Section 2.2

day 1 of 3

Density Curves

Exploring Quantitative Data

Adding the same positive number a to (subtracting a from) each observation:

1. Always plot your data: make a graph, usually a dotplot, stemplot, or histogram.
2. Look for the overall pattern (shape, center, and variability) and for striking departures such as outliers.
3. Calculate numerical summaries to describe center and variability.
4. When there's a regular overall pattern, use a simplified model called a density curve to describe it.

A lot of information quickly. Take notes later.

Goal 1

Use density curves to model distributions

A **density curve** is a curve that

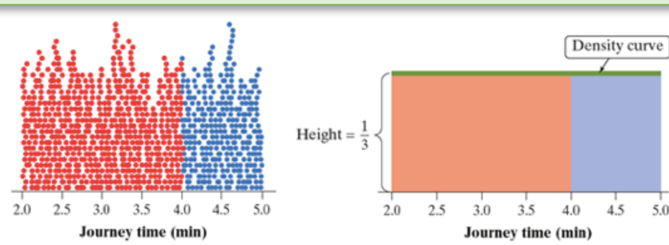
- Is always on or above the horizontal axis
- Has area exactly 1 underneath it

The area under the curve and above any interval of values on the horizontal axis estimates the proportion of all observations that fall in that interval.

A **density curve** is a curve that

- Is always on or above the horizontal axis
- Has area exactly 1 underneath it

The area under the curve and above any interval of values on the horizontal axis estimates the proportion of all observations that fall in that interval.



The overall pattern of this dotplot of the amount of time it has taken Selena to get to the bookstore by train each day for the last 1000 days she worked can be described by a horizontal line.

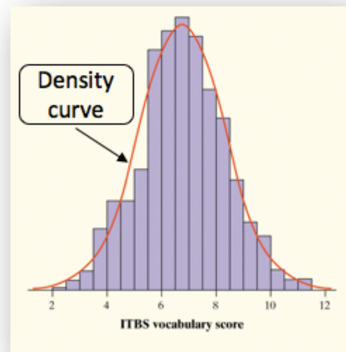
theoretical

A **density curve** is a curve that

- Is always on or above the horizontal axis
- Has area exactly 1 underneath it

The area under the curve and above any interval of values on the horizontal axis estimates the proportion of all observations that fall in that interval.

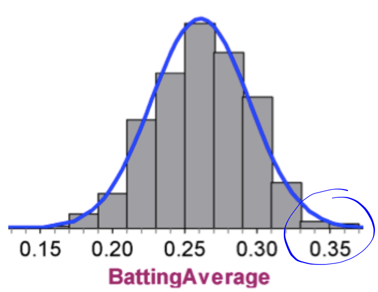
The overall pattern of this histogram of the scores of all 947 seventh-grade students in Gary, Indiana, on the vocabulary part of the Iowa Test of Basic Skills (ITBS) can be described by a smooth curve drawn through the tops of the bars.



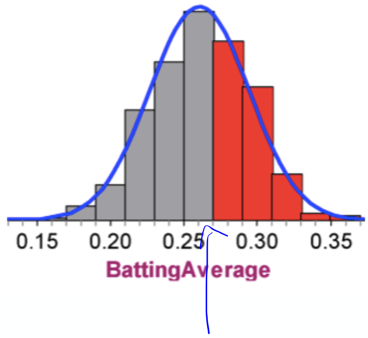
Density Curves

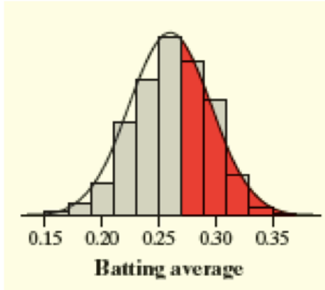
Example: Batting Averages

The histogram below shows the distribution of batting average (proportion of hits) for the 432 Major League Baseball players with at least 100 plate appearances in the 2009 season. The smooth curve shows the overall shape of the distribution.



In the first graph below, the bars in red represent the proportion of players who had batting averages of at least 0.270. There are 177 such players out of a total of 432, for a proportion of 0.410. In the second graph below, the area under the curve to the right of 0.270 is shaded. This area is 0.391, only 0.019 away from the actual proportion of 0.410.





1. Suppose you use a calculator or computer random number generator to produce a number between 0 and 2 (like 0.84522 or 1.1111119). The random number generator will spread its output uniformly across the entire interval from 0 to 2 as we allow it to generate a long sequence of random numbers.

   a) Draw a density curve to model this distribution of random numbers. Be sure to include scales on both axes.

   b) About what percent of the randomly generated numbers will fall between 0.87 and 1.55 ?

   c) Estimate the 65th percentile of this distribution of random numbers.

1. Suppose you use a calculator or computer random number generator to produce a number between 0 and 2 (like 0.84522 or 1.1111119). The random number generator will spread its output uniformly across the entire interval from 0 to 2 as we allow it to generate a long sequence of random numbers.

   a) Draw a density curve to model this distribution of random numbers. Be sure to include scales on both axes.

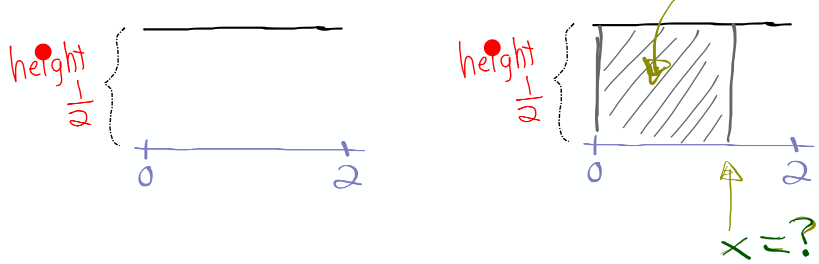
Area $= \frac{1}{2}(1.55 - .87)$

$= .34 = 34\%$

   b) About what percent of the randomly generated numbers will fall between 0.87 and 1.55 ?

c) Estimate the 65th percentile of this distribution of random numbers.

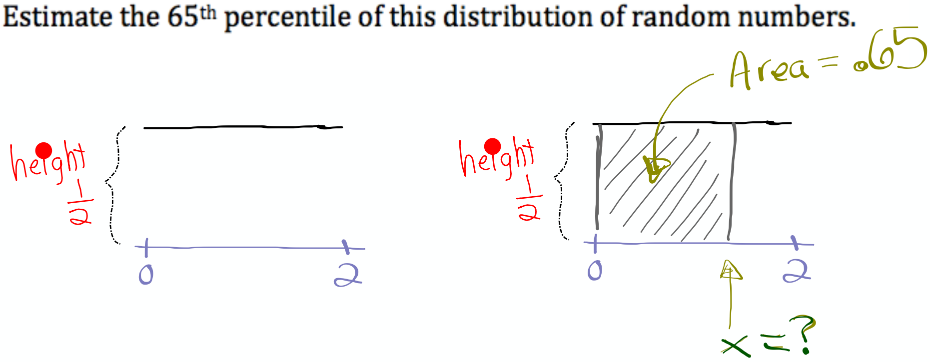
Area = .65



$\frac{x}{2} = .65$

$2(0.65)$

c) Estimate the 65th percentile of this distribution of random numbers.



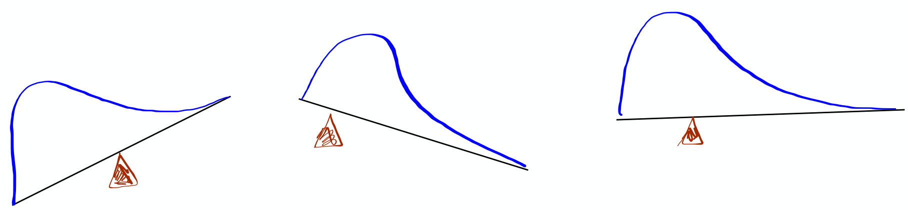
$$.65 = (x-0)\cdot\frac{1}{2}$$

$$x = 1.30$$

Goal 2

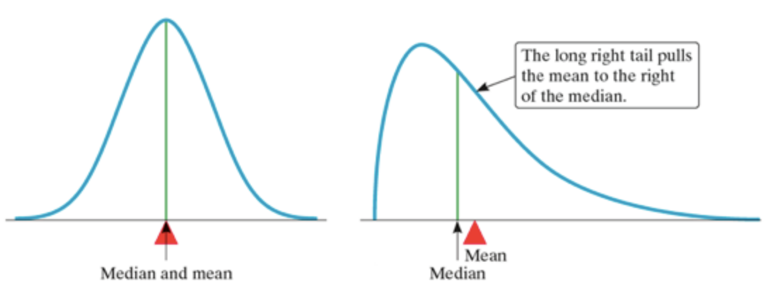
Identify relative locations of mean and median of a distribution.

Measures of Center apply to density curves as well

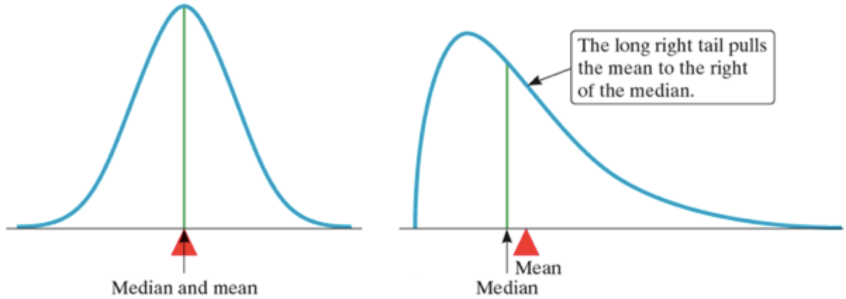


Mean (△) is a balance point at which the curve was made of solid material. (like wood)

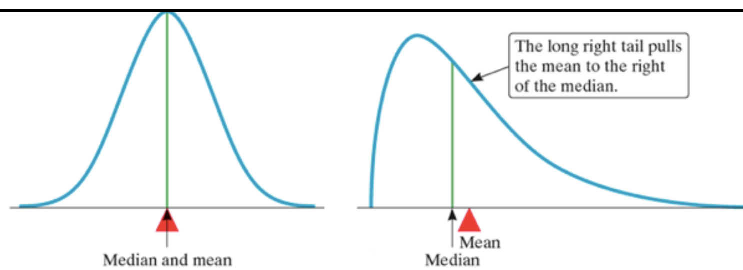
The median of a density curve is the curves equal areas point, the point that divides the area in half.



The median and the mean are the same for a symmetric density curve. They both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.



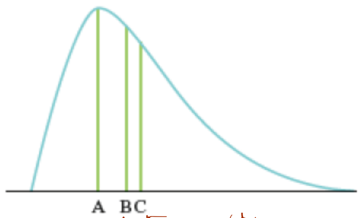
- A density curve is an idealized description of a distribution of data.



- A density curve is an idealized description of a distribution of data.
- We distinguish between the mean and standard deviation of the density curve and the mean and standard deviation computed from the actual observations.
- The usual notation for the mean of a density curve is $\boldsymbol{\mu}$ (the Greek letter mu). We write the standard deviation of a density curve as $\boldsymbol{\sigma}$ (the Greek letter sigma).

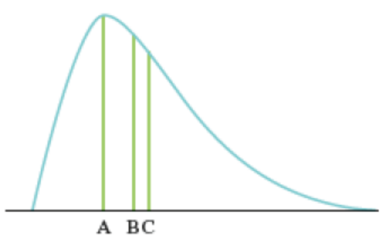
2. What does the right skew do? *Mean versus median*

The density curve that models a distribution of quantitative data is shown. Identify the location of the mean and median by letter. Justify your answers.



mean

median



SOLUTION:

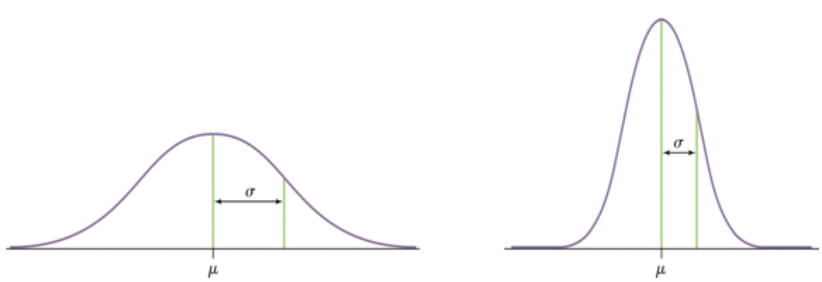
Median = B, mean = C; B is the equal-areas point of the distribution. The mean will be greater than the median due to the right-skewed shape. Even though A is directly under the peak of the curve, more than half of the area is to its right, so it cannot be the median.

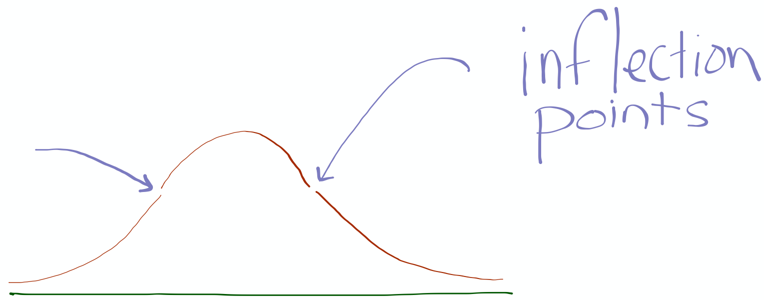
Normal Distributions

One particularly important family of density curves are the **Normal curves,** which describe **Normal distributions.**

One particularly important family of density curves are the **Normal curves,** which describe **Normal distributions.**

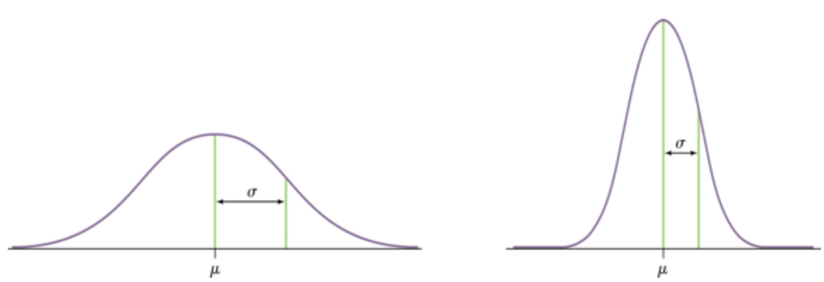
- Shape: All Normal distributions have the same overall shape: symmetric, single-peaked, and bell-shaped.





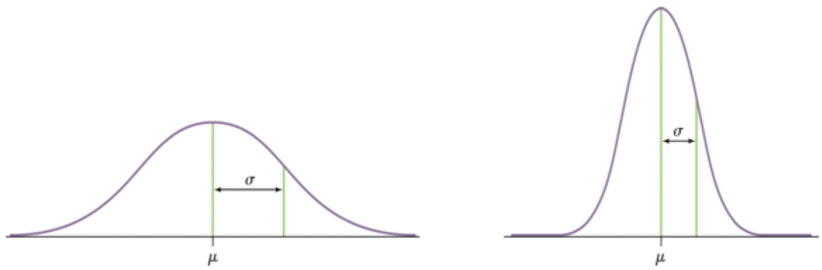
One particularly important family of density curves are the **Normal curves,** which describe **Normal distributions.**

- Shape: All Normal distributions have the same overall shape: symmetric, single-peaked, and bell-shaped.
- Center: The mean μ is located at the midpoint of the symmetric density curve and is the same as the median.



One particularly important family of density curves are the **Normal curves,** which describe **Normal distributions.**

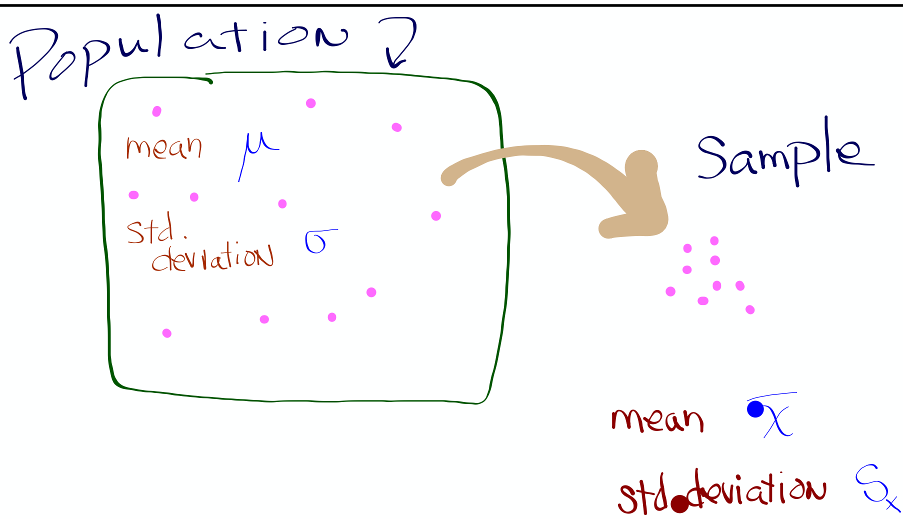
- **Shape:** All Normal distributions have the same overall shape: symmetric, single-peaked, and bell-shaped.
- **Center:** The mean μ is located at the midpoint of the symmetric density curve and is the same as the median.
- **Variability:** The standard deviation σ measures the variability (width) of a Normal distribution.



Population

mean μ

std. deviation σ

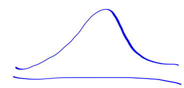


A **Normal distribution** is described by a symmetric, single-peaked, bell-shaped density curve called a **Normal curve**. Any Normal distribution is completely specified by two numbers: its mean μ and standard deviation σ.

Why are Normal distributions important in statistics? Here are three reasons.

Why are Normal distributions important in statistics? Here are three reasons.

1. Normal distributions are good descriptions for some distributions of real data. Distributions that are often close to Normal include:
   - Scores on tests taken by many people (such as SAT exams and IQ tests)
   - Repeated careful measurements of the same quantity (like the diameter of a tennis ball)
   - Characteristics of biological populations (such as lengths of crickets and yields of corn)



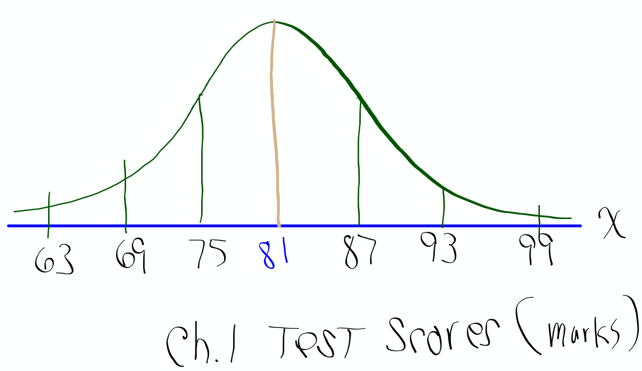
2. Normal distributions are good approximations to the results of many kinds of chance outcomes, like the proportion of heads in many tosses of a fair coin.
3. Many of the inference methods in Chapters 8–12 are based on Normal distributions.

3. Chapter 1 Test scores *-Sketching a Normal distribution*

Chapter 1 test scores from Mrs. Gallas's first-hour class follow an approximately Normal distribution with a mean of 81 and standard deviation of 6. Sketch the Normal curve that approximates the distribution of Chapter 1 test scores. Label the mean and the points that are 1, 2, and 3 standard deviations from the mean.

$\mu = 81$ points

$\sigma = 6$ points



What is so special about Normal Distributions?

page 116

Activity with laptops

cell phones ???

Write answers down in your notes

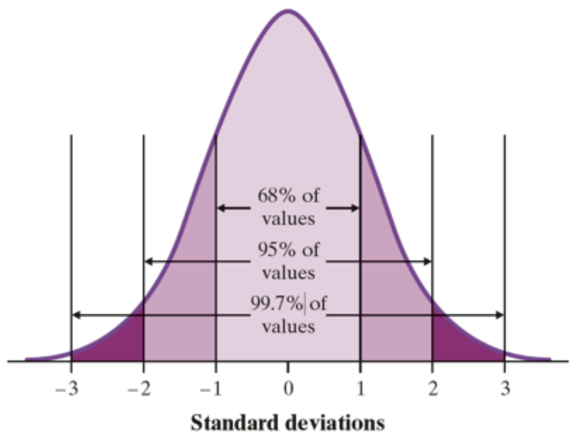
The 68–95–99.7 Rule

Although there are many Normal curves, they all have properties in common.

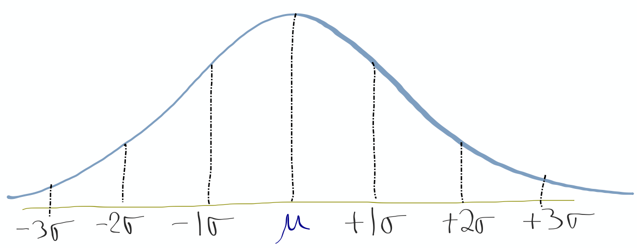
The 68-95-99.7 Rule

In a Normal distribution with mean μ and standard deviation σ:

- Approximately **68%** of the observations fall within σ of the mean μ.
- Approximately **95%** of the observations fall within 2σ of the mean μ.
- Approximately **99.7%** of the observations fall within 3σ of the mean μ.



Nice and loud



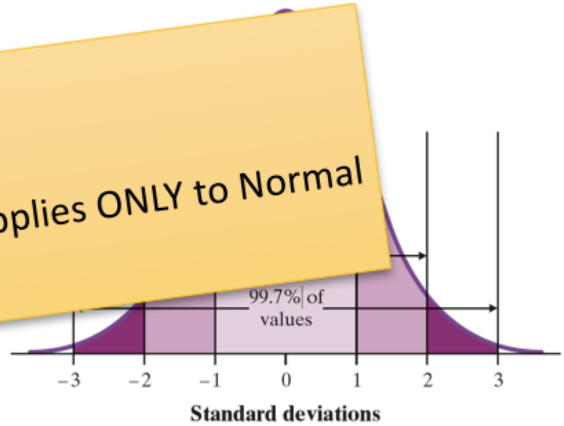
The 68-95-99.7 Rule

In a Normal distribution with mean μ and standard deviation σ:

- A
  o
  σ
- Ap
  ob
  2σ
- App...ximately **99.7%** of the observations fall within 3σ of the mean μ.

CAUTION:

The 68–95–99.7 Rule applies ONLY to Normal distributions.



4. *Using the 68–95–99.7 rule*

Chapter 1 test scores from Mrs. Gallas's first-hour class follow an approximately Normal distribution with a mean of 81 and standard deviation of 6.

What percent of the scores are greater than 81?

What percent of the scores are between 63 and 99?

If there are 50 students in the class, approximately how many students have a score within one standard deviation of the mean ?

Assignment

I suggest you read pp. 109-119

2.2... 41, 45, 47, 49, 51