Lecture 1

Introduction to sampling distributions (distribuciones de muestreo)

7.1 Sampling Error: What It Is and Why It Happens

- The sample may not be a perfect representation of the population
- Sampling Error
  - The difference between a measure computed from a sample (a statistic) and the corresponding measure computed from the population (a parameter)

$$\text{Sampling error} = \bar{x} - \mu$$

$\bar{x}$ - Sample mean
μ - Population mean

Sampling Error

- Parameter
  - A measure computed from the entire population. As long as the population does not change, the value of the parameter will not change. It can also be interpreted as a constant of the mathematical model used to study the population or system.
- Simple Random Sample
  - A sample selected in such a manner that each possible sample of a given size has an equal chance of being selected.

Population and Sample Mean

Population Mean

$$\mu = \frac{\sum x}{N}$$

μ - Population mean
x - Values in the population
N - Population size

Sample Mean

$$\bar{x} = \frac{\sum x}{n}$$

$\bar{x}$ - Sample mean
x - Sample values selected from the population
n - Sample size

Example: If the population mean is μ = 98.6 degrees and a sample of n = 5 temperatures yields a sample mean of $\bar{x}$ = 99.2 degrees, then the sampling error is:

$$\bar{x} - \mu = 99.2 - 98.6 = 0.6 \text{ degrees}$$

Population and Sample Mean

- The population mean previously defined can also be seen as the mean or expected value of discrete random variable.
- The sample mean previously defined is called the arithmetic mean of a finite set of numbers.

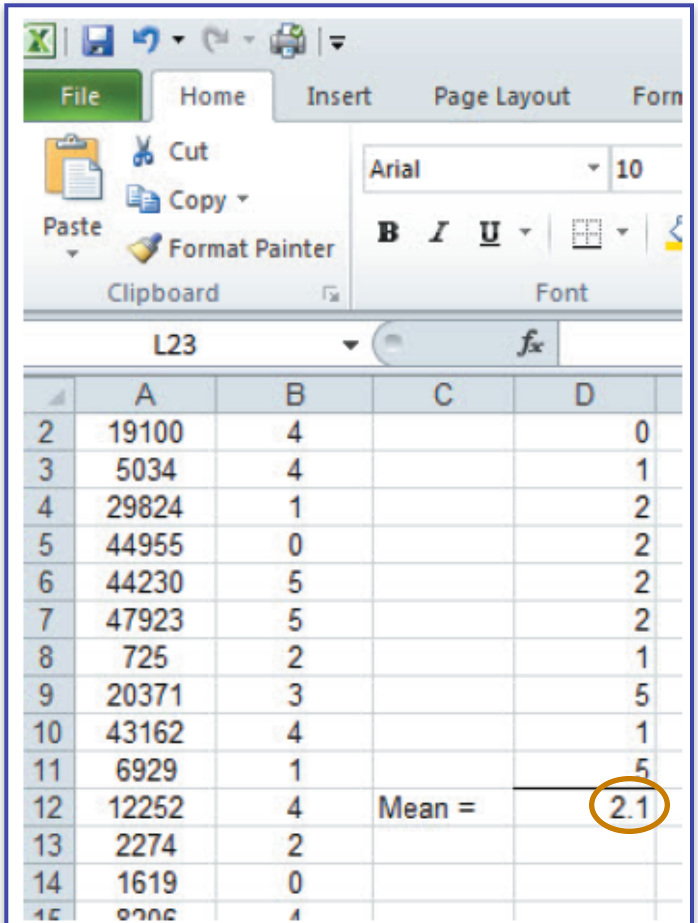
Sampling Error

- The size of the sampling error depends on which sample is selected.
- The sampling error may be positive or negative.
- There is potentially a different $\bar{x}$ for each possible sample.

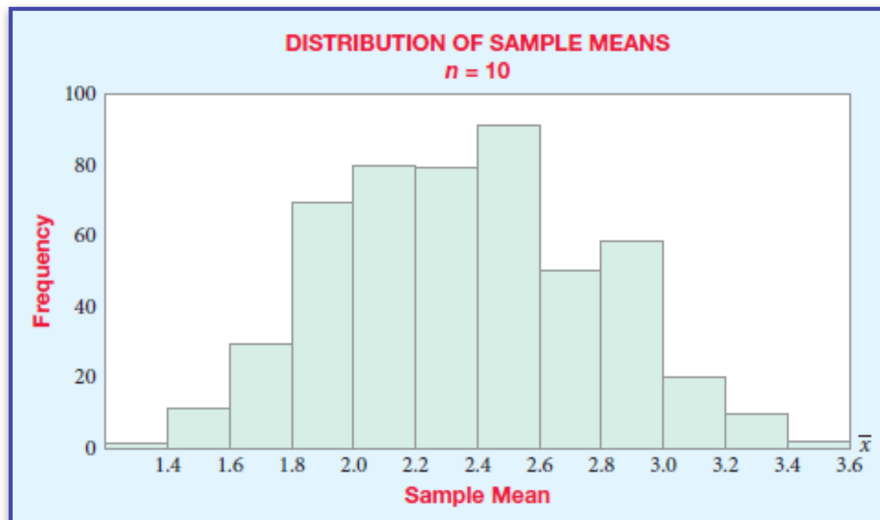
7.2 Sampling Distribution of the Mean

- Sampling Distribution
  - The distribution of all possible values of a statistic for a given sample size that has been randomly selected from a population
- Excel tool can be used to build sampling distribution

Using Excel for Sampling Distribution



1. Open file.
2. Select **Data > Data Analysis**
3. Select **Sampling**
4. Define the population data range.
5. Select **Random, Number of Samples**: 10.
6. Select **Output Range**: D2.
7. Compute sample mean using Average function.



Average Value of Sample Means

- Theorem 1:
  - For any population, the average value of all possible sample means computed from all possible random samples of a given size from the population will equal the population mean

$$\mu_{\bar{x}} = \mu$$

  - Unbiased Estimator: A characteristic of certain statistics in which the average of all possible values of the sample statistic equals parameter

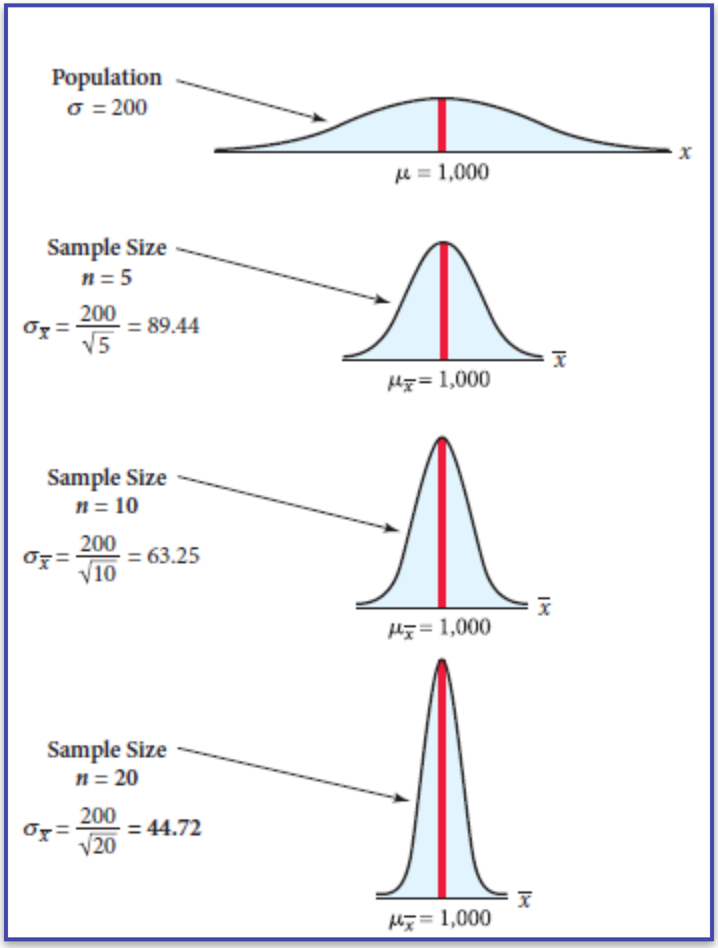
Standard Deviation of Sample Means

- Theorem 2:
  For any population, the standard deviation of the possible sample means computed from all possible random samples of size *n* is equal to the population standard deviation divided by the square root of the sample size. Also called standard error.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Consistent Estimator

- An unbiased estimator is said to be a consistent estimator if the difference between the estimator and the parameter tends to become smaller as the sample size becomes larger.
- Recall definition of the standard deviation or variance.



z-Value for Sampling Distribution of $\bar{x}$

- The relative distance that a given sample mean is from the center can be determined by *standardizing* the sampling distribution
- A *z-value* measures the number of standard deviations a value is from the mean

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{x}$ - Sample mean
μ - Population mean
σ - Population standard deviation
n - Sample size

z-Value Corrected for Finite Population

- The sample is large relative to the size of the population (greater than 5% of the population size), and the sampling is being done without replacement
- Use the finite population correction factor to calculate z-value

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

N- Population size

n - Sample size

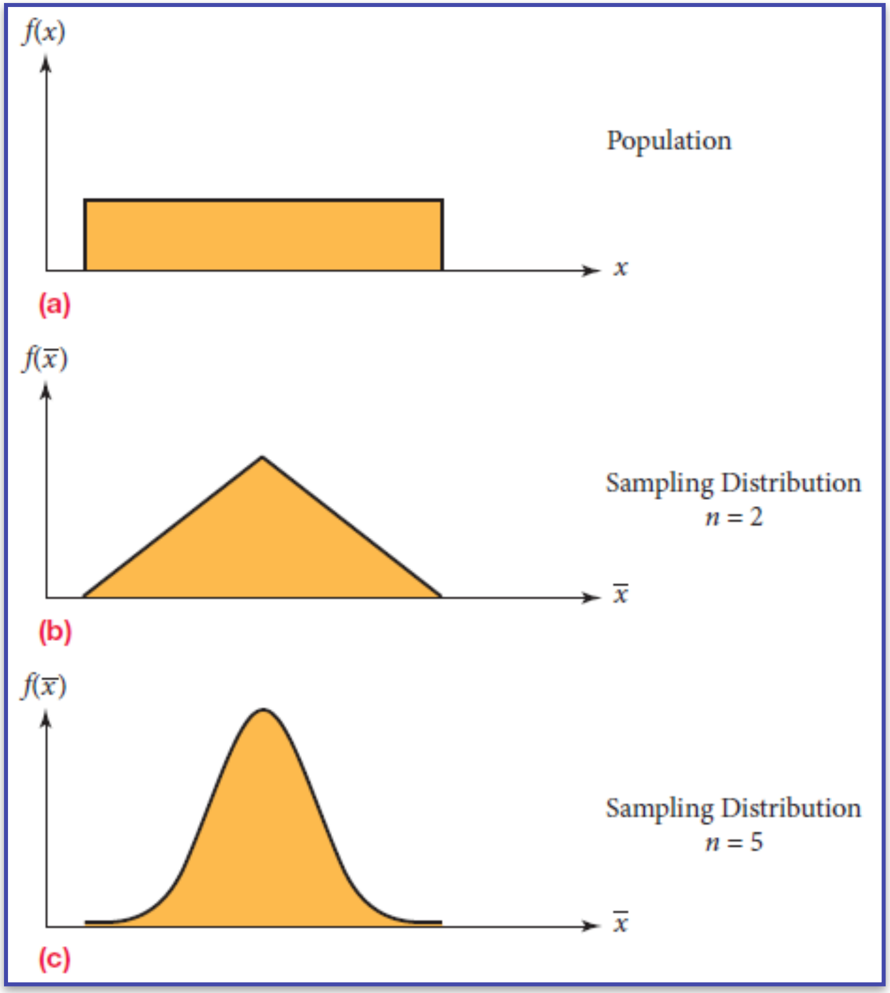
$\sqrt{\frac{N-n}{N-1}}$ - Finite population correction factor

The Central Limit Theorem

- Population may not be normally distributed
- Theorem 4: For simple random samples of n observations taken from a population with mean μ and standard deviation σ, regardless of the population's distribution, provided the sample size is sufficiently large, the distribution of the sample means, $\bar{x}$, will be approximately normal with a mean equal to the population mean ($\mu_{\bar{x}} = \mu$) and a standard deviation equal to the population standard deviation divided by the square root of the sample size ($\sigma_{\bar{x}} = \sigma/\sqrt{n}$).
- The larger the sample size, the better the approximation to the normal distribution.

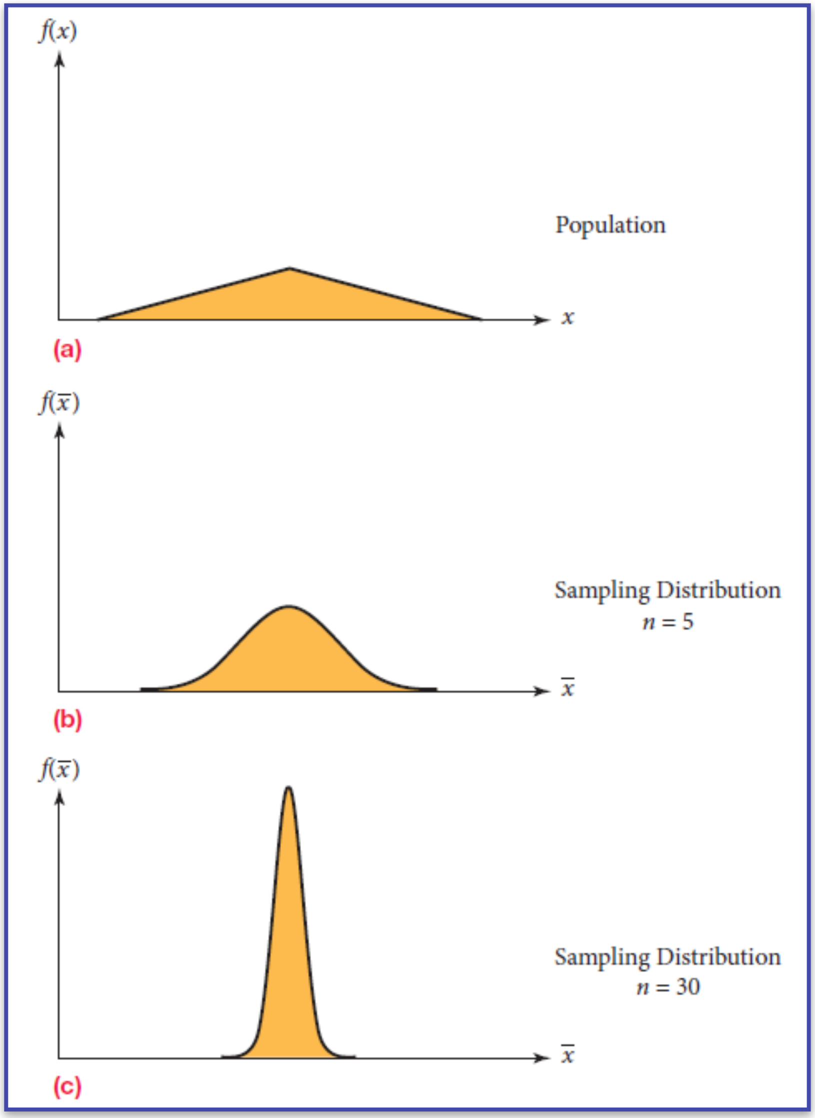
The Central Limit Theorem

- Example:
  Uniform Population Distribution



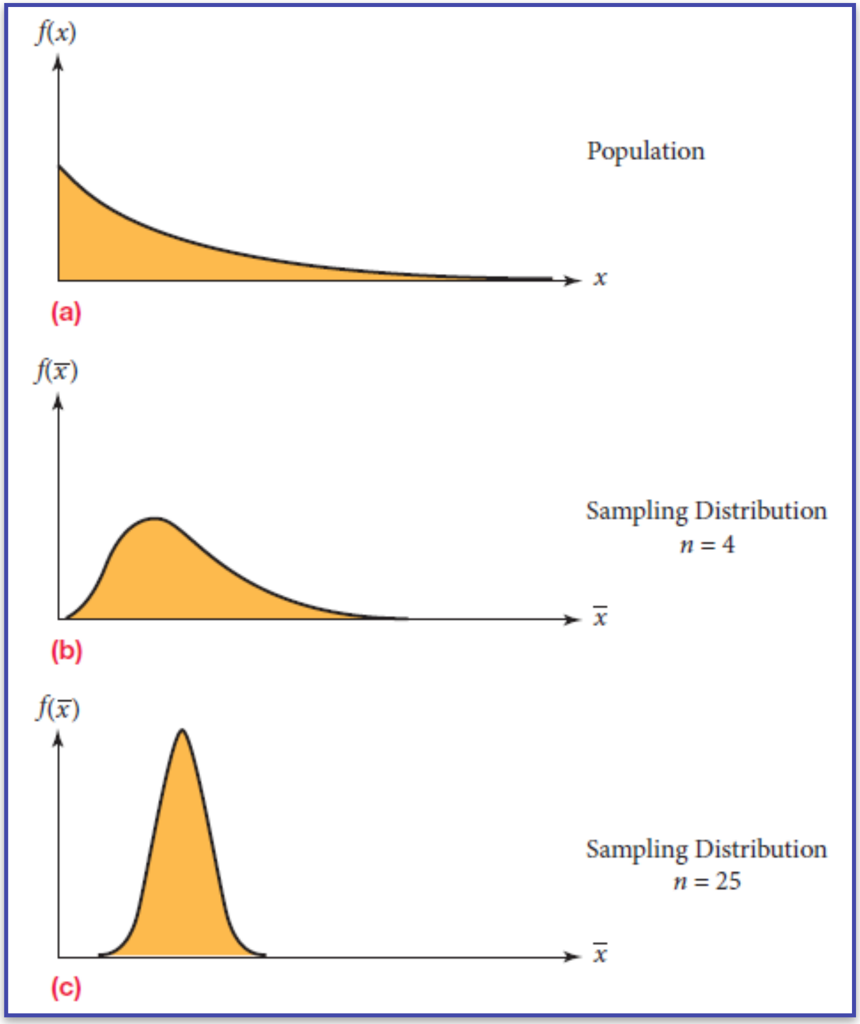
The Central Limit Theorem

- Example:
  Triangular Population Distribution



The Central Limit Theorem

- Example:
  Skewed Population Distribution



The Central Limit Theorem

- The sample size must be "sufficiently large"
- If the population is quite symmetric, then sample sizes as small as 2 or 3 can provide a normally distributed sampling distribution
- If the population is highly skewed or otherwise irregularly shaped, the required sample size will be larger
- A conservative definition of a sufficiently large sample size is $n \geq 30$

The Central Limit Theorem

Example…

7.3 Sampling Distribution of a Proportion

- Population Proportion
  - The fraction of values in a population that have a specific attribute

$$p = \frac{X}{N}$$

p - Population proportion
X - Number of items in the population having the attribute of interest
N - Population size

- Sample Proportion
  - The fraction of items in a sample that have the attribute of interest

$$\bar{p} = \frac{x}{n}$$

$\bar{p}$ - Sample proportion
x - Number of items in the sample with the attribute of interest
n - Sample size

Sampling Error for a Proportion

$$\text{Sampling Error} = \bar{p} - p$$

p - Population proportion
$\bar{p}$ - Sample proportion

- Step 1: Determine the population proportion
- Step 2: Compute the sample proportion
- Step 3: Compute the sampling error

Sampling Distribution of $\bar{p}$

- The best estimate of the population proportion will be $\bar{p}$, the sample proportion
- Any inference about how close your estimate is to the true population value will be based on the distribution of this sample proportion, $\bar{p}$, whose underlying distribution is the binomial
- If the sample size is sufficiently large such that $np \geq 5$ and $n(1 - p) \geq 5$ then the normal distribution can be used as a reasonable approximation to the discrete binomial distribution

Sampling Distribution of $\bar{p}$

$$\text{Mean} = \mu_{\bar{p}} = p$$

$$Standard\ error = \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

p - Population proportion
n - Sample size
$\bar{p}$ - Sample proportion

Sampling Distribution of $\bar{p}$

- Theorem 5: Regardless of the value of the population proportion, p, (with the obvious exceptions of p = 0 and p = 1), the sampling distribution for the sample proportion, $\bar{p}$, will be approximately normally distributed with $\mu_{\bar{p}} = p$ and $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$ providing $np \geq 5$ and $n(1 - p) \geq 5$. The approximation to the normal distribution improves as the sample size increases and p approaches 0.50.

z-Value for Sampling Distribution of $\bar{p}$

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}}$$

Z - Number of standard errors $\bar{p}$ is from p

$\bar{p}$ - Sample proportion

$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$ - Standard error of the sampling distribution

p - Population proportion

If the sample size *n* is greater than 5% of the population size, the standard error of the sampling distribution should be computed using the finite population correction factor

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Using the Sample Distribution for Proportion

- Steps to find probabilities associated with a sampling distributions for proportions:
  - Step 1: Determine the population proportion
  - Step 2: Calculate the sample proportion
  - Step 3: Determine the mean and standard deviation of the sampling distribution
  - Step 4: Define the event of interest
  - Step 5: If np and $n(1 - p)$ are both ≥ 5, then convert $\bar{p}$ to a standardized *z*-value
  - Step 6: Use the standard normal distribution table in Appendix D to determine the required probability.

Using the Sample Distribution for Proportion - Example

- Given a simple random sample: $n = 100 \quad x = 73$
- Step 1: Determine the population proportion $p = 0.80$
- Step 2: Calculate the sample proportion $\bar{p} = \frac{73}{100} = 0.73$
- Step 3: Determine the mean and standard deviation of the sampling distribution $\mu_{\bar{p}} = 0.80 \quad \sigma_{\bar{x}} = \sqrt{\frac{0.80(1-0.80}{100}} = 0.04$
- Step 4: Define the event of interest $P(\bar{p} \leq 0.73) = ?$
- Step 5: Convert the sample proportion to a standardized *z*-value $np = 80 \geq 5 \quad n(1-p) = 20 \geq 5$

$$z = \frac{0.73 - 0.80}{0.04} = -1.75$$

- Step 6: Use the normal distribution table to determine the probability $P(\bar{p} \leq 0.73) = 0.0401$