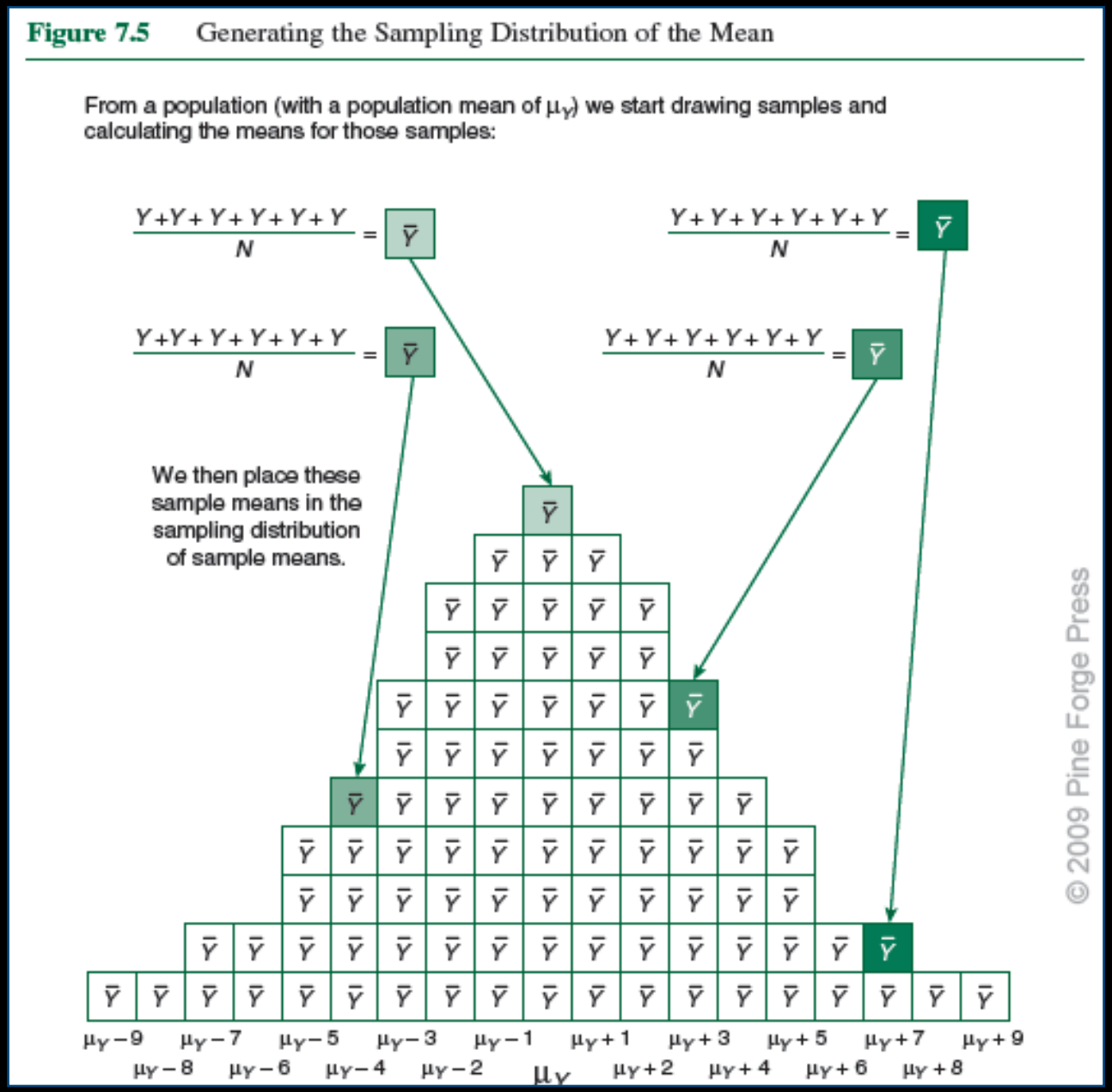
- **Sampling error** – The difference between a sample estimate of a population parameter and the real population parameter.
- *But, we don't usually know the population parameter, so how do we know how much confidence we should place in the estimate?*
- **Sampling distribution** – A theoretical distribution of all possible sample values (from all possible random samples) for the statistic (e.g., mean) in which we are interested.

Sampling Distribution of the Mean

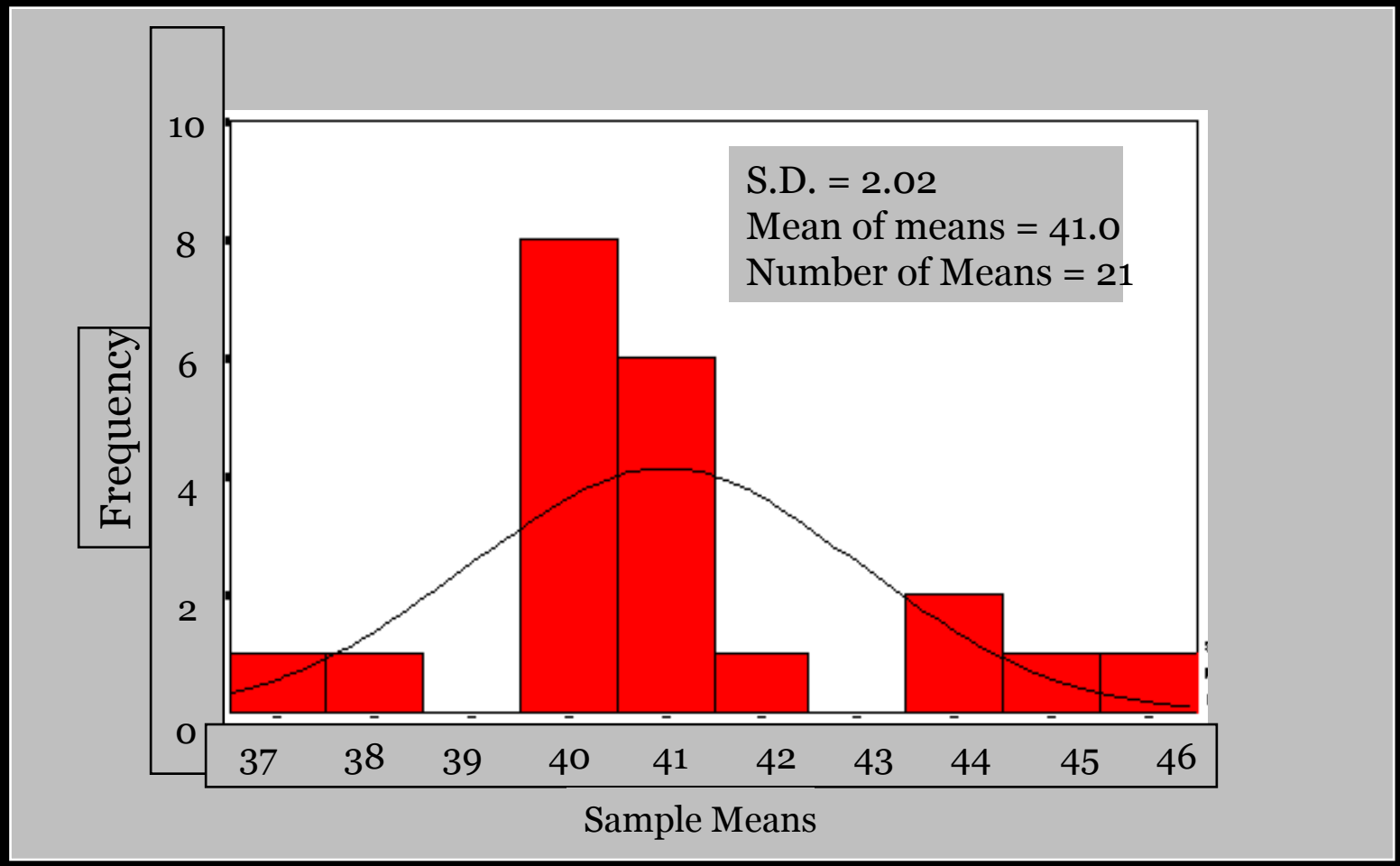
- **Sampling distribution of the mean** – A theoretical probability distribution of sample means that would be obtained by drawing all possible samples of the same size from the population .

- **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 (how much variability there is in the value of the mean from sample to sample).

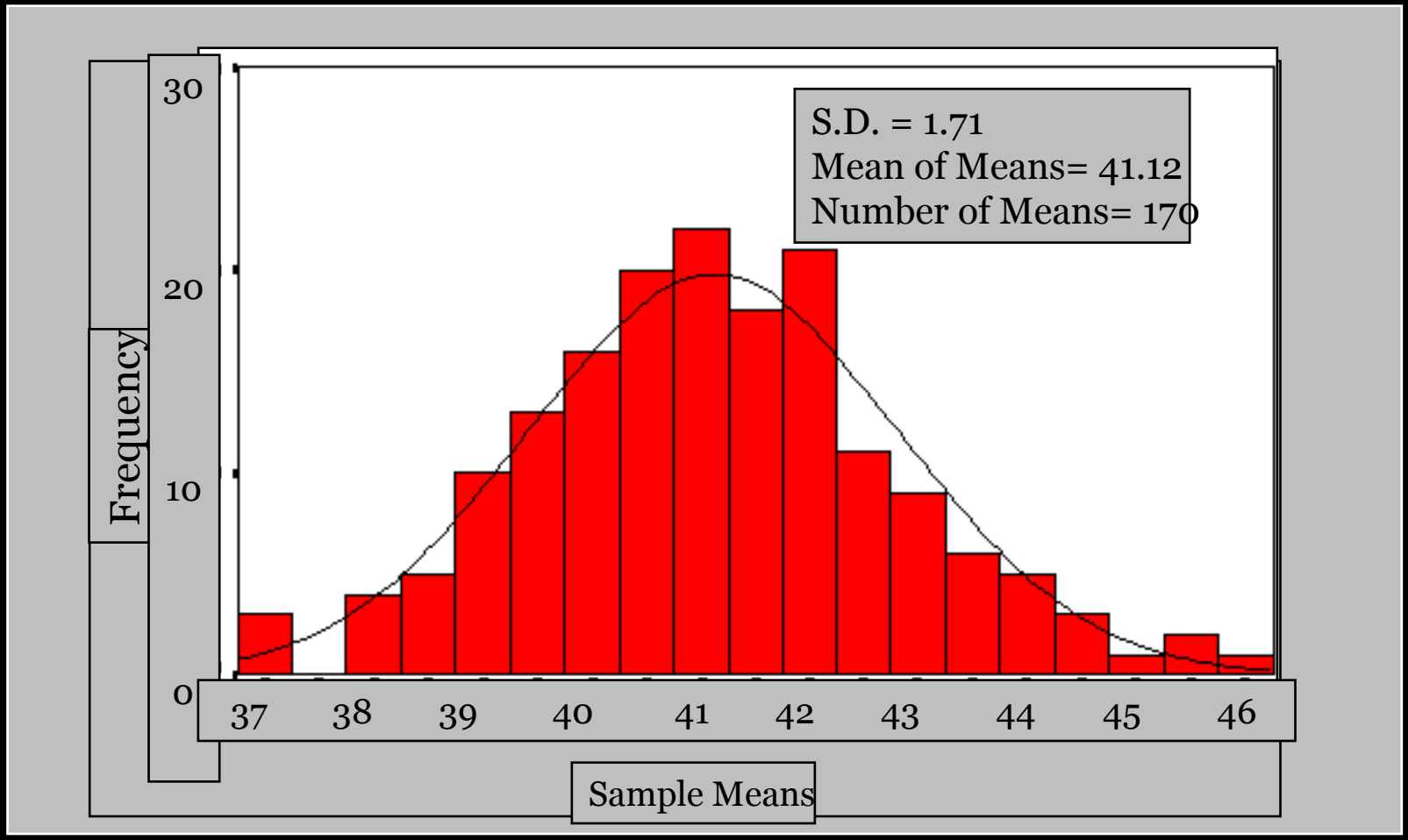
Sampling Distributions



Distribution of Sample Means with 21 Samples



Distribution of Sample Means with 170 Samples



The Central Limit Theorem

- As sample size increases, the sampling distribution of sample means becomes approximately normal.
- With a sufficient sample size, we can use the sampling distribution (which has the same properties as the normal distribution) to test hypotheses (our next class topic…)

See You Wednesday!