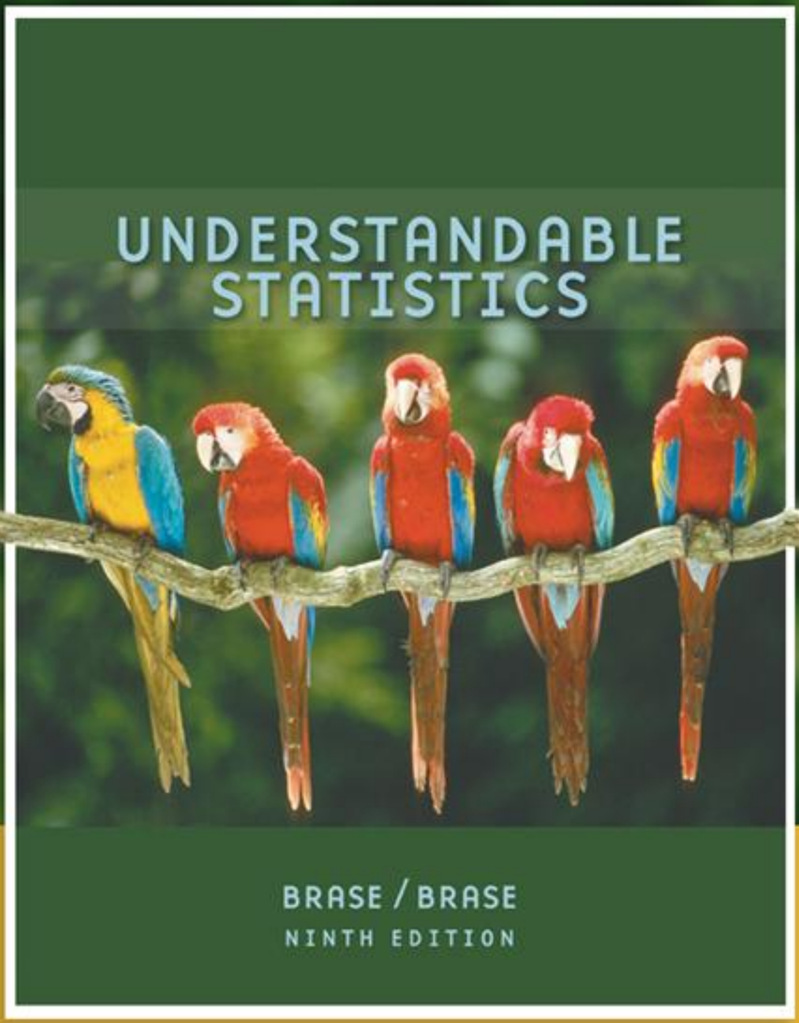


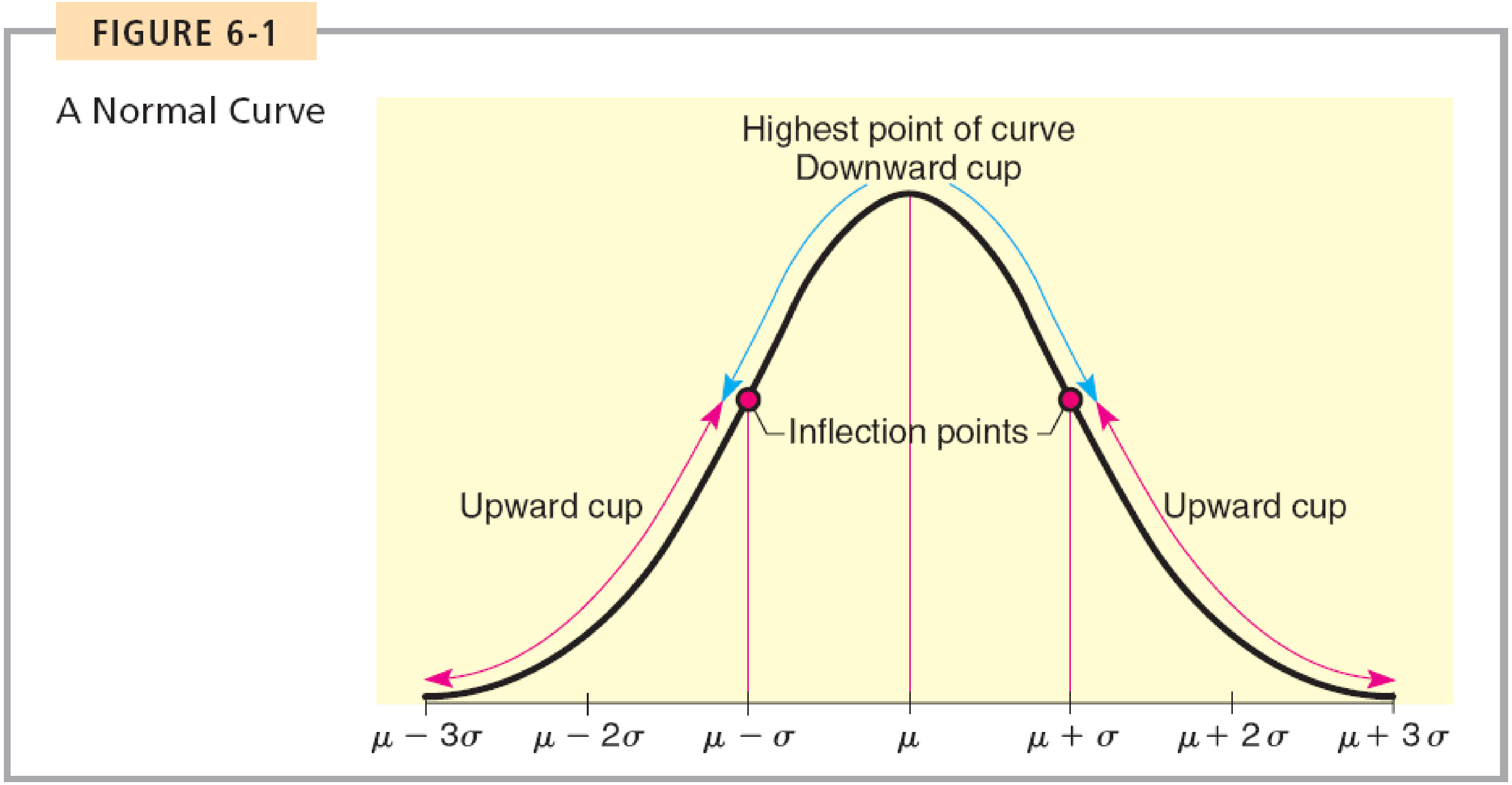

# Chapter 6
## Normal Distributions

**Understandable Statistics**
**Ninth Edition**
By Brase and Brase
Prepared by Yixun Shi
Bloomsburg University of Pennsylvania
Edited by José Neville Díaz Caraballo
University of Puerto Rico at Aguadilla

# The Normal Distribution

- A continuous distribution used for modeling many natural phenomena.
- Sometimes called the Gaussian Distribution, after Carl Gauss.
- The defining features of a Normal Distribution are the mean, $\mu$, and the standard deviation, $\sigma$.

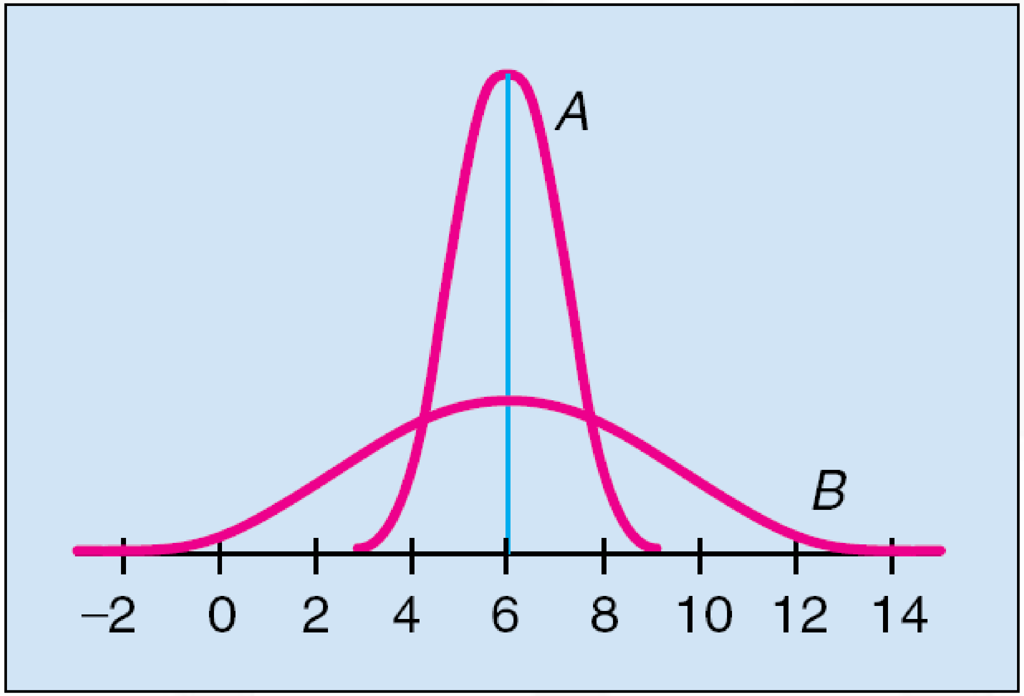
# The Normal Curve



# Features of the Normal Curve

- Smooth line and symmetric around $\mu$.
- Highest point directly above $\mu$.
- The curve never touches the horizontal axis in either direction.
- As $\sigma$ increases, the curve spreads out.
- As $\sigma$ decreases, the curve becomes more peaked around $\mu$.
- Inflection points at $\mu \pm \sigma$.

# Two Normal Curves




**Both curves have the same mean, $\mu = 6$.**

**Curve A has a standard deviation of $\sigma = 1$.**

**Curve B has a standard deviation of $\sigma = 3$.**

# Normal Probability

- The area under any normal curve will always be 1.
- The portion of the area under the curve within a given interval represents the probability that a measurement will lie in that interval.

# The Empirical Rule

**Empirical rule**

For a distribution that is symmetrical and bell-shaped (in particular, for a normal distribution):

Approximately 68% of the data values will lie within one standard deviation on each side of the mean.

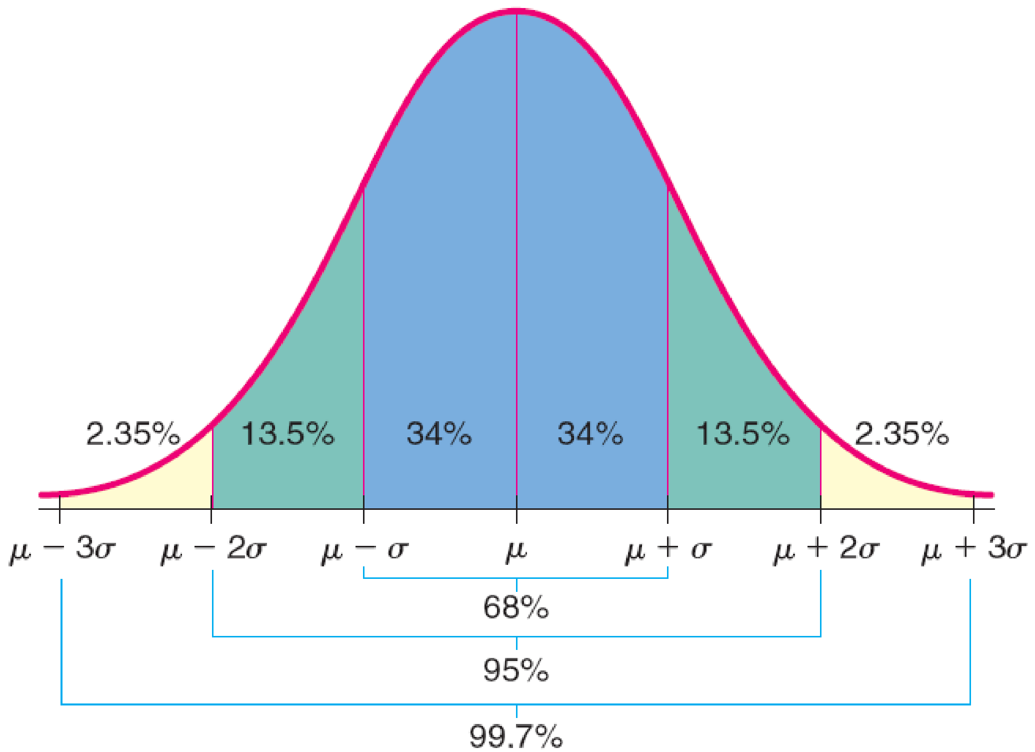
Approximately 95% of the data values will lie within two standard deviations on each side of the mean.

Approximately 99.7% (or almost all) of the data values will lie within three standard deviations on each side of the mean.

# The Empirical Rule

FIGURE 6-5

Area Under a Normal Curve



# Control Charts

- A graph to examine data over equally spaced time intervals.
- Used to determine if a variable is in statistical control.
  - Statistical Control: A variable $x$ is in statistical control if it can be described by the same probability distribution over time.

PROCEDURE

**How to make a control chart for the random variable *x***

A control chart for a random variable $x$ is a plot of observed $x$ values in time sequence order.

1. Find the mean $\mu$ and standard deviation $\sigma$ of the $x$ distribution by
   (a) using past data from a period during which the process was "in control" or
   (b) using specified "target" values for $\mu$ and $\sigma$.
2. Create a graph where the vertical axis represents $x$ values and the horizontal axis represents time.
3. Draw a horizontal line at height $\mu$ and horizontal, dashed control-limit lines at $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$.
4. Plot the variable $x$ on the graph in time sequence order. Use line segments to connect the points in time sequence order.

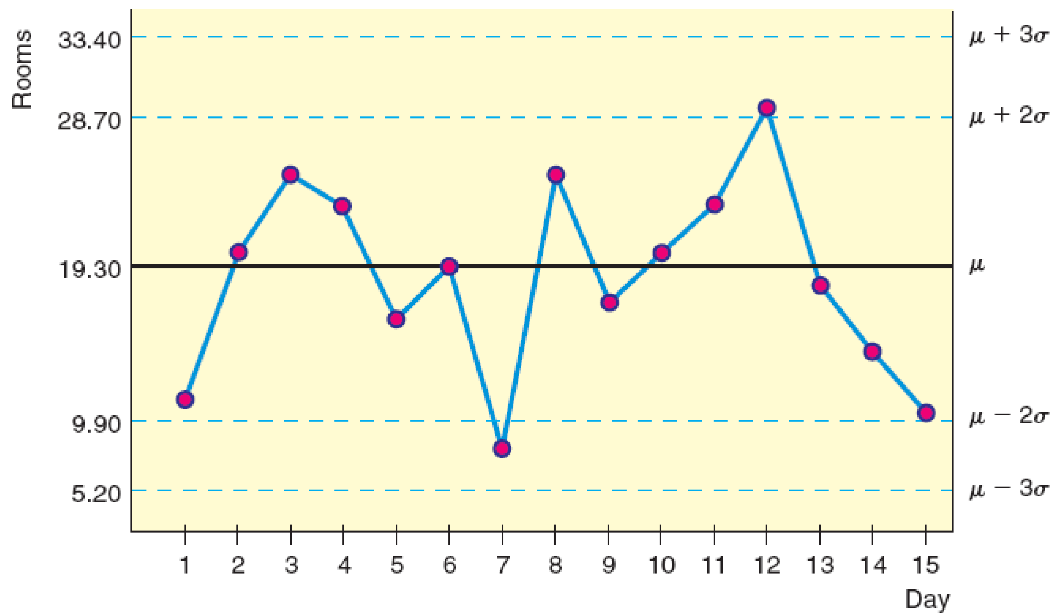
# Control Chart Example

**TABLE 6-1 Number of Rooms *x* Not Made Up by 3:30 P.M.**

| Day | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 11 | 20 | 25 | 23 | 16 | 19 | 8 | 25 | 17 | 20 | 23 | 29 | 18 | 14 | 10 |

FIGURE 6-9

Number of Rooms Not Made Up by 3:30 P.M.



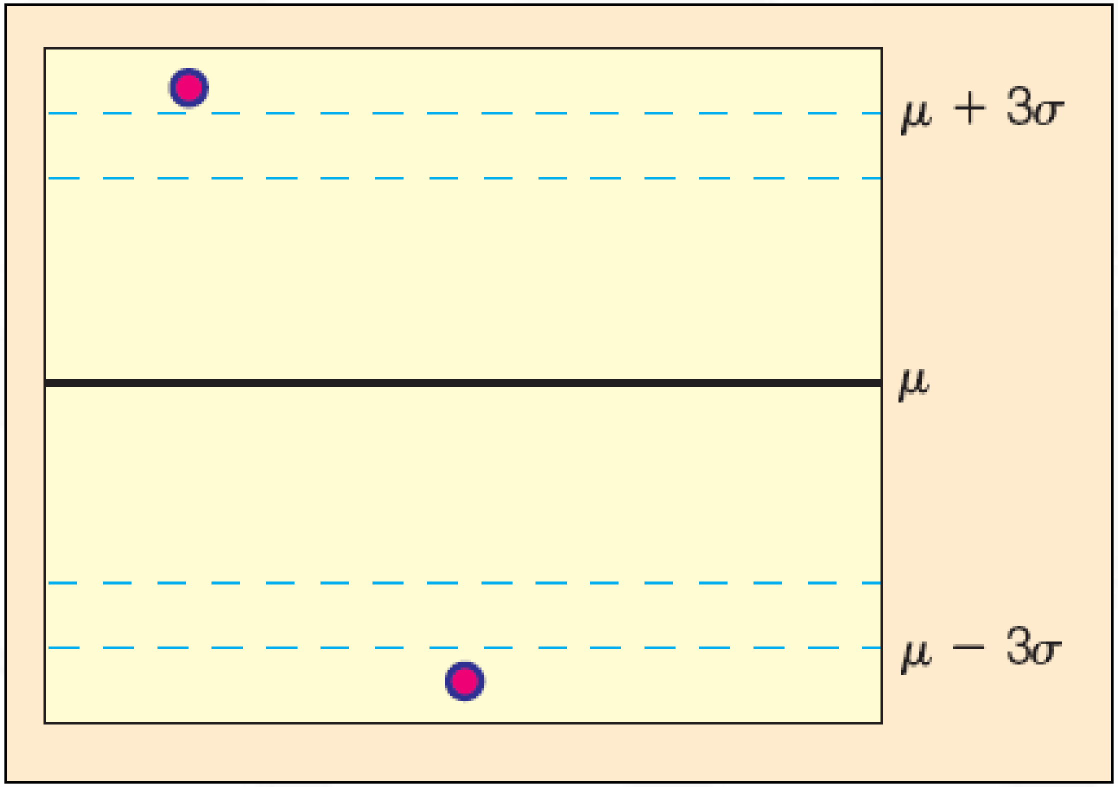
# Determining if a Variable is Out of Control

1) One point falls beyond the $3\sigma$ level.

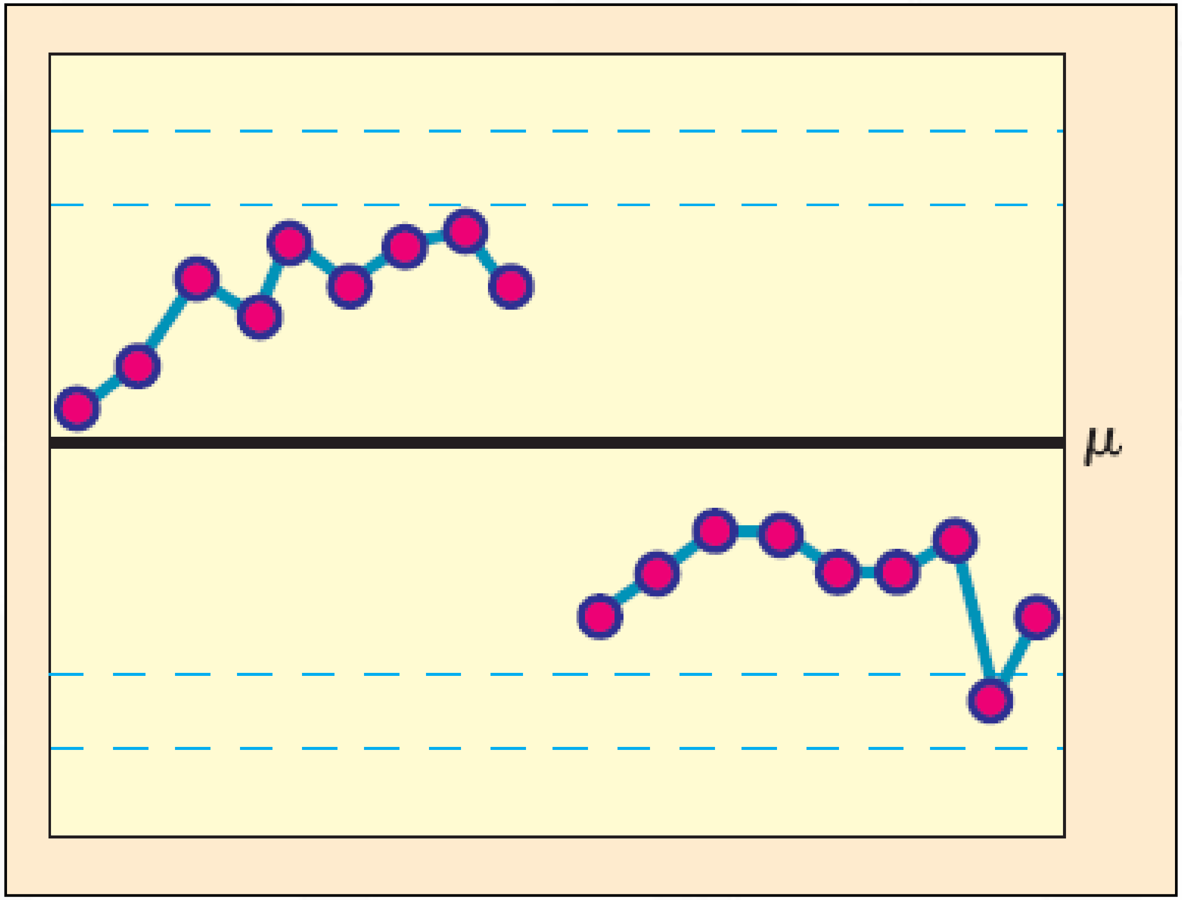
2) A run of nine consecutive points on one side of the center line.

3) At least two of three consecutive points lie beyond the $2\sigma$ level on the same side of the center line.

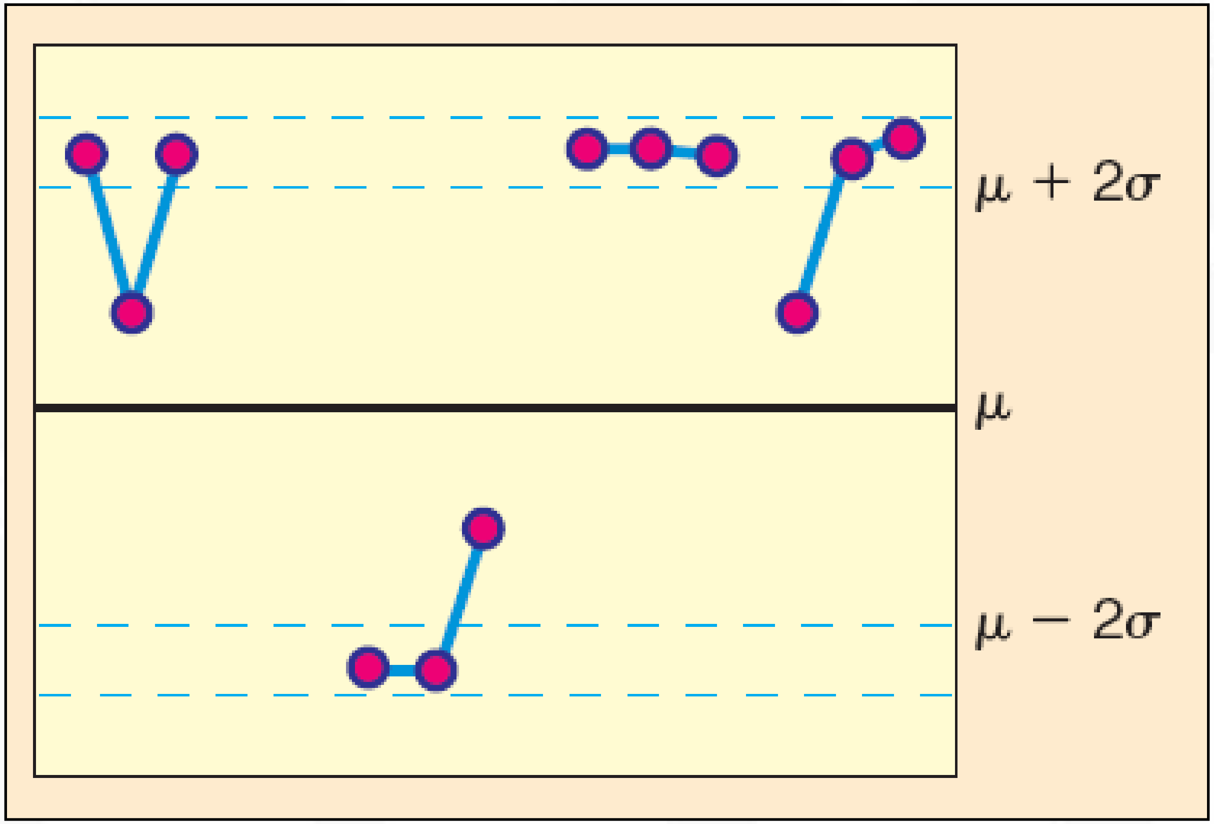
# Out of Control Signal I
# Probability = 0.0003



# Out of Control Signal II
# Probability = 0.004



# Out of Control Signal III
# Probability = 0.002



# Computing z Scores

The $z$ **value** or $z$ **score** gives the number of standard deviations between the original measurement $x$ and the mean $\mu$ of the $x$ distribution.

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

# Work With General Normal Distributions

Given an $x$ distribution with mean $\mu$ and standard deviation $\sigma$, the raw score $x$ corresponding to a $z$ score is

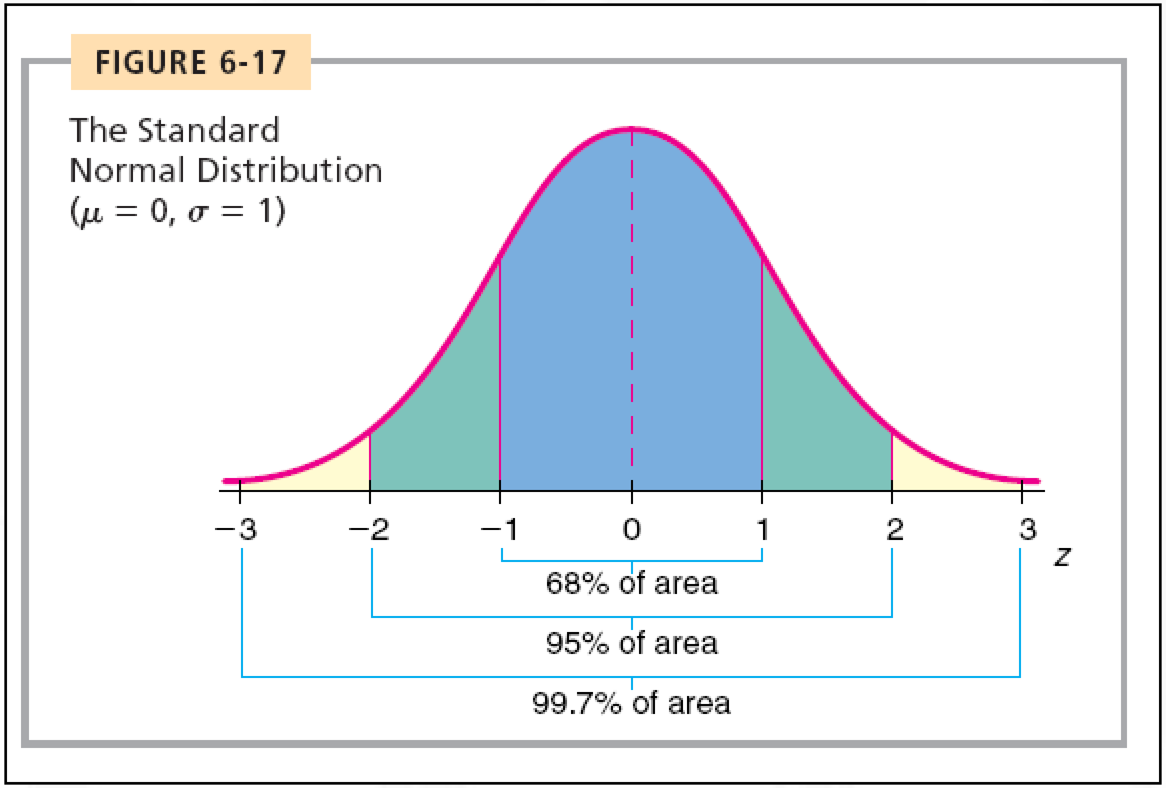
$$x = z\sigma + \mu$$

Or equivalently,

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

# The Standard Normal Distribution

- Z scores also have a normal distribution
  - $\mu = 0$
  - $\sigma = 1$



# Using the Standard Normal Distribution

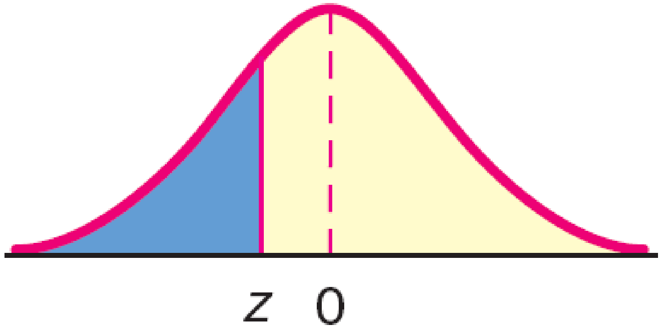
There are extensive tables for the Standard Normal Distribution.

- We can determine probabilities for normal distributions:
  1) Transform the measurement to a *z* Score.
  2) Utilize Table 5 of Appendix II.

# Using the Standard Normal Table

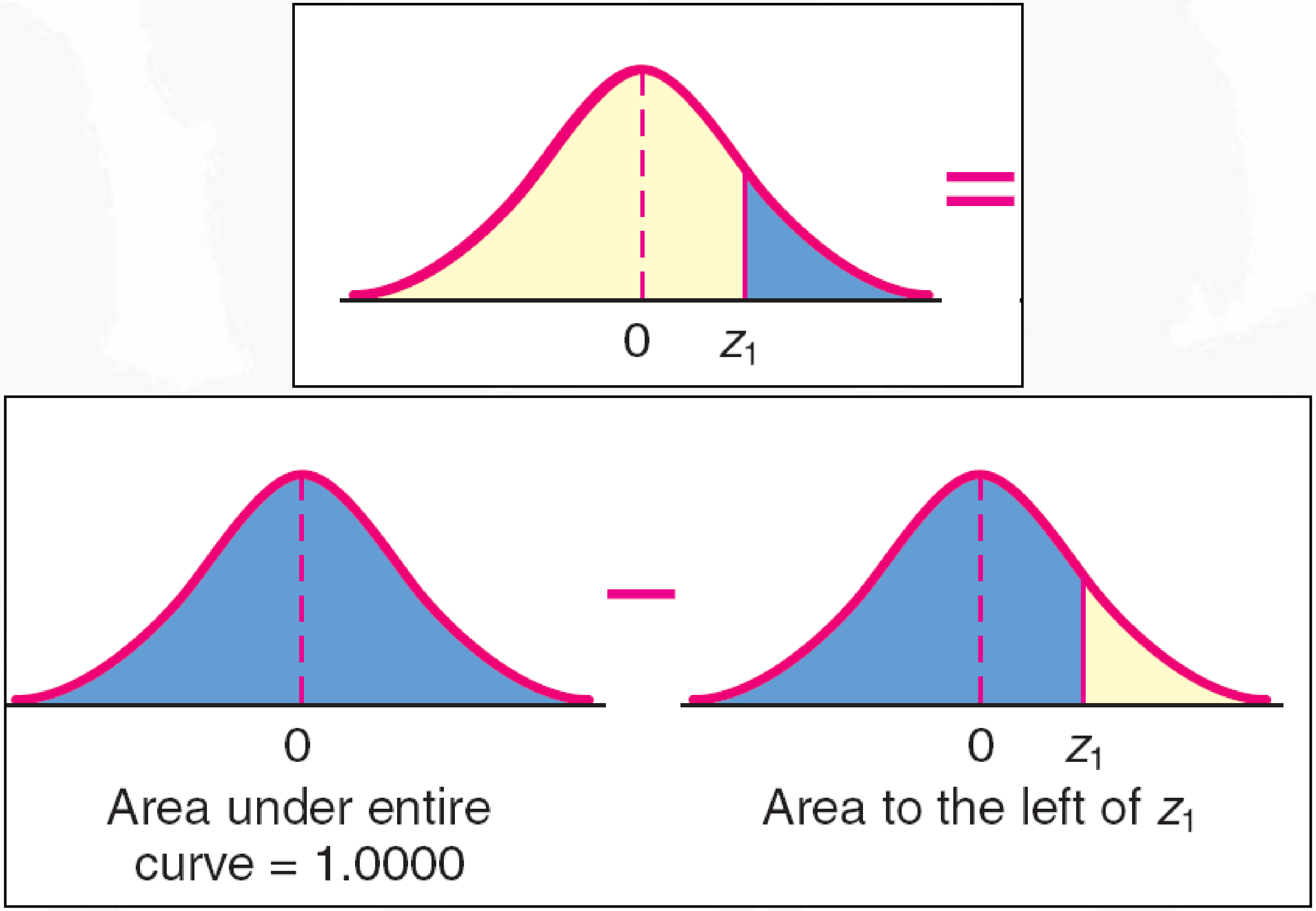
- Table 5(a) gives the cumulative area for a given *z* value.
- When calculating a *z* Score, round to 2 decimal places.
- For a *z* Score less than -3.49, use 0.000 to approximate the area.
- For a *z* Score greater than 3.49, use 1.000 to approximate the area.

# Area to the Left of a Given *z* Value

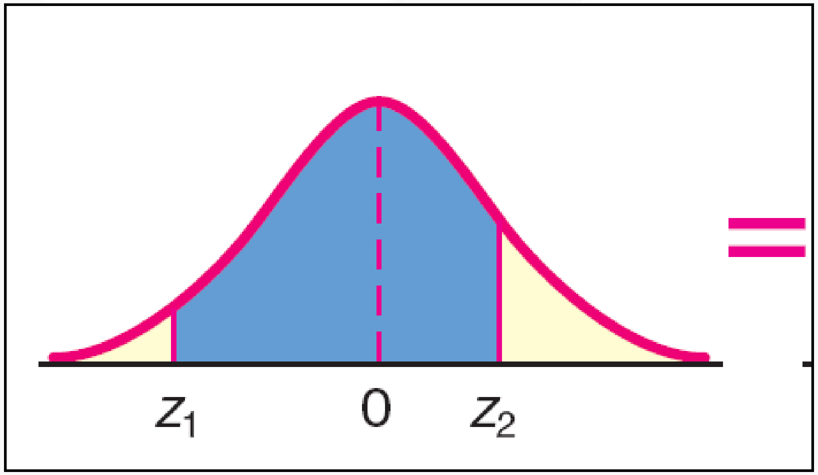


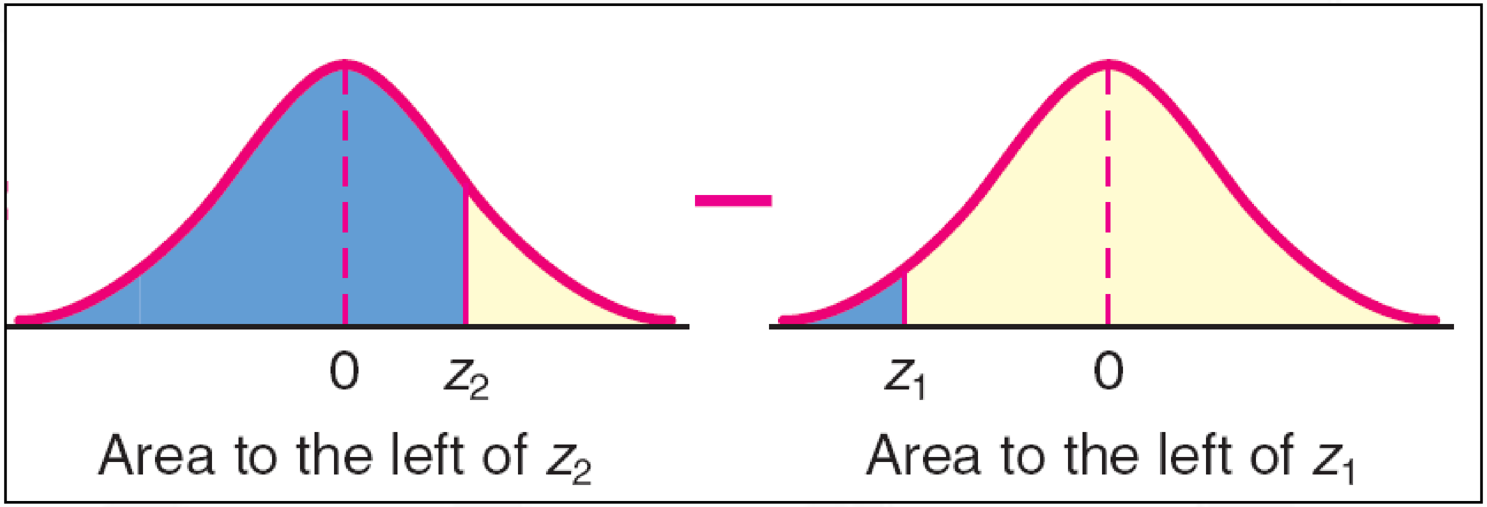

Use Table 5 of Appendix II directly.

# Area to the Right of a Given $z$ Value



# Area Between Two $z$ Values



Example

Some doctors believe that a person can lose 5 pounds, on the average, in a month by reducing his/her fat intake and by exercising consistently. Suppose weight loss has a normal distribution. Let  = the amount of weight lost (in pounds) by a person in a month.

Use a standard deviation of 2 pounds X ~N(5,2) . Fill in the blanks.

Problem 1

Suppose a person lost 10 pounds in a month. The z-score when  pounds is

(verify). This z-score tells you that  is _________ standard deviations to the _________

(right or left) of the mean _____ (What is the mean?).

Problem 2

Suppose a person gained 3 pounds (a negative weight loss). Then  = ____________. This z-score tells you that is __________ standard deviations to the ____________ (right or left) of the mean.

Suppose the random variables and have the following normal distributions: $X \sim N(5,6)$ and $Y \sim N(2,1)$. If $X=17$ and $Y=4$ , then what is z?

The z-score allows us to compare data that are scaled differently.

# Normal Probability Final Remarks

- The probability that $z$ equals a certain number is always 0.
  - $P(z = a) = 0$
- Therefore, < and ≤ can be used interchangeably. Similarly, > and ≥ can be used interchangeably.
  - $P(z < b) = P(z \leq b)$
  - $P(z > c) = P(z \geq c)$

# Example 1

An average light bulb manufactured by the Acme Corporation lasts 300 days with a standard deviation of 50 days. Assuming that bulb life is normally distributed, what is the probability that an Acme light bulb will last at most 365 days?

*Solution*

# Example 2

Suppose scores on an IQ test are normally distributed. If the test has a mean of 100 and a standard deviation of 10, what is the probability that a person who takes the test will score between 90 and 110?

*Solution:*

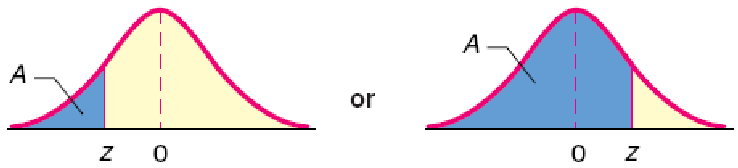
# Inverse Normal Distribution

- Sometimes we need to find an *x* or *z* that corresponds to a given area under the normal curve.
  - In Table 5, we look up an area and find the corresponding *z.*

**FIGURE 6-31**

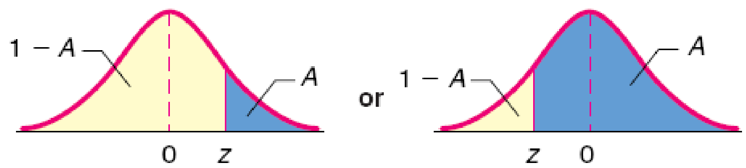
Inverse Normal: Use Table 5 of Appendix II to Find $z$ Corresponding to a Given Area $A$ $(0 < A < 1)$

(a) **Left-tail case:**
The given area $A$ is to the left of $z$.



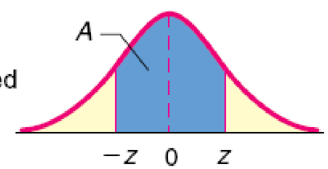

For the left-tail case, look up the number $A$ in the body of the table and use the corresponding $z$ value.

(b) **Right-tail case:**
The given area $A$ is to the right of $z$.




For the right-tail case, look up the number $1 - A$ in the body of the table and use the corresponding $z$ value.

(c) **Center case:**
The given area $A$ is symmetric and centered above $z = 0$. Half of $A$ lies to the left and half lies to the right of $z = 0$.




For the center case, look up the number $\frac{1 - A}{2}$ in the body of the table and use the corresponding $\pm z$ value.

The average life of a certain type of motor is 10 years, with a standard deviation of 2 years. If the manufacturer is willing to replace only 3% of the motors that fail, how long a guarantee should he offer? Assume that the lives of the motors follow a normal distribution.

# Critical Thinking – How to tell if data follow a normal distribution?

- Histogram – a normal distribution's histogram should be roughly bell-shaped.
- Outliers – a normal distribution should have no more than one outlier

# Critical Thinking – How to tell if data follow a normal distribution?

- Skewness –normal distributions are symmetric. Use the Pearson's index:

  Pearson's index = $$\frac{3(\bar{x} - median)}{s}$$

  A Pearson's index greater than 1 or less than -1 indicates skewness.
- Normal quantile plot – using a statistical software (see the Using Technology feature.)

# Normal Approximation to the Binomial

**Normal approximation to the binomial distribution**

Consider a binomial distribution where

$n$ = number of trials

$r$ = number of successes

$p$ = probability of success on a single trial

$q = 1 - p$ = probability of failure on a single trial

If $np > 5$ and $nq > 5$, then $r$ has a binomial distribution that is approximated by a normal distribution with

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{npq}$$

*Note:* As $n$ increases, the approximation becomes better.

# Continuity Correction

**PROCEDURE**

**How to make the continuity correction**

Convert the discrete random variable $r$ (number of successes) to the continuous normal random variable $x$ by doing the following:

1. If $r$ is a **left-point** of an interval, subtract 0.5 to obtain the corresponding normal variable $x$; that is, $x = r - 0.5$.
2. If $r$ is a **right-point** of an interval, add 0.5 to obtain the corresponding normal variable $x$: that is, $x = r + 0.5$.

For instance, $P(6 \leq r \leq 10)$, where $r$ is a binomial random variable, is approximated by $P(5.5 \leq x \leq 10.5)$, where $x$ is the corresponding normal random variable (see Figure 6-38).