Business Statistic

Week 6

Continuous Probability

Normal Distribution

Agenda

Time	Activity
70 minutes	Normal Probability Distribution
30 minutes	Evaluating Normality
100 minutes	Exercise

Learning Objectives

In this chapter, you learn:

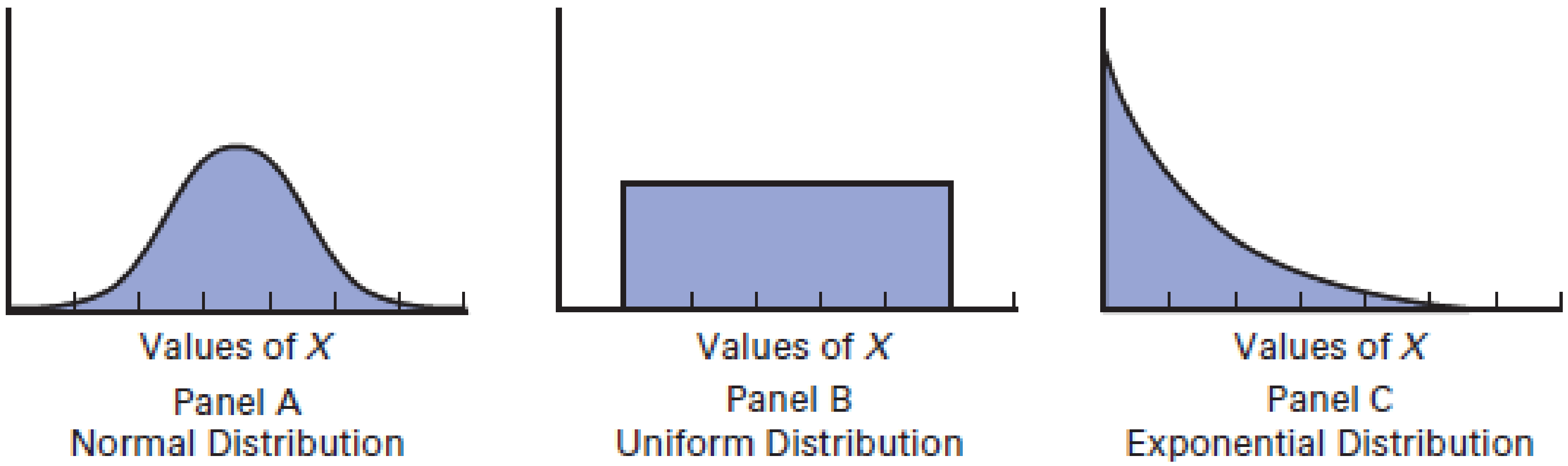
- To compute probabilities from the normal distribution
- To use the normal probability plot to determine whether a set of data is approximately normally distributed

Continuous Probability Distributions

- A continuous random variable is a variable that can assume any value on a continuum (can assume an uncountable number of values)
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
  - height, in inches

- These can potentially take on any value depending only on the ability to precisely and accurately measure

Continuous Probability Distributions



The Normal Distribution

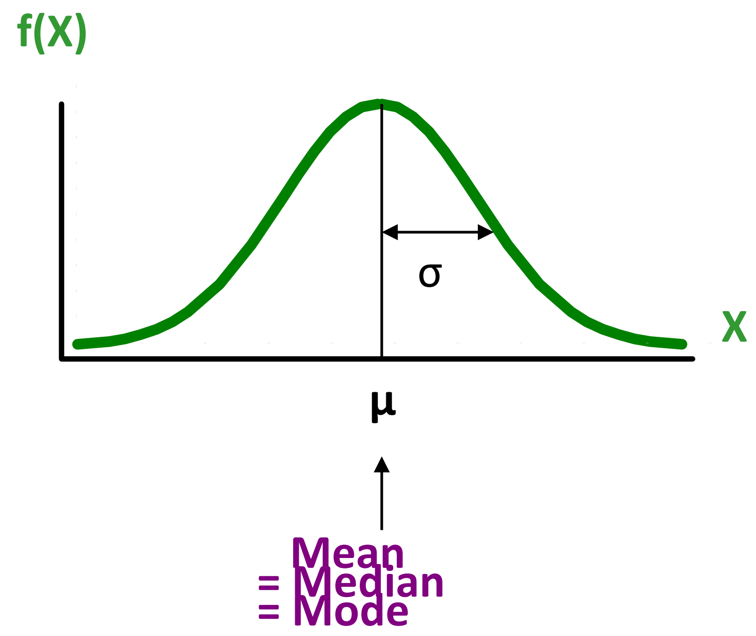
- 'Bell Shaped'
- Symmetrical
- Mean, Median and Mode are Equal

Location is determined by the mean, μ

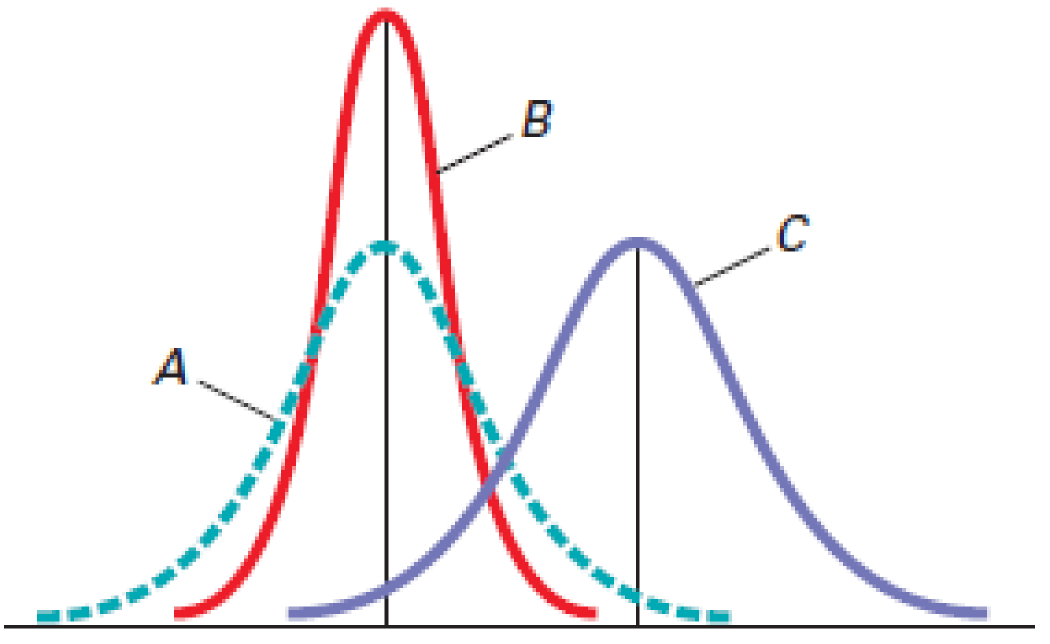
Spread is determined by the standard deviation, σ

The random variable has an infinite theoretical range:

$+\infty$ to $-\infty$



Many Normal Distributions



By varying the parameters μ and σ, we obtain different normal distributions

The Standardized Normal

- Any normal distribution (with any mean and standard deviation combination) can be transformed into the standardized normal distribution (Z)
- Need to transform X units into Z units
- The standardized normal distribution (Z) has a mean of 0 and a standard deviation of 1

Translation to the Standardized Normal Distribution

Translate from X to the standardized normal (the "Z" distribution) by subtracting the mean of X and dividing by its standard deviation:

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

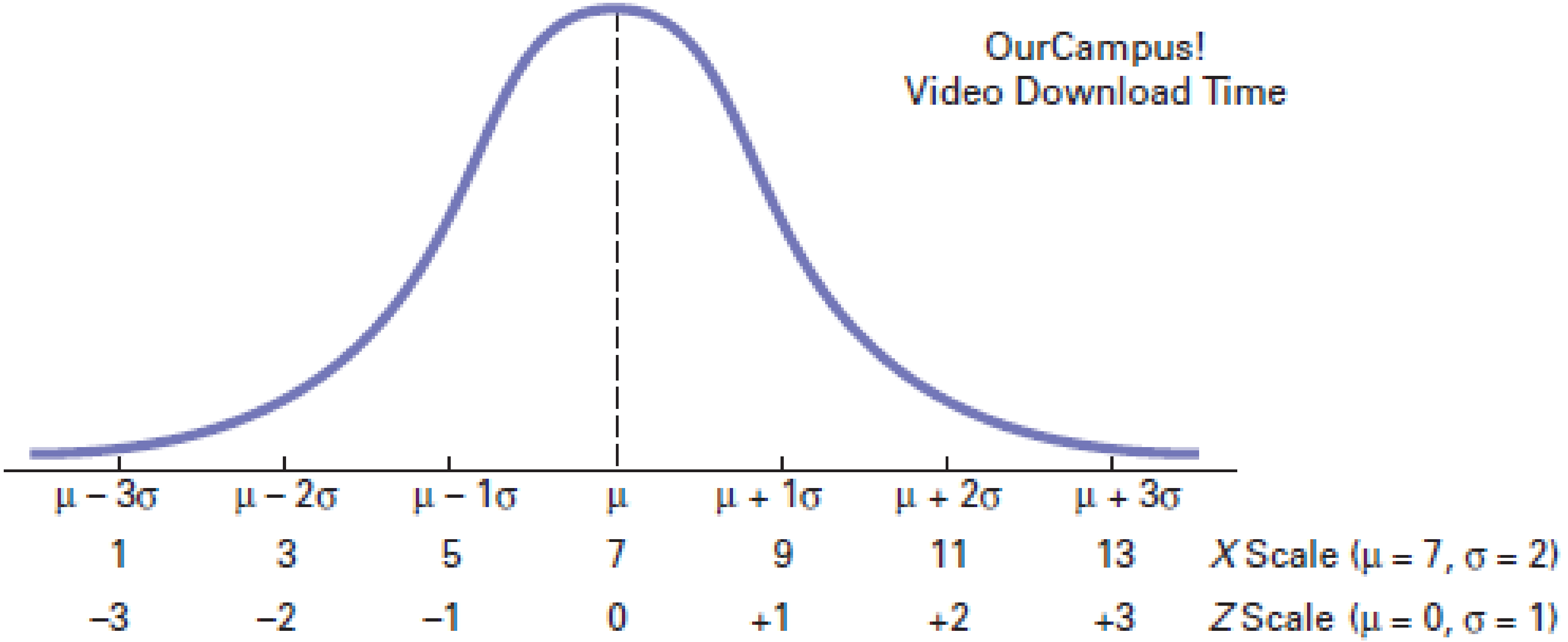
The Z distribution always has mean = 0 and standard deviation = 1

Case

- OurCampus! Website
- Past data indicate that the mean download time is 7 seconds, and that the standard deviation is 2 seconds
- The download times are distributed as a bell-shaped curve, with a clustering around the mean of 7 seconds.

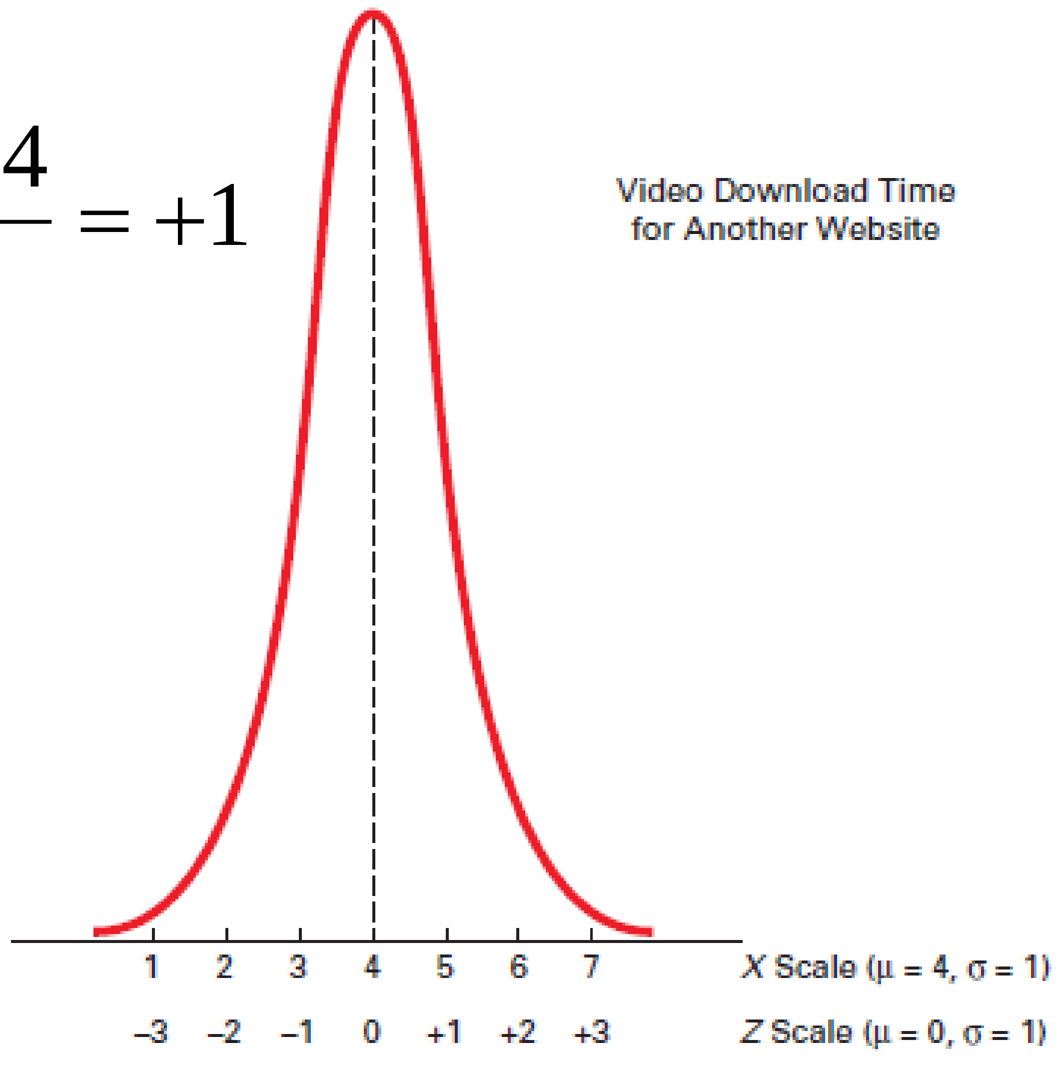
The Standardized Normal OurCampus!

$$Z = \frac{X - \mu}{\sigma} = \frac{9 - 7}{2} = +1$$

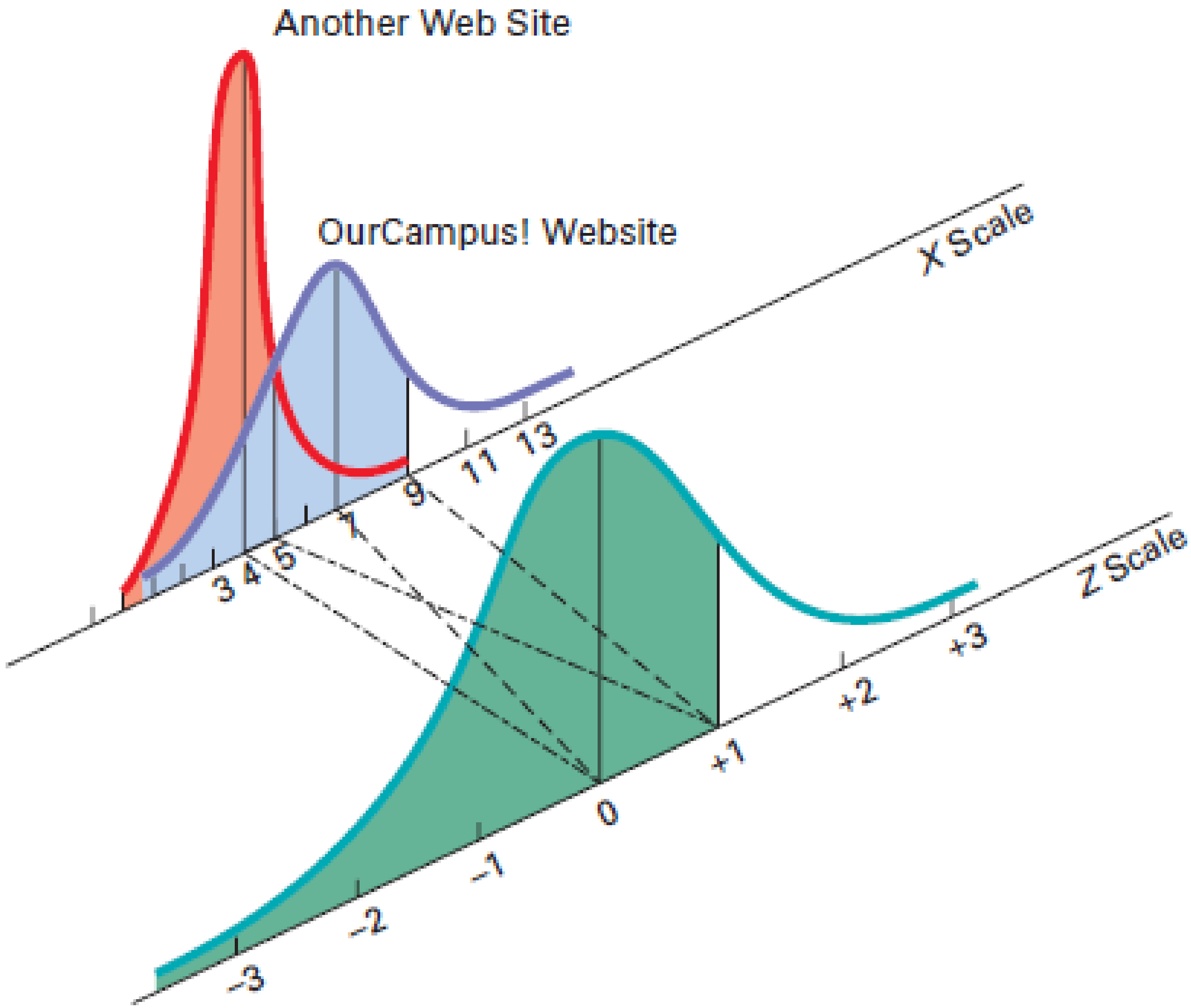


The Standardized Normal

$$Z = \frac{X - \mu}{\sigma} = \frac{5 - 4}{1} = +1$$



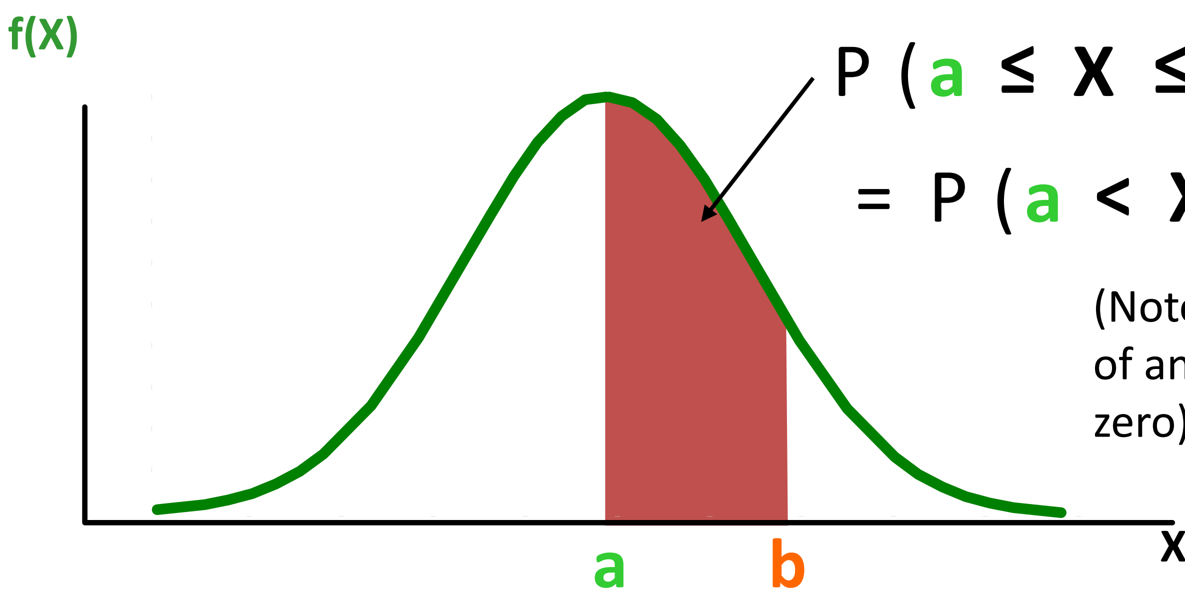
Comparing Two Normal



FINDING NORMAL PROBABILITIES

Finding Normal Probabilities

Probability is measured by the area under the curve



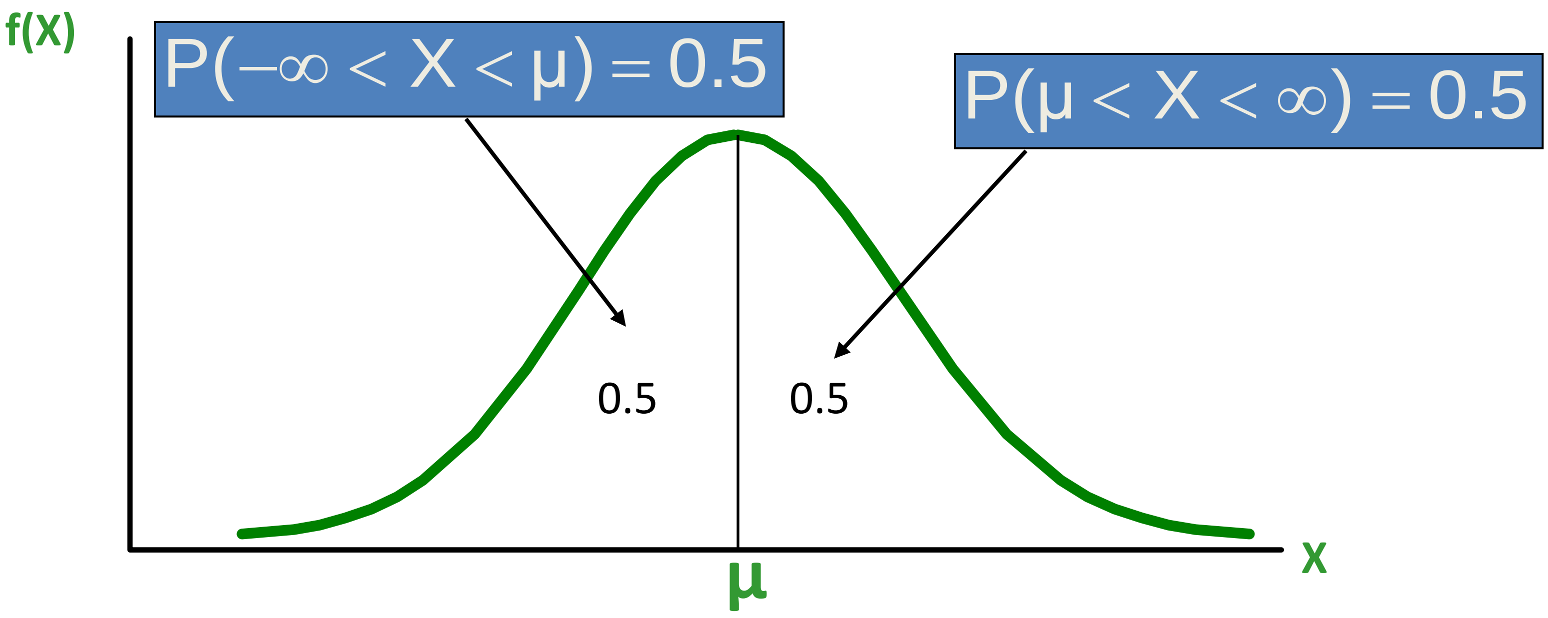
$P(a \leq X \leq b)$

$= P(a < X < b)$

(Note that the probability of any individual value is zero)

Probability as Area Under the Curve

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below



$$P(-\infty < X < \infty) = 1.0$$

The Standardized Normal Table

(continued)

The column gives the value of Z to the second decimal point

Z	0.00	0.01	0.02 …
0.0			
0.1			
⋮			
2.0	**.9772**		

The row shows the value of Z to the first decimal point

The value within the table gives the probability from $Z = -\infty$ up to the desired Z value

P(Z < 2.00) = 0.9772

General Procedure for Finding Normal Probabilities

To find $P(a < X < b)$ when X is distributed normally:

- Draw the normal curve for the problem in terms of X
- Translate X-values to Z-values
- Use the Standardized Normal Table

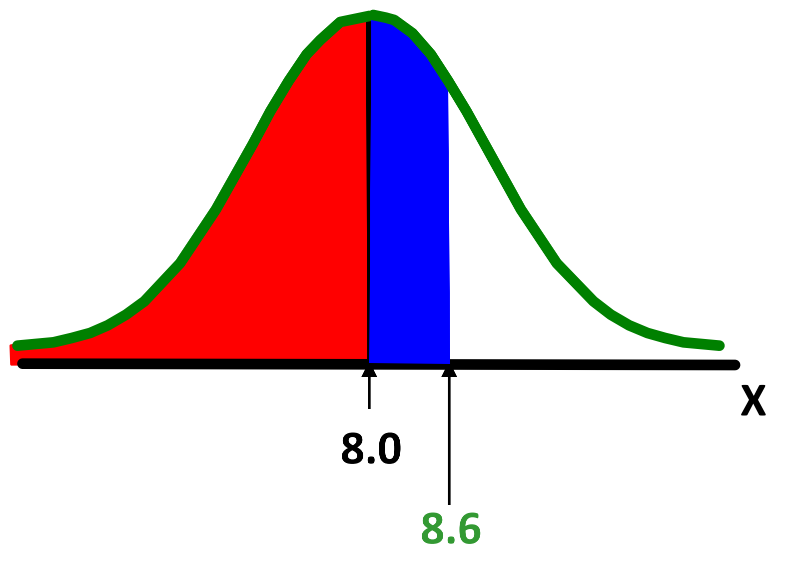
Example

Let X represent the time it takes (in seconds) to download an image file from the internet. Suppose X is normal with a mean of 8.0 seconds and a standard deviation of 5.0 seconds.

a. Find $P(X < 8.6)$
b. Find $P(X > 8.6)$
c. Find $P(8 < X < 8.6)$
d. Find $P(7.4 < X < 8)$
e. Find X such that 20% of download times are less than X.

Finding Normal Probabilities

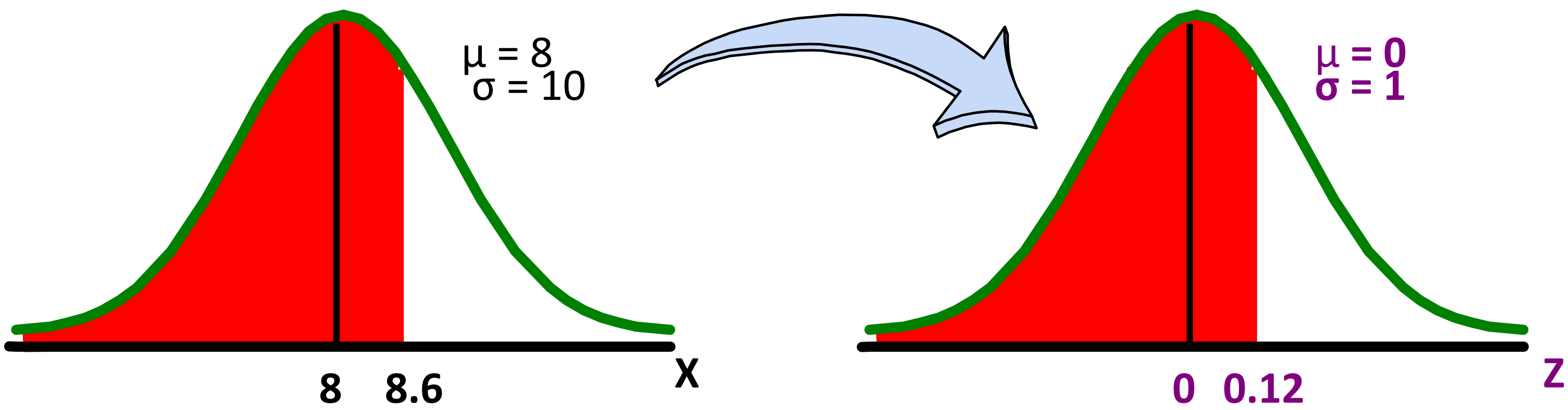
- Let X represent the time it takes to download an image file from the internet.
- Suppose X is normal with mean 8.0 and standard deviation 5.0. Find P(X < 8.6)



Finding Normal Probabilities

(continued)

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$



P(X < 8.6)

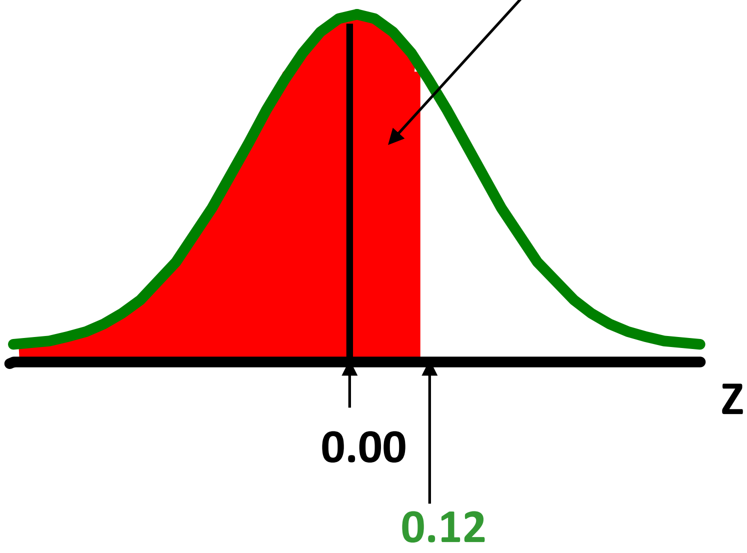
P(Z < 0.12)

Solution: Finding P(Z < 0.12)

Standardized Normal Probability Table (Portion)

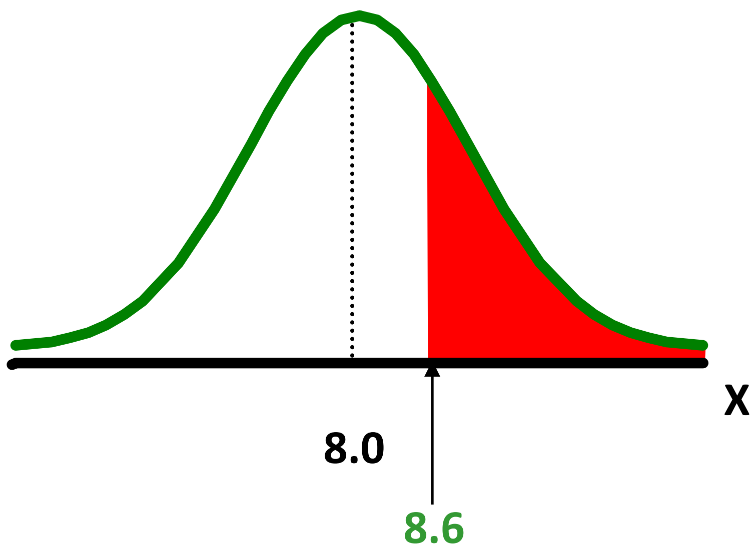
Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

P(X < 8.6)
= P(Z < 0.12)
.5478



Finding Normal Upper Tail Probabilities

- Suppose X is normal with mean 8.0 and standard deviation 5.0.
- Now Find P(X > 8.6)



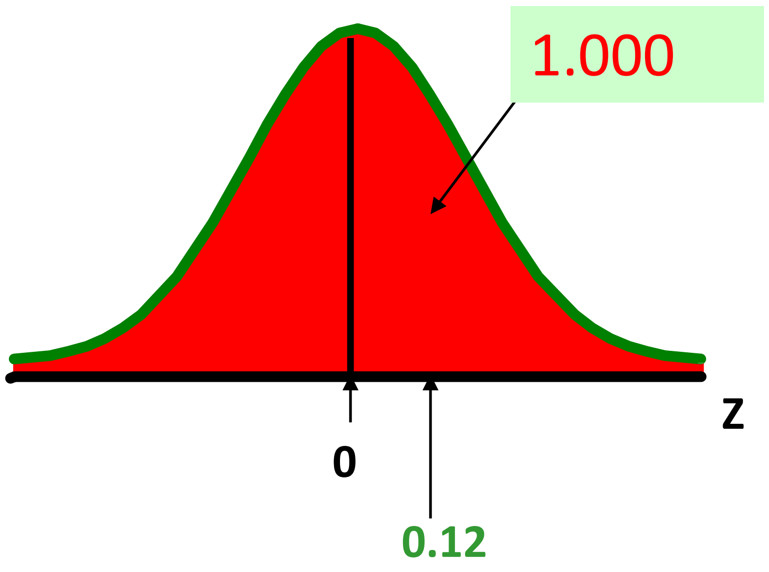
Finding Normal Upper Tail Probabilities

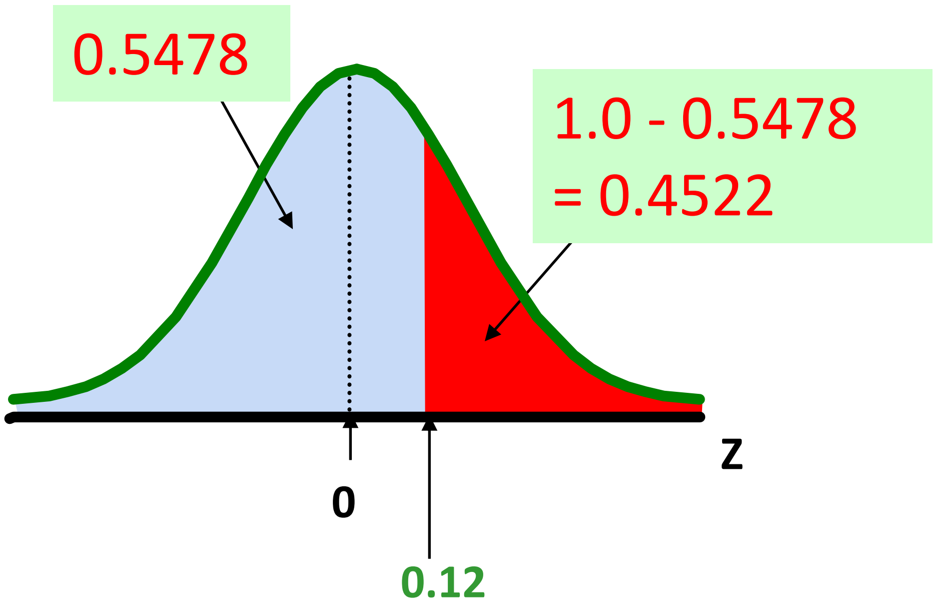
(continued)

- Now Find P(X > 8.6)…

$$P(X > 8.6) = P(Z > 0.12) = 1.0 - P(Z \le 0.12)$$

$$= 1.0 - 0.5478 = 0.4522$$





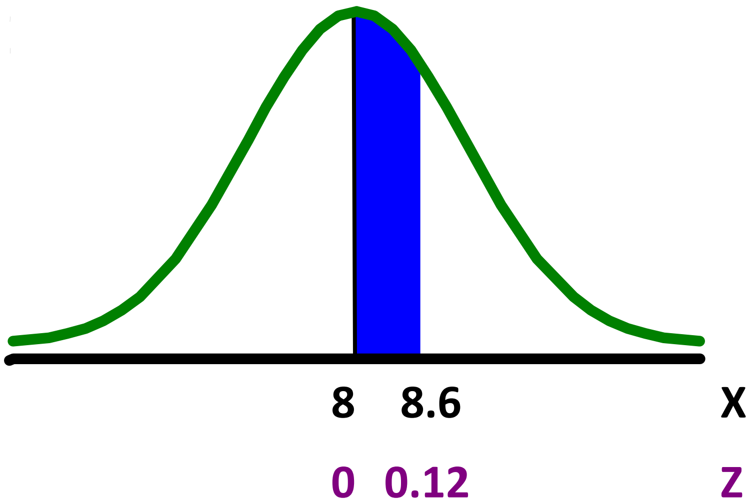
Finding a Normal Probability Between Two Values

Suppose X is normal with mean 8.0 and standard deviation 5.0. Find P(8 < X < 8.6)

Calculate Z-values:

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 8}{5} = 0$$

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8}{5} = 0.12$$



$P(8 < X < 8.6)$
$= P(0 < Z < 0.12)$

Solution: Finding P(0 < Z < 0.12)

Standardized Normal Probability Table (Portion)

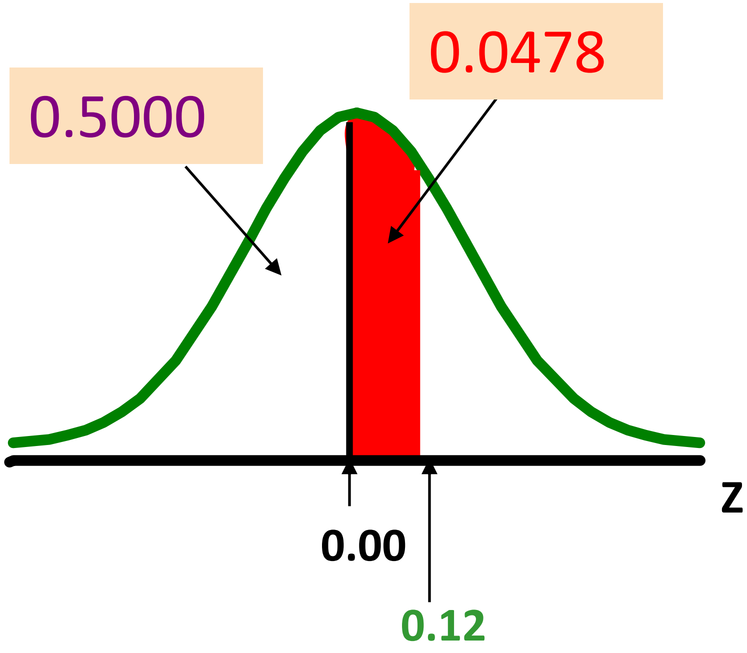
Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$P(8 < X < 8.6)$

$= P(0 < Z < 0.12)$

$= P(Z < 0.12) - P(Z \leq 0)$

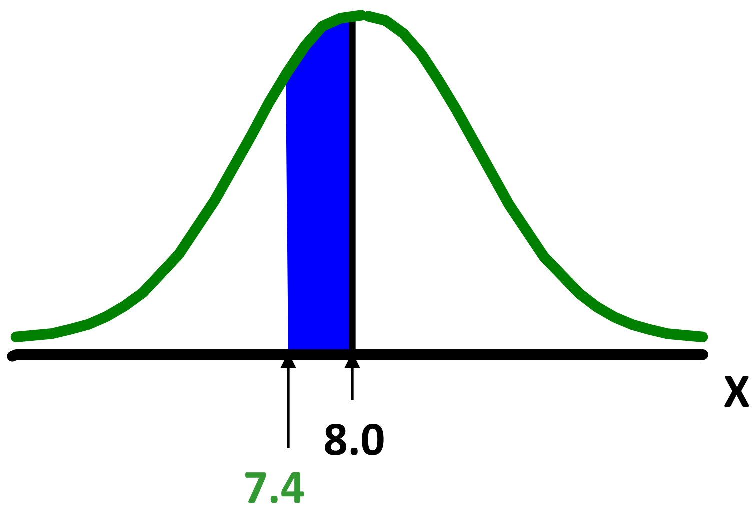
$= 0.5478 - .5000 = 0.0478$



Probabilities in the Lower Tail

Suppose X is normal with mean 8.0 and standard deviation 5.0.

Now Find P(7.4 < X < 8)



Probabilities in the Lower Tail

(continued)

Now Find P(7.4 < X < 8)…

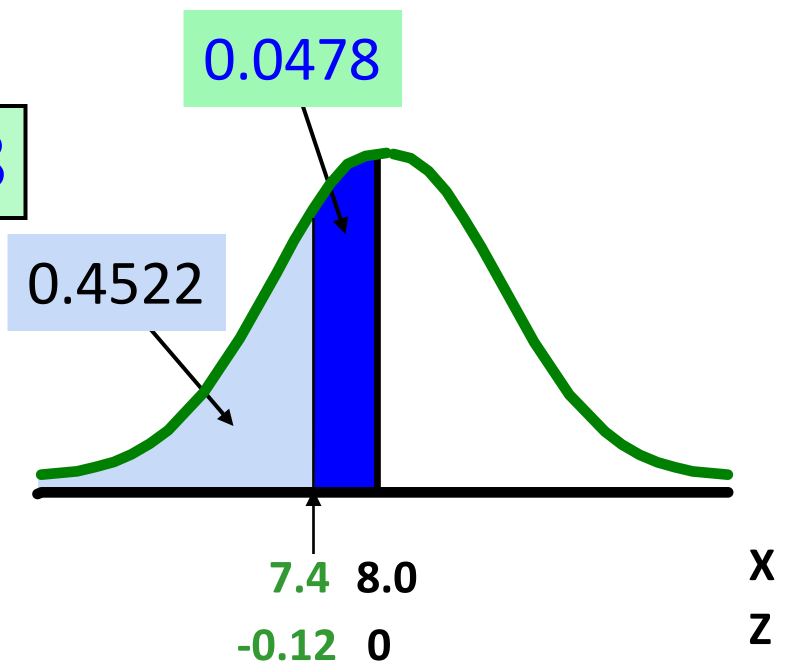
$P(7.4 < X < 8)$

$= P(-0.12 < Z < 0)$

$= P(Z < 0) – P(Z \le -0.12)$

$= 0.5000 - 0.4522 = 0.0478$

The Normal distribution is symmetric, so this probability is the same as $P(0 < Z < 0.12)$



FINDING THE X VALUE

Given a Normal Probability
Find the X Value

Steps to find the X value for a known probability:

1. Find the Z value for the known probability
2. Convert to X units using the formula:

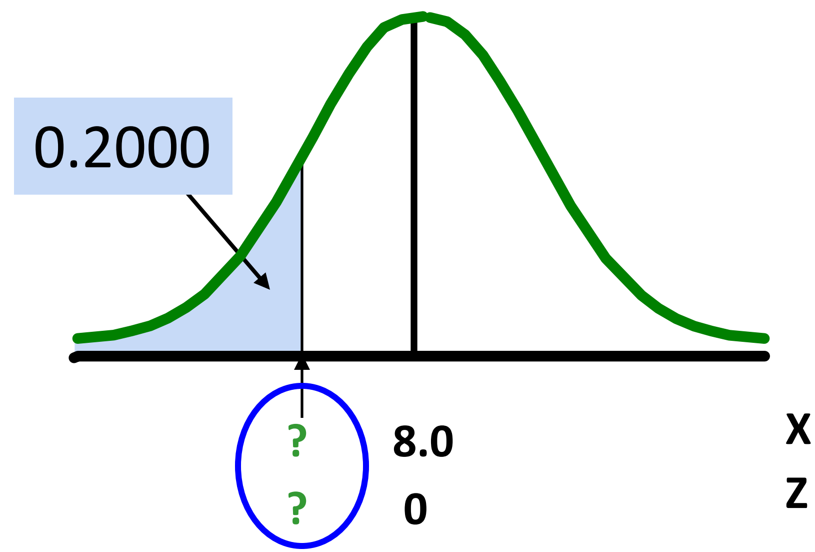
$$X = \mu + Z\sigma$$

Finding the X value for a Known Probability

(continued)

Example:

- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find X such that 20% of download times are less than X.



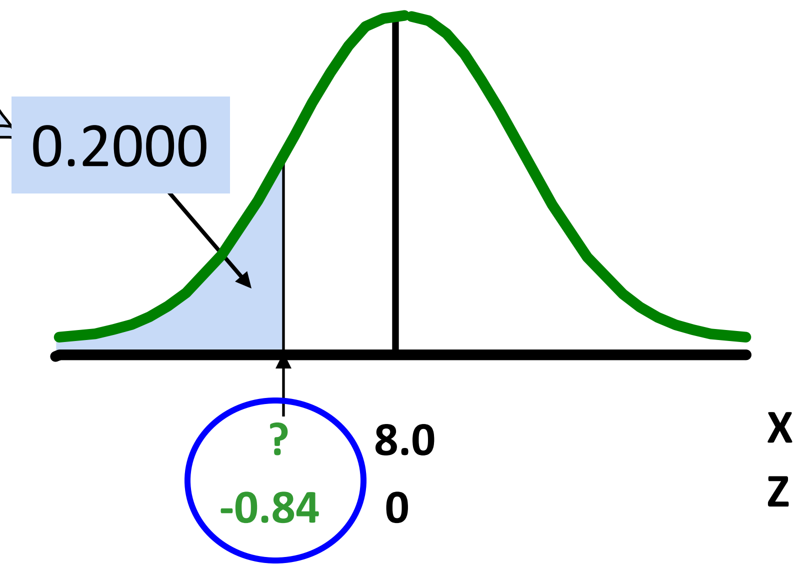
Find the Z value for 20% in the Lower Tail

1. Find the Z value for the known probability

Standardized Normal Probability Table (Portion)

Z	…	.03	.04	.05
-0.9	…	.1762	.1736	.1711
-0.8	…	.2033	.2005	.1977
-0.7	…	.2327	.2296	.2266

20% area in the lower tail is consistent with a Z value of -0.84



Finding the X value

2. Convert to X units using the formula:

$$X = \mu + Z\sigma$$

$$= 8.0 + (-0.84)5.0$$

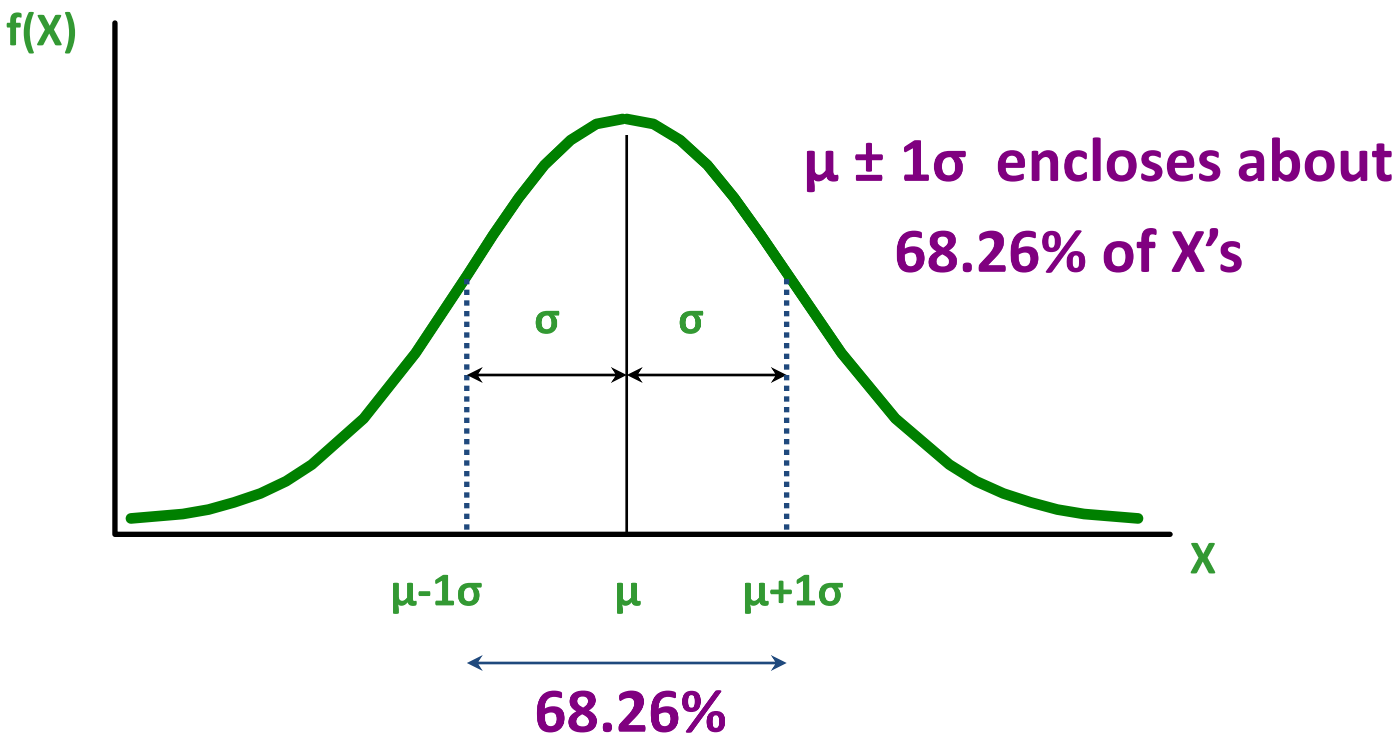
$$= 3.80$$

So 20% of the values from a distribution with mean 8.0 and standard deviation 5.0 are less than 3.80

EMPIRICAL RULES

Empirical Rules

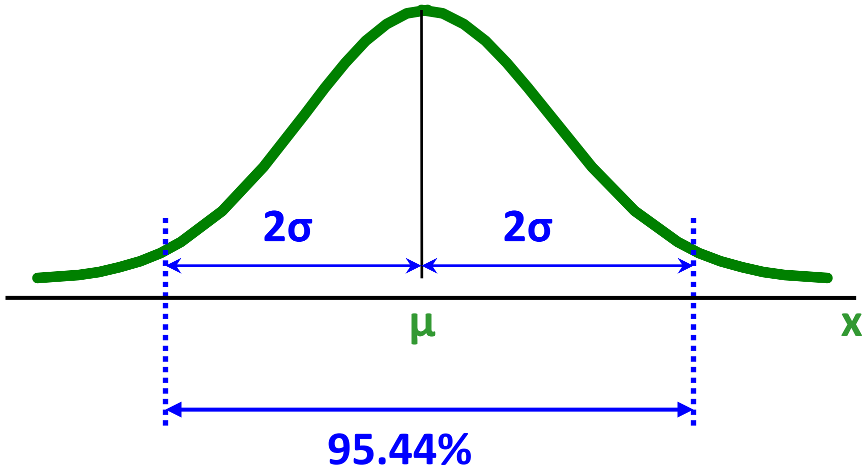
What can we say about the distribution of values around the mean? For any normal distribution:

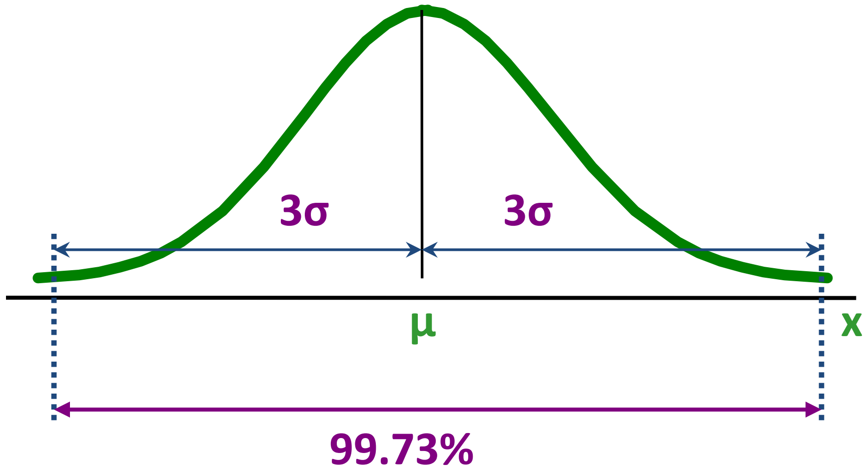


The Empirical Rule

(continued)

- **μ ± 2σ covers about 95% of X's**
- **μ ± 3σ covers about 99.7% of X's**





EVALUATING NORMALITY

Evaluating Normality

- Not all continuous distributions are normal
- It is important to evaluate how well the data set is approximated by a normal distribution.
- Normally distributed data should approximate the theoretical normal distribution:
  - The normal distribution is bell shaped (symmetrical) where the mean is equal to the median.
  - The empirical rule applies to the normal distribution.
  - The interquartile range of a normal distribution is 1.33 standard deviations.

Evaluating Normality

(continued)

Comparing data characteristics to theoretical properties

- Construct charts or graphs
  - For small- or moderate-sized data sets, construct a stem-and-leaf display or a boxplot to check for symmetry
  - For large data sets, does the histogram or polygon appear bell-shaped?
- Compute descriptive summary measures
  - Do the mean, median and mode have similar values?
  - Is the interquartile range approximately 1.33 σ?
  - Is the range approximately 6 σ?

Evaluating Normality

(continued)

Comparing data characteristics to theoretical properties

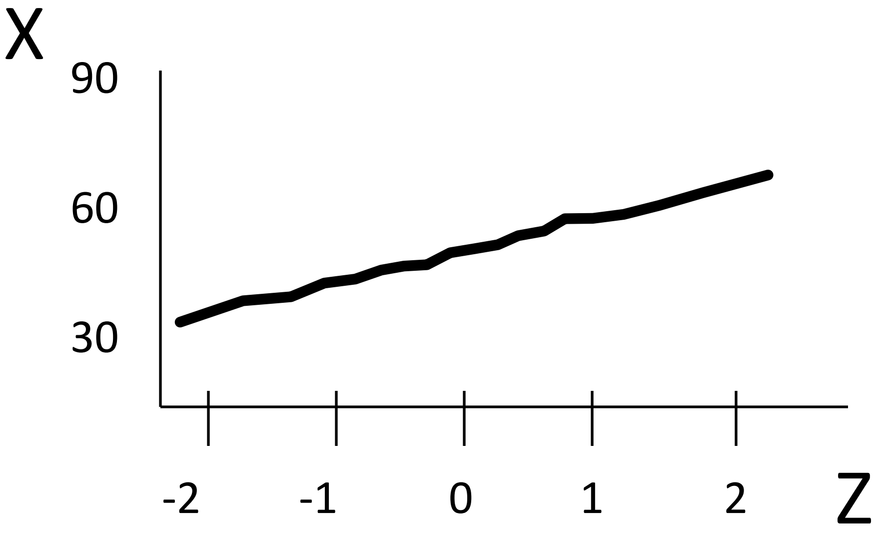
- Observe the distribution of the data set
  - Do approximately 2/3 of the observations lie within mean ± 1 standard deviation?
  - Do approximately 80% of the observations lie within mean ± 1.28 standard deviations?
  - Do approximately 95% of the observations lie within mean ± 2 standard deviations?
- Evaluate normal probability plot
  - Is the normal probability plot approximately linear (i.e. a straight line) with positive slope?

Constructing A Normal Probability Plot

- Normal probability plot
  - Arrange data into ordered array
  - Find corresponding standardized normal quantile values (Z)
  - Plot the pairs of points with observed data values (X) on the vertical axis and the standardized normal quantile values (Z) on the horizontal axis
  - Evaluate the plot for evidence of linearity

The Normal Probability Plot Interpretation

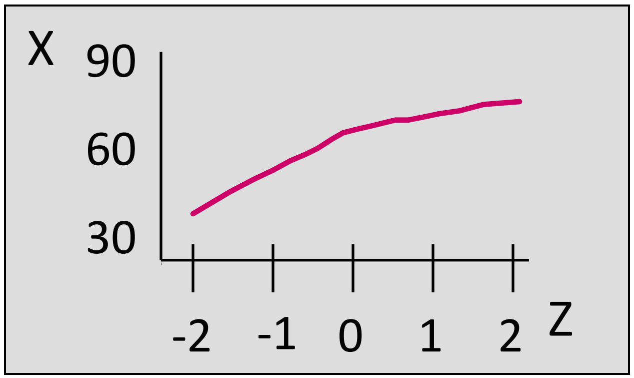
A normal probability plot for data from a normal distribution will be approximately linear:



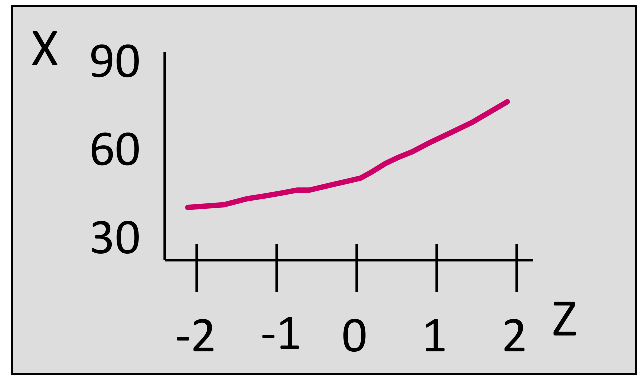
Normal Probability Plot Interpretation

(continued)

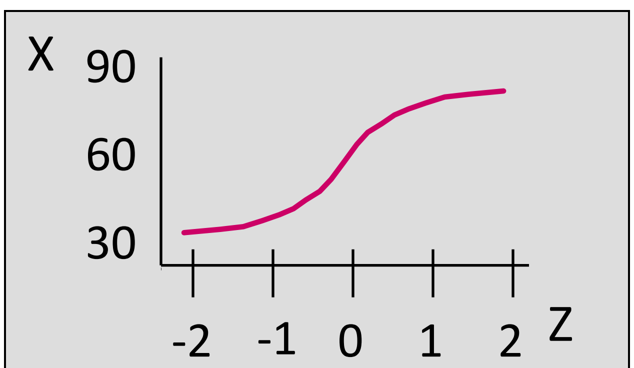
Left-Skewed



Right-Skewed



Rectangular



Nonlinear plots indicate a deviation from normality

EXERCISE

Exercise 1

Toby’s Trucking Company determined that the distance traveled per truck per year is normally distributed, with a mean of 50 thousand miles and a standard deviation of 12 thousand miles.

a. What proportion of trucks can be expected to travel between 34 and 50 thousand miles in a year?
b. What percentage of trucks can be expected to travel either below 30 or above 60 thousand miles in a year?
c. How many miles will be traveled by at least 80% of the trucks?

Exercise 2

A set of final examination grades in an statistics for business course is normally distributed, with a mean of 73 and a standard deviation of 8.

a. What is the probability that a student scored below 91 on this exam?
b. What is the probability that a student scored between 65 and 89?
c. The probability is 5% that a student taking the test scores higher than what grade?

ANSWER

Exercise 1

μ = 50

σ = 12

a. P(34 < X < 50) **= 0.5000 - 0.0918 = 0.4082**

$$Z_{34} = \frac{X - \mu}{\sigma} = \frac{34 - 50}{12} = -1.33$$

P(Z<1.55) = 0.0918

$$Z_{50} = \frac{X - \mu}{\sigma} = \frac{50 - 50}{12} = 0$$

P(Z<0) = 0.5000

Exercise 1

b. P(X<30 or X>60) `= **0.0475 + (1-0.7967)**

= 0.2508

$$Z_{34} = \frac{X-\mu}{\sigma} = \frac{30-50}{12} = -1.67$$

P(Z<-1.67) = 0.0475

$$Z_{50} = \frac{X-\mu}{\sigma} = \frac{60-50}{12} = 0.83$$

P(Z<0.83) = 0.7967

Exercise 1

a. P(X > ?) = 0.8

P(Z < ?) = 0.2 → Z = - 0.84

$$Z_{?} = \frac{X - \mu}{\sigma}$$

$$-0.84 = \frac{X - 50}{12}$$

$$X = -0.84(12) + 50 = 39.92$$

Exercise 1

d. $\sigma = 10$

a) P(34 < X < 50) = P(-1.60<Z<0) = 0.4452

b) P(X < 30 or X > 60) = P(Z<-2.00 or Z>1.00)

= 0.0228 + (1-0.8413) = 0.1815

c) P(X > ?) = 0.8

$$X = -0.84(10) + 50 = 41.6$$

Exercise 2

$\mu = 75$

$\sigma = 8$

a. $P(X < 91) = 0.9878$

b. $P(65 < X < 89) = 0.8185$

c. $P(X > ?) = 0.05$

$X = 86.16$

THANK YOU