EMA 351SW: STATISTICS AND PROBABILITY

Unit 3a: The Normal Distribution

TOPIC OUTLINE

- The Normal Distribution
  1) Introduction
  2) Definition of Terms and Statistical Symbols Used
  3) How To Find Areas Under the Normal Curve
  4) Finding the Unknown Z represented by Z_o
  5) Examples
- Hypothesis Testing

The Normal Distribution

- **Introduction**

Before exploring the complicated Standard Normal Distribution, we must examine how the concept of Probability Distribution changes when the Random Variable is Continuous.

The Normal Distribution

- Introduction

A Probability Distribution will give us a Value of P(x) = P(X=x) to each possible outcome of x. For the values to make a Probability Distribution, we needed two things to happen:

1. $\Sigma P(x) = P(X = x)$
2. $0 \leq P(x) \leq 1$

For a *Continuous* Random Variable, a Probability Distribution must be what is called a **<u>Density Curve.</u>** This means:

1. The Area under the Curve is 1.
2. $0 \leq P(x)$ for all outcomes x.

The Normal Distribution

- **Introduction: Example:**

Suppose the temperature of a piece of metal is always between 0°F and 10°F. Furthermore, suppose that it is equally likely to be any temperature in that range. Then the graph of the probability distribution for the value of the temperature would look like the one below:

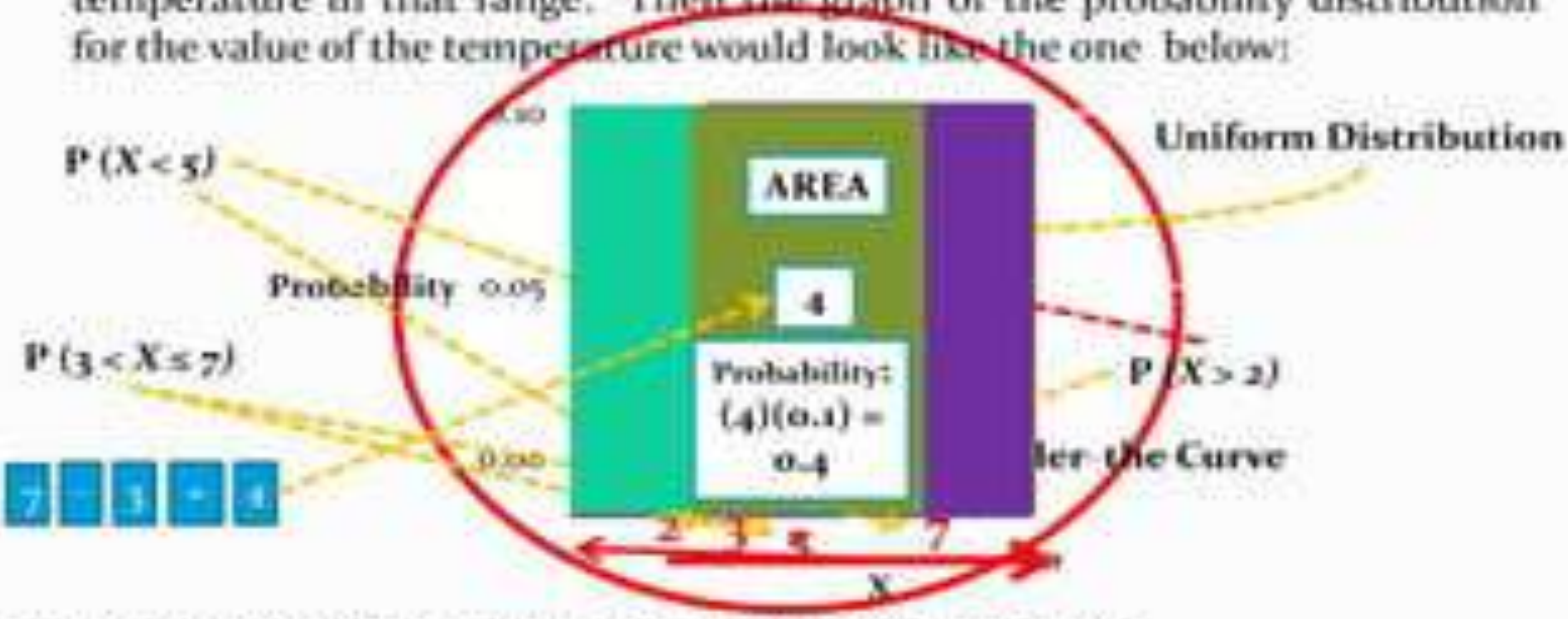


Illustration of the fundamental fact about DENSITY CURVE

The Normal Distribution: Definition of Terms and Symbols Used

- Normal Distribution Definition:

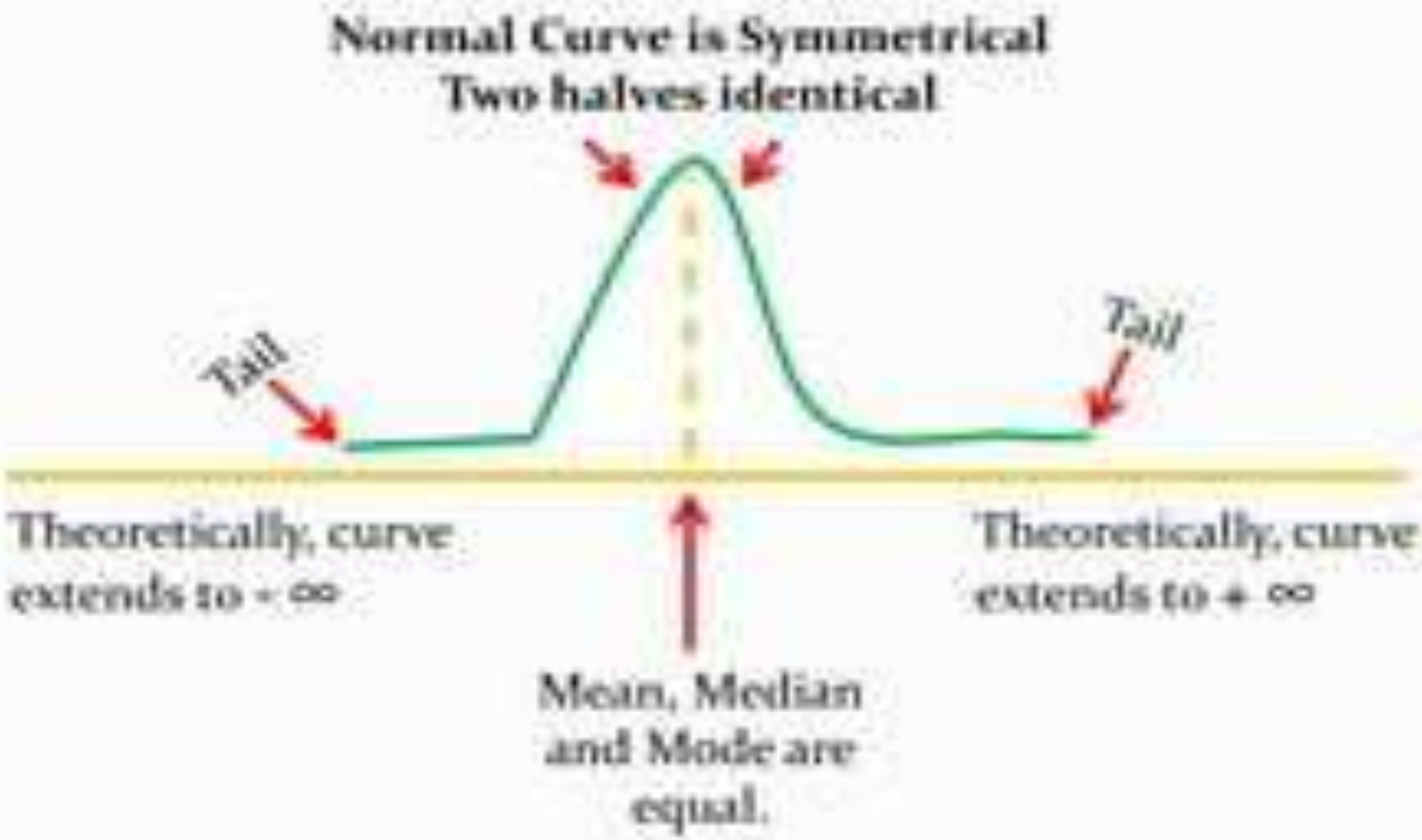
  1) A continuous variable X having the symmetrical, bell shaped distribution is called a Normal Random Variable.
  2) The *normal probability distribution (Gaussian distribution)* is a continuous distribution which is regarded by many as the most significant probability distribution in statistics particularly in the field of statistical inference.

- Symbols Used:
  - "z" – z-scores or the standard scores. The table that transforms every normal distribution to a distribution with mean 0 and standard deviation 1. This distribution is called the ***standard normal distribution*** or simply ***standard distribution*** and the individual values are called ***standard scores*** or the ***z-scores.***
  - "μ" – the Greek letter "mu," which is the Mean, and
  - "σ" – the Greek letter "sigma," which is the Standard Deviation

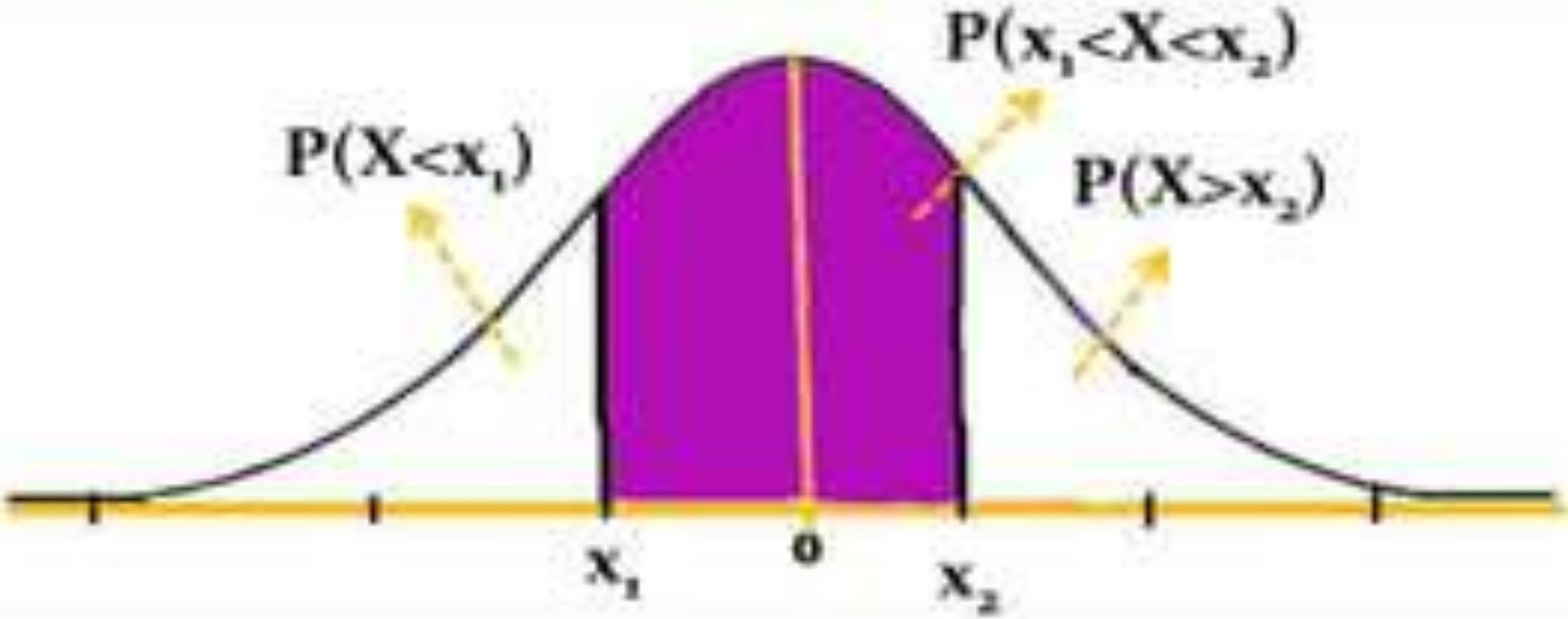
The Normal Distribution: Definition of Terms and Symbols Used

- Characteristics of Normal Distribution:
  1) It is "**Bell-Shaped**" and has a single peak at the center of the distribution,
  2) The arithmetic Mean, Median and Mode are equal.
  3) The total area under the curve is 1.00; half the area under the normal curve is to the right of this center point and the other half to the left of it,
  4) It is **Symmetrical** about the mean,
  5) It is **Asymptotic:** The curve gets closer and closer to the X – axis but never actually touches it. To put it another way, the tails of the curve extend indefinitely in both directions.
  6) The location of a normal distribution is determined by the Mean, **μ**, the Dispersion or spread of the distribution is determined by the Standard Deviation, σ.

The Normal Distribution: Graphically



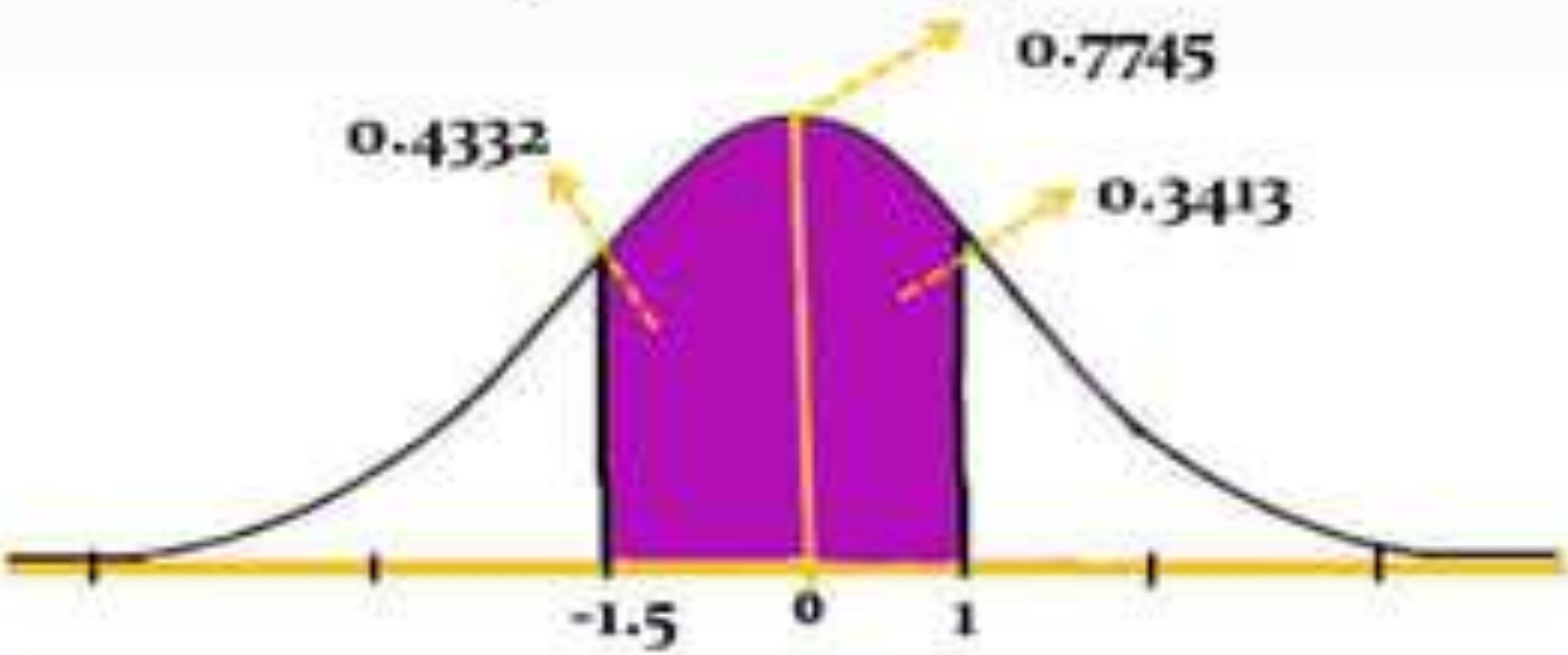
AREA UNDER THE NORMAL CURVE



AREA UNDER THE NORMAL CURVE

Let us consider a variable X which is normally distributed with a mean of 100 and a standard deviation of 10. We assume that among the values of this variable are x1= 110 and x2 = 85.

$$z_1 = \frac{110 - 100}{10} = 1.00 \qquad z_2 = \frac{85 - 100}{10} = -1.50$$



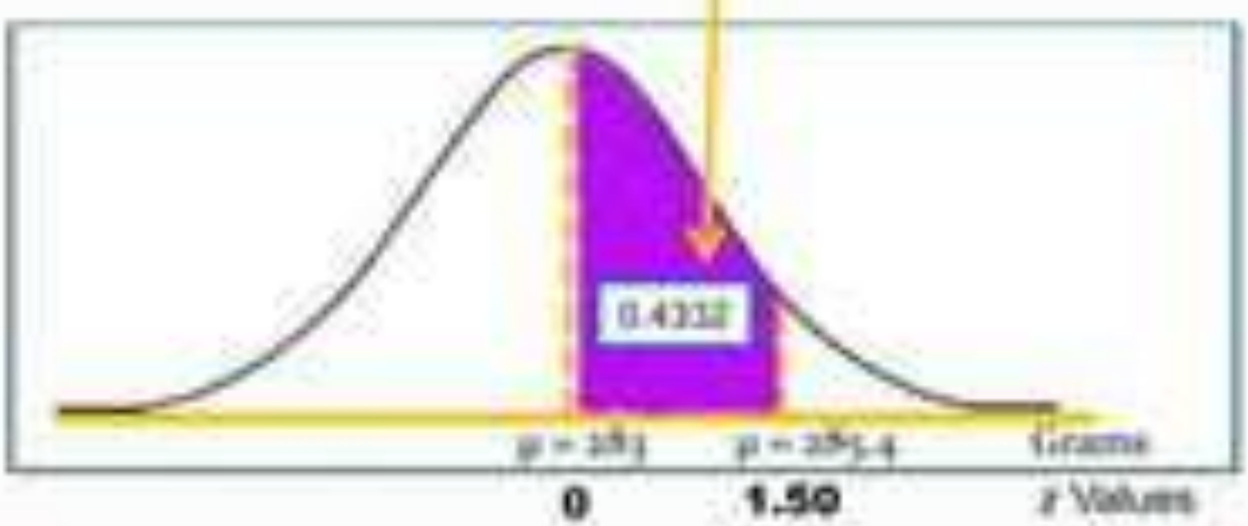
The Standard Normal Probability Distribution

- The Standard Normal Distribution is a Normal Distribution with a Mean of **0** and a Standard Deviation of **1**.
- It is also called the *z* distribution
- A ***z* -value** is the distance between a selected value , designated *X*, and the population Mean **μ**, divided by the Population Standard Deviation, **σ**.
- The formula is :

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

Areas Under the Normal Curve

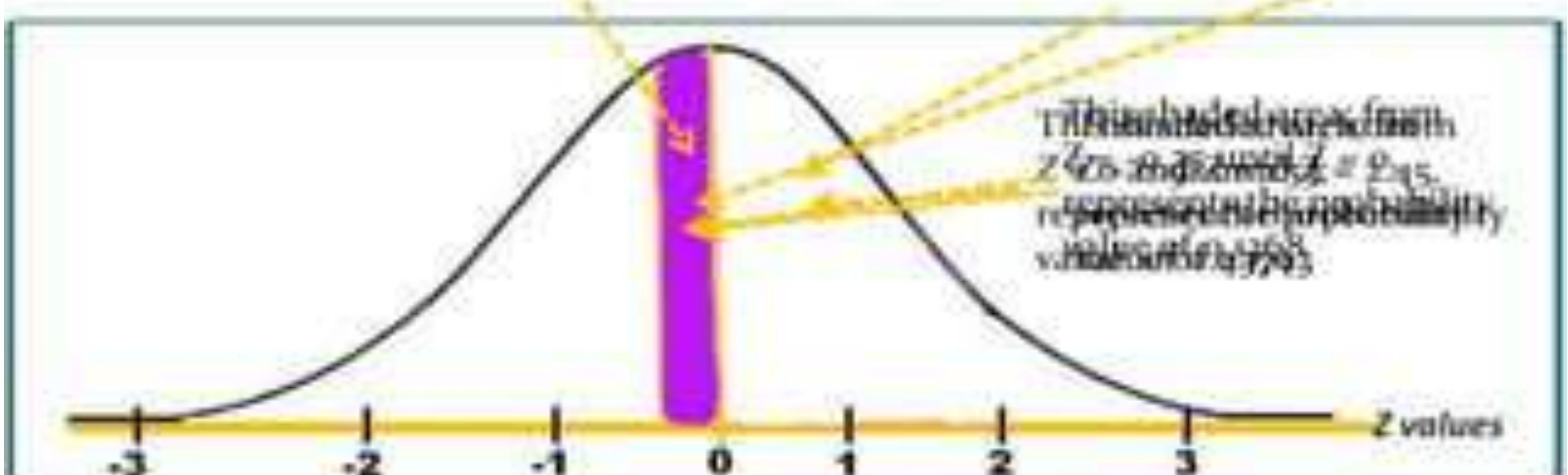
z	0	1	2	3	4	5	6	7	8	9
0	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0.1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.5	**0.4332**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



How to Find Areas Under the Normal Curve

Let
: **Z** be a standardized random variable
: **P** stands for Probability
: **Φ(z)** indicates the area co[illegible] under the Normal Curve.

b.) P ([illegible]) = Φ([illegible]) − Φ (0.01)
= [illegible] − 0.3186
= 0.1743



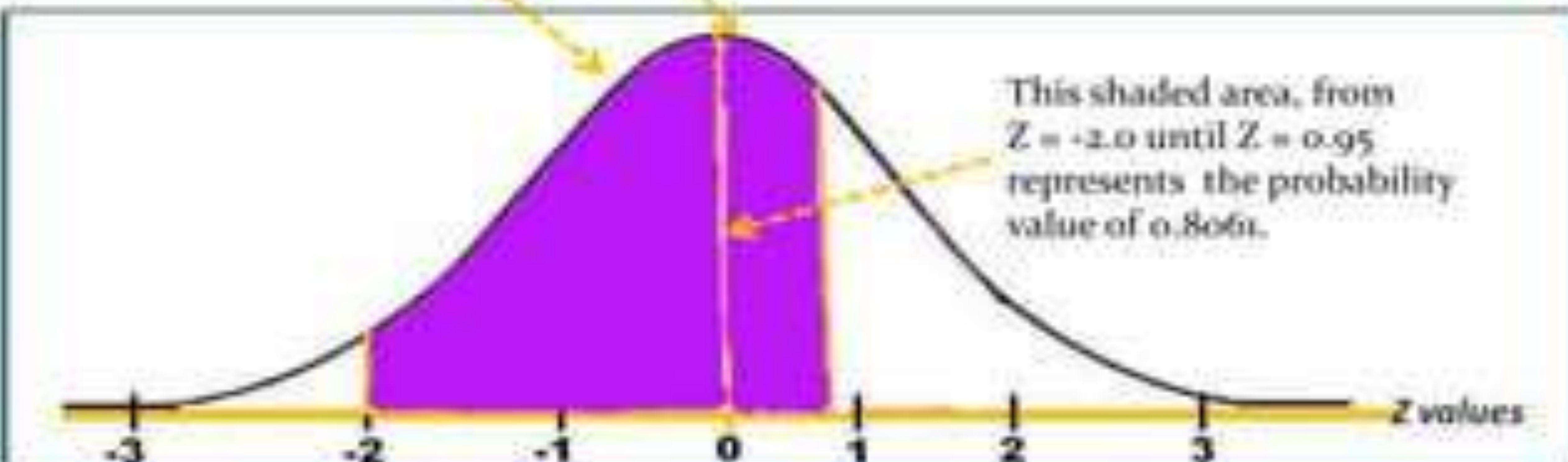
How to Find Areas Under the Normal Curve

Let
- Z be a standardized random variable
- **P** stands for Probability
- $\Phi(z)$ indicates the area covered under the Normal Cu

d.) $P(-2.0 \le Z \le 0.95) = \Phi(-2.0) + \Phi(0.95)$

$= 0.4772 + 0.3289$

$= 0.8061$



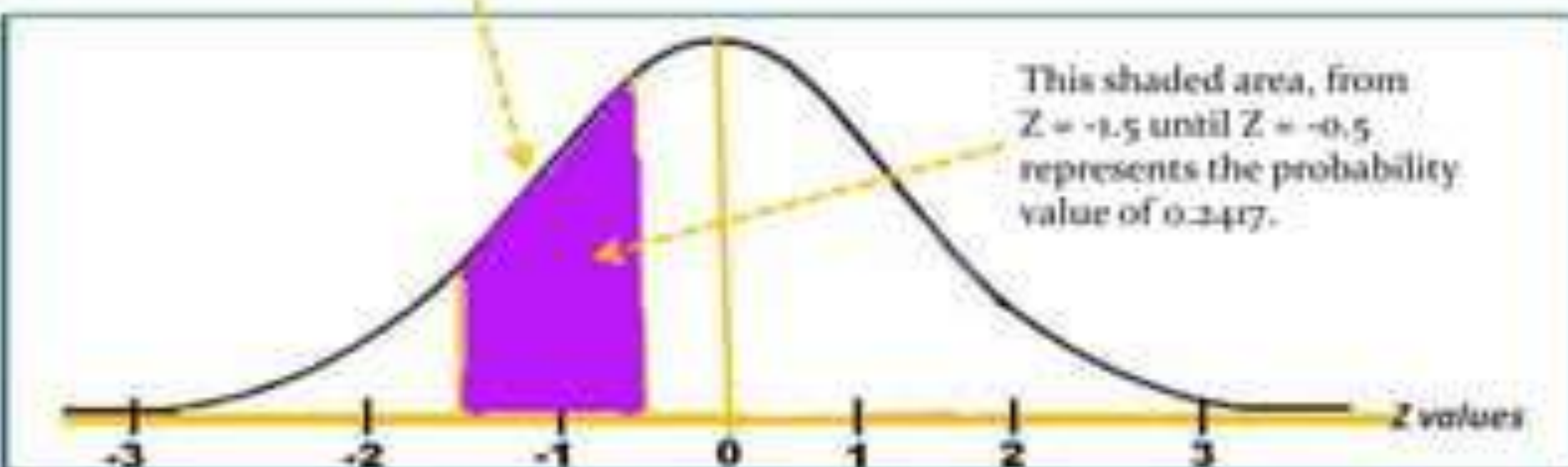
How to Find Areas Under the Normal Curve

Let Z be a standardized random variable
P stands for Probability
$\Phi(z)$ indicates the area covered under the Normal Curve

e.) $P(-1.5 \le Z \le -0.5) = \Phi(-1.5) - \Phi(-0.5)$

$= 0.4332 - 0.1915$

$= 0.2417$

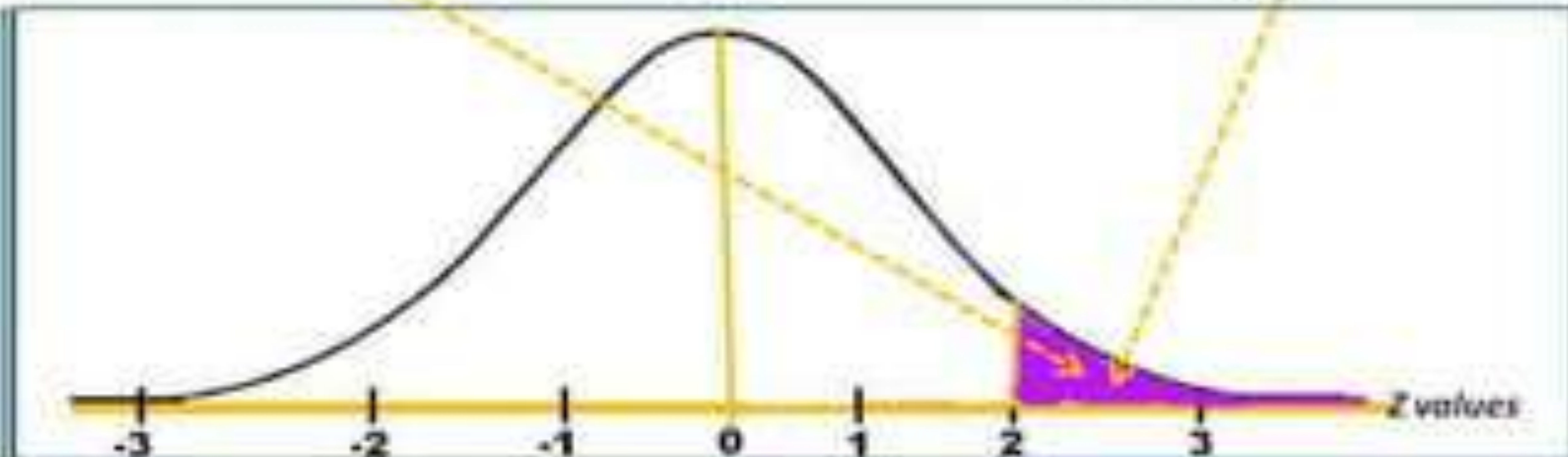


How to Find Areas Under the Normal Curve

Let Z be a standardized random variable
P stands for Probability
Φ(z) indicates the area covered under the Normal Curve

(f.) $P(Z \geq 2.0) = 0.5 - \Phi(2.0)$
$= 0.5 - 0.4772$
$= 0.0228$

This shaded area, from Z = 2.0 until beyond Z = 3 represents the probability value of 0.0228.



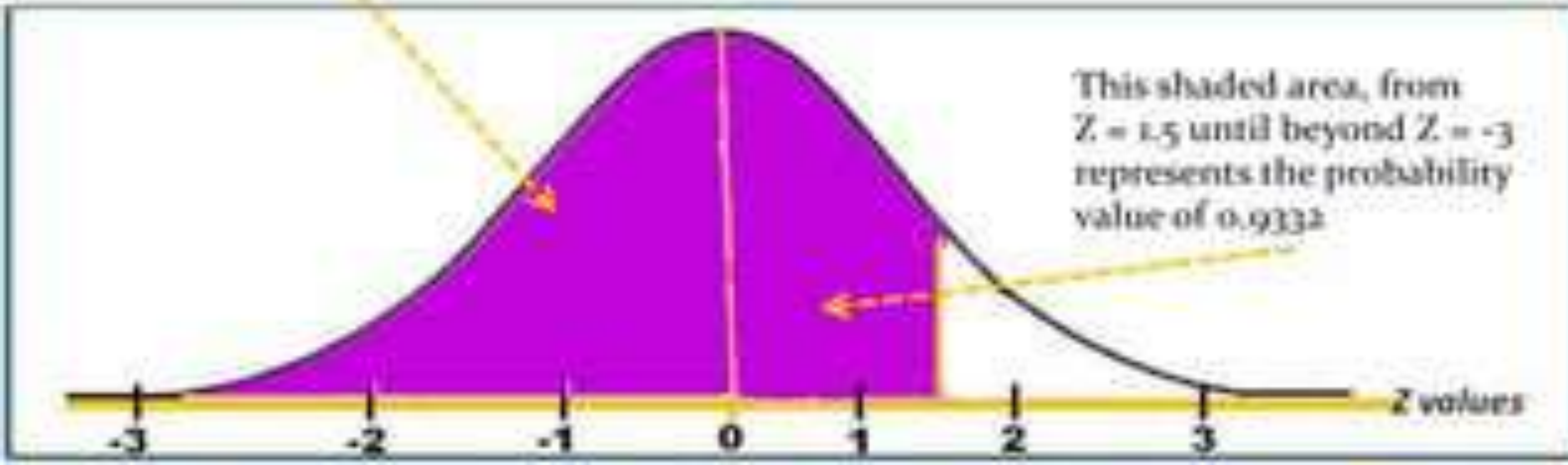
How to Find Areas Under the Normal Curve

Let
- **Z** be a standardized random variable
- **P** stands for Probability
- **Φ(z)** indicates the area covered under the Normal Curve.

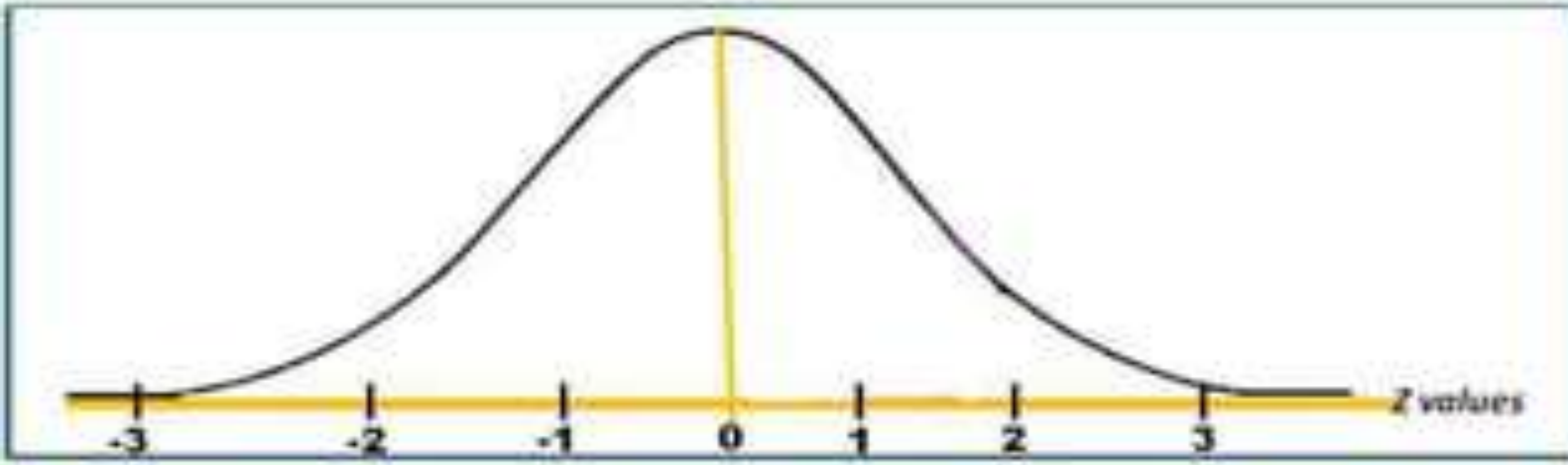
g.) $P(Z \leq 1.5) = 0.5 + \Phi(1.5)$

$= 0.5 + 0.4332$

$= 0.9332$



Finding the unknown Z represented by Z_o



Finding the unknown Z represented by Z_0

- $P(Z \leq Z_0) = 0.8461$

  $0.5 + X = 0.8461$

  $X = 0.8461 - 0.5$

  $X = 0.3461$

  $Z_0 = 1.02$ ans.

- $P(-1.72 \leq Z \leq Z_0) = 0.9345$
  - $\Phi(-1.72) + X = 0.9345$
  - $X = 0.9345 - 0.4573$
  - $X = 0.4772$
  - $Z_0 = 2.0$

Finding the unknown Z represented by Z_o

CASE	MNEMONICS
$0 \leq Z \leq Z_o$	$\Phi(Z_o)$
$(-Z_o \leq Z \leq 0)$	$\Phi(Z_o)$
$Z_1 \leq Z \leq Z_2$	$\Phi(Z_2) - \Phi(Z_1)$
$(-Z_1 \leq Z \leq Z_2)$	$\Phi(Z_1) + \Phi(Z_2)$
$(-Z_1 \leq Z \leq -Z_2)$	$\Phi(Z_1) - \Phi(Z_2)$
$Z \geq Z_o$	$0.5 - \Phi(Z_o)$
$Z \leq Z_o$	$0.5 + \Phi(Z_o)$
$Z \leq -Z_o$	$0.5 - \Phi(Z_o)$
$Z \geq -Z_o$	$0.5 + \Phi(Z_o)$

The event X has a normal distribution with mean $\mu = 10$ and Variance = 9. Find the probability that it will fall:

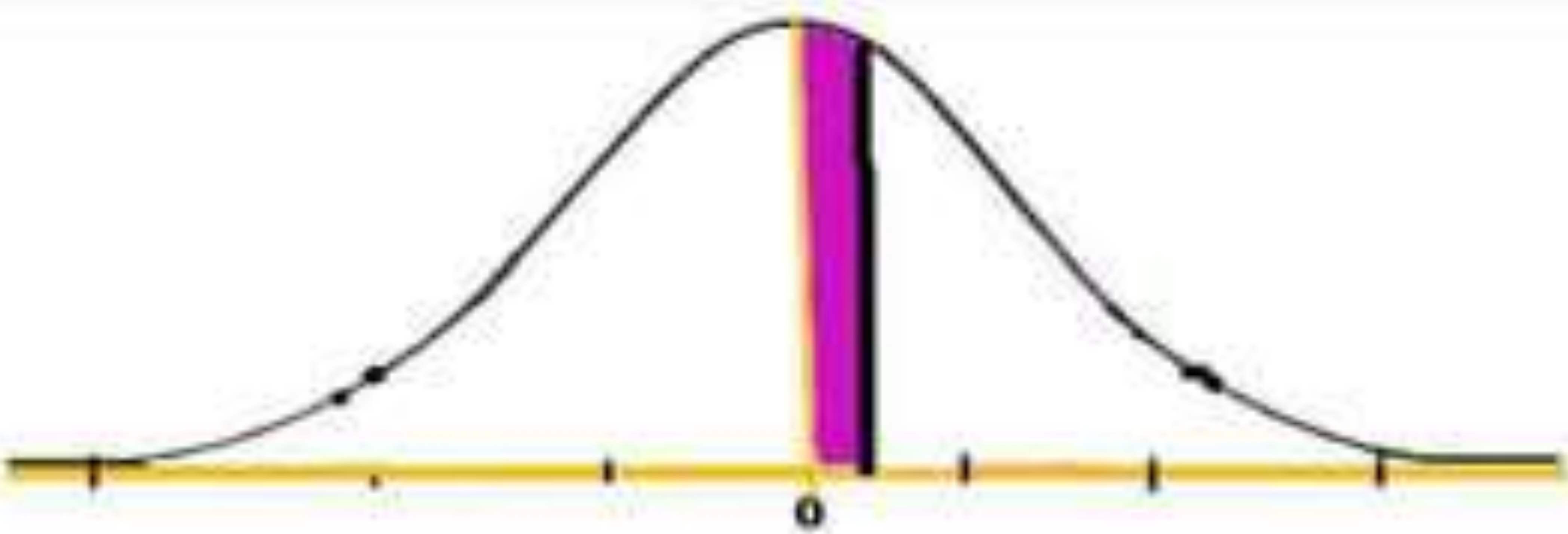
a.) between 10 and 11
b.) between 12 and 19
c.) above 13
d.) at $x = 11$
e.) between 8 amd 12

$$a.)\ Z = \frac{x-\mu}{\sigma} = \frac{10-10}{3} = 0$$

$$Z = \frac{x-\mu}{\sigma} = \frac{11-10}{3} = \frac{1}{3} = 0.33$$

$$P(10 \le X \le 11) = P(0 \le Z \le 0.33)$$

$$= \phi(0.33) = 0.1293$$

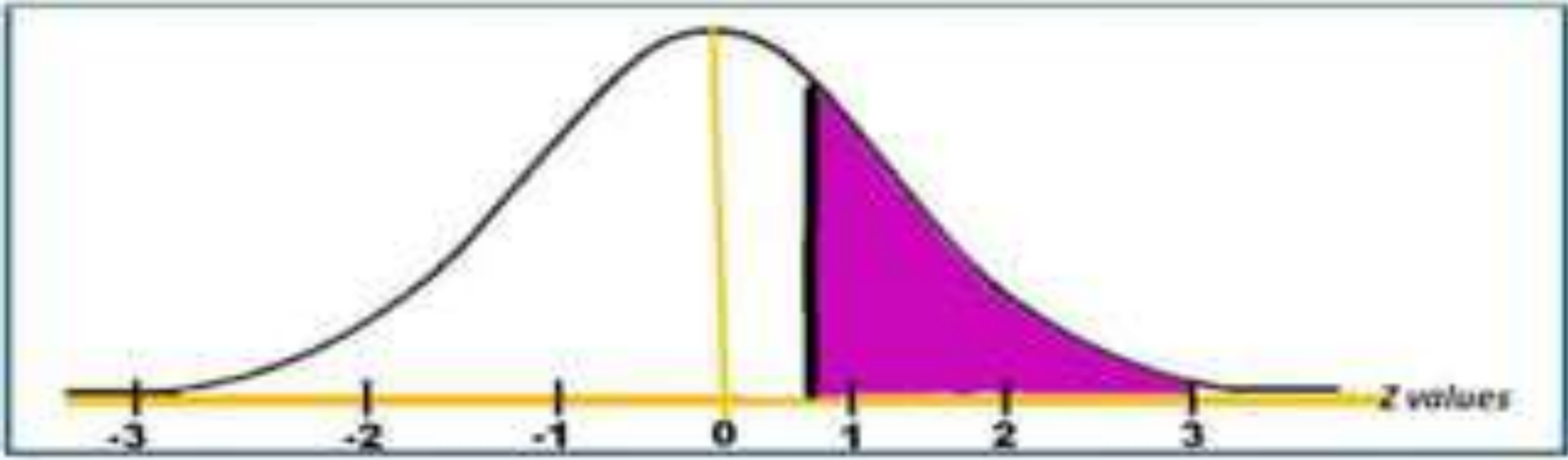


$$b.)\ Z = \frac{x-\mu}{\sigma} = \frac{12-10}{3} = \frac{2}{3} = 0.67$$

$$Z = \frac{x-\mu}{\sigma} = \frac{19-10}{3} = \frac{9}{3} = 3$$

$$P(12 \le X \le 19) = P(0.67 \le Z \le 3)$$

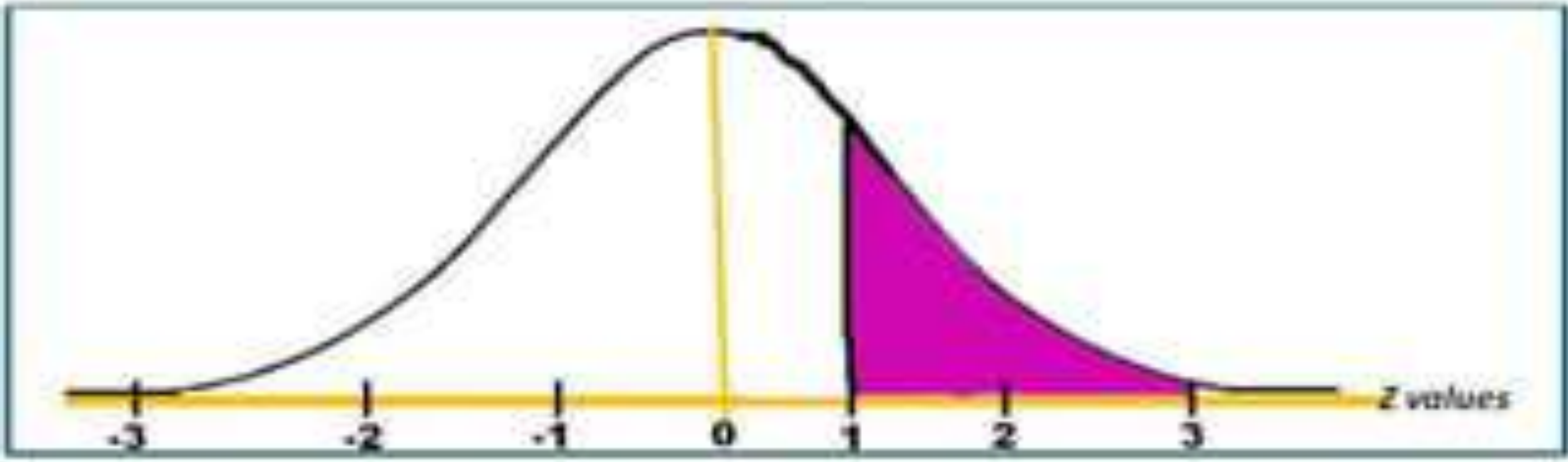
$$= \phi(3) - \phi(0.67) = 0.4987 - 0.2486 = 0.251$$



$$c.)\ Z = \frac{x - \mu}{\sigma} = \frac{13 - 10}{3} = \frac{3}{3} = 1$$

$$P(X \geq 13) = P(Z \geq 1) = 0.5 - \phi(1)$$

$$= 0.5 - 0.3413 = 0.1587$$

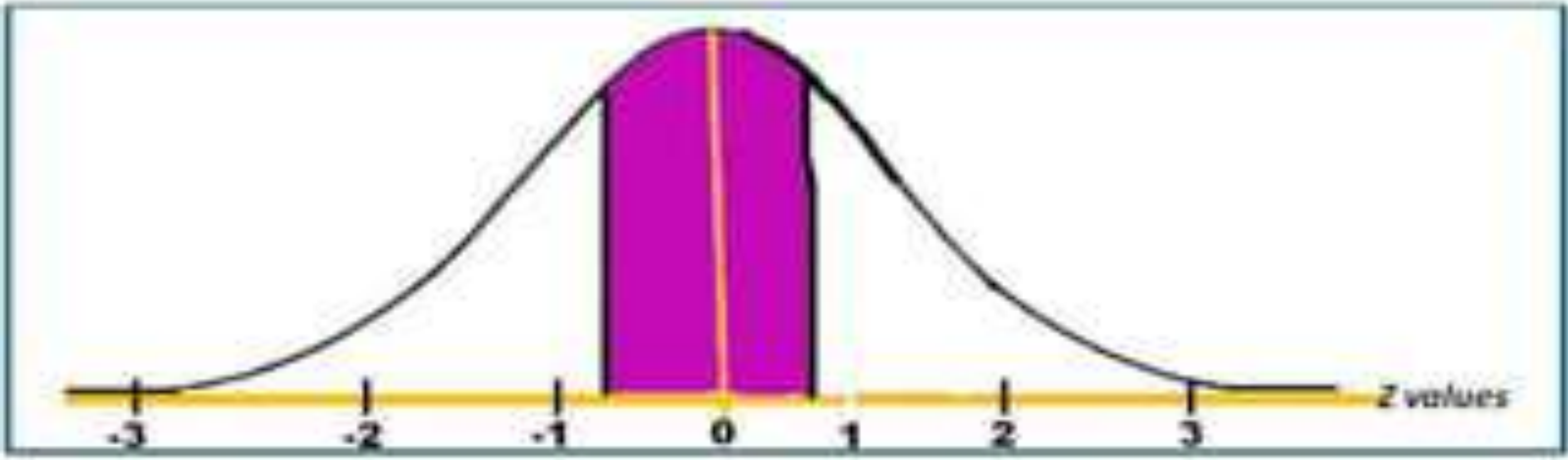


d.) $P(X = 11) = 0$

e.) $Z = \frac{x - \mu}{\sigma} = \frac{8 - 10}{3} = \frac{-2}{3} = -0.67$

$P(8 \le X \le 12) = P(-0.67 \le Z \le 0.67)$

$= \phi(0.33) = 0.1293$



2. A random variable X has a normal distribution with mean 5 and variance 16.

a.) Find an interval (b,c) so that the probability of X lying in the interval is 0.95.
b.) Find d so that the probability that $X \geq d$ is 0.05.

1. $z = \frac{X - \mu}{\sigma} = \frac{X - 5}{4}$

2. $\frac{0.95}{2} = 0.475$

Solution: A.

$P(b \leq X \leq c) = P(Z_b \leq Z \leq Z_c)$

$= P(-1.96 \leq \frac{X - 5}{4} \leq 1.96)$

$= P(-1.96(4) \leq X - 5 \leq 1.96(4))$

$= P(-7.84 + 5 \leq X \leq 7.84 + 5)$

$P(b \leq X \leq c) = P(-2.84 \leq X \leq 12.84)$

thus: $b = -2.84$ and $c = 12.84$

2. A random variable X has a normal distribution with mean 5 and variance 16.

a.) Find an interval (b,c) so that the probability of X lying in the interval is 0.95.
b.) Find d so that the probability that $X \geq d$ is 0.05.

Solution B:

$P(X \geq d) = P(Z \geq Z_d) = 0.052$

$= P\left(\frac{X-5}{4}\right) \geq 1.64$

$= P[X - 5 \geq (1.64)(4)]$

$= P(X - 5 \geq 6.56)$

$= P(X \geq 6.56 + 5)$

$P(X \geq d) = P(X \geq 11.56)$

thus: d = 11.56

1 $0.5 - 0.05 = 0.45$

from the table

$0.45 \rightarrow Z = 1.64$

3. A certain type of storage battery last on the average 3.0 years, with a standard deviation σ of 0.5 year. Assuming that the battery are normally distributed, find the probability that a given battery will last less than 2.3 years.

Solution:

$$z = \frac{X - \mu}{\sigma} = \frac{2.3 - 3}{0.5} = \frac{-0.7}{0.5} = -1.4$$

$$\begin{aligned} P(X < 2.3) &= P(Z < -1.4) \\ &= 0.5 - \Phi(-1.4) \\ &= 0.5 - 0.4192 \\ P(X < 2.3) &= 0.0808 \end{aligned}$$