Some Important Sampling Distribution (I)

Huining Kang

HuKang@salud.unm.edu

August 12, 2020

Outline

- Content
  - Population and sample
  - Population parameters and sample statistics
  - Population distribution and sampling distribution
  - Distribution of the sample mean for the finite population
- Objectives
  - Understand the concepts of the population and the sample
  - Understand sampling with or without replacement
  - Understand the difference and relationship between population parameters and sample statistics
  - Understand the distribution of sample mean and be able to construct the distribution of sample means for a simple finite population.

Population and Sample

- Population
  - The largest collection of entities for which we have an interest at a particular time
- Population of values
  - The largest collection of values of a random variable for which we have an interest at a particular time
  - May be finite or infinite
  - The members of population is usually denoted by $x_1, x_2, \cdots, x_N$
- Sample
  - A part of a population, usually denoted by $X_1, X_2, \cdots, X_n$
  - The observation of a sample is usually denoted by $x_1, x_2, \cdots, x_n$
- Simple Random Sample (SRS)
  - A sample of size n is drawn from a population of size of N in such a way that every possible sample of size n has the same chance of being selected.
  - Sampling with replacement
  - Sampling without replacement
  - Each member of the SRS has an equal probability of being chosen
  - An SRS is meant to be an unbiased representation of the population

Population Parameters and Sample statistics

- Population parameters
  - Descriptive measures computed from all the data of a population
  - Suppose $x_1, x_2, \cdots, x_N$ are all the values of a (finite) population
    - Population mean: $\mu = \frac{x_1+x_2+\cdots+x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$
    - Population variance: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$
- Sample statistics
  - Descriptive measures computed from the data of a sample
  - Suppose $X_1, X_2, \cdots, X_n$ is a sample of size n drawn from a population
    - Sample mean: $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$
    - Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
    - A statistic is a function of the sample $X_1, X_2, \cdots, X_n$.
- Difference
  - Population parameters are fixed numbers and often unknown.
  - Sample statistics are random variables
  - Given the observations of a sample $x_1, x_2, \cdots, x_n$, we call $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ the observed value of the sample mean $\bar{X}$. The same to other statistics such the sample variance S^2.
- Note
  - Out textbook uses the lower case characters to denote the samples and sample statistics.

Population Distribution and Sampling Distribution

- Population distribution
  - Suppose X is a random sample of size one drawn from a population, then the probability distribution of X is defined as the distribution of the population
  - Example
    - The population of heights of adult male in the US. $X \sim N(\mu, \sigma^2)$ with $\mu = 70.9$ and variance $\sigma^2 = 2.75^2$. The population distribution is a normal distribution.
- Sampling Distribution
  - A statistic T (e.g. sample mean $\bar{X}$, sample variance S^2) is a random variable.
  - The probability distribution of T is called the sampling distribution (of T).
  - Example
    - Suppose $X_1, X_2, \cdots, X_n$ is a sample of size n drawn from the above population. Later we will see that $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ also follows a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, i.e. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
  - Definition from the text book
    - The distribution of all possible values that can be assumed by some statistic, computed from samples of the same size randomly drawn from the same population, is called the sampling distribution of that statistic.

Sampling Distribution

- Purposes
  - Allow us to answer probability questions about sample statistics
  - Provide theory for making statistical inference (estimation, hypothesis testing)
- Importance
  - Bridge between probability theory and statistical inference
  - Key to understanding statistical inference
- Construction
  - From a finite population of size *N*, randomly draw all possible samples of size *n*.
  - Compute the statistic of interest for each sample.
  - List in one column the different distinct observed values of the statistic, and in another column list the corresponding frequency of occurrence of each distinct observed value of the statistic.
- Important Characteristics
  - Mean
  - Variance
  - Functional form

Height example

$$E(\bar{X}) = \mu, Var(\bar{X}) = \frac{\sigma^2}{n},\ \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Distribution of sample mean

- Example 5.3.1 (page 136, 10th Ed)
  - A population of size $N = 5$

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

  - Population distribution

X	x				
	6	8	10	12	14
$p(x) = \mathrm{P(X = x)}$	1/5	1/5	1/5	1/5	1/5

  - Population Mean
    - $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{6+8+10+12+14}{5} = 10,$
    - $\mu = E(X) = \sum_{i=1}^{N} x_i p(x_i) = 6 \times \frac{1}{5} + 8 \times \frac{1}{5} + 10 \times \frac{1}{5} + 12 \times \frac{1}{5} + 14 \times \frac{1}{5} = 10$
  - Population variance
    - $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{(6-10)^2+(8-10)^2+(10-10)^2+(12-10)^2+(14-10)^2}{5} = 8$
    - $\sigma^2 = Var(X) = \sum_{i=1}^{N}(x_i - \mu)^2 p(x_i) = 8$

Distribution of sample mean

- Example 5.3.1 (Cont.)
  - A population of size $N = 5$

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

  - Problem
    - Draw a sample of size $n = 2$ (with replacement) X_1, X_2 and find the distribution of sample mean $\bar{X} = (X_1 + X_2)/2$.
  - Solution

All possible samples of size $n = 2$

		Second Draw (X_2)				
		6	**8**	**10**	**12**	**14**
First Draw	**6**	6, 6	6, 8	6, 10	6, 12	6, 14
		(6)	(7)	(8)	(9)	(10)
	8	8, 6	8, 8	8, 10	8, 12	8, 14
		(7)	(8)	(9)	(10)	(11)
(X_1)	**10**	10, 6	10, 8	10, 10	10, 12	10, 14
		(8)	(9)	(10)	(11)	(12)
	12	12, 6	12, 8	12, 10	12, 12	12, 14
		(9)	(10)	(11)	(12)	(13)
	14	14, 6	14, 8	14, 10	14, 12	14, 14
		(10)	(11)	(12)	(13)	(14)

Sampling distribution of $\bar{X}$

$\bar{x}$	**Frequency**	**Relative Frequency**
6	1	1/25
7	2	2/25
8	3	3/25
9	4	4/25
10	5	5/25
11	4	4/25
12	3	3/25
13	2	2/25
14	1	1/25
Total	25	25/25

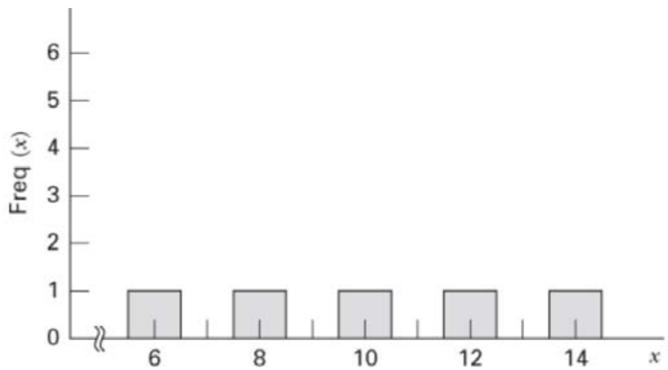
Distribution of sample mean

- Example 5.3.1 (Cont.)
  - A population of size $N = 5$

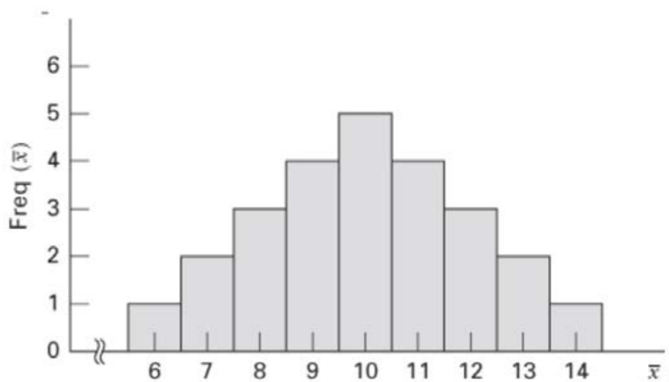
x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

  - Population distribution

X	x				
	6	8	10	12	14
$p(x) = \mathrm{P}(\mathrm{X} = \mathrm{x})$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$



Distribution of population



Sampling distribution of $\bar{X}$

Sampling distribution of $\bar{X}$

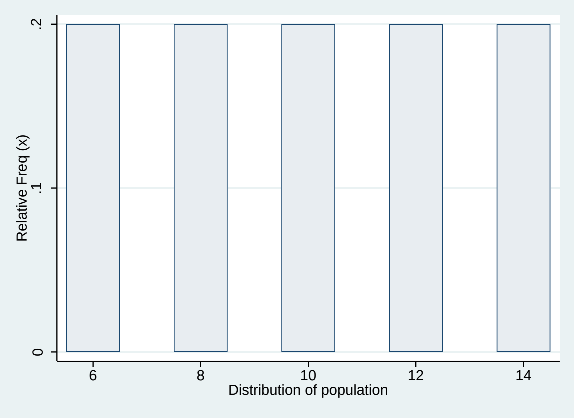
$\bar{x}$	**Frequency**	**Relative Frequency**
6	1	1/25
7	2	2/25
8	3	3/25
9	4	4/25
10	5	5/25
11	4	4/25
12	3	3/25
13	2	2/25
14	1	1/25
Total	25	25/25

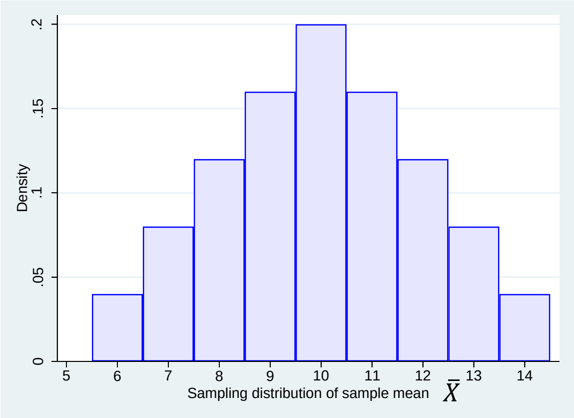
Distribution of sample mean

- Example 5.3.1 (Cont.)
  - A population of size $N = 5$
  - Population distribution

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

X	x				
	6	8	10	12	14
$p(x) = \mathrm{P}(\mathrm{X} = \mathrm{x})$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$





Sampling distribution of $\bar{X}$

$\bar{x}$	Frequency	Relative Frequency
6	1	1/25
7	2	2/25
8	3	3/25
9	4	4/25
10	5	5/25
11	4	4/25
12	3	3/25
13	2	2/25
14	1	1/25
Total	25	25/25

Distribution of sample mean

- Example 5.3.1 (Cont.)
  - A population of size $N = 5$

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

  - Population distribution

X	x				
	6	8	10	12	14
$p(x) = \mathrm{P(X = x)}$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$

  - Characteristics of population distribution
    - Mean: $\mu = E(X) = 10$
    - Variance: $\sigma^2 = Var(X) = 8$
  - Characteristics of sampling distribution
    - Mean: $\mu_{\bar{X}} = E(\bar{X}) = \frac{6\times1+7\times2+\cdots+14\times1}{25} = 10$
    - Variance: $\sigma_{\bar{X}}^2 = Var(\bar{X}) = (6-10)^2 \times \frac{1}{25} + \cdots$
      $+(14-10)^2 \times \frac{1}{25} = 4$
  - Relationship
    - Mean of the sample mean is equal to the population mean, i.e. $\mu_{\bar{X}} = \mu$
    - Variance of the sample mean is equal to the population variance divided by the sample size, i.e. $\sigma_{\bar{X}}^2 = \sigma^2/n$
    - Standard error (of the mean): $se = \sigma/\sqrt{n}$.

Sampling distribution of $\bar{X}$

$\bar{x}$	**Frequency**	**Relative Frequency**
6	1	1/25
7	2	2/25
8	3	3/25
9	4	4/25
10	5	5/25
11	4	4/25
12	3	3/25
13	2	2/25
14	1	1/25
Total	25	25/25

Distribution of sample mean

- Example 5.3.1 (Cont.)
  - A population of size $N = 5$
  - Problem
    - Draw a sample of size $n = 2$ <u>without replacement</u> X_1, X_2 and find the distribution of sample mean $\bar{X} = (X_1 + X_2)/2$.
  - Solution

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

All possible samples of size $n = 2$

		Second Draw (X_2)				
		6	**8**	**10**	**12**	**14**
First Draw (X_1)	**6**		6, 8 (7)	6, 10 (8)	6, 12 (9)	6, 14 (10)
	8			8, 10 (9)	8, 12 (10)	8, 14 (11)
	10				10, 12 (11)	10, 14 (12)
	12					12, 14 (13)
	14					

Sampling distribution of $\bar{X}$

$\bar{X}$	Freq	Relative Freq
7	1	1/10
8	1	1/10
9	2	2/10
10	2	2/10
11	2	2/10
12	1	1/10
13	1	1/10
Total	10	1

- Mean: $\mu_{\bar{X}} = E(\bar{X}) = \frac{(7+8+12+13)\times 1+(9+10+11)\times 2}{10} = 10$
- Variance: $\sigma_{\bar{X}}^2 = Var(\bar{X}) = \frac{[(7-10)^2+(8-10)^2+(12-10)\text{^}2+(13-10)^2]\times 1+[(9-10)^2+(10-10)^2+(11-10)^2]\times 2}{10} = 3$

Distribution of sample mean

- Example 5.3.1 (Cont.)
  - A population of size $N = 5$
  - Problem
    - Draw a sample of size $n = 2$ without replacement X_1, X_2 and find the distribution of sample mean $\bar{X} = (X_1 + X_2)/2$.
  - Solution
    - Mean $\mu_{\bar{X}} = E(\bar{X})$
$= \frac{(7+8+12+13)\times 1 + (9+10+11)\times 2}{10} = 10$
    - Variance: $\sigma_{\bar{X}}^2 = Var(\bar{X})$
$= (7-10)^2 \times \frac{1}{10} + \cdots + (13-10)^2 \times \frac{1}{10} = 3$
    - $\sigma_{\bar{X}}^2 \neq \sigma^2/n$

x_1	x_2	x_3	x_4	x_5
6	8	10	12	14

Sampling distribution of $\bar{X}$

$\bar{X}$	Freq	Relative Freq
7	1	1/10
8	1	1/10
9	2	2/10
10	2	2/10
11	2	2/10
12	1	1/10
13	1	1/10
Total	10	1

Distribution of sample mean

- Sampling with replacement
  - Mean of the sample mean is equal to the population mean, i.e. $\mu_{\bar{X}} = \mu$
  - Variance of the sample mean is equal to the population variance divided by the sample size, i.e. $\sigma_{\bar{X}}^2 = \sigma^2/n$
  - Standard error (of the mean): $se = \sigma/\sqrt{n}$.
- Sampling without replacement
  - Mean of the sample mean is equal to the population mean, i.e. $\mu_{\bar{X}} = \mu$
  - Variance of the sample mean is not equal to the population variance divided by the sample size, i.e. $\sigma_{\bar{X}}^2 \neq \sigma^2/n$
  - $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$
- The finite population correction
  - $\frac{N-n}{N-1}$ is called finite population correction factor
  - It can be ignored if N is large (e.g. $\frac{n}{N} \leq .05$)
  - It is seldom used in practice.

Summary

- We reviewed concepts related to a population and a sample
- We learned population distribution and sample distribution
- We learned the distribution of sample mean and how to construct the sampling distribution of a sample mean for a simple finite population. The purpose is to enable use to understand the concepts better.
- It is very important to have a clear understanding of the sampling distributions

Homework 08

- Due date: Tuesday, August 24 (same as the previous homework)
- Exercises 5.3.8 (page 145, 10th Ed)