Objectives:

At the end of this handout you should be able to:

1) describe how adding or subtracting the same value to every number on a list will change the median, mean, max, min, quartiles, IQR and the standard deviation.

2) derive from the original values of the mean, median, max, min, quartiles, IQR and standard deviation the new values of these summary statistics for a list of numbers that was obtained by adding or subtracting the same value to every number on the original list.

3) summarize how multiplying or dividing every number on a list by the same value will change the median, mean, max, min, quartiles, IQR and the standard deviation.

4) derive from the original values of the mean, median, max, min, quartiles, IQR and standard deviation the new values of these summary statistics for a list of numbers that was obtained by multiplying or dividing every number on the original list by the same value.

5) List the key characteristics of the normal model:
   * The normal model is symmetric around its mean.
   * The mean and median of a normal model are the same.
   * The normal model is "bell-shaped".

6) Apply the 68-95-99.7 rule to summarize the behavior of observations whose values can be described using the normal model.

7) Calculate z-scores.

8) Interpret z-scores.

9) Use z-scores and technology (or normal tables) to calculate proportions from a normal model when you are given information about the mean and standard deviation.

10) Identify a percentile of the normal model given the mean and standard deviation.

Comparing Apples and Oranges: z-scores and the Standard Deviation

standardizing with z-scores:

\[ z = \frac{(y - \bar{y})}{s} \]

called standardized values or z-scores

* z-scores have no units

* z-scores measure the distance of each data value from the mean in standard deviation units

Example (SAT vs ACT)

National results for the SAT test show that for college-bound seniors the average combined SAT Writing, Math and Verbal score is 1500 and the standard deviation is 250.

National results for the ACT test show that for college-bound seniors the average composite ACT score is 20.8 and the standard deviation is 4.8.

Your SAT score: 1860. Your neighbor's ACT: 29.

Who did better on their respective test?

SAT z-score =

ACT z-score = INTERPRET:
Shifting Data
In general, adding a constant to every data value adds the same constant to measures of center and percentiles, but leaves measures of spread unchanged.

weights of 80 men age 19 to 24 of average height (5'8" to 5'10"); \( \bar{y} = 82.36 \) kg

NIH recommends maximum healthy weight of 74 kg. To compare their weights to the recommended maximum, subtract 74 kg from each weight; new \( \bar{y} = 82.36 - 74 = 8.36 \) kg (overweight)

What happens to the SHAPE of the distribution?

Rescaling Data
When we divide or multiply all the data values by any constant value:

i) measures of location (e.g., mean and median) are divided and multiplied by the same value.

ii) measures of spread (e.g., range, IQR, standard deviation) are divided and multiplied by the absolute value of the constant.

To change kilograms to pounds, multiply by 2.2

\[
\begin{align*}
\text{original data} & \quad \text{new data} \\
\bar{y} = 82.36 \text{ kg} & \quad \bar{y} = 82.36 \times 2.2 = 181.19 \text{ lb} \\
\text{min} = 54.30 \text{ kg} & \quad \text{min} = 54.30 \times 2.2 = 119.46 \text{ lb} \\
\text{max} = 161.50 \text{ kg} & \quad \text{max} = 161.50 \times 2.2 = 355.30 \text{ lb} \\
\text{range} = 107.20 \text{ kg} & \quad \text{range} = 107.20 \times 2.2 = 235.84 \text{ lb}
\end{align*}
\]

Back to z-scores

i) Standardizing data into z-scores shifts the data by subtracting the mean and rescales the values by dividing by their standard deviation.

ii) Standardizing into z-scores does not change the shape of the distribution.

iii) Standardizing into z-scores changes the center by making the mean 0.

iv) Standardizing into z-scores changes the spread by making the standard deviation 1
Example (Old) SAT score from ACT score
\[ \text{SAT} = (40 \times \text{ACT}) + 150 \]

ACT results for 2355 college applicants:
- \( \text{min} = 19 \)
- \( \text{mean} = 27 \)
- \( \text{Stan dev} = 3 \)
- \( Q_3 = 30 \)
- \( \text{Median} = 28 \)
- \( \text{IQR} = 6 \)

summarizes of equivalent SAT scores:

---

**When is a z-score big? The Normal Model**

- The larger a z-score is (negative or positive), the more unusual it is.
- There is no universal standard for z-scores, but there is a **model** that shows up over and over in Statistics.

**The Normal Model** (bell-shaped curves)

- Normal models are appropriate for distributions whose shapes are unimodal and roughly symmetric.
- The Normal model provides a measure of how extreme a z-score is.
- There is a Normal model for every possible combination of mean and standard deviation.

**NOTATION:** We write \( N(\mu, \sigma) \) to represent a Normal model with a mean of \( \mu \) and a standard deviation of \( \sigma \)

We use **Greek letters** because this mean and standard deviation do not come from data—they are numbers (called parameters) that specify the model.
When we standardize a Normal model, we still call the standardized value a z-score, and we write
\[
z = \frac{y - \mu}{\sigma}
\]

Once we have standardized, we need only one model:
The N(0,1) model is called the standard Normal model (or the standard Normal distribution).

**Be careful!** — don’t use a Normal model for just any data set, since standardizing does not change the shape of the distribution.

**Nearly Normal Condition:**
- The shape of the data’s distribution is unimodal and symmetric.
- This condition can be checked with a histogram or a Normal probability plot (to be explained later).

**The 68-95-99.7% Rule**

- Draw a picture!
- Draw a picture!
- Draw a picture!
Example
Suppose it takes you 20 minutes, on average, to drive to school, with a standard deviation of 2 minutes. Suppose a Normal model is appropriate for the distribution of driving times.

1) How often will you arrive at school in less than 22 minutes?

2) How often will it take you more than 24 minutes?

3) 97.5% of the time you will arrive at school in less than how many minutes?

Finding Normal Percentiles

• To find normal percentiles using a TI 83/84, see p. 13 in the Statistics on TI Graphing Calculators (pdf) link on our class web page http://www.stat.ncsu.edu/people/reiland/courses/st372/.

• You can also use a z-table, Excel, or statcrunch.

Example (SAT scores)
SAT Math scores can be approximated by a $N(500, 100)$ model. If you score 680, what proportion of scores on the SAT Math test were below your score?

$$z - score = \frac{680 - 500}{100} = \frac{180}{100} = 1.8$$
From Table Z, the proportion of scores below 680 is 0.9641.

Finding z-scores from percentiles: “reverse” problems

To find z-scores from percentiles using a TI 83/84, type 2nd VARS 3: invNorm.

Example: Cholesterol levels in American women

Cholesterol levels in American women can be described by a Normal model with mean \( \mu = 188 \) mg/dL and a standard deviation of 24 mg/dL. A family physician decides to prescribe a cholesterol reducing drug to female patients who are above the 90th percentile of the cholesterol level for American women. Above what cholesterol level will the physician prescribe the drug?

Are You Normal? How Can You Tell?

NEVER just assume your data can be approximated by a Normal model.

ALWAYS check your data:

\[ \rightarrow \text{ histogram} \]

\[ \rightarrow \text{ normal probability plot} \] (another graphical display to assist you in deciding whether data is approximately normal)
The CEO compensation data represented by the above histogram should NOT be described by a Normal model.

→ **normal probability plot**
A more specialized graphical display that can help you decide whether a Normal model is appropriate

If the distribution of the data is roughly Normal, the Normal probability plot approximates a diagonal straight line. Deviations from a straight line indicate that the distribution is not Normal.

A normal probability plots the ordered data values against the values that would be expected in that order position for a sample from a standard Normal.

Normal probability plots can be produced using TI 83/84 calculators, statcrunch, and other software.
Example MPG ratings for several Nissan automobiles

Example Weights of a group of men

Example Student heights ST 101