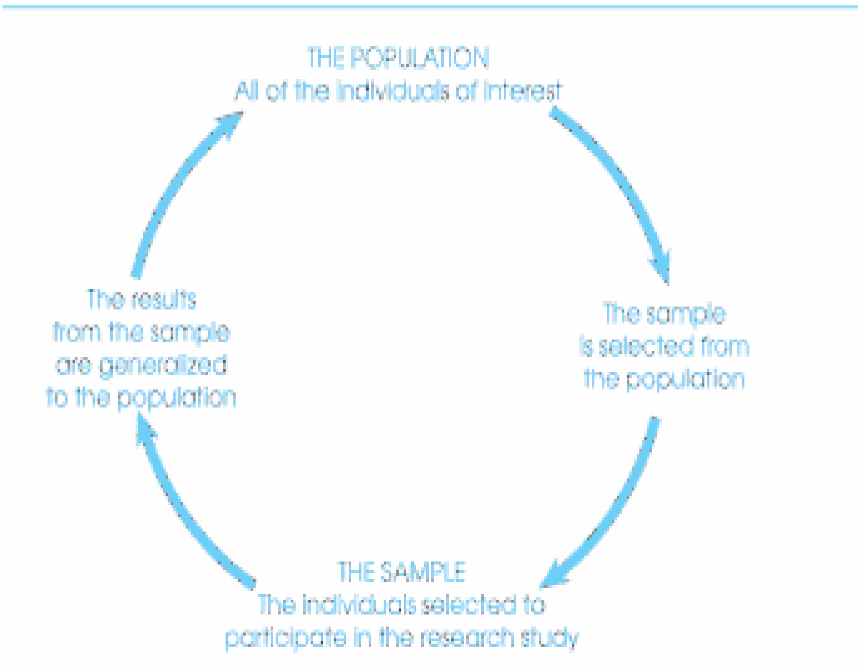
SAMPLING

IMPORTANT TERMS

Population:

Sample:



Sampling error:

Sampling variability:

Error variation:

SAMPLING TECHNIQUES

Nonprobability Sampling

Characteristics:

-
-

Types of Nonprobability Sampling:

- Accidental Sampling / Haphazard Sampling
- Purposive Sampling
- Quota Sampling

Advantages of Nonprobability Sampling:

-
-
-

Disadvantages:

-

Probability Sampling/Random Sampling:

Characteristics:

-
-
-
-

Random Number Table:

Types of Probability Sampling:

- Simple Random Sampling
- Stratified Random Sampling

Advantages:

-
-

Disadvantages:

-
-
-

SAMPLING EXAMPLE:

The National Sleep Foundation commissioned WB & A Research to conduct the 2001 Sleep in America poll.

Sampling technique:

- Used a random list of telephone numbers.
- Quotas were established by region based on the 2000 U.S. Census household data.
- Interviewers telephoned an equal number of men and women (over 18 years) between 5:00 p.m. and 9:00 p.m. on weekdays and 10:00 a.m. to 4:00 p.m. on weekends.

How representativeness was the sample?

THE SAMPLING DISTRIBUTION OF SAMPLE MEANS

Sampling Distribution:

The Sampling Distribution of Sample Means:

Characteristics of the Sampling Distribution of Sample Means:

1. The sample means tend to pile up around/approximate the population mean.

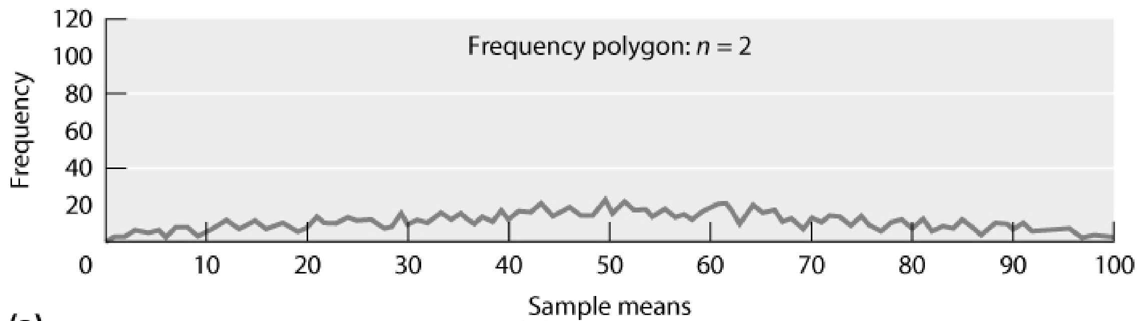
2. The distribution of sample means is approximately normal in shape.

3. We can use the distribution of sample means to answer probability questions about sample means

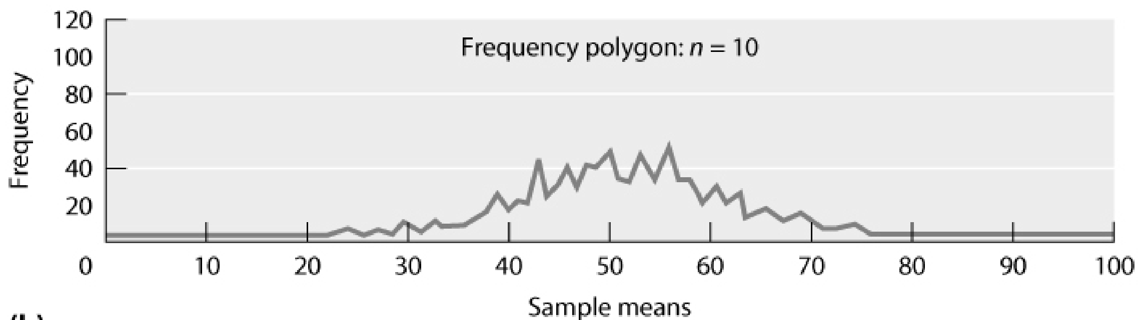
CENTRAL LIMIT THEOREM

- If repeated random samples of size *n* are taken from a population with a mean of μ and standard deviation of σ, the sampling distribution of sample means will have a mean equal to μ and standard error equal to $\sigma/\sqrt{n}$.

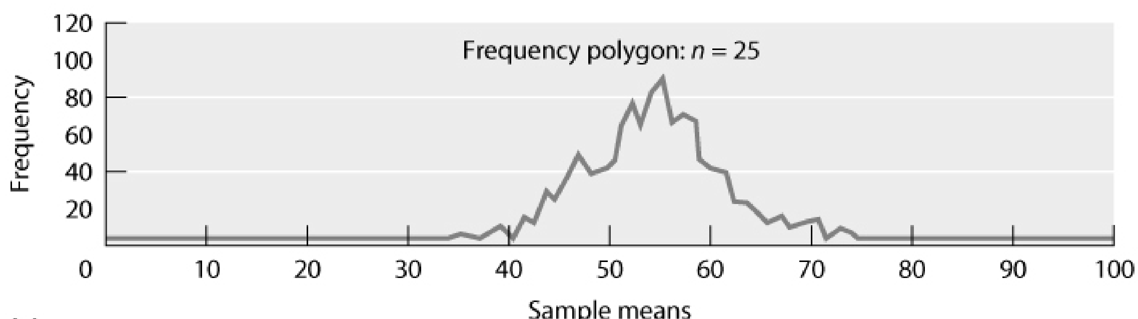
- As *n* increases, the sampling distribution will approach a normal distribution.


Frequency polygon: n = 2
Frequency
120
100
80
60
40
20
0
10
20
30
40
50
60
70
80
90
100
Sample means

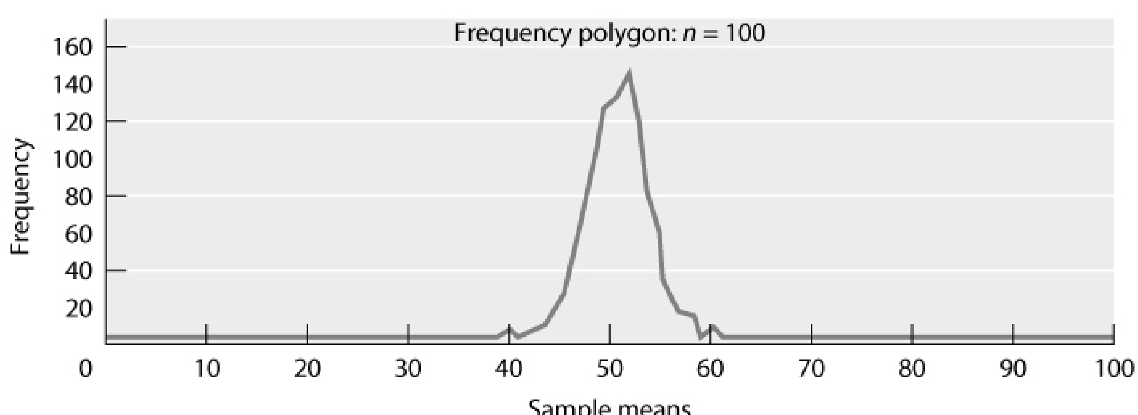
(a)


Frequency polygon: n = 10
Frequency
120
100
80
60
40
20
0
10
20
30
40
50
60
70
80
90
100
Sample means

(b)


Frequency polygon: n = 25
Frequency
120
100
80
60
40
20
0
10
20
30
40
50
60
70
80
90
100
Sample means

(c)


Frequency polygon: n = 100
Frequency
160
140
120
100
80
60
40
20
0
10
20
30
40
50
60
70
80
90
100
Sample means

(d)

The central limit theorem describes the distribution of sample means in terms of the three basic characteristics of any distribution:

- ⇨ **Shape:**

- ⇨ **Central tendency:**

- ⇨ **Variability**:

Standard Error of the Mean = the standard deviation of the sampling distribution of sample means

The magnitude of the standard error of the mean is determined by 2 factors:

1) The size of the sample

2) The standard deviation of the population

   standard error = $\sigma_M = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}}$

Example:
A population of scores is normal with $\mu = 80$ and $\sigma = 20$.

Would the distribution of sample means be normal?

Why or why not?

What is the mean and standard error of the distribution of sample means for the above population for samples of size $n = 16$?
mean = *standard error =*

What is the mean and standard error of the distribution of sample means for the above population for samples of size n = 100?
mean = *standard error =*

Example: M&M Demonstration

- Take a sample of 10 M&M's from the bag.
- Code the M & M's candy as follows: Brown = 1 point, Yellow = 2 points, Red = 3 points, Green = 4 points, Blue = 5 points, Orange = 6 points.
- Compute the mean of the coded M&Ms
- Write your mean on the board.
- Compute the mean of the sample means.
- Compute the standard deviation of the sample means.

Using data from the M&M website, this population data are as follows:
M&M milk chocolate candies: $\mu = 3.84$, $\sigma = 1.44$
(brown 13%, red 13%, yellow 14%, green 16%, orange 20%, blue 24%)

Compare the mean of the sample means to the population mean.

Use the population standard deviation to compute the standard error of the mean. *Compare the standard error of the mean computed from the population value to the standard deviation of the means computed from the class samples.*

The least predominant colors of M&Ms milk chocolate candies are brown and red and the most prominent are orange, green, and blue. How might the results differ if only 3 M&Ms were sampled?

Would they be more or less likely to select brown and red M&Ms?

Would the standard error be larger or smaller?

PROBABILITY AND THE DISTRIBUTION OF SAMPLE MEANS

Primary Use of the Sampling Distribution of Sample Means:

***z* Test for Sample Means**

Uses:

-
-

Formula:

$$z = \frac{M - \mu}{\sigma_M}$$

e.g. The population of scores on the SAT forms a normal distribution with μ = 500 and σ = 100. If you take a random sample of n = 25 students, what is the probability that the sample mean will be greater than M = 540? i.e., out of all possible sample means, what proportion have values greater than 540?

e.g. The distribution of SAT scores forms a normal distribution with a mean of μ = 500 and a standard deviation of σ = 100. If n = 25, what is the exact range of values that is expected for our sample mean 80% of the time?

Probability and Sleep Demonstration

According to statistics from the National Sleep Foundation (2001), Americans are sleeping less and working more hours. It seems Americans tend to add more responsibilities and activities to their daily lives by reducing sleep time. Note that the responses to various reasons people reported that they would "try" to sleep longer are not mutually exclusive.

National Sleep Foundation – Sleep in America Poll (2001)
Random Sample of 1004
(Quotas by Region, Random List of Telephone Numbers)

Gender:			**Driven While Drowsy:**	
Men	500		Yes	532
Women	504		No	472
Hours Slept on Weekday:			**Dozed Off While Driving:**	
8 hours or more	372		Yes	192
8 hours or less	632		No	812
Hours Slept on Weekend:			**Time Spent Sleeping Compare to 5 years ago:**	
8 hours or more	607		More	186
8 hours or less	397		Same	428
			Less	390
Would Try To Sleep More Hours If....			**Days Per Week Need Alarm Clock to Wake Up**	
They could look better	713		0 days	361
It would improve sex drive	703		1 to 3 days	160
They could lose weight	643		4 to 5 days	261
Could slow the aging process	783		6 to 7 days	222

Class Example:

- Collect data from the class on the number of hours students sleep workdays and weekends.
- Separate data by gender.
- Use the *2001 Sleep in America* data (listed below) as population sleep statistics to determine if the sample data from this class comes from the same population (i.e., conduct a *z* test).

Sleep in America Data:

Adult Women

Mean hrs on workdays = 7.1 hrs, standard deviation = 1.4 hrs

Mean hrs on weekends = 7.7 hrs, standard deviation = 1.5 hrs

Adult Men

Mean hrs on workdays = 6.9 hrs, standard deviation 1.5

Mean hrs on weekends = 7.8 hrs, standard deviation = 1.3

Conclusion:

- In most research, the population mean is unknown and the researcher selects a sample to help obtain information about the unknown population.
- The sample mean provides information about the value of the unknown population mean.
- The sample mean is not expected to give a perfectly accurate representation of the population mean; there will be some error.
- The standard error tells exactly how much error, on average, should exist between the sample mean and the unknown population mean.

Concept Map

Sampling distributions and the central limit theorem provide the basis for making statistical inferences about a population from a sample.