

# Central Measures

When dealing with numerical data, it is often useful to summarize the data by a single number:

- ▶ class average on an exam
- ▶ average time for waiting in line at the store
- ▶ the most popular choice for our Movie Night

The three most popular ways to summarize data are:

- ▶ the mean/average
- ▶ the median
- ▶ the mode

Each of these offers a different **measure of the center** of the data set we are working with.

# The Mean

For a data set  $\{x_1, x_2, x_3, \dots, x_n\}$ , the **mean** (denoted by the Greek letter  $\mu$  miu), can be calculated by

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

where  $n$  is the number of data points.

**Example:** The following data represents the height of 16 individuals in centimeters. Calculate the average value.

201	190	196	180	171	157	190	157
153	189	167	198	142	123	151	182

The sum of all the data points is 2747, so the average is  $2747/16 = 171.69$  cm.

# The Mean

**Example:** The average predicted daily temperatures (in degrees Fahrenheit) for the next two weeks is given below:

Week 1		88	75	73	72	84	81	88
Week 2		85	87	93	80	84	84	80

Compute the average temperature for each week separately, and the average temperature for both weeks combined. Interpret your results.

For week one, the sum of the values is 561, and the average temperature is  $561/7 = 80.1$  degrees.

For week two, the sum of the values is 593, and the average temperature is  $593/7 = 84.7$  degrees.

For both weeks combined, the sum is  $561 + 593 = 1154$ , and the average temperature is  $1154/14 = 82.4$  degrees.

# The Mean

**Example:** In a finite math class, the scores for Quiz 5 and Quiz 6 are given below. Both quizzes are out of 10 points.

Quiz 5		3	5	1	4	8	3	3	1	1	7
Quiz 6		7	7	3	8	10	7	4	4	4	10

Calculate the class average for each individual quiz and interpret your results.

For Quiz 5, the class average is  $41/10 = 4.1$  points.

The Quiz 6, the class average is  $64/10 = 6.4$  points.

**Interpretation:** We could infer that Quiz 5 was harder than Quiz 6. However, it is also possible that the students started studying more after seeing their scores on Quiz 5, and their grades improved!

# Estimating The Mean

**Example:** The GPA of 11 students is summarized by the following frequency table:

GPA	Freq
2-2.5	3
2.5-3	4
3-3.5	2
3.5-4	2
Total	11

Give an estimate of the average GPA for the students.

For each bin, we pick the **middle point**, then multiply it by the frequency to obtain an estimate for the GPA overall sum. Then we divide this by the number of students:

$$\mu \approx \frac{(2.25)(3) + (2.75)(4) + (3.25)(2) + (3.75)(2)}{11} = 2.89$$

# The Median

Another central measure is the **median**. This can be computed by **sorting the data** in order, then picking the **middle number**.

**Example:** Compute the median of the following data set:

69   81   83   41   103   94   53

The sorted data is: 41, 53, 69, **81**, 83, 94, 103, so the median is 81.

**Example:** Compute the median of the following data set:

2.3   4   2   3.3   3.2   2.1

Sort: 2   2.1   **2.3**   **3.2**   3.3   4. We have two candidates for the median, so we average:  $\frac{2.3+3.2}{2} = 2.75$ .

# The Mode

The **mode** of a data set is the value that occurs most often (highest frequency). It is possible to have multiple modes, or no mode (if no value occurs more than once).

**Example:** The temperature forecast (in degrees Fahrenheit) for the next week is:

88   75   73   72   84   81   88

What is the mode of this data set? 88 degrees.

# The Mode

**Example:** The scores on two quizzes are given below.  
Calculate the mode for each quiz.

Quiz 5	3	5	1	4	8	3	3	1	1	7
Quiz 6	7	7	3	8	10	7	4	4	4	10

Below are the frequency tables for each quiz:

(Q5) Score	Freq	(Q6) Score	Freq
1	3	3	1
3	3	4	3
4	1	7	3
5	1	8	1
7	1	10	2
8	1		

For Quiz 5, we have 2 modes: 1 and 3.

For Quiz 6, we also have 2 modes: 4 and 7.



# A Visual Interpretation Of The Mode

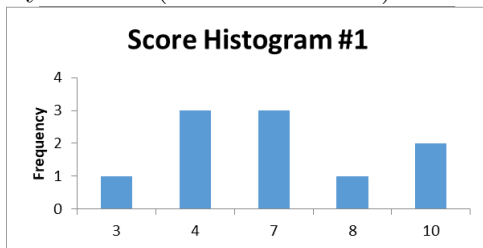
When describing the data with a histogram, the mode will usually appear as the **tallest bar/peak** in the picture.

**Warning:** This varies wildly depending on how the bins are chosen! It is not always the case that the mode matches the highest peak in the histogram.

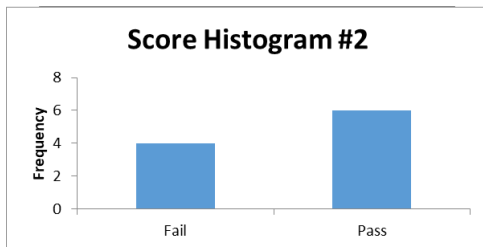
# A Visual Interpretation Of The Mode

For the Quiz 6 scores  $\{6, 7, 7, 3, 8, 10, 7, 4, 4, 4, 10\}$ , we can construct different frequency tables. (Modes: 4 and 7)

(Q6) Score	Freq
3	1
4	3
7	3
8	1
10	2



(Q6) Bin	Freq
Fail ( $< 6$ )	4
Pass ( $\geq 6$ )	6



# Which is better?

## Facts about **the mean**:

- ▶ easy to compute for any numerical data set
- ▶ accurate if not too many extreme values occur
- ▶ not good if there are a lot of **outliers**

## Facts about **the median**:

- ▶ divides the data into two equal parts
- ▶ good for ranking things
- ▶ better than the mean if there are outliers

## Facts about **the mode**:

- ▶ only takes into account a few values (with highest frequency) rather than the entire data set
- ▶ good for locating preferences, as it picks the most frequent category