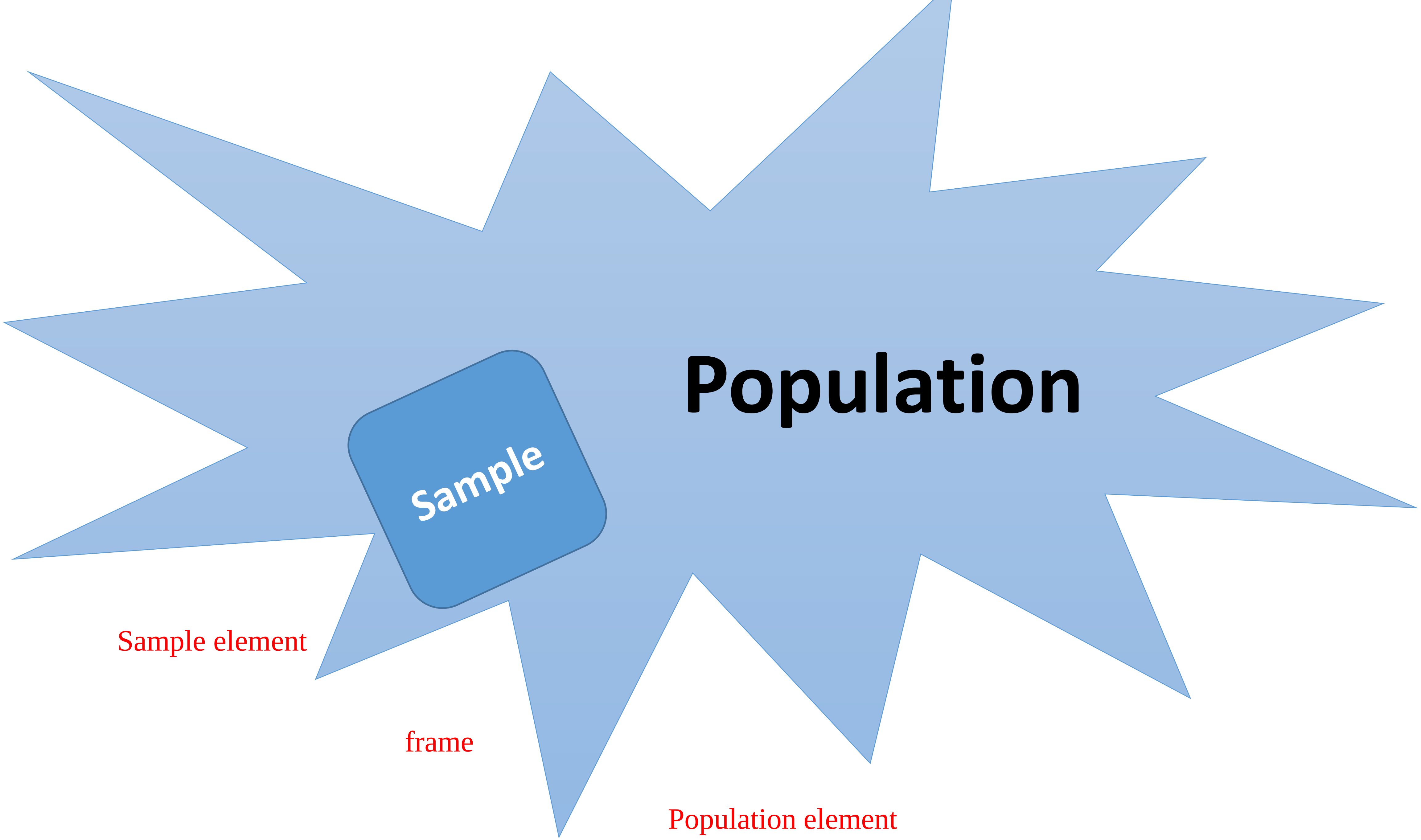
Sampling and Sampling Distributions

- An element is the entity on which data are collected.
- A population is the collection of all the elements of interest.
- A sample is a subset of the population.

The sampled population is the population from which the sample is drawn, and a frame is a list of the elements that the sample will be selected from.


Population
Sample
Sample element
frame
Population element

Point Estimator

if the measures are computed for data from a population, they are called **population parameters**.

Parameter

Example

POPULATION MEAN

$$\mu = \frac{\Sigma x_i}{N}$$

if the measures are computed for data from a sample, they are called **sample statistics**.

Statistic

Example

SAMPLE MEAN

$$\bar{x} = \frac{\Sigma x_i}{n}$$

in statistical inference:

a **sample statistic** is referred to as the point estimator of the corresponding population parameter.

The target population is the population we want to make inferences about,

while the sampled population is the population from which the sample is actually taken.

point estimator

Population	**Sample**
Parameter	**Statistic**
Example: POPULATION MEAN $\mu = \frac{\Sigma x_i}{N}$	Example: SAMPLE MEAN $\bar{x} = \frac{\Sigma x_i}{n}$

Parameter		Point estimator	Statistic	
$Mean:$	μ	$\xrightarrow{\hat{\mu}}$	$Mean:$	$\bar{x}$
$Variance:$	σ^2	$\xrightarrow{\hat{\sigma}^2}$	$Variance:$	S^2
$Standard\ Deviation::$	σ	$\xrightarrow{\hat{\sigma}}$	$Standard\ Deviation:$	S
$Proportion:$	p	$\xrightarrow{\hat{p}}$	$Proportion:$	$\bar{p}$

Example

The following marks are from a simple random **sample** for student : 2, 7, 10, 9, 6, 8

a. What is the point estimate of the population mean?
b. What is the point estimate of the population variance?
c. What is the point estimate of the population standard deviation?
d. What is the point estimate of the population proportion of failed student, assuming the student failed if his/her mark is less than 5 ?

Answer:

a. What is the point estimate of the population mean? $\hat{\mu} = \overline{x} = \frac{\sum x}{n}$

b. What is the point estimate of the population variance? $\hat{\sigma}^2 = S^2 = \frac{\sum(x-\overline{x})^2}{n-1}$

c. What is the point estimate of the population standard deviation? $\hat{\sigma}$ = $S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

d. What is the point estimate of the population proportion of failed student, assuming the student failed if his/her mark is less than 5 ? $\hat{p}$ = $\overline{p} = \frac{\sum I_i}{n}$,
where: I_i=1: if the student mark is less than 5, and zero otherwise

2 , 7 , 10 , 9 , 6 , 8

Sampling Distributions

The sample mean $\bar{x}$ is the point estimator of the population mean μ
The sample proportion $\bar{p}$ is the point estimator of the population proportion p.

Example:
Assuming we have a simple random sample of $n = 30$ managers,
The point estimate of μ is $\bar{x}$ = $51,814 and
the point estimate of p is $\bar{p}$ = 0.63.

Suppose we select another simple random sample of $n = 30$ managers and obtain the following point estimates:
The point estimate of μ is $\bar{x}$ = $52,670 and
the point estimate of p is $\bar{p}$ = 0.70.
……
Now, suppose we repeat the process of selecting a simple random sample of $n = 30$ managers and obtain the point estimates of each μ and p.

TABLE 7.4

VALUES OF $\bar{x}$ AND $\bar{p}$ FROM 500 SIMPLE RANDOM SAMPLES OF 30 EAI MANAGERS

Sample Number	Sample Mean ($\bar{x}$)	Sample Proportion ($\bar{p}$)
1	51,814	.63
2	52,670	.70
3	51,780	.67
4	51,588	.53
⋮	⋮	⋮
500	51.752	.50

Sampling Distributions

The probability distribution of any particular sample **statistic** is called the sampling distribution of the **statistic**

TABLE 7.5
FREQUENCY AND RELATIVE FREQUENCY DISTRIBUTIONS OF $\bar{x}$ FROM 500 SIMPLE RANDOM SAMPLES OF 30 MANAGERS

Mean Annual Salary ($)	Frequency	Relative Frequency
49,500.00–49,999.99	2	.004
50,000.00–50,499.99	16	.032
50,500.00–50,999.99	52	.104
51,000.00–51,499.99	101	.202
51,500.00–51,999.99	133	.266
52,000.00–52,499.99	110	.220
52,500.00–52,999.99	54	.108
53,000.00–53,499.99	26	.052
53,500.00–53,999.99	6	.012
Totals	500	1.000

FIGURE 7.1

RELATIVE FREQUENCY HISTOGRAM OF $\bar{x}$ VALUES FROM 500 SIMPLE RANDOM SAMPLES OF SIZE 30 EACH

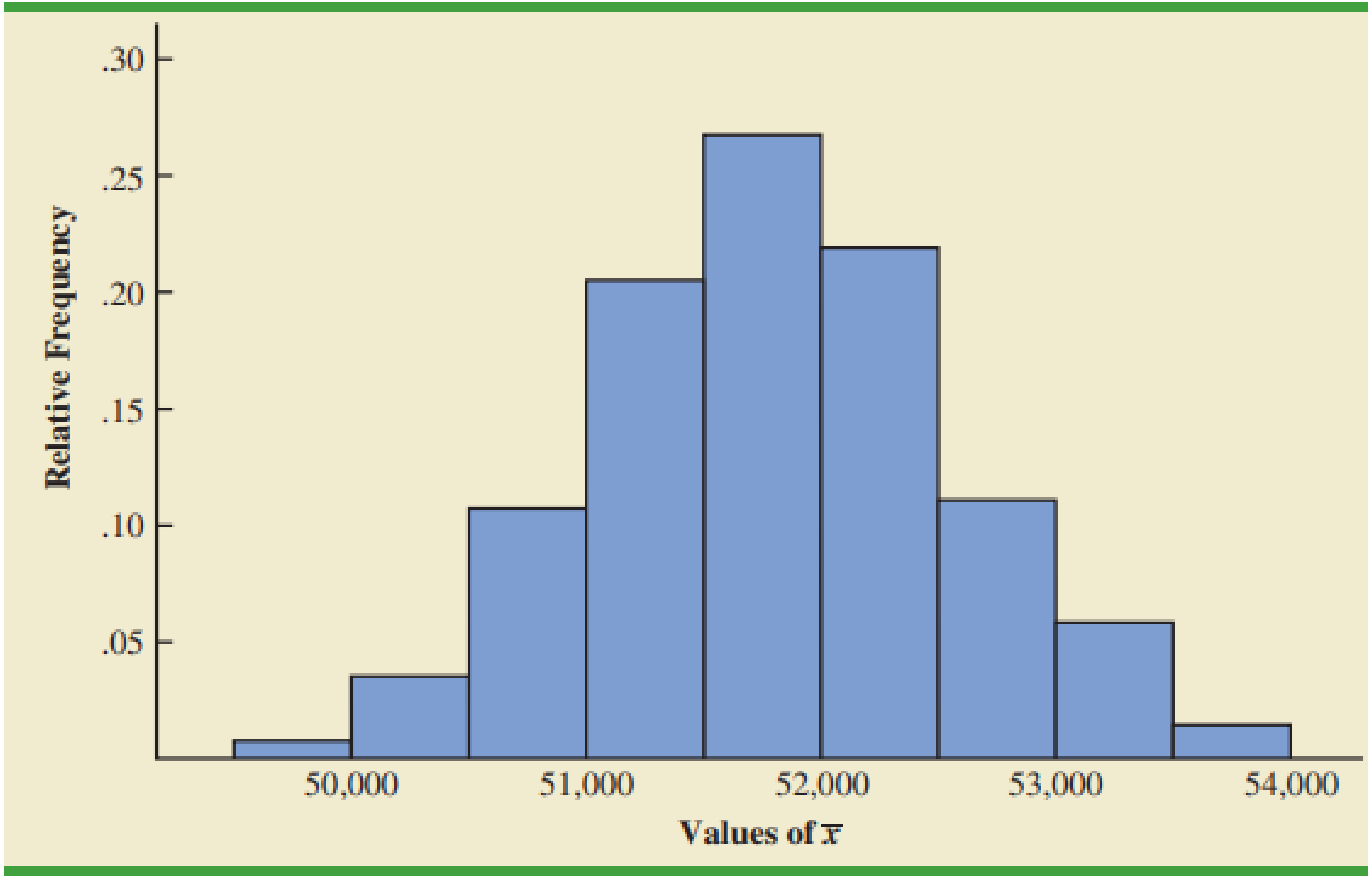
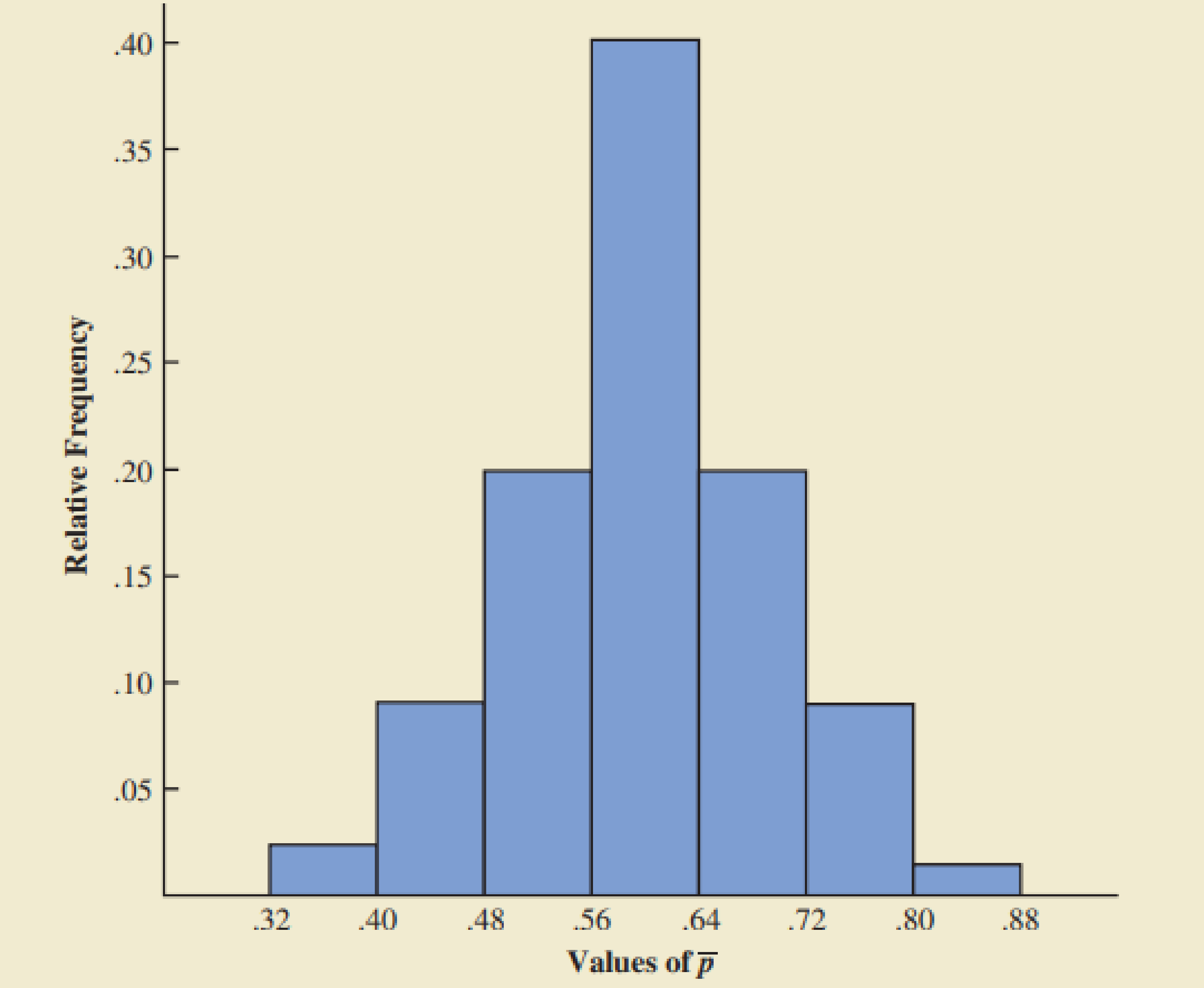


FIGURE 7.2

RELATIVE FREQUENCY HISTOGRAM OF $\bar{p}$ VALUES FROM 500 SIMPLE RANDOM SAMPLES OF SIZE 30 EACH



The sampling distribution of $\bar{x}$ is the probability distribution of all values of the sample mean $\bar{x}$.

E($\bar{x}$): The Expected value of $\bar{x}$ is the mean of $\bar{x}$ (where $\bar{x}$ is a random variable).

$$\mathbf{E}(\bar{\boldsymbol{x}}) = \boldsymbol{\mu}$$, where μ *is the population mean*

Standard Error of ($\bar{x}$) ↔ Standard deviation of ($\bar{x}$):

$$\boldsymbol{\sigma}_{\bar{\boldsymbol{x}}} = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$$, where σ *is the population* standard deviation, and n is the sample size

Note:

- If the population of x is Normally distributed, then sampling distribution of $\bar{\boldsymbol{x}}$ is Normally distributed.
- If the sample size is large (n>30), we assume the measure is Normally distributed.
- If (Expected value of a statistic)=parameter, then:
  The statistic is an unbiased estimator of the parameter. (Example: $\mathrm{E}(\bar{x}) = \mu$).

The sampling distribution of $\bar{\boldsymbol{x}}$ is Normally distributed,

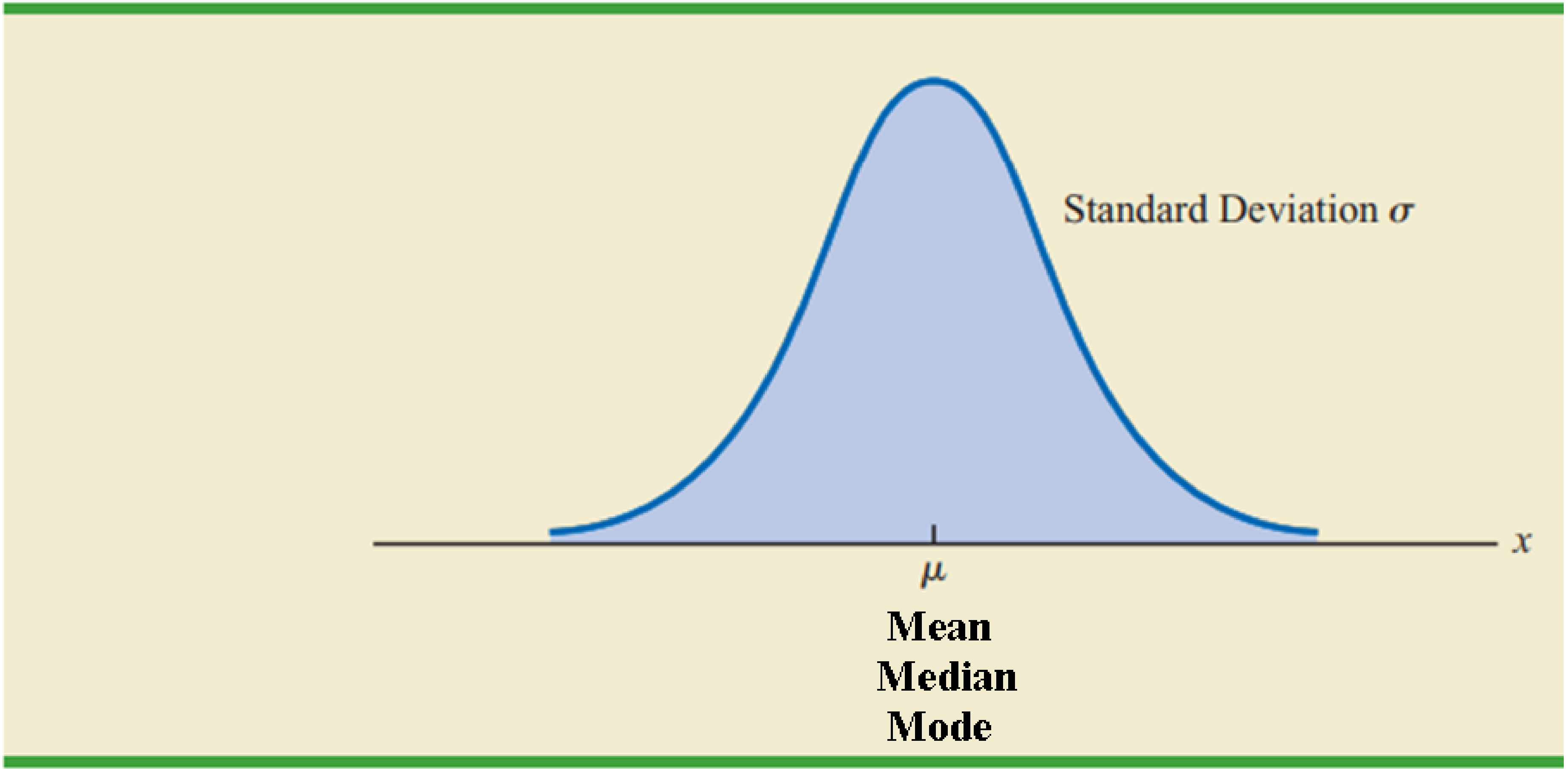
and you can compute any probability of $\bar{x}$ using the normal distribution.

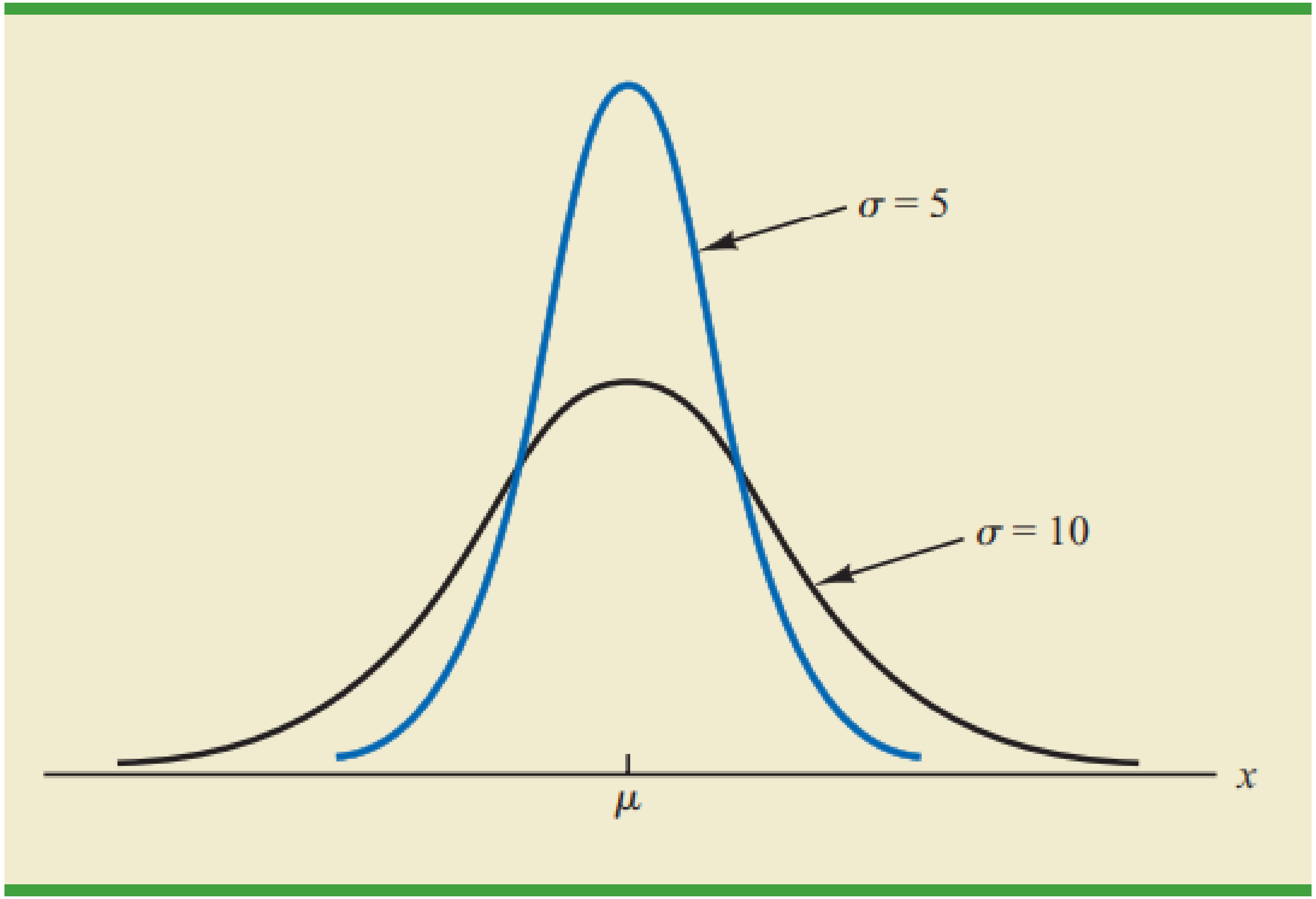
Normal Probability Distribution

The most important probability distribution for describing a continuous random variable is the normal probability distribution.

The normal distribution has the bell shaped curve and symmetric around mean (where mean=median=mode).

FIGURE 6.3 BELL-SHAPED CURVE FOR THE NORMAL DISTRIBUTION



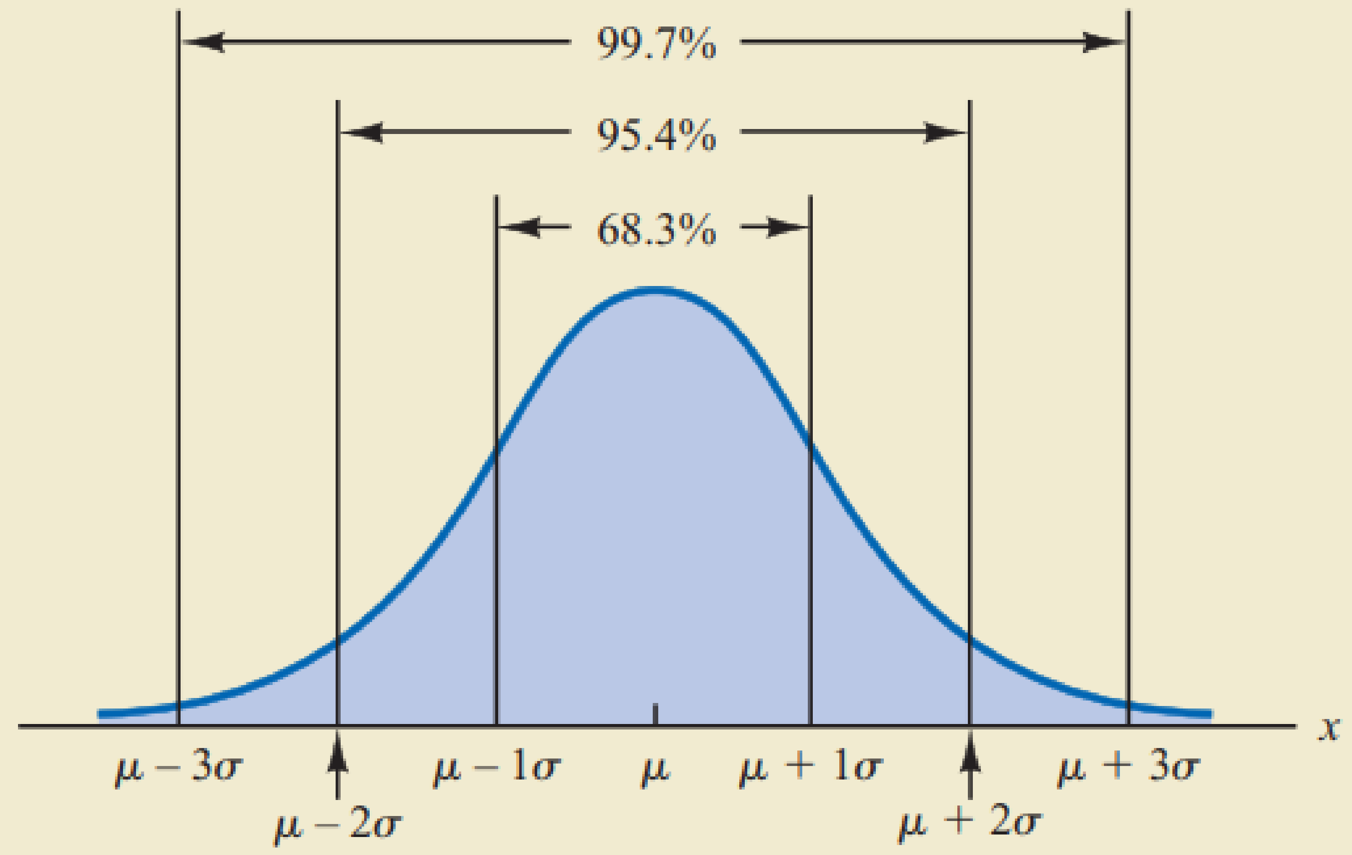


σ = 5
σ = 10
x
μ

Normal distribution properties:

1. The normal distributions is based on two parameters: the mean μ and the variance σ^2.
2. The highest point on the normal curve is at the mean, which is also the median and mode of the distribution.
3. The mean of the distribution can be any numerical value: negative, zero, or positive.
4. The normal distribution is symmetric, with the shape of the normal curve to the left of the mean a mirror image of the shape of the normal curve to the right of the mean. The tails of the normal curve extend to infinity in both directions and theoretically never touch the horizontal axis. Because it is symmetric, the normal distribution is not skewed; its skewness measure is zero.
5. The standard deviation determines how flat and wide the normal curve is. Larger values of the standard deviation result in wider, flatter curves, showing more variability in the data.
6. Probabilities for the normal random variable are given by areas under the normal curve. The total area under the curve for the normal distribution is (1). Because the distribution is symmetric, the area under the curve to the left of the mean is (0.50) and the area under the curve to the right of the mean is (0.50).
7. The percentage of values in some commonly used intervals are:
   a. 68.3% of the values of a normal random variable are within (+) or (-) one standard deviation of its mean.
   b. 95.4% of the values of a normal random variable are within (+) or (-) two standard deviations of its mean.
   c. 99.7% of the values of a normal random variable are within (+) or (-) three standard deviations of its mean

FIGURE 6.4 AREAS UNDER THE CURVE FOR ANY NORMAL DISTRIBUTION



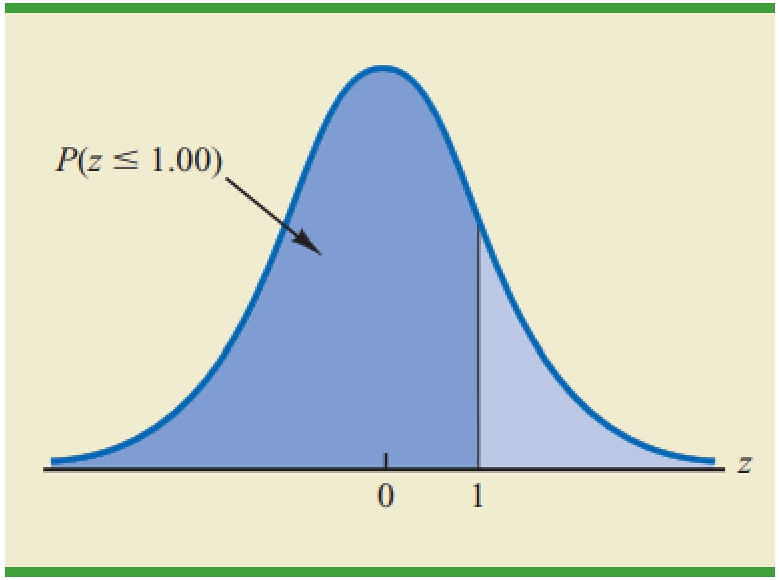
Standard Normal Probability Distribution (Z)

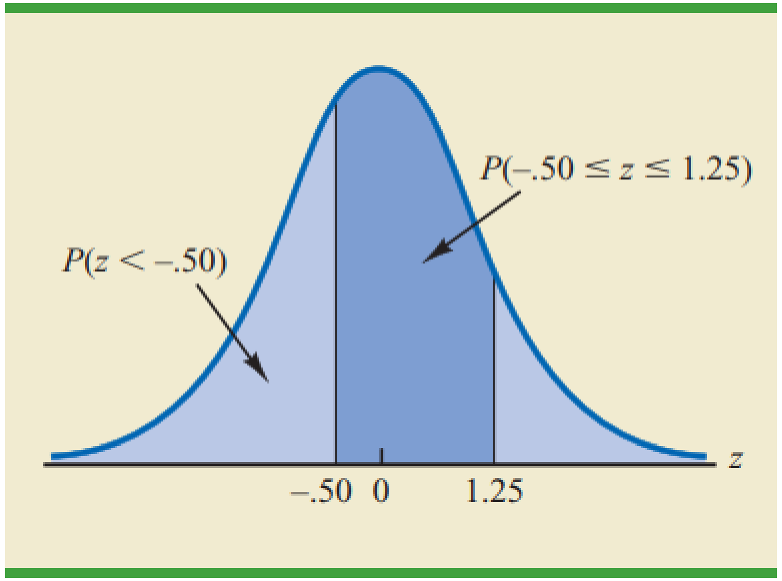
Standard Normal Probability Distribution: Is a normal distributions with mean $\mu = 0$ and variance $\sigma^2 = 1$.

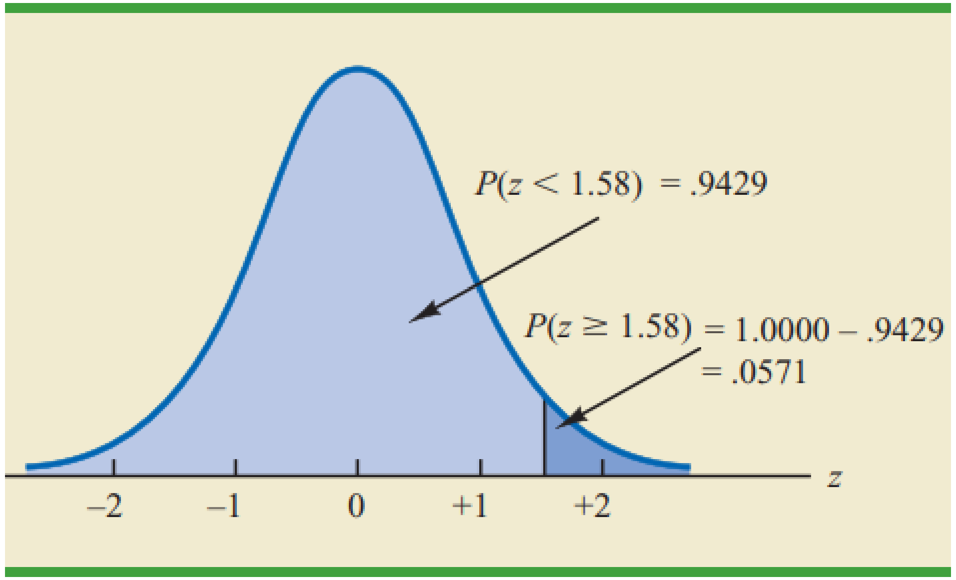
You can **convert any normal distribution of (x)** with **mean μ** and **variance σ^2** to **standard Normal distribution (z)** by:

$$z = \frac{x - \mu}{\sigma}$$

Probabilities for the normal random variable are given by **areas under the normal curve**.
You can use standard Normal table (z).







Example:

Based on the standard Normal table (z), compute:

1. $P(Z < 0)$
2. $P(Z < 1.0)$
3. $P(Z < 1.65)$
4. $P(Z < 0.8)$
5. $P(Z < -1.28)$
6. $P(Z > 0)$
7. $P(Z > -1.34)$
8. $P(Z > 2.0)$
9. $P(-0.32 < Z < 0.5)$
10. $P(-1.28 < Z < 1.28)$
11. $P(-1.65 < Z < 1.65)$
12. $P(-1.96 < Z < 1.96)$
13. What is the value of Z that have 40% of chance (probability) less than Z, i.e. $P(Z < z) = 0.40$
14. What is the value of Z that have 80% of chance (probability) greater than Z, i.e. $P(Z > z) = 0.80$

An Application using the Normal Distribution:

Example:

Assuming a population of 1000 students, if it mean of student marks is 70 and variance is 144. if the student marks is normally distributed.

a) Compute the probability that the student mark is greater than 79? $P(X > 79)$?
b) Compute the probability that the student mark is between 64 and 85?
c) How many students can there marks between 64 and 85?
d) How many students can there marks more than 85?
e) What is the mark that a 330 students will take less than it?
f) If you drawn a random sample from the population. if the sample size is 36. Compute the probability that the sample mean ($\bar{X}$) is greater than 75?

Example:

Assuming a population of 1000 student, if it mean of student marks is 70 and variance is 144. if the student marks is normally distributed.

$$N = 1000\ , \qquad \mu = 70\ , \qquad \sigma^2 = 144 \ \rightarrow \ \sigma = 12,$$

Note:

$$Z = \frac{X - \mu}{\sigma}\ ,$$

a) Compute the probability that the student mark is greater than 79? $P(X > 79)$?

$$P(X > 75) = P\left(Z > \frac{X - \mu}{\sigma}\right) = P\left(Z > \frac{79 - 70}{12}\right) = P(Z > 0.75) = 0.2266$$

$$\boldsymbol{N = 1000}, \quad \boldsymbol{\mu = 70}, \quad \boldsymbol{\sigma^2 = 144 \rightarrow \sigma = 12},$$

Note:

$$\boldsymbol{Z = \frac{X-\mu}{\sigma}},$$

b) Compute the probability that the student mark is between 64 and 85?

$P(64 < X < 85)$?

$$P(64 < X < 85) = P\left(\frac{\boldsymbol{X_1 - \mu}}{\boldsymbol{\sigma}} < \boldsymbol{Z} < \frac{\boldsymbol{X_2 - \mu}}{\boldsymbol{\sigma}}\right)$$

$$= P\left(\frac{64-70}{12} < Z < \frac{85-70}{12}\right)$$

$$= P(-0.5 < Z < 1.25)$$

$$= T(1.25) - T(-0.5)$$

$$= 0.8944 - 0.3085 = 0.5859$$

$$\boldsymbol{N = 1000}, \quad \boldsymbol{\mu = 70}, \quad \boldsymbol{\sigma^2 = 144 \rightarrow \sigma = 12},$$

Note: $$\boldsymbol{Z = \frac{X-\mu}{\sigma}},$$

c) How many students can there marks between 64 and 85?

Note: Number of student is = Population size * Probability

$= \mathrm{N} * P(64 < X < 85)$?

$$= 1000*0.5859$$

= "about 586 from 1000 “ will take marks between than 64 and 85

d) How many students can there marks more than 85?

$= \mathrm{N} * P(X > 85)$?

$$= 1000*(1-0.8944)=1000*0.1056$$

= "about 106 from 1000 “ will take marks more than 85

$$N = 1000, \quad \mu = 70, \quad \sigma^2 = 144 \rightarrow \sigma = 12,$$

Note: $$Z = \frac{X-\mu}{\sigma},$$

e) What is the mark that a 330 students will take less than it?

$330 = 1000 * P(X < x_{330}) \rightarrow P(X < x_{330}) = 0.33?$

From Z table, find the z values that have probability (area) less than it= 0.33:

$$P(Z < -0.44) = 0.33 \quad , z = -0.44$$

$$Z = \frac{X - \mu}{\sigma}$$

$$-0.44 = \frac{x_{330} - \mu}{\sigma}$$

$$-0.44 = \frac{x_{330} - 70}{12}$$

$$x_{330} = 64.72$$

→ There are 330 students will take less than the mark 64.72

Sampling distribution of Mean (μ):

Assuming X distributed Normal (mean=μ, Variance=σ_X^2).

Then $\bar{X}$ is distributed Normal (mean=μ, Variance=$\sigma_{\bar{X}}^2$).

Such that: $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$

Then: We can convert $\bar{X}$ to standard Normal distribution using:

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

f) If you drawn a random sample from the population. if the sample size is 36. Compute the probability that the sample mean ($\bar{X}$) is greater than 75?

Note: $N = 1000$, $\mu = 70$, $\sigma^2 = 144$, $n = 36$, $P(\bar{X} > 75)$?

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

$$P(\bar{X} > 75) = P\left(Z > \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) = P\left(Z > \frac{75 - 70}{\frac{12}{\sqrt{36}}}\right) = P(Z > 2.50) = 0.0062$$

g) If you drawn a random sample from the population. if the sample size is 25. Compute the probability that the sample mean ($\bar{X}$) is greater than 76?

$$N = 1000\,, \qquad \mu = 70\,, \qquad \sigma^2 = 144, \qquad n = 25 \quad , \qquad P(\bar{X} > 78)\,?$$

Note:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \;, \mu_{\bar{X}} = \mu \;, \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$P(\bar{X} > 75) = P\left(Z > \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) = P\left(Z > \frac{76 - 70}{\frac{12}{\sqrt{25}}}\right) = P(Z > 0.42) =$$

$$= 1 - P(Z < 0.42) = 1 - 0.6628 = 0.3372$$

sampling distribution of proportion

The sampling distribution of P is the probability distribution of all possible values of the sample proportion $\bar{p}$.

The sampling distribution of P is the probability distribution of all possible values of the sample proportion $\bar{p}$.

$E(\bar{p})$: The Expected value of $\bar{p}$ is the mean of $\bar{p}$ (where $\bar{p}$ is a random variable).

$$E(\bar{p}) = p$$, where p *is the population* proportion

Standard Error of ($\bar{p}$) ↔ Standard deviation of ($\bar{p}$):

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$, where p *is the population* proportion , and n is the sample size

Note:

- **The population of $\bar{p}$ can be approximated by a normal distribution** whenever **$np \geq 5$** and **$n(1-p) \geq 5$**

<u>In this course</u>

we assuming the sampling distribution of $\bar{p}$ is Normally distributed,

and you can compute any probability of $\bar{p}$ using the normal distribution.

Example:

Assuming a sample of 30 students, if the population proportion of failed student is 0.2.

Compute the Expected value of the sample proportion $\bar{p}$. And the Standard Error of ($\bar{p}$):

Answer:

$$\mathbf{E(\bar{p}) = p = 0.2}$$, where p *is the population* proportion

Standard Error of ($\bar{p}$) ↔ Standard deviation of ($\bar{p}$):

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2(1-0.2)}{30}} = 0.073$$