Module 12: Populations, Samples and Sampling Distributions

This module provides basic information about the statistical concepts of populations and samples, selecting samples from population and the critical issue of sampling distributions.

Populations and Samples

- **Population:** The entire group about which information is desired.
- **Sample:** A proportion or part of the population - usually the proportion from which information is gathered.

Target Population

- The participants to whom the answer to the question pertains.

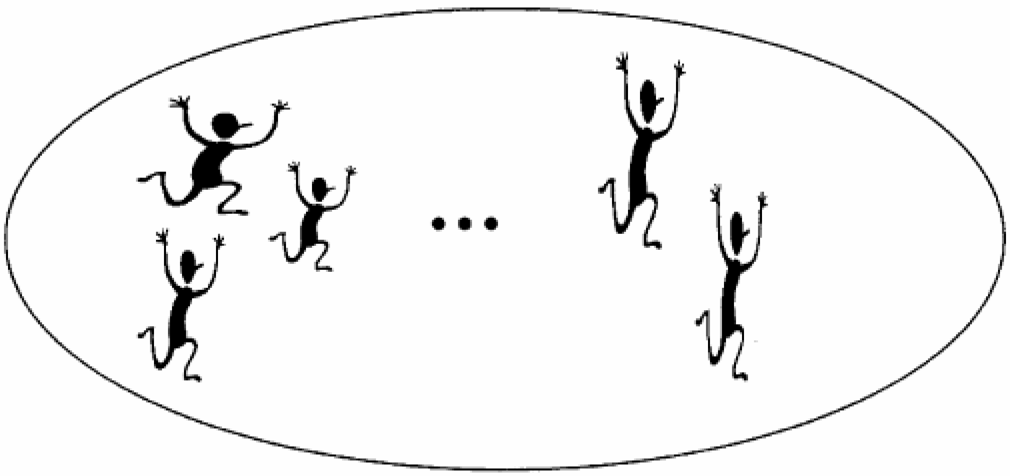
- The target population definition has two aspects:
    - Conceptual
    - Operational

Population Definition

A ***population definition*** gives a clear statement of those included. The following are some examples:

- Adults and children 10-59 years of age residing in four census tracts in Richfield, a suburb of Minneapolis
- Adults 25-59 years of age residing in Cedar County, Iowa and certain rural townships in neighboring counties on July 1, 1973
- Employees of Pacific Northwest Bell Telephone Company working in King County

Population



Population of Persons

Person	Population of Cholesterol values (mg/dl)
1	201
2	182
3	199
.	.
.	.
.	.
128	124
129	180

Population of Cholesterol values

1	201	**26**	172	**51**	162	**76**	127	**101**	206	**126**	172
2	182	**27**	164	**52**	151	**77**	147	**102**	159	**127**	276
3	199	**28**	136	**53**	197	**78**	164	**103**	141	**128**	124
4	136	**29**	161	**54**	206	**79**	161	**104**	166	**129**	180
5	152	**30**	160	**55**	160	**80**	178	**105**	154		
6	195	**31**	165	**56**	131	**81**	177	**106**	111		
7	162	**32**	169	**57**	193	**82**	176	**107**	228		
8	206	**33**	159	**58**	235	**83**	146	**108**	95		
9	138	**34**	168	**59**	192	**84**	179	**109**	188		
10	190	**35**	185	**60**	221	**85**	185	**110**	134		
11	152	**36**	189	**61**	194	**86**	155	**111**	198		
12	120	**37**	174	**62**	153	**87**	150	**112**	140		
13	169	**38**	114	**63**	168	**88**	167	**113**	188		
14	136	**39**	161	**64**	162	**89**	154	**114**	154		
15	141	**40**	153	**65**	162	**90**	159	**115**	191		
16	194	**41**	165	**66**	158	**91**	187	**116**	169		
17	173	**42**	142	**67**	143	**92**	164	**117**	156		
18	158	**43**	173	**68**	184	**93**	151	**118**	141		
19	181	**44**	138	**69**	133	**94**	155	**119**	172		
20	247	**45**	174	**70**	180	**95**	159	**120**	206		
21	192	**46**	186	**71**	165	**96**	261	**121**	145		
22	192	**47**	175	**72**	149	**97**	169	**122**	138		
23	123	**48**	164	**73**	155	**98**	137	**123**	170		
24	149	**49**	220	**74**	129	**99**	154	**124**	151		
25	158	**50**	150	**75**	217	**100**	189	**125**	154		

Sampling

- In its broadest sense, sampling is a procedure by which one or more members of a population are picked from the population.
- The objective is to make certain observations upon the members of the sample and then, on the basis of these results, to draw conclusions about the characteristics of the entire population.

Selecting a Sample

- **Haphazard Sample:** Haphazard samples are constructed by arbitrarily selecting individual sample members.

- **Random Sample:** There are several methods for constructing random samples—we consider only simple random samples. This process operates so that each member of the population has an equal chance of being selected into the sample.

Selecting a Sample

The selection process:

- Assign to each member of the population the equivalent of sequential ID number;
- Use a random number table or computer generated numbers;
- For computer generated numbers, generate one for each ID number, sort the ID numbers in order according to the random number and take the first on the list up to the point when you have the sample size you need
- For a table, haphazardly select a starting point and then
    - Ignore numbers that are too large
    - Ignore a number after it appears the first time

Fundamental and Important Concept

We now begin the discussion of perhaps the most important concept in biostatistics. It is fundamental to understanding and thus interpreting correctly the use of the many statistical tools we will cover in this course. The concept is not complex, in fact, it is rather simple. It does require, however, thinking about issues in a manner that may initially appear somewhat different and unusual.

Looking at the Process

When we randomly select a sample from a population, we can use the mean for the sample as an estimate or guess as to the value for the mean of the population. This should bring up the question as to how good is this sample mean or sample statistic as a guess for the value of the population mean or population parameter.

The essence of this question has to do with how well this process works—the process of using a sample to make guesses about the population.

Understanding the Process

Two important aspects of this fundamental process:

FIRST: It is critical to recognize that it is a process

SECOND: It is important to understand how and how well the process works

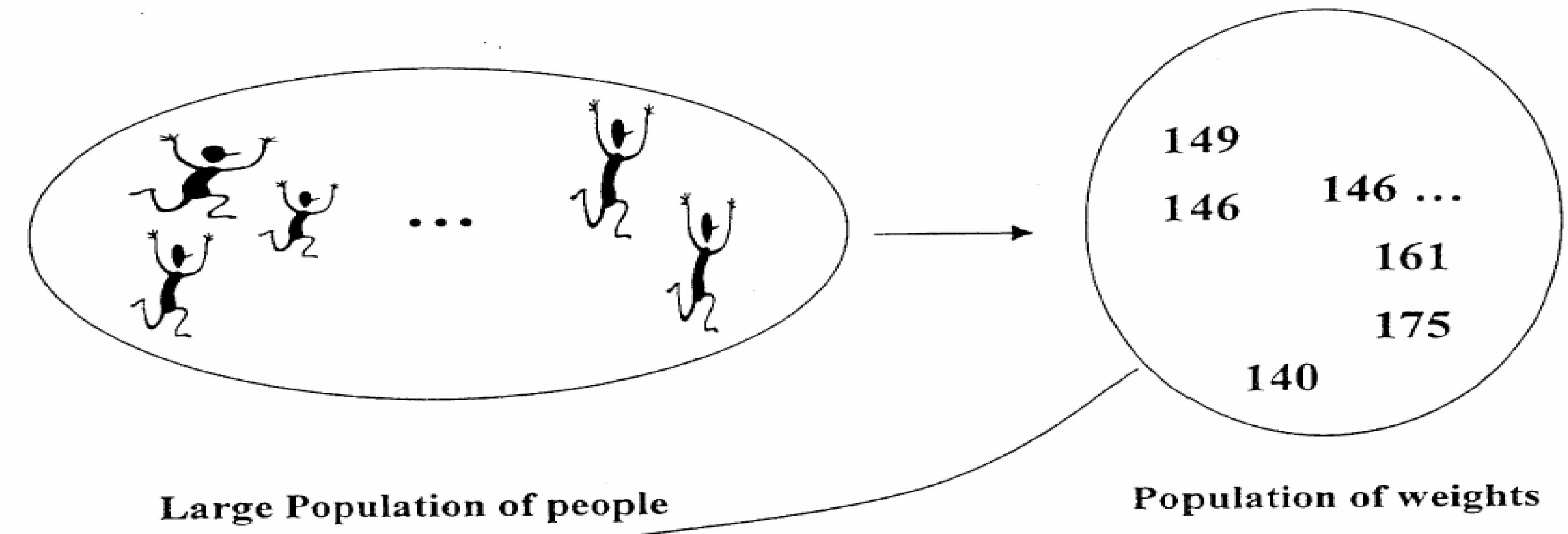
How Good is a Sample Mean

The essential question is *"How good is a sample mean as an estimate of the population mean?"*

One way to examine this question is to understand that we used a process that involved randomly selecting a sample from the population and then calculating the mean for the values of the observations in the sample.

We can repeat this process as many times as we wish and examine what it produces.

Sampling Distributions

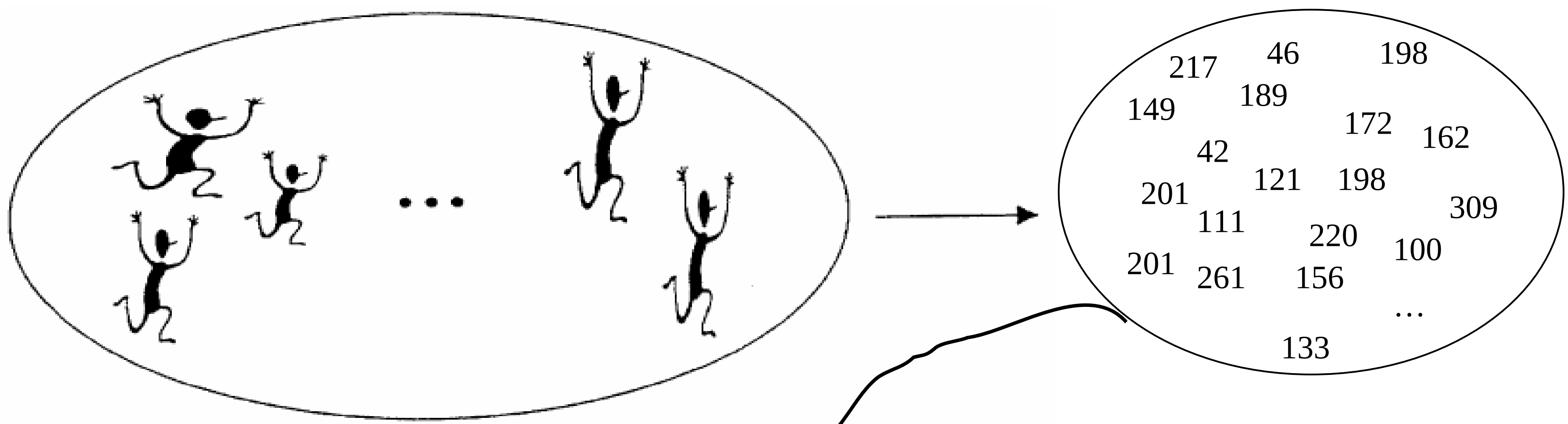


Individual Observations
149
146
132
.
.
.

$n = 1$, $\mu = 150$lbs

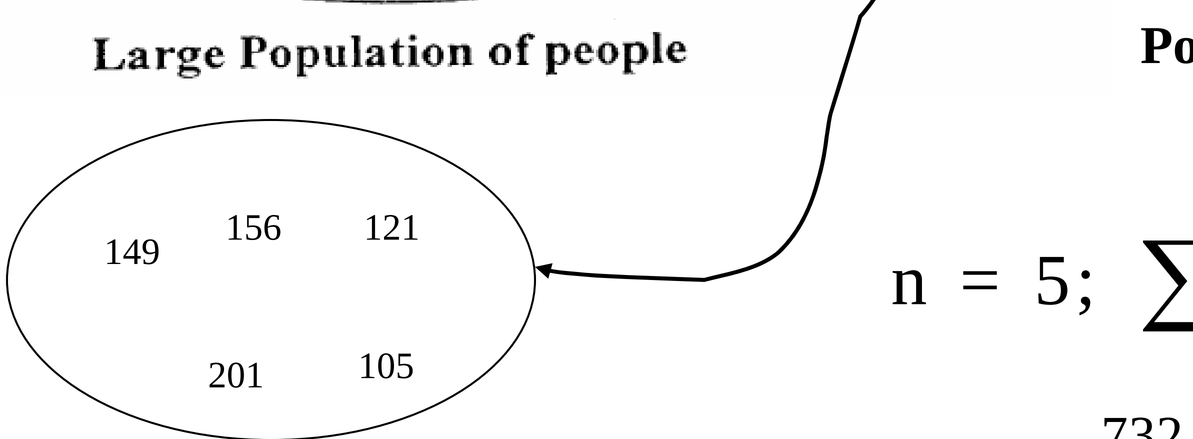
$\sigma^2 = 100$lbs, $\sigma = 10$lbs

Sample with n = 5



Large Population of people

Population of weights



Sample of 5 weights

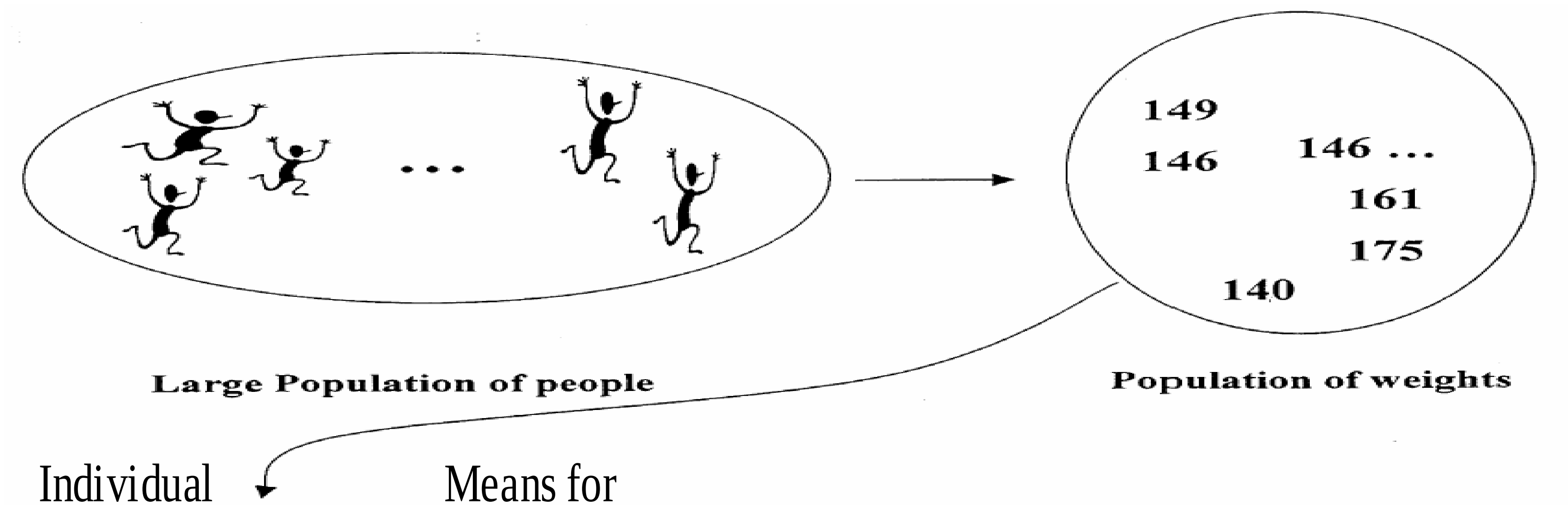
$$n = 5;\ \sum x = 732$$

$$\bar{x} = \frac{732}{5} = 146.4$$

Ten Different Samples, n = 5

Sample	n	Mean	s^2	s
1	5	147.43	88.14	9.39
2	5	153.98	117.91	10.86
3	5	146.50	103.66	10.18
4	5	155.53	91.99	9.59
5	5	147.87	149.65	12.23
6	5	143.60	66.76	8.17
7	5	146.87	64.23	8.01
8	5	149.19	280.88	16.76
9	5	150.05	200.28	14.15
10	5	146.92	173.36	13.17
Average		148.79	133.69	11.25

Sampling Distributions



Individual Observations	Means for n = 5
149	153.0
146	146.4
:	:
n = 1	n = 5
$\mu = 150$ Ibs	$\mu = 150$ Ibs
$\sigma^2 = 100$ Ibs^2	$\sigma^2_{\bar{x}} = \dfrac{\sigma^2}{n} = 20\text{Ibs}^2$
$\sigma = 10$ Ibs	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = 4.47\text{Ibs}$

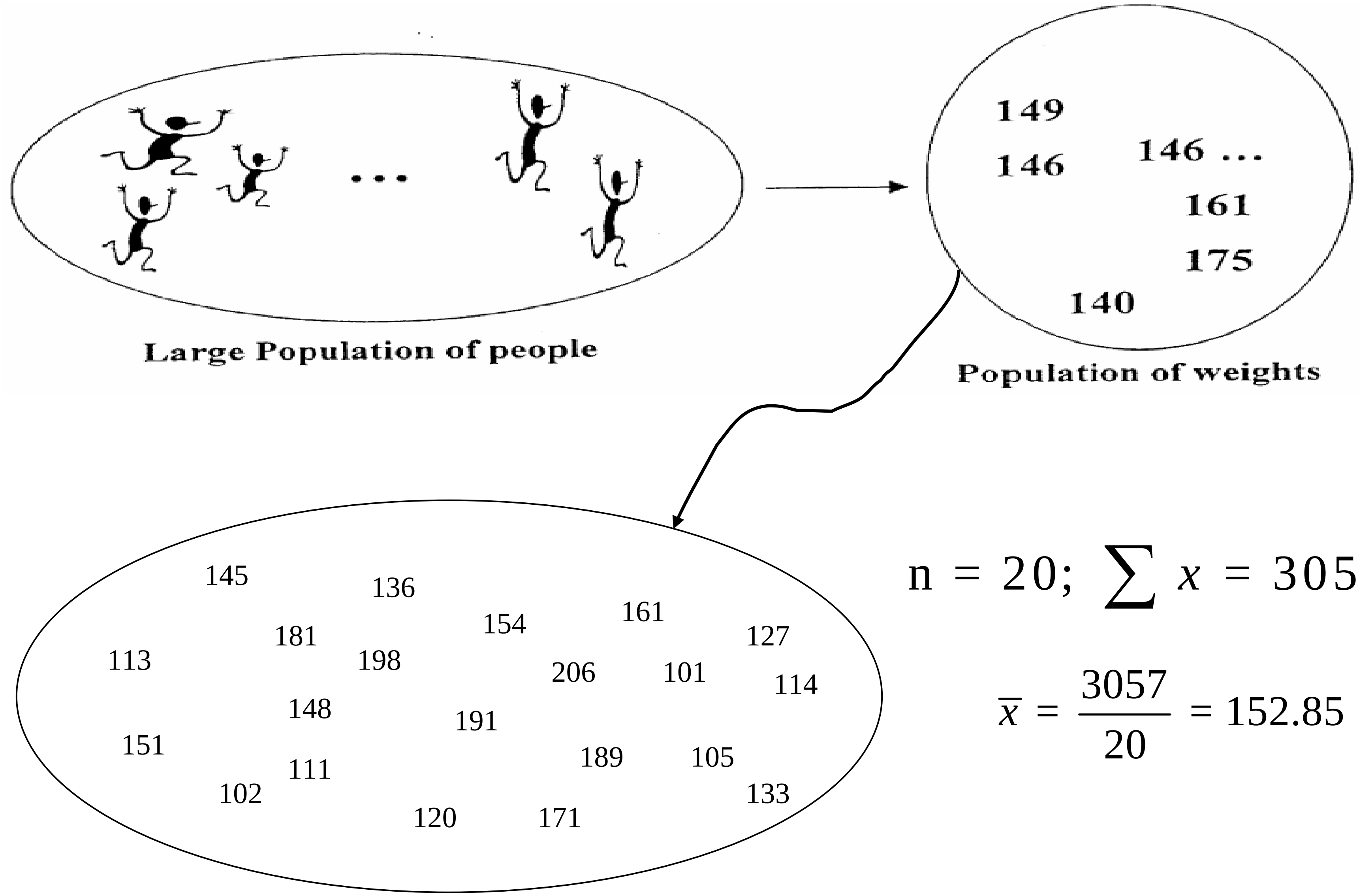
Standard Error of the Mean

The population that includes all possible samples of size n is a long list of numbers and the variance for these numbers can, in theory, be calculated.

The square root of this variance is called the standard error of the mean. It is simply the standard deviation for this population of means.

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

Sample with n = 20



$$n = 20; \ \sum x = 3057$$

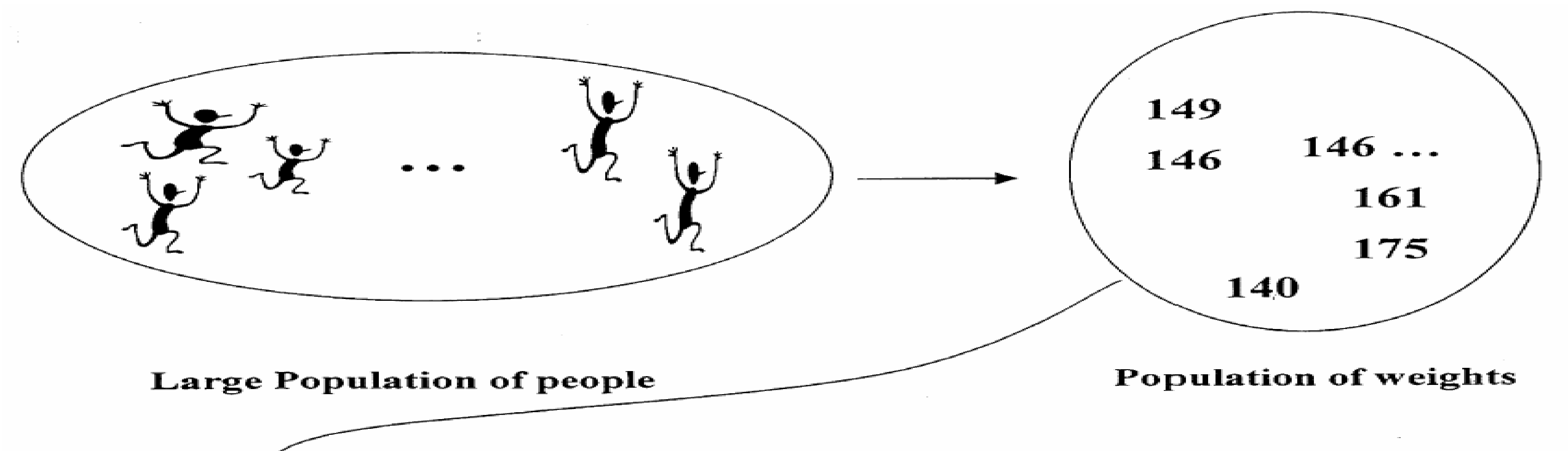
$$\bar{x} = \frac{3057}{20} = 152.85$$

Sample of 20 weights

Ten Different Samples, n = 20

Sample	n	Mean	s^2	s
1	20	150.86	100.96	10.05
2	20	146.88	122.70	11.08
3	20	147.65	119.51	10.93
4	20	149.37	51.07	7.15
5	20	153.30	109.54	10.47
6	20	152.83	111.96	10.58
7	20	148.62	91.94	9.59
8	20	152.16	140.83	11.87
9	20	154.40	179.56	13.40
10	20	151.43	115.85	10.76
Average		150.75	114.39	10.59

Sampling Distributions



Individual observations	Means for n = 5	Means for n = 20
149	153.0	151.6
146	146.4	151.3
. . .	. . .	. . .
μ = 150 lbs	μ = 150 lbs	μ = 150 lbs
$\sigma^2 = 100$lbs	$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = 20 \text{ lbs}^2$	$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = 5 \text{ lbs}^2$
σ = 10 lbs	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 4.47 \text{ lbs}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 2.23 \text{ lbs}$