# Chapter 4: Numerical Methods for Distributions of Data

## Interpreting Center & Variability in a Distribution

Adapted from Statistics and Data Analysis, 5th edition - For AP*

PECK, OLSEN, & DEVORE

# AGENDA: Oct 11, 2022

- Review three common measures of position for observations in a data set
- Introduce the Normal distribution
- Introduce *z-scores*
- *Introduce density curves as a model for the normal distribution*

# Warm-UP: Oct 6, 2022

- 1. “Sara is in the $84^{th}$ percentile of heights” means that she is ***as tall or taller*** than 84 percent of the girls her same age.
- 2. Define the standard deviation of a sample:

$$s_x = sample\ standard\ deviation$$

A **statistic** that measures the typical distance from the mean for values (observations) in a distribution. It is calculated by finding the “average” of the squared distances, and then taking the square root

# Review from last class

What are the three common measures of position for observations in a data set?

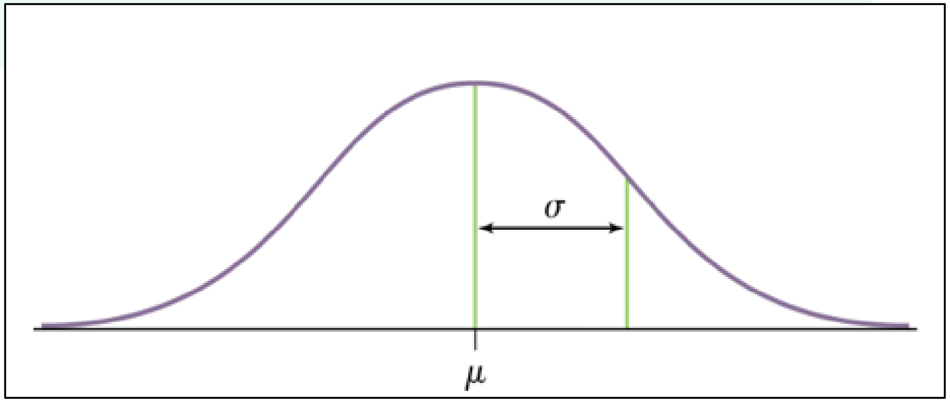
1. Percentiles
2. Quartiles
3. Standard scores (or *z-scores*)

**Note**: a z-score (or standard score) is a measure of position for an observation within a data set that provides a "standardized" measure of *distance* and *direction* in relation to The mean of the data, in terms of standard deviation

# Warm-UP: Oct, 2022

- 4. Normal is what you are accustomed to experiencing. Maybe eating eggs and bacon every morning is “normal” for you. Maybe you normally eat cereal with almond milk. Maybe normal breakfast is a bowl of rice and fried fish.

- A Normal distribution is the commonly referred parametric distribution in statistics. It is symmetric, bell-shaped, and has equivalent measures of center (mean = median= mode). Every Normal distribution is clearly defined by the value of it’s mean and its standard deviation.



# Modeling Distributions of Data

- **D**escribing Location in a Distribution
- Density Curves
- **Normal Distributions**
- **The Empirical Rule**
- **Calculating z Scores**

# Normal Distributions & z-scores
# Describing Location in a Distribution

## Learning Objectives

After this section, you should be able to…

- ✓ MEASURE position using percentiles
- ✓ MEASURE position using *z*-scores
- ✓ TRANSFORM data (z-scores)
- ✓ DEFINE and DESCRIBE density curves

## Measuring Position: Percentiles

- One way to describe the location of a value in a distribution is to tell what percent of observations are less than it.

**Definition:**

The $p^{th}$ **percentile** of a distribution is the value with $p$ percent of the observations less than or equal to it.

**Example**

Jenny earned a score of 86 on her test. How did she perform relative to the rest of the class?

| | |
|---|---|
| 6 | 7 |
| 7 | 2334 |
| 7 | 5777899 |
| 8 | 00123334 |
| 8 | 569 |
| 9 | 03 |

Her score was greater than 21 of the 25 observations. Since 22 of the 25, or 88%, of the scores are less than or equal to hers, Jenny is at the 88$^{th}$ percentile in the class's test score distribution.

## Measuring Position: *Z*-Scores

- A *z*-score tells us how many standard deviations from the mean an observation falls, and in what direction.

**Definition:**

If *x* is an observation from a distribution that has known mean and standard deviation, the **standardized value** of *x* is:

$$z = \frac{x - \text{mean}}{\text{standard deviation}}$$

A standardized value is often called a ***z*-score.**

Jenny earned a score of 86 on her test. The class mean is 80 and the standard deviation is 6.07. What is her standardized score?

$$z = \frac{x - \text{mean}}{\text{standard deviation}} = \frac{86 - 80}{6.07} = 0.99$$

## Using *z*-scores for Comparison

We can use *z*-scores to compare the position of individuals in different distributions.

**Example**

Jenny earned a score of 86 on her statistics test. The class mean was 80 and the standard deviation was 6.07. She earned a score of 82 on her chemistry test. The chemistry scores had a fairly symmetric distribution with a mean 76 and standard deviation of 4. On which test did Jenny perform better ***relative to the rest of her class***?

$$z_{stats} = \frac{86 - 80}{6.07}$$

$$z_{stats} = 0.99$$

$$z_{chem} = \frac{82 - 76}{4}$$

$$z_{chem} = 1.5$$

# Z-score WS practice: $z = \frac{obsev. - mean}{S.D.}$

1. A normal distribution of scores has a standard deviation of 10. Find the z-scores corresponding to each of the following values:

   a) A score that is 20 points above the mean.

   b) A score that is 10 points below the mean.

   c) A score that is 15 points above the mean

   d) A score that is 30 points below the mean.

# *Z-score* WS practice: $z = \frac{obsev. - mean}{S.D.}$

The Welcher Adult Intelligence Test Scale is composed of a number of subtests. On one subtest, the raw scores have **a mean of 35** and a **standard deviation of 6.** Assuming these raw scores form a normal distribution:

a) What number represents the 65$^{th}$ percentile (what number separates the lower 65% of the distribution)?

| Percentile | Quartile | z-score |
|---|---|---|
| 65$^{th}$ | $> \boxed{\phantom{0}}$ ***but*** $< \boxed{\phantom{0}}$ | **?** |

## Density Curves

- In Chapter 1, we developed a kit of graphical and numerical tools for describing distributions. Now, we'll add one more step to the strategy.

**Exploring Quantitative Data**

1. Always plot your data: make a graph.
2. Look for the overall pattern (shape, center, and spread) and for striking departures such as outliers.
3. Calculate a numerical summary to briefly describe center and spread.
4. Sometimes the overall pattern of a large number of observations is so regular that we can describe it by a smooth curve.

## Density Curve

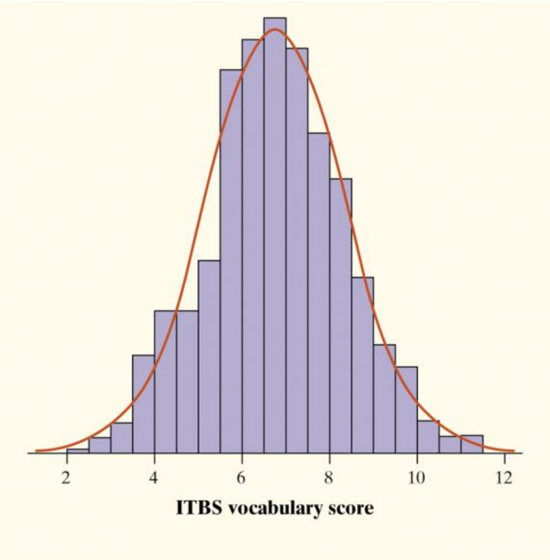
**Definition:**

A **density curve** is a curve that

- is always on or above the horizontal axis, and
- has area of *exactly 1* underneath it.

A density curve describes the overall pattern of a distribution. The area under the curve and above any interval of values on the horizontal axis is 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.



## Describing Density Curves

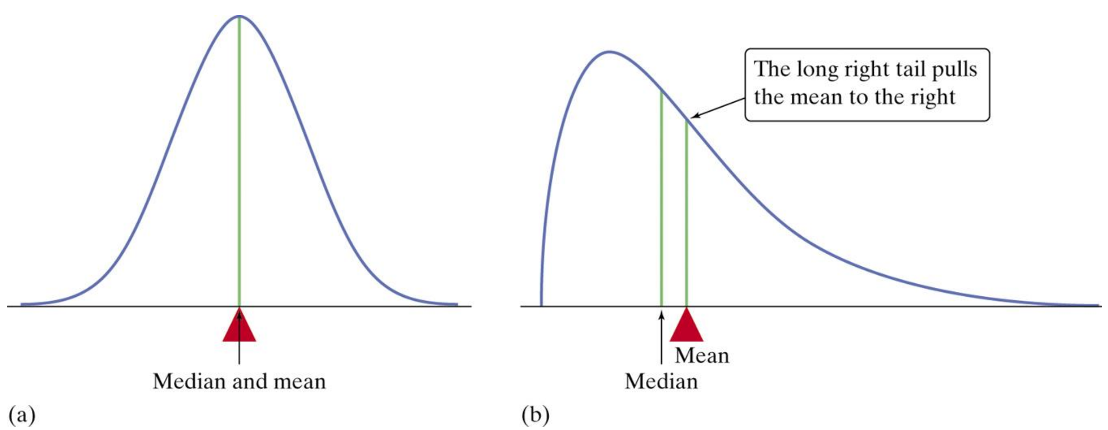
- Our measures of center and spread apply to density curves as well as to actual sets of observations.

**Distinguishing the Median and Mean of a Density Curve**

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

The **mean** of a density curve is the balance point, at which the curve would balance if made of solid material.

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.



# Review of Position
# Describing Location in a Distribution

## Summary

In this section, we learned that…

- ✓ There are two ways of describing an individual's location within a distribution – the **percentile** and ***z*-score**..
- ✓ It is common to **transform data**, especially when changing units of measurement. Transforming data can affect the shape, center, and spread of a distribution.
- ✓ We can sometimes describe the overall pattern of a distribution by a **density curve** (an idealized description of a distribution that smooths out the irregularities in the actual data).

# Looking Ahead…

**<u>In the next Section…</u>**

We'll learn about one particularly important class of density curves – the **Normal Distributions**

We'll learn

- ✓**The 68-95-99.7 Rule**
- ✓**The Standard Normal Distribution**
- ✓**Normal Distribution Calculations, and**
- ✓**Assessing Normality**

# Normal Distributions (cont.)

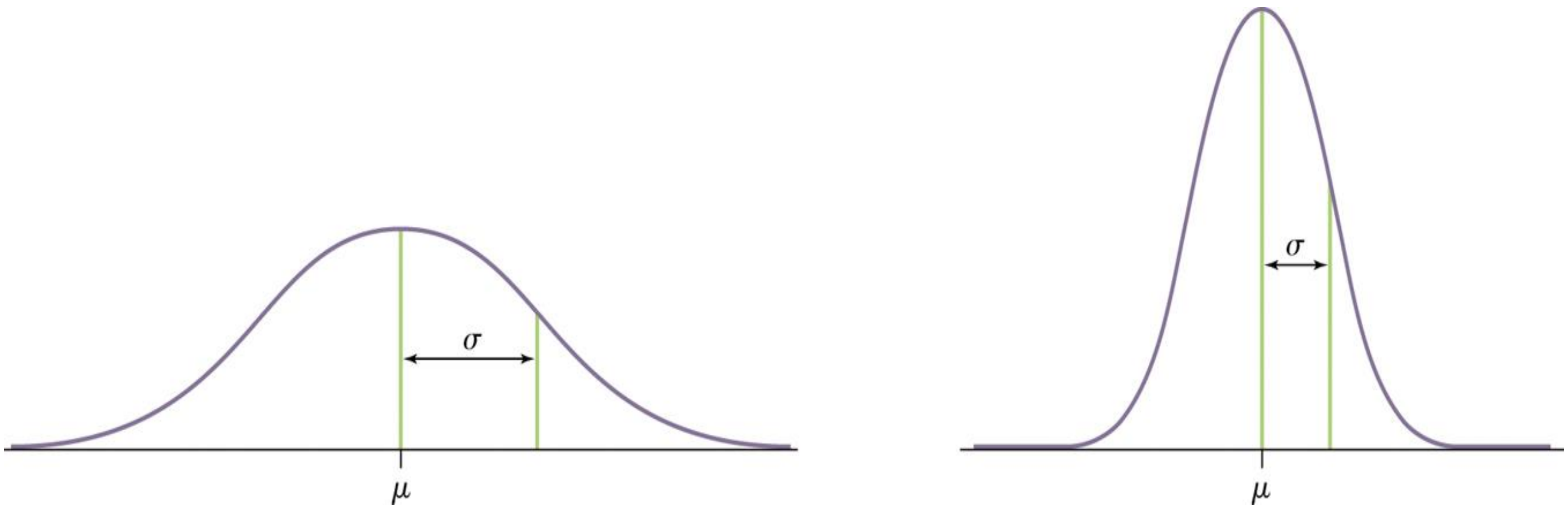
## Learning Objectives

After this section, you should be able to…

- ✓ DESCRIBE and APPLY the Empirical Rule (68-95-99.7 Rule)
- ✓ DESCRIBE the standard Normal Distribution
- ✓ PERFORM Normal distribution calculations
- ✓ ASSESS Normality

## Normal Distributions

- One particularly important class of density curves are the **Normal curves**, which describe **Normal distributions**.
  - All Normal curves are symmetric, single-peaked, and bell-shaped
  - Any  Specific Normal curve is described by giving its mean $\mu$ *("mu")* and standard deviation σ ("sigma").




Two Normal curves, showing the mean $\mu$ and standard deviation σ.

# Normal Distributions



**Definition:**

A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by two parameters: its **mean $\mu$** *("mu")* and standard deviation σ ("sigma").

- The **mean** *($\mu$)* of a Normal distribution is the center of the symmetric **Normal curve**.
- The standard deviation (σ) is the distance from the center to the change-of-curvature points (*points of inflection*) on either side.
- We abbreviate the ***N***ormal distribution with mean and standard deviation as: $\boldsymbol{N}\,(\mu,\sigma)$.

Normal distributions are good descriptions for some distributions of *real data.*

Normal distributions are good approximations of the results of many kinds of *chance outcomes*.

Most of our *statistical inference* procedures are based on Normal distributions.

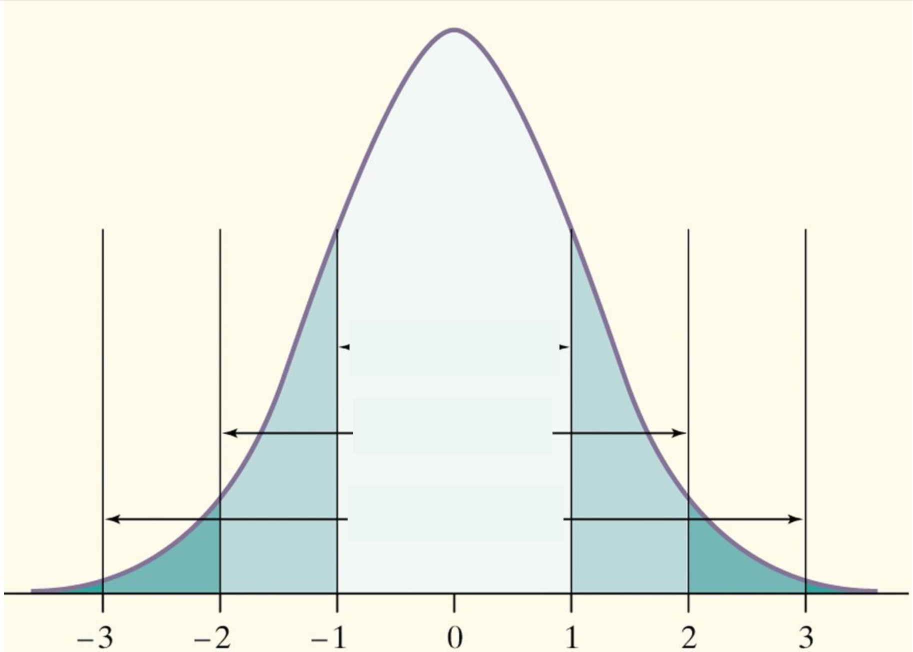
# The Empirical Rule (68-95-99.7 Rule)

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

**Definition:** **The 68-95-99.7 Rule ("The Empirical Rule")**

In the Normal distribution with mean $\mu$ and standard deviation σ:

- Approximately **68%** of the observations fall within 1σ of $\mu$.
- Approximately **95%** of the observations fall within 2σ of $\mu$.
- Approximately **99.7%** of the observations fall within 3σ of $\mu$.



## The Standard Normal Distribution

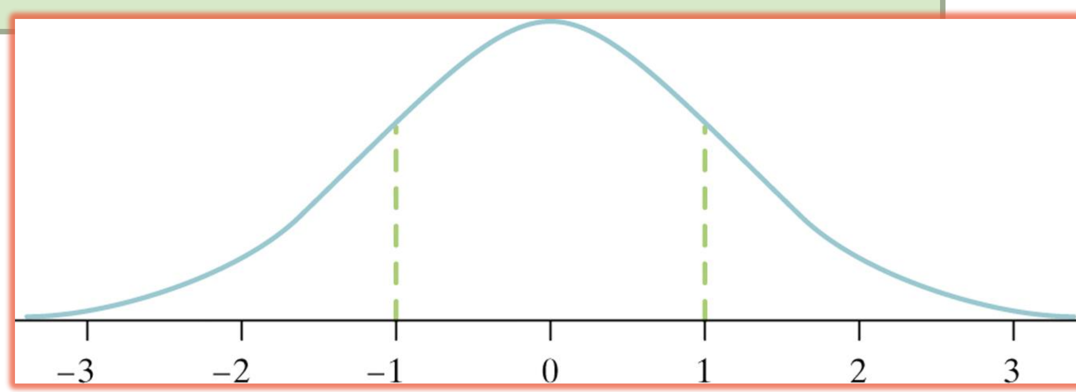
- All Normal distributions are the same if we measure in units of size σ from the mean *μ* as center.

**Definition:**

The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1.
If a variable *x* has any Normal distribution *N*(*μ*,*σ*) with mean *μ* and standard deviation σ, then the standardized variable

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

has the standard Normal distribution, *N*(0,1).



## Normal Distribution Calculations

**How to Solve Problems Involving Normal Distributions**

**State:** Express the problem in terms of the observed variable *x*.

**Plan:** Draw a picture of the distribution and shade the area of interest under the curve.

**Do:** Perform calculations.

- ***Standardize*** *x* to restate the problem in terms of a standard Normal variable *z*.
- ***Use Standard Table*** and the fact that the total area under the curve is 1 to find the required area under the standard Normal curve.

**Conclude:** Write your conclusion in the context of the problem.

**Example**

The distribution of Iowa Test of Basic Skills (ITBS) vocabulary scores for 7th grade students in Gary, Indiana, is close to Normal. Suppose the distribution is $N(6.84, 1.55)$.

$N\,(\mu, \sigma)$

a) Sketch the Normal density curve for this distribution.

b) What percent of ITBS vocabulary scores are less than 3.74?

c) What percent of the scores are between 5.29 and 9.94?