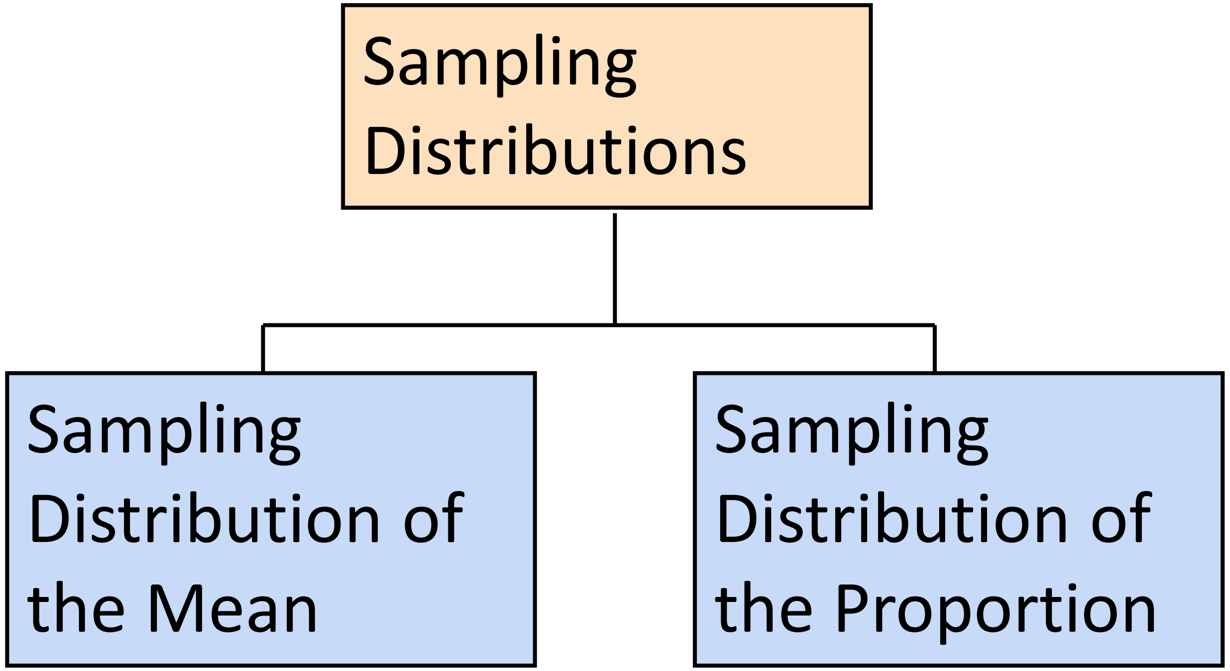
Unit 3 - Estimation

Sampling Distributions

In this unit, you learn:

- The concept of the sampling distribution
- To compute probabilities related to the sample mean
- The importance of the Central Limit Theorem
- To distinguish between different survey sampling methods
- To evaluate survey worthiness and survey errors


Sampling Distributions
Sampling Distribution of the Mean
Sampling Distribution of the Proportion

- A sampling distribution is a distribution of all of the possible values of a statistic for a given size sample selected from a population

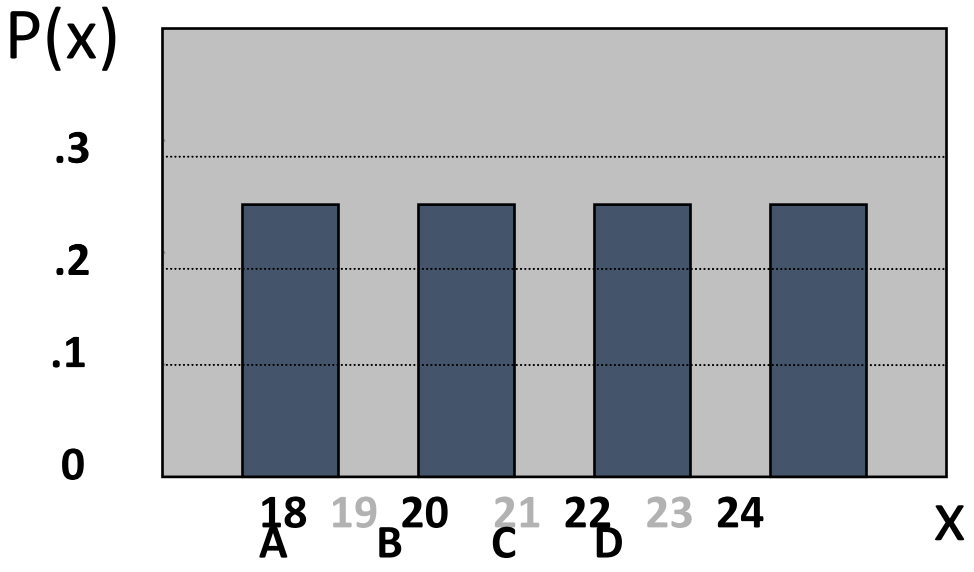
- **Assume there is a population …**
- Population size N=4
- Random variable, X, is age of individuals
- Values of X: 18, 20, 22, 24 (years)

(continued)

Summary Measures for the Population Distribution:

$$\mu = \frac{\sum X_i}{N} = \frac{18+20+22+24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$



Uniform Distribution

UNIVERSITI TEKNOLOGI MARA

(continued)

Now consider all possible samples of size n=2

1st Obs	2nd Observation			
	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

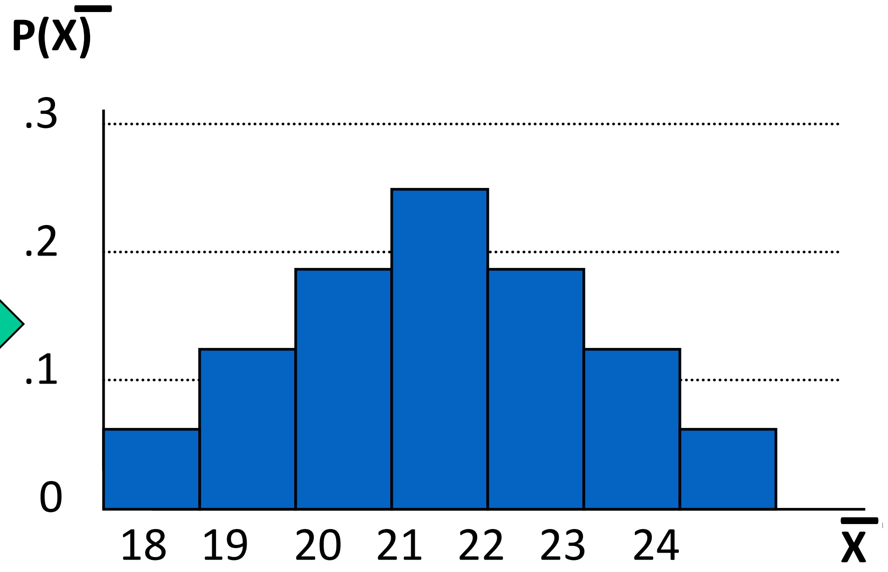
(continued)

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample Means Distribution



(no longer uniform)

UNIVERSITI TEKNOLOGI MARA

(continued)

Summary Measures of this Sampling Distribution:

$$\mu_{\overline{X}} = \frac{\sum \overline{X}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21$$

$$\sigma_{\overline{X}} = \sqrt{\frac{\sum (\overline{X}_i - \mu_{\overline{X}})^2}{N}}$$

$$= \sqrt{\frac{(18 - 21)^2 + (19 - 21)^2 + \cdots + (24 - 21)^2}{16}} = 1.58$$

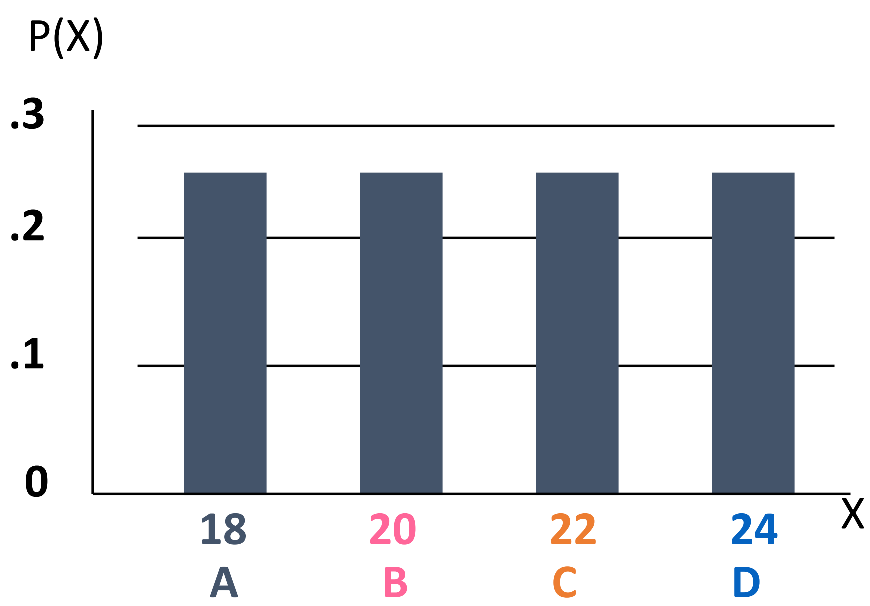
Comparing the Population with its Sampling Distribution

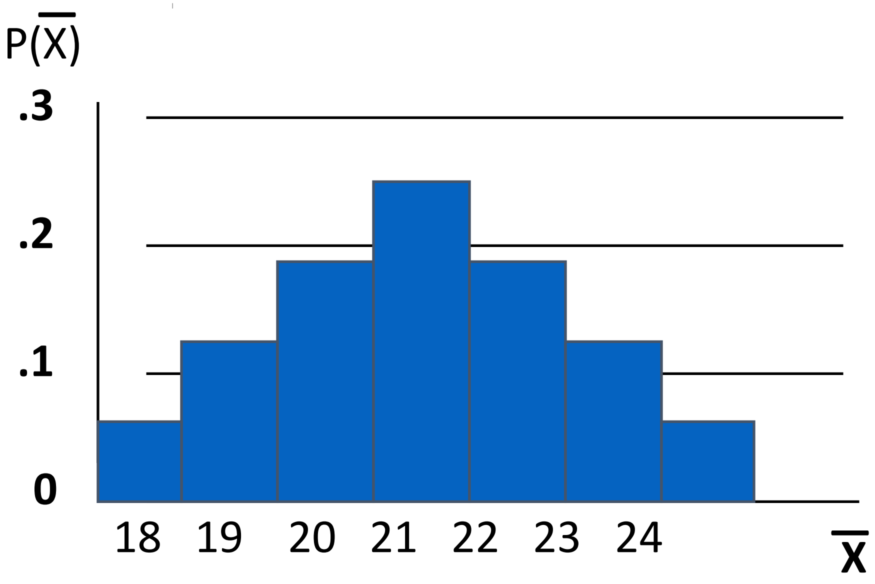
Population
N = 4

$\mu = 21 \qquad \sigma = 2.236$

Sample Means Distribution
n = 2

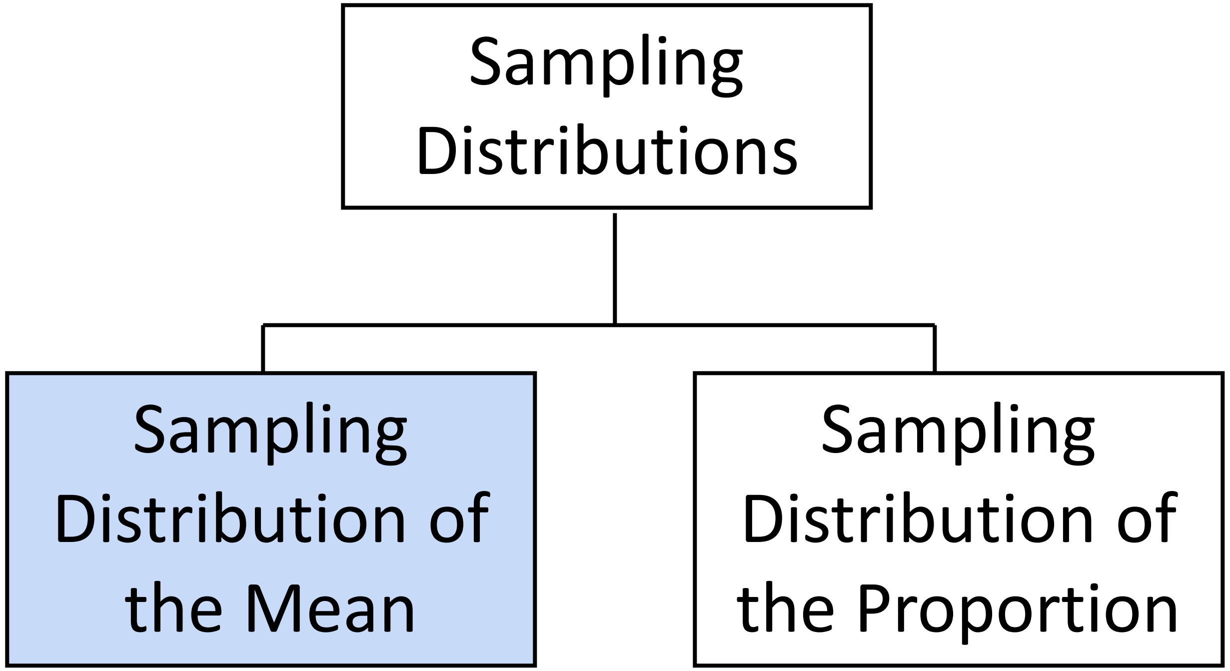
$\mu_{\bar{X}} = 21 \qquad \sigma_{\bar{X}} = 1.58$





Sampling Distributions

Sampling Distribution of the Mean



UNIVERSITI TEKNOLOGI MARA

- Different samples of the same size from the same population will yield different sample means
- A measure of the variability in the mean from sample to sample is given by the Standard Error of the Mean:

  (This assumes that sampling is with replacement or sampling is without replacement from an infinite population)

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

- Note that the standard error of the mean decreases as the sample size increases

- If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with

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

and

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

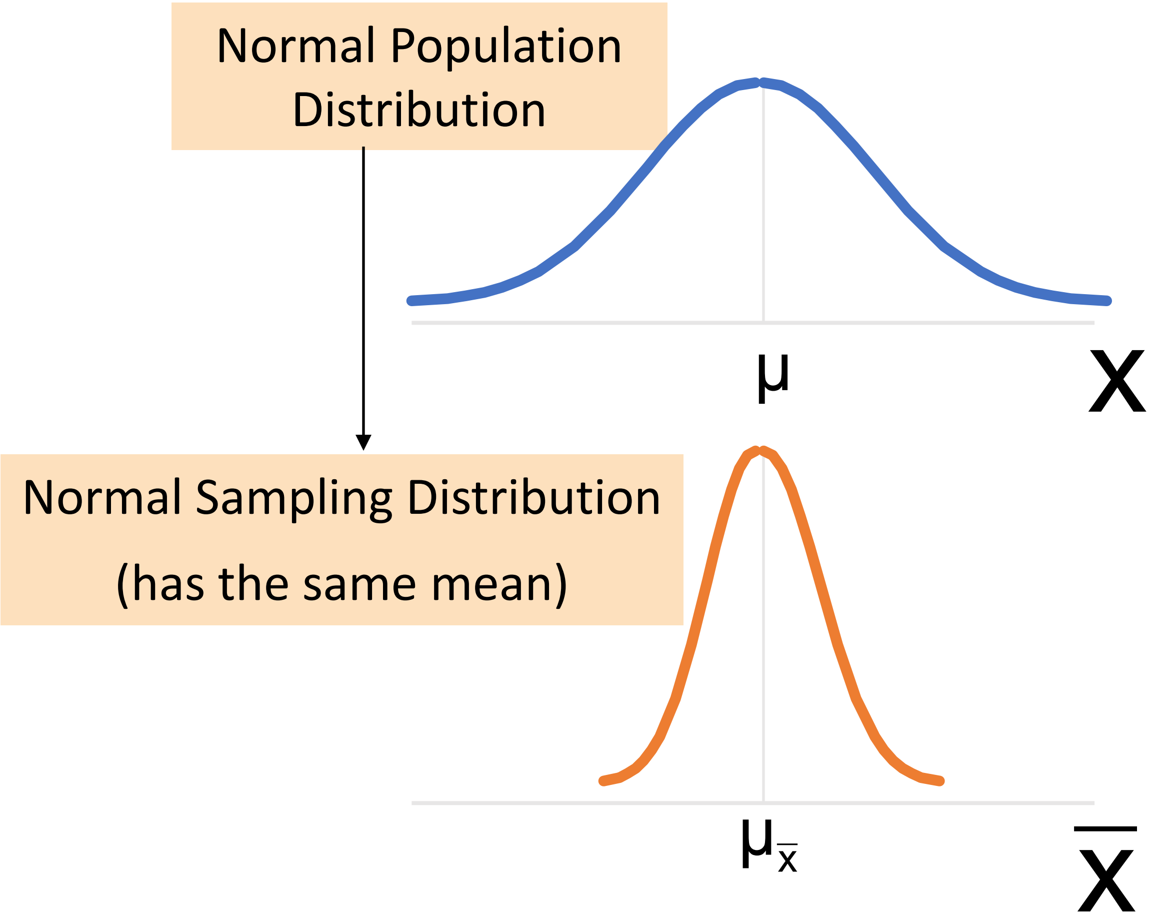
- Z-value for the sampling distribution of : $\bar{X}$

$$Z = \frac{(\bar{X} - \mu_{\bar{X}})}{\sigma_{\bar{X}}} = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

where:
$\bar{X}$ = sample mean
μ = population mean
σ = population standard deviation
n = sample size

- $$\mu_{\bar{X}} = \mu$$

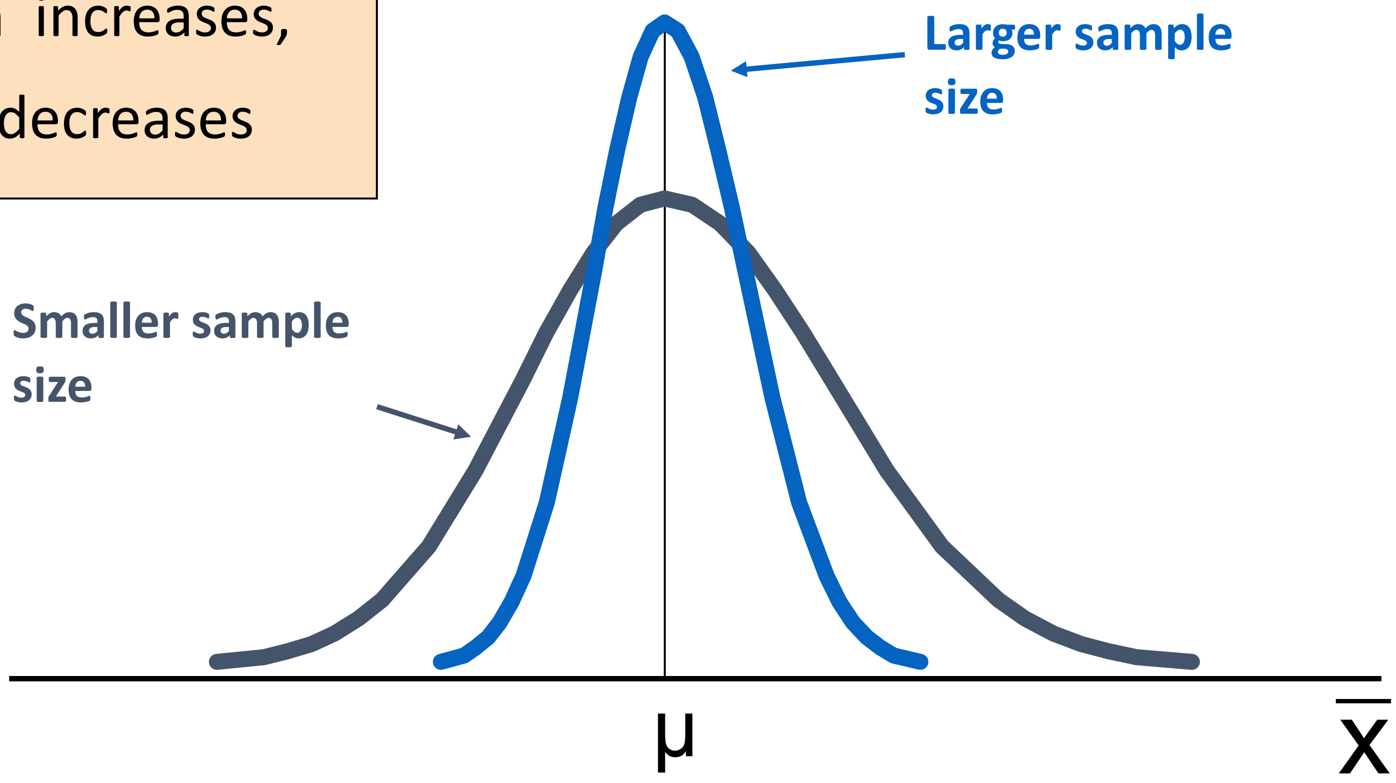
(i.e. $\bar{X}$ is unbiased)



(continued)

As n increases,

$\sigma_{\bar{x}}$ decreases

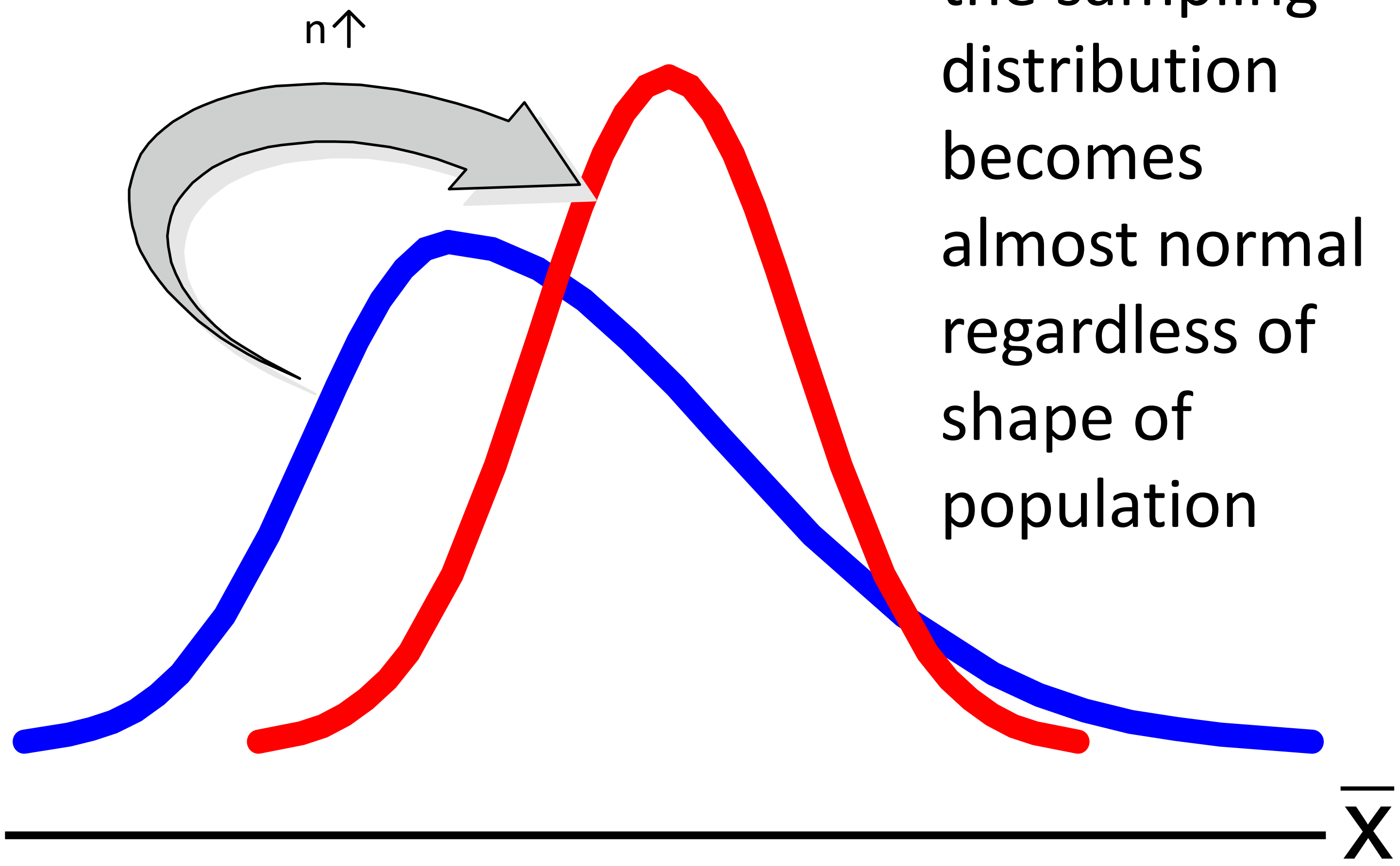


- We can apply the Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means from the population will be approximately normal as long as the sample size is large enough.

Properties of the sampling distribution:

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

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

(continued)

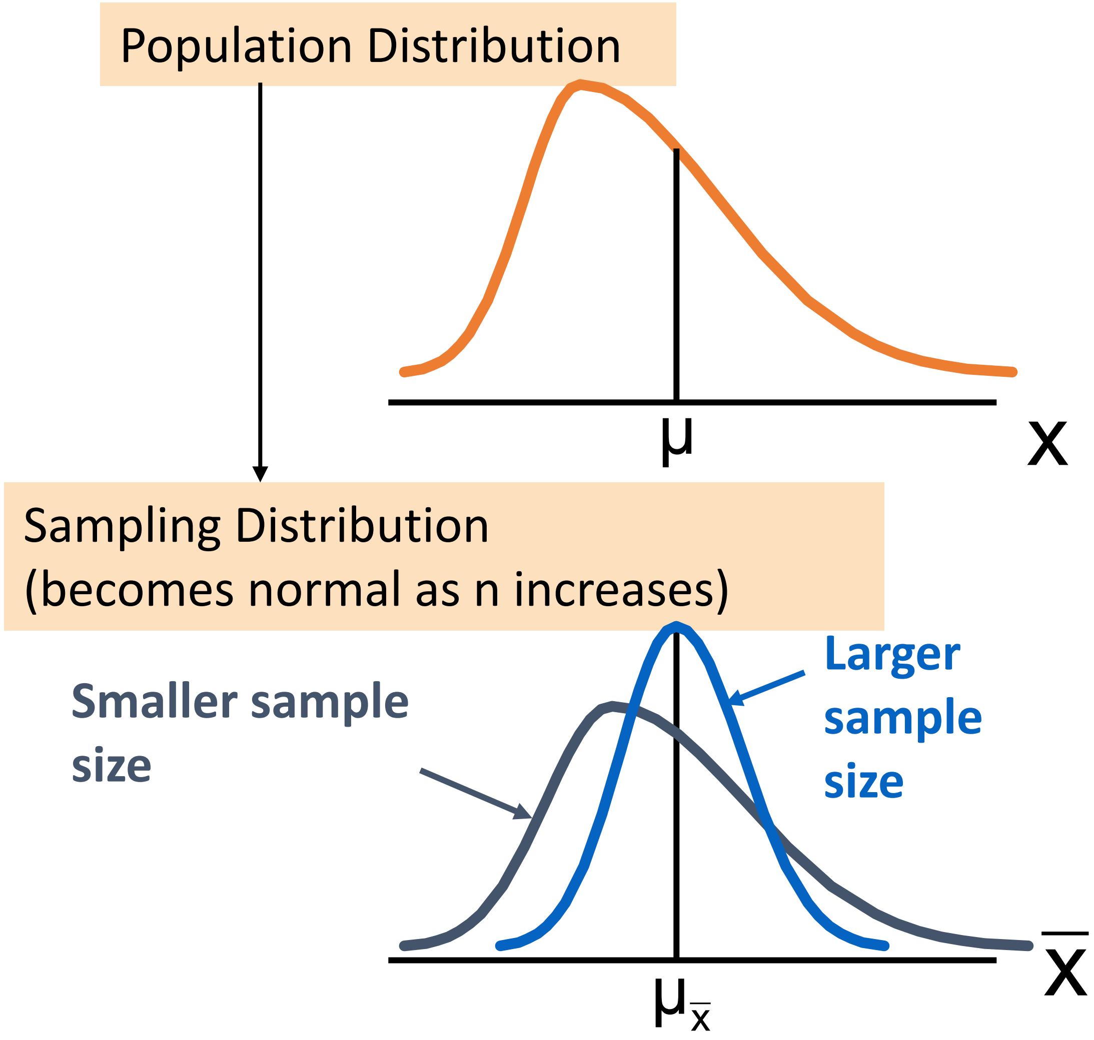
Sampling distribution properties:

Central Tendency

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

Variation

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



How Large is Large Enough?

- For most distributions, $n > 30$ will give a sampling distribution that is nearly normal
- For fairly symmetric distributions, $n > 15$
- For normal population distributions, the sampling distribution of the mean is always normally distributed

- Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.

- What is the probability that the sample mean is between 7.8 and 8.2?

(continued)

Solution:

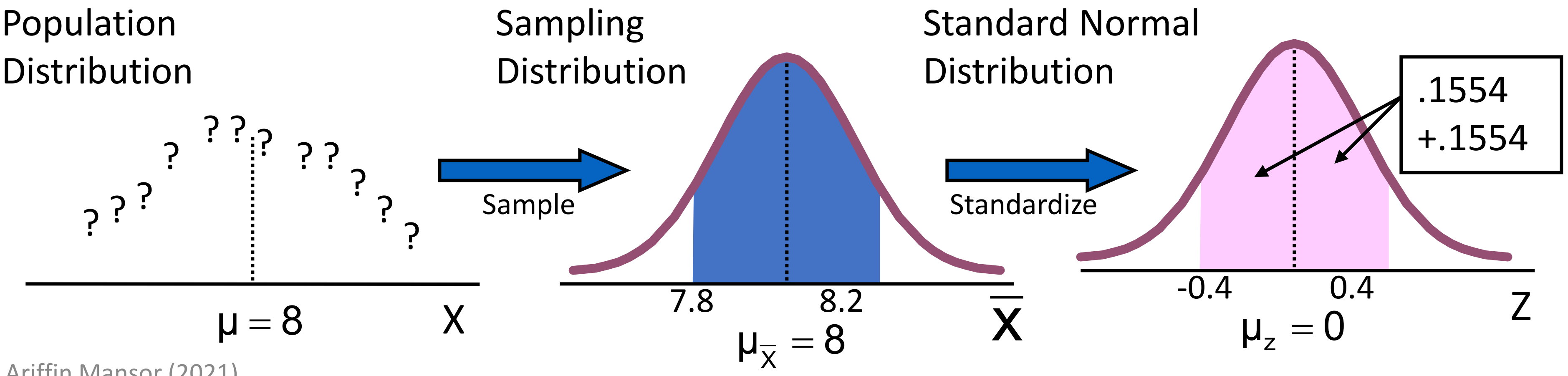
- Even if the population is not normally distributed, the central limit theorem can be used (n > 30)
- … so the sampling distribution of $\overline{X}$ is approximately normal
- … with mean $\mu_{\bar{X}} = 8$
- …and standard deviation $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{3}{\sqrt{36}} = 0.5$

(continued)

Solution (continued):

$$P(7.8 < \bar{X} < 8.2) = P\left(\frac{7.8-8}{3/\sqrt{36}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}}\right)$$

$$= P(-0.4 < Z < 0.4) = \boxed{0.3108}$$



- Introduced sampling distributions
- Described the sampling distribution of the mean
  - For normal populations
  - Using the Central Limit Theorem
- Calculated probabilities using sampling distributions