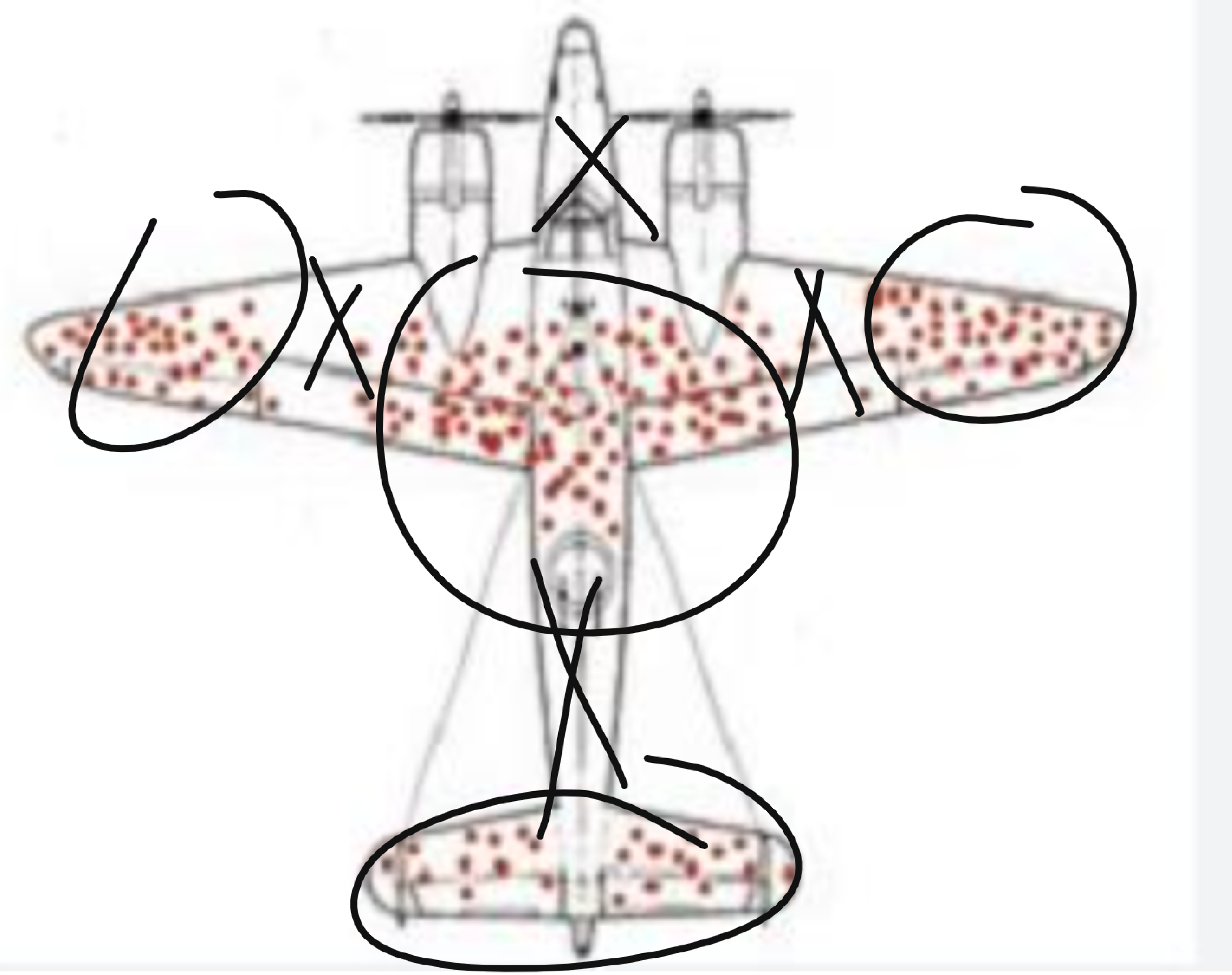
"By a small sample, we may judge the whole piece."

— Miguel de Cervantes from *Don Quixote de la Mancha*

9 CHAPTER

Samples and Sampling Distributions



Biased Sample

A **sample is biased** if it overrepresents or underrepresents some segment(s) of the population.

DEFINITION

Sampling Frame

A **sampling frame** is a list which identifies all members of the population.

DEFINITION

Simple Random Sample

A **simple random sample** from a finite population is one in which every possible sample of the same size n has the same probability of being selected.

DEFINITION

Systematic Sample

A sample in which you choose a starting point and then every k^{th} member of the population is included in the sample.

DEFINITION

Cluster Sampling

Cluster sampling involves dividing the population into clusters, and randomly selecting a sample of clusters to represent the population. Cluster sampling is used when "natural" groupings are evident in the population.

DEFINITION

Stratified Sampling

In **stratified sampling,** the population is divided into **strata**, which are sub-populations. A **strata** can be any identifiable characteristic that can be used to classify the population. If the population consists of people, then strata could be sex, income, political party, religion, education, race, or location.

DEFINITION

Sampling Distribution of a Statistic

The **sampling distribution of a statistic** (such as the sample mean or sample proportion) is the probability distribution of all values of the statistic when all possible samples of size *n* are taken from a population.

DEFINITION

Point Estimator

A **point estimator** is a single-valued estimate calculated from the sample data, which is intended to be close to the true population value.

DEFINITION

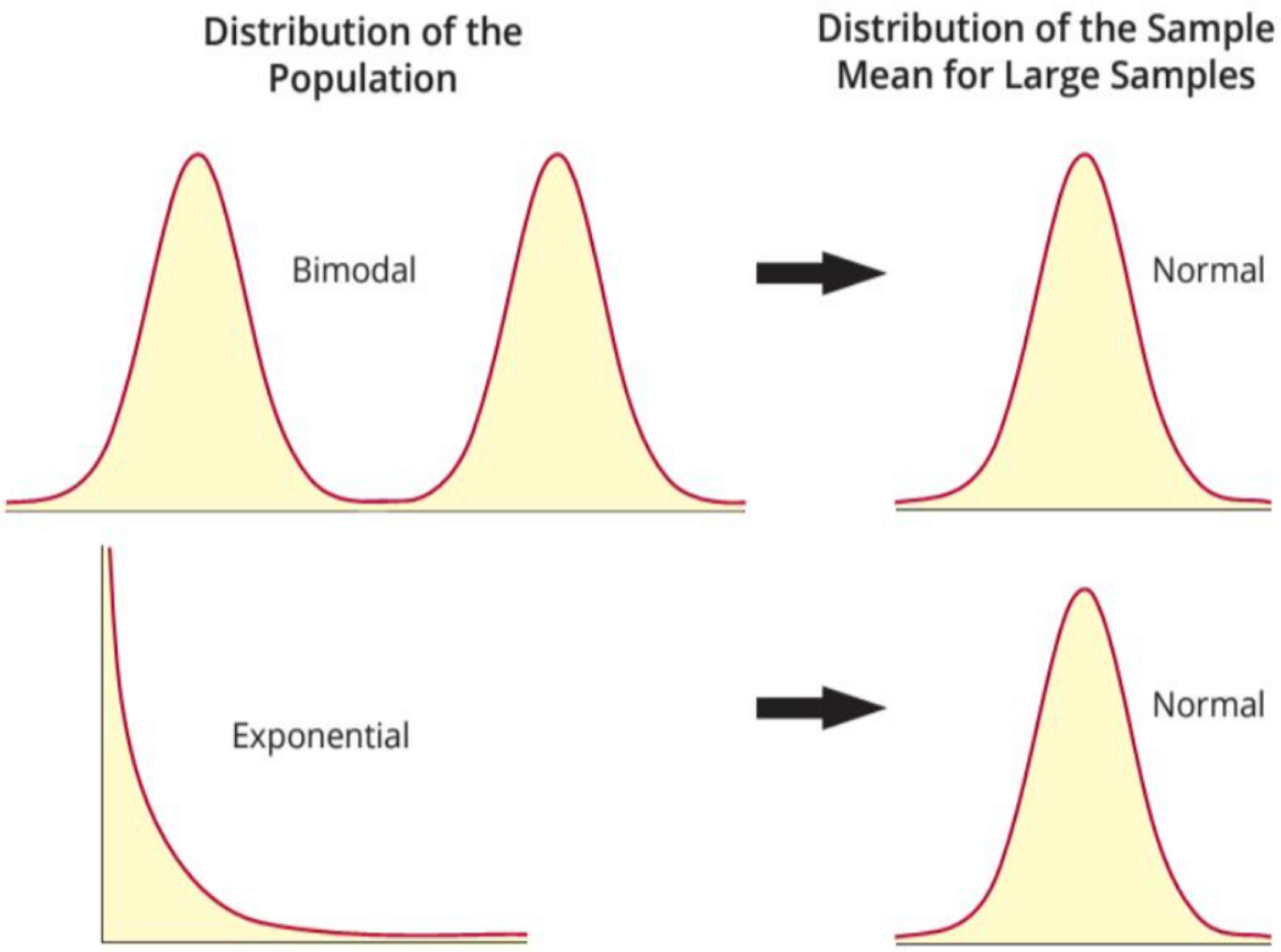


Figure 9.3.3

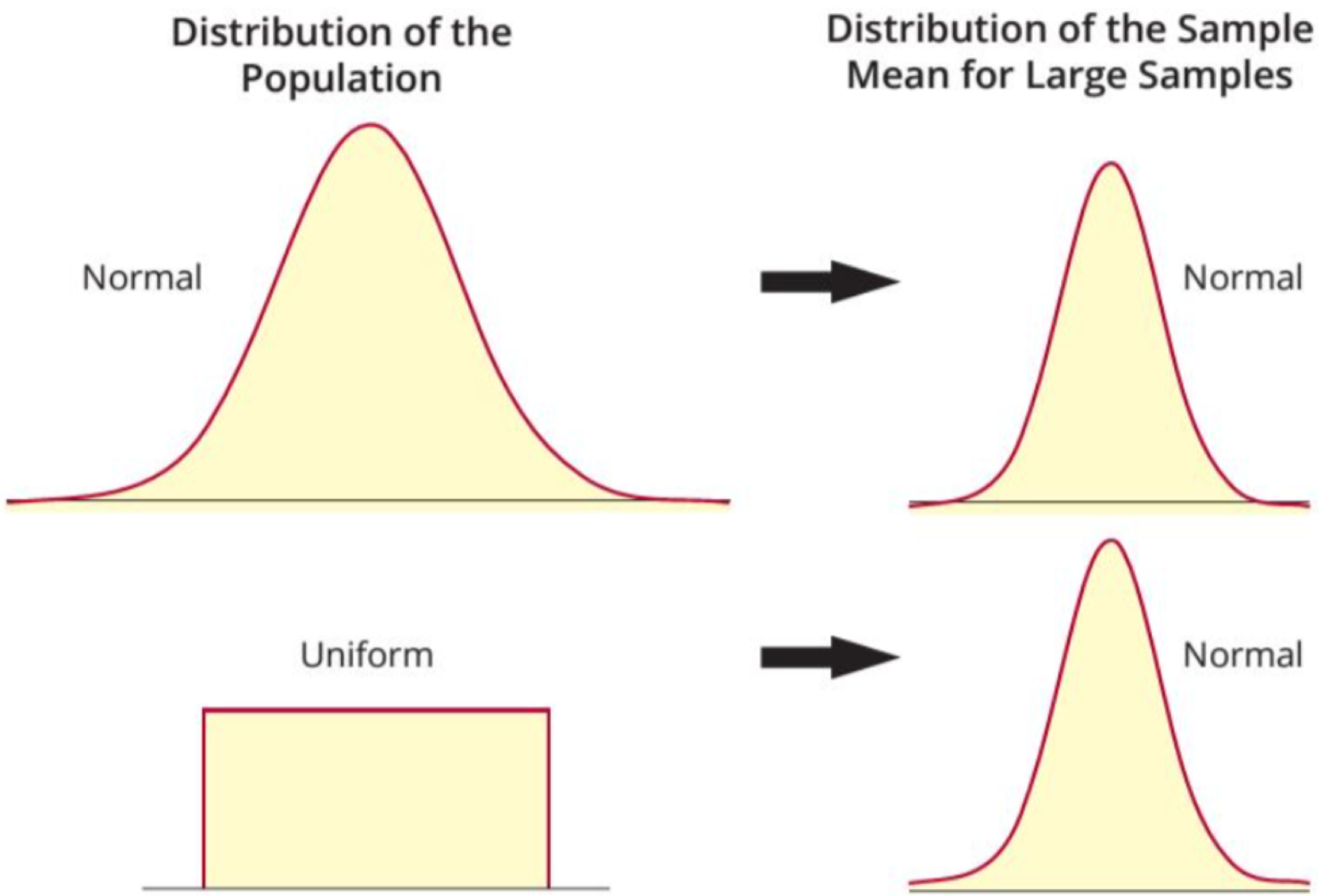


Figure 9.3.3 (cont.)

The Central Limit Theorem

If a sufficiently large random sample (i.e., $n > 30$) is drawn from a population with mean μ and standard deviation σ, the distribution of the sample mean will have the following characteristics.

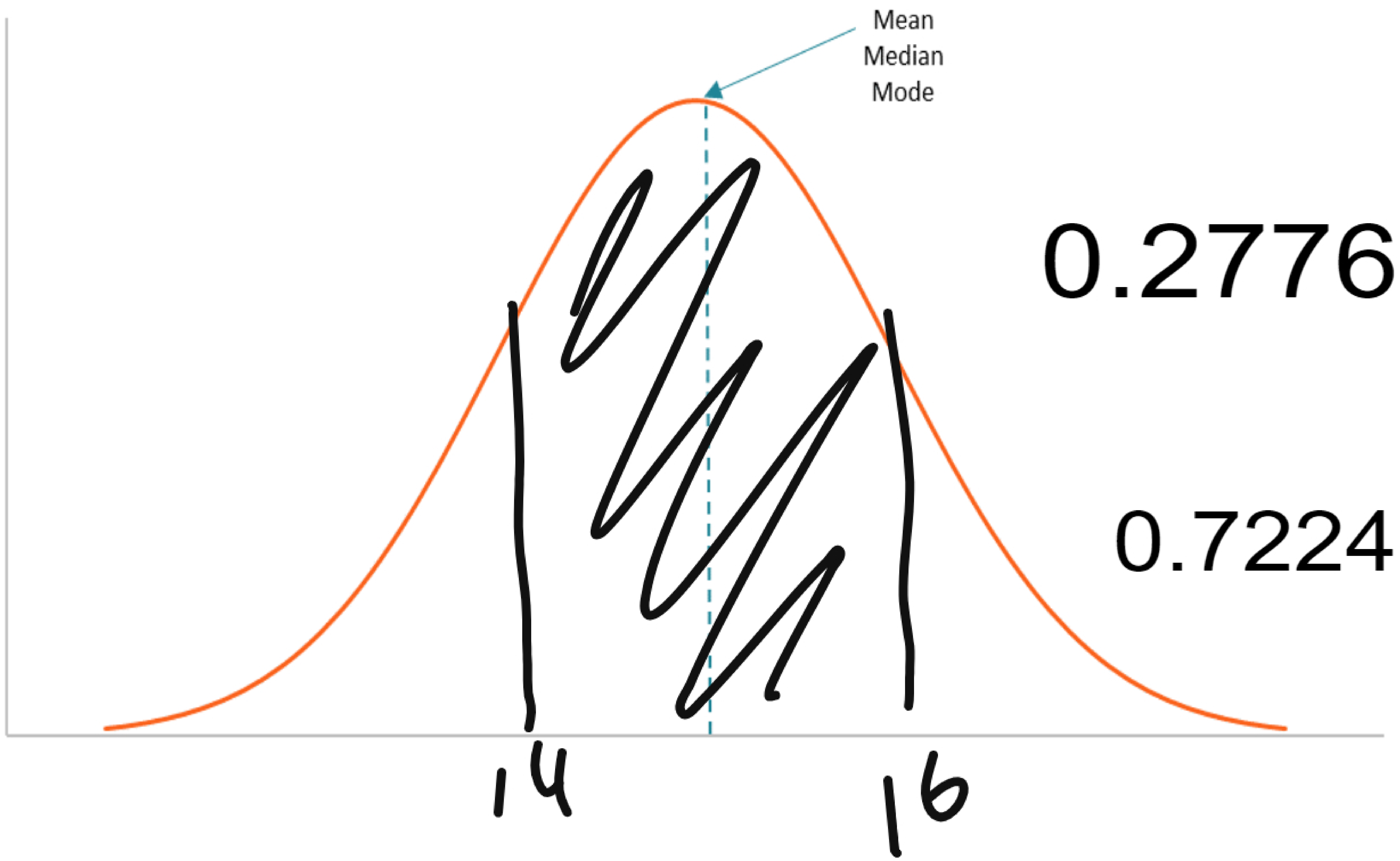
1. An approximately normal distribution regardless of the distribution of the underlying population.
2. $\mu_{\bar{x}} = E(\bar{x}) = \mu$ (The mean of the sample means equals the population mean.)
3. $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ (The standard deviation of the sample means equals the standard deviation of the population divided by the square root of the sample size.)

DEFINITION

A company fills bags with fertilizer for retail sale. The weights of the bags of fertilizer have a normal distribution with a mean weight of 15 lb and standard deviation of 1.70 lb.

0.4448

a. What is the probability that a randomly selected bag of fertilizer will weigh between 14 and 16 pounds?



$\mu = 15 \qquad \sigma = 1.70$

0.2776 $X = 14 \qquad z = \frac{14 - 15}{1.70} = -0.59$

0.7224 $X = 16 \qquad z = \frac{16 - 15}{1.70} = 0.59$

a. What is the probability that a randomly selected bag of fertilizer will weigh between 14 and 16 pounds?

b. If 35 bags of fertilizer are randomly selected, find the probability that the average weight of the 35 bags will be between 14 and 16 pounds.

$\mu = 15 \qquad \sigma = 1.70$

$\sigma_{\bar{x}} = \frac{1.70}{\sqrt{35}} = 0.287$

.99950

0.00025 $\quad x = 14 \quad z = \frac{14-15}{0.287} = -3.48$

0.99975 $\quad x = 16 \quad z = \frac{16-15}{0.287} = 3.48$

The average score for a water safety instructor (WSI) exam is 75 with a standard deviation of 12. Fifty scores for the WSI exam are randomly selected.

a. Find the probability that the average of the fifty scores is at least 80.

b. Find the probability that the average of the fifty scores is at most 70.

c. Find the probability that the average of the fifty scores is between 72 and 78.

$\mu = 75$ $\sigma = 12$ $\sigma_{\bar{x}} = \frac{12}{\sqrt{50}} = 1.697$

0.00164

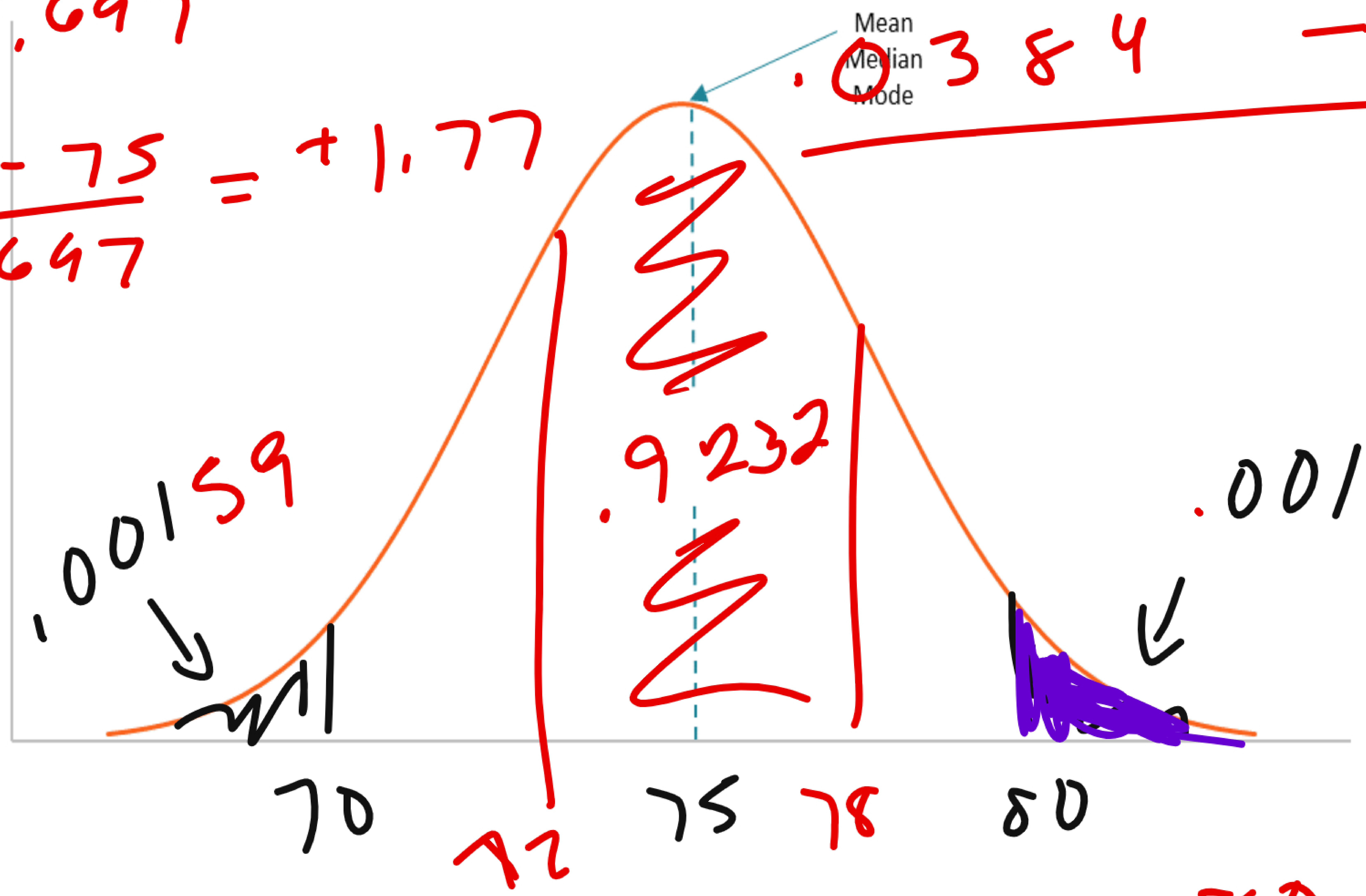
$z = \frac{80-75}{1.697} = \frac{5}{1.697} = 2.95$ -2.95

$X = 72 \quad \frac{72-75}{1.697} = -1.77$

$X = 78 \quad \frac{78-75}{1.697} = +1.77$

$.96164 \; +$

$.0384 \; -$



$80 \rightarrow$

Sampling Distribution of the Sample Proportion

If the population is infinite and the sample is sufficiently large, the distribution of $\hat{p}$ has the following characteristics:

1. An approximately normal distribution.
2. $\mu_{\hat{p}} = E\left(\hat{p}\right) = p$. (The mean of the sample proportions equals the population proportion.)
3. $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}} \approx \sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n}}$.

Sample Proportion

The sample proportion is given by

$$\hat{p} = \frac{x}{n},$$

where x is the number of observations in the sample possessing the characteristic of interest and n is the total number of observations in the sample.

FORMULA

Suppose a sample of 400 persons is used to perform a taste test. If the true fraction in the population that prefers Pepsi is really 0.5, what is the probability that less than 0.44 of the persons in the sample will prefer Pepsi?

$n = 400$ $\quad p = 0,5$ $\quad \hat{p} = 0,44$

$\mu = 0.5$ $\quad \sigma = \sqrt{\frac{(0,5)(1-0.5)}{400}} = 0.025$

$\sqrt{\frac{0,25}{400}}$

$\hat{p} \leq 0.44$

$z = \frac{0.44 - 0.5}{0.025} = -2.4$

0.0082
.82%