Sampling and Sampling Distributions

Carlos Hurtado
churtado@richmond.edu

Robins School of Business
University of Richmond

Sep 30, 2019

On the Agenda

1. Selecting a Sample
2. Point Estimation
3. Sampling Distribution of $\bar{x}$
4. Sampling Distribution of $\bar{p}$

On the Agenda

1. Selecting a Sample
2. Point Estimation
3. Sampling Distribution of $\bar{x}$
4. Sampling Distribution of $\bar{p}$

Selecting a Sample

- The reason we select a sample is to collect data to answer a research question about a population
- The sample results provide only estimates of the values of the population characteristics
- The reason is simply that the sample contains only a portion of the population
- With proper sampling methods, the sample results can provide "good" estimates of the population characteristics

Selecting a Sample

- Sampling from a Finite Population

- Sampling from an Infinite Population

Selecting a Sample

- Finite populations are often defined by lists such as
  - Credit card account numbers
  - Students in a university
- A simple random sample of size n from a finite population of size N
  - is a sample selected such that each possible sample of size n has the same probability of being selected

Example: Selecting a Sample

- National Baseball League teams:
  - There are 16 teams that played in 2012 national baseball league
  - We want to select a simple random sample of 5 teams to conduct interviews about how they manage their legal franchises

Example: Selecting a Sample

- National Baseball League teams:

  - Step 1: Assign a random number to each of the 16 teams in the population

  - Step 2: Select the five teams corresponding to the 5 smallest random numbers as the sample

Selecting a Sample

- Sometimes we want to select a sample, but find it is not possible to obtain a list of all elements in the population
- Hence, we cannot use the random number selection procedure
- Most often this situation occurs in infinite population cases

Selecting a Sample

- Example: Populations are often generated by an ongoing process where there is no upper limit on the number of units that can be generated
  - Parts being manufactured on a production line
  - Transactions occurring at a bank
  - Telephone calls arriving at a technical help desk
  - Customers entering a store

Selecting a Sample

- We must select a random sample in order to make valid statistical inferences
- A random sample from an infinite population is a sample selected such that the following conditions are satisfied:
    1. Each element selected comes from the population of interest
    2. Each element is selected independently

On the Agenda

1. Selecting a Sample
2. Point Estimation
3. Sampling Distribution of $\bar{x}$
4. Sampling Distribution of $\bar{p}$

Point Estimation

- Point estimation is a form of statistical inference
- In point estimation we use the data from the sample to compute a value of a sample statistic that serves as an estimate of a population parameter
- We refer to $\bar{x}$ as the point estimator of the population mean μ
- We refer to s as the point estimator of the population standard deviation σ
- We refer to $\bar{p}$ as the point estimator of the population proportion p

Point Estimation

Example: Point Estimation

- EAI Employee Data
  - Out of a total of 2,500 employees, a simple random sample of 30 employees and corresponding data was taken
  - We denote by $x_1, x_2, \cdots, x_n$ the annual salary of the employees
  - We indicate the participation in the management training program as yes/no

Example: Point Estimation

Annual Salary ($)	Management Training Program	Annual Salary ($)	Management Training Program	Annual Salary ($)	Management Training Program
X_1 = 49094.30	YES	45,922.60	YES	45,120.90	YES
X_2 = 53263.90	YES	57,268.40	NO	51,753.00	YES
49,643.50	YES	55,688.40	YES	54,391.80	NO
49,894.90	YES	51,564.70	NO	50,164.20	NO
47,621.60	NO	56,188.20	NO	52,973.60	NO
55,924.00	YES	51,766.00	YES	50,241.30	NO
49,092.30	YES	52,541.30	NO	52,793.90	NO
51,404.40	YES	44980.00	YES	50,979.40	YES
50,957.70	YES	51,932.60	YES	55,860.90	YES
55,109.70	YES	52,973.00	YES	57,309.10	NO

Example: Point Estimation

- EAI Employee Data
  - To estimate the value of population parameter, we can compute the corresponding characteristic of the sample
  - We refer to the characteristic of the sample as <u>sample statistic</u>
  - Different random samples will result in different point estimates!

Example: Point Estimation

- EAI Employee Data
  - Estimate of the mean annual salary: $\bar{x}$ as point estimator of μ

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1,554,420}{30} = \$51,814$$

  - Estimate of the standard deviation of the annual salary: s as point estimator of σ

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{325,009,260}{n-1}} = \$3,348$$

  - Estimate of the proportion of workers with management training: $\bar{p}$ as point estimator of p

$$\bar{p} = \frac{\sum q_i}{n} = \frac{19}{30} = 0.63$$

Example: Point Estimation

Population Parameter	Parameter value	Point estimator	Point estimate
μ = Population mean annual salary	$ 51,800	$\bar{x}$ = Sample mean annual salary	$ 51,814
σ = Population standard deviation for annual salary	$ 4,000	s = Sample standard deviation for annual salary	$ 3,348
p = Population proportion having completed MTP	.60	$\bar{p}$ = Sample proportion having completed the MTP	.63

Point Estimation

- The target population is the population we want to make inferences about

- The sampled population is the population from which the sample is actually taken

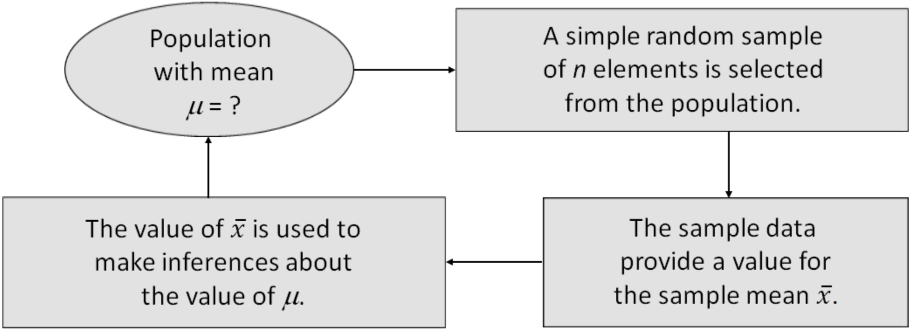
- Whenever a sample is used to make inferences about a population:
  - we should make sure that the targeted population and the sampled population are in close agreement

On the Agenda

1. Selecting a Sample
2. Point Estimation
3. Sampling Distribution of $\bar{x}$
4. Sampling Distribution of $\bar{p}$

Sampling Distribution of $\bar{x}$

- Process of Statistical Inference



Sampling Distribution of $\bar{x}$

- The sampling distribution of $\bar{x}$ is the probability distribution of all possible values of the sample mean

- Expected value of $\bar{x}$

$$E(\bar{x}) = \mu$$

  where $\mu =$ the population mean

- When the expected value of the point estimator equals the population parameter, we say the point estimator is unbiased

Sampling Distribution of $\bar{x}$

- We will use the following notations to define the standard deviation of the sampling distribution of $\bar{x}$:

$\sigma_{\bar{x}}=$ the standard deviation of $\bar{x}$

$\sigma=$ the standard deviation of the population

$n=$ the sample size

$N=$ the population size

Sampling Distribution of $\bar{x}$

- Variance of $\bar{x}$:

Finite Population

$$\sigma_{\bar{x}}^2 = \frac{N-n}{N-1}\left(\frac{\sigma^2}{n}\right)$$

Infinite Population

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

- A finite population is treated as being infinite if $\frac{n}{N} \leq 5\%$
- The term $\frac{N-n}{N-1}$ is the finite population correction factor for the variance of the mean
- $\sigma_{\bar{x}}$ is referred to as the standard error of the mean

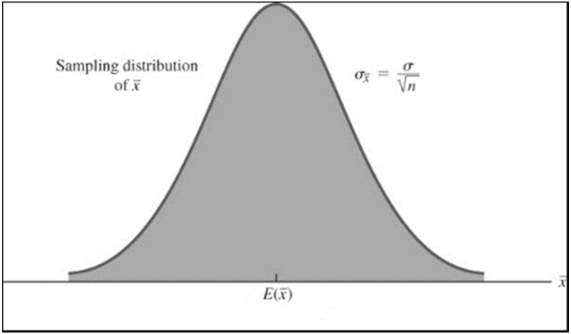
Sampling Distribution of $\bar{x}$

- When the population has a normal distribution, the sampling distribution of $\bar{x}$ is normally distributed for any sample size
- The sampling distribution of $\bar{x}$ can be approximated by a normal distribution whenever the sample is size 30 or more
- In cases where the population is highly skewed or outliers are present, samples of size 50 may be needed

Sampling Distribution of $\bar{x}$

- The sampling distribution of $\bar{x}$:
  - can be used to provide probability information about

    how close the sample mean $\bar{x}$ is to the population mean μ



Sampling Distribution of $\bar{x}$

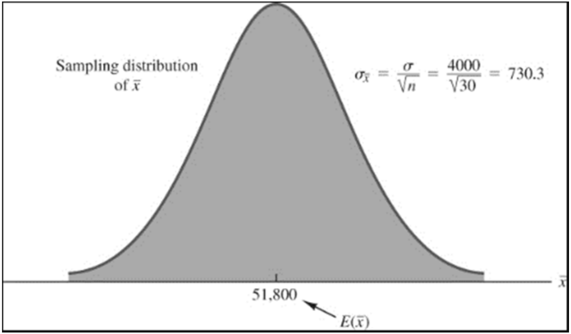
- CENTRAL LIMIT THEOREM:
  - In selecting random samples of size n from a population:

    the sampling distribution of the sample mean $\bar{x}$ can be approximated by a normal distribution as the sample size becomes large

- When the population from which we are selecting a random sample does not have a normal distribution:
  - the central limit theorem is helpful in identifying the shape of the sampling distribution of $\bar{x}$

Example: Sampling Distribution of $\bar{x}$

- EAI Employee Data
- Note that $n/N = 30/2,500 = 1.2\%$
- Because sample is less than 5% of the population size, we ignore the finite population correction factor



Example: Sampling Distribution of $\bar{x}$

- EAI Employee Data
- The director believes the sample mean will be an acceptable estimate of population mean if the sample mean is within $500 of the population mean
- What is the probability that the sample mean computed using a simple random sample of 30 employees will be within $500 of population mean?

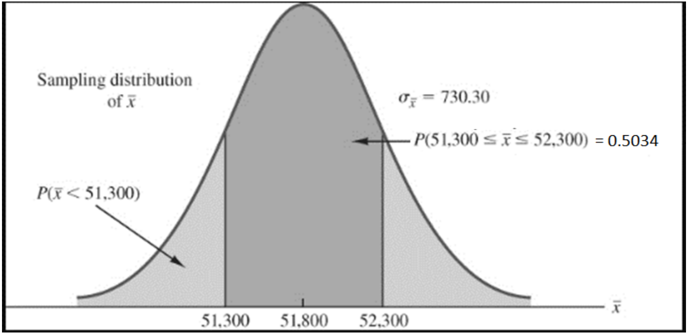
Example: Sampling Distribution of $\bar{x}$

- EAI Employee Data
- The question in mathematical terms:

$$P(|\bar{x} - E(\bar{x})| \leq \$500)$$

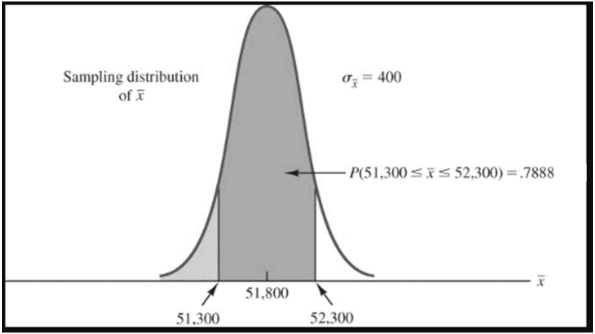
- A little bit of algebra to get:

$$P(51,300 \leq \bar{x} \leq 52,300)$$



Example: Sampling Distribution of $\bar{x}$

- Suppose that the random sample had 100 employees instead of the 30
- And $E(\bar{x})$ remains at 51,800
- Whenever the sample size is increased, the standard error of the mean $\sigma_{\bar{x}}$ is decreased

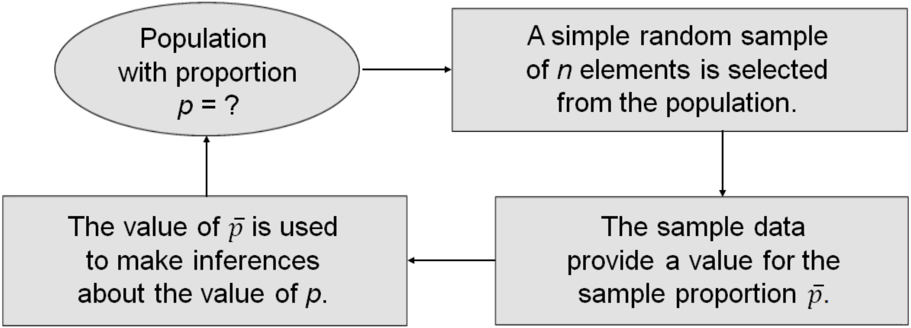


On the Agenda

1. Selecting a Sample
2. Point Estimation
3. Sampling Distribution of $\bar{x}$
4. Sampling Distribution of $\bar{p}$

Sampling Distribution of $\bar{p}$

- Making inferences about a population proportion



Sampling Distribution of $\bar{p}$

- The sampling distribution of $\bar{p}$ is the probability distribution of all possible values of the sample proportion

- Expected Value of $\bar{p}$:

$$E(\bar{p}) = p$$

where: $p=$ the population proportion

Sampling Distribution of $\bar{p}$

- Variance of $\bar{p}$:

Finite Population

$$\sigma^2_{\bar{p}} = \frac{N-n}{N-1}\left(\frac{p(1-p)}{n}\right)$$

Infinite Population

$$\sigma^2_{\bar{p}} = \frac{p(1-p)}{n}$$

- The term $\frac{N-n}{N-1}$ is the finite population correction factor
- $\sigma_{\bar{p}}$ is referred to as the standard error of the proportion

Sampling Distribution of $\bar{p}$

- The sampling distribution of $\bar{p}$ can be approximated by a normal distribution when the sample size is large enough

$$np \geq 5 \quad \text{and} \quad n(1-p) \geq 5$$

- When these conditions are satisfied the sampling distribution of $\bar{p}$ can be approximated by a normal distribution

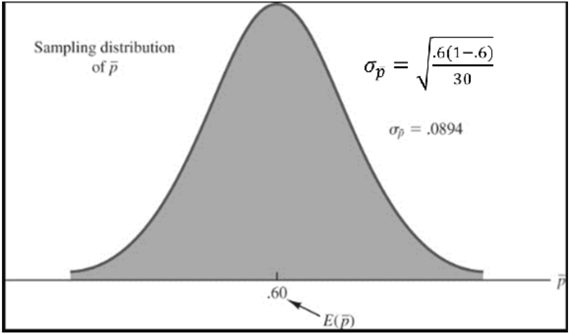
Example: Sampling Distribution of $\bar{p}$

- EAI Employee Data
- We know that the population proportion of employees with management training is $p = 0.6$
- What is the probability that a random sample of 30 employees will provide an estimate of the population proportion that is within plus or minus 0.05 of the actual population proportion?

Example: Sampling Distribution of $\bar{p}$

- EAI Employee Data
- For our example, with n = 30 and p = 0.6, the normal distribution is an acceptable approximation because

$$np = 30 \times 0.6 = 18 > 5 \quad \text{and} \quad n(1-p) = 30 \times 0.4 = 12 > 5$$



Example: Sampling Distribution of $\bar{p}$

- EAI Employee Data
- The question in mathematical terms:

$$P(|\bar{p} - E(\bar{p})| \leq 0.05)$$

- A little bit of algebra to get:

$$P(0.55 \leq \bar{p} \leq 0.65)$$