11-2 Measures of Central Tendency

Dec 12-11:00 AM

- The **median** is the number that divides the (ordered) data in half. At least half the data is equal to or smaller than the median, and at least half the data is equal to or greater than the median. (In a histogram, the median is that middle value that divides the histogram into two equal areas.)
- The **mode** of a set of data is the most common value among the data.
- The **mean** (more precisely, the arithmetic mean) is commonly called the average. It is the sum of the data, divided by the number of data:

$$\text{mean} = \frac{\text{sum of data}}{\text{number of data}} = \frac{\text{total}}{\text{number of data}}$$

THE MODE

For discrete numerical data, the **mode** is the most frequently occurring value in the data set.

For continuous numerical data, we cannot talk about a mode in this way because no two data values will be *exactly* equal. Instead we talk about a **modal class**, which is the class or group that occurs most frequently.

Nov 5-8:09 AM

The **arithmetic mean** or **average** of a set of n measurements (data set) is equal to the sum of the measurements divided by n.

Notation

The sample mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$, where n is the sample size.

This is a **statistic**.

The population mean: $\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$, where N is the population size. This is a **parameter**.

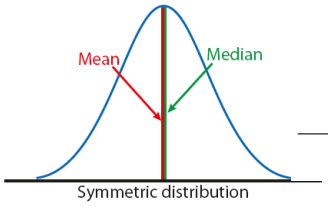
It is important to observe that you normally do not know the mean of the population μ and that you usually estimate it with the sample mean $\bar{x}$.

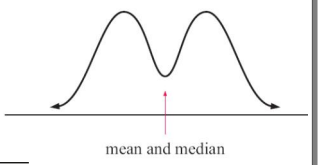
The **median** of a set of n measurements is the value of x that falls in the middle position when the data is sorted in ascending order.

Nov 5-8:11 AM

Symmetry

The shape of a distribution is said to be **symmetric** if the observations are balanced, or evenly distributed, about the mean. In a symmetric distribution, *the mean and the median are equal.*

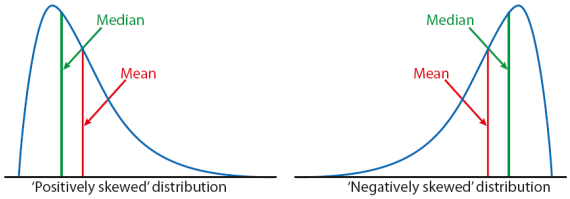




Nov 5-8:12 AM

Skewness

A distribution is **skewed** if the observations are not symmetrically distributed above and below the mean.



A **positively skewed** (or skewed to the right) distribution has a tail that extends to the right in the direction of positive values. A **negatively skewed** (or skewed to the left) distribution has a tail that extends to the left in the direction of negative values.

Nov 5-8:12 AM

CASE STUDY **DRIVING A GOLF BALL**

While attending a golf championship, I measured how far Ethan, a professional golfer, hit 30 drives on the practice fairway. The results are given below in metres:

244.6	245.1	248.0	248.8	250.0	251.1
251.2	253.9	254.5	254.6	255.9	257.0
260.6	262.8	262.9	263.1	263.2	264.3
264.4	265.0	265.5	265.6	266.5	267.4
269.7	270.5	270.7	272.9	275.6	277.5

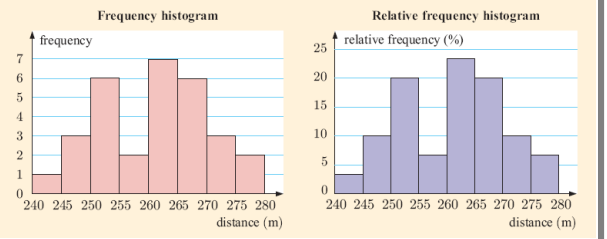
To organise the data, we sort it into **groups** in a frequency table.

Ethan's 30 drives

Distance (m)	*Tally*	*Frequency* (f)	*% relative frequency*	
$240 \leqslant d < 245$	|	1	3.3	$\approx \frac{1}{30} \times 100$
$245 \leqslant d < 250$	|||	3	10.0	$= \frac{3}{30} \times 100$
$250 \leqslant d < 255$	𝍸 |	6	20.0	
$255 \leqslant d < 260$	||	2	6.7	
$260 \leqslant d < 265$	𝍸 ||	7	23.3	
$265 \leqslant d < 270$	𝍸 |	6	20.0	
$270 \leqslant d < 275$	|||	3	10.0	
$275 \leqslant d < 280$	||	2	6.7	
	Totals	30	100.0	

Nov 5-8:03 AM

From this table we can draw both a **frequency histogram** and a **relative frequency histogram**:



Nov 5-8:05 AM

What to do:

1 Enter the data into your graphics calculator as a list, or use the **statistics package** supplied.

- **a** Produce a frequency histogram of the data. Set the X values from 240 to 280 with an increment of 5. Set the Y values from 0 to 30.
- **b** Comment on the shape of the distribution.
- **c** Find the mean and median of the data.
- **d** Compare the mean and the median. Is the mean an accurate measure of the centre?

2 Since we have continuous numerical data, we have a modal class rather than an individual mode.

- **a** What is the modal class?
- **b** What would the modal class be if our intervals were 2 m wide starting at 240 m?

Nov 5-8:15 AM

3 Now suppose Ethan had hit a few very bad drives. Let us say that his three shortest drives were very short!

- **a** Change the three shortest drives to 82.1 m, 103.2 m, and 111.1 m.
- **b** Repeat **1 a**, **b**, **c**, and **d** but set the X values from 75 to 300 with an increment of 25 for the frequency histogram.
- **c** Describe the distribution as symmetric, positively skewed, or negatively skewed.
- **d** What effect have the changed values had on the mean and median as measures of the centre of the data?

4 What would have happened if Ethan had hit a few really long balls in addition to the very bad ones? Let us imagine that the longest balls he hit were very long indeed!

- **a** Change the three longest drives to 403.9 m, 415.5 m, and 420.0 m.
- **b** Repeat **1 a**, **b**, **c**, and **d** but set the X values from 50 to 450 with an increment of 50 for the frequency histogram.
- **c** Describe the distribution as symmetric, positively skewed, or negatively skewed.
- **d** What effect have the changed values had on the mean and median as measures of the centre of the data?

Nov 5-8:16 AM

pg 585 #1, 4, 5, 6, 8, 16, 17

Dec 17-10:59 AM