A **variable** is a characteristic that varies among individuals in a population or in a sample (a subset of a population).

Example: age, height, blood pressure, ethnicity, leaf length, first language

Variables can be either **quantitative**

- Something that can be counted or measured for each individual and then added, subtracted, averaged, etc. across individuals in the population.
- Example: How tall you are, your age, your blood cholesterol level, the number of credit cards you own

… or **categorical.**

- Something that falls into one of several categories. What can be counted is the count or proportion of individuals in each category.
- Example: Your blood type (A, B, AB, O), your hair color, your ethnicity, whether you paid income tax last tax year or not

Histogram

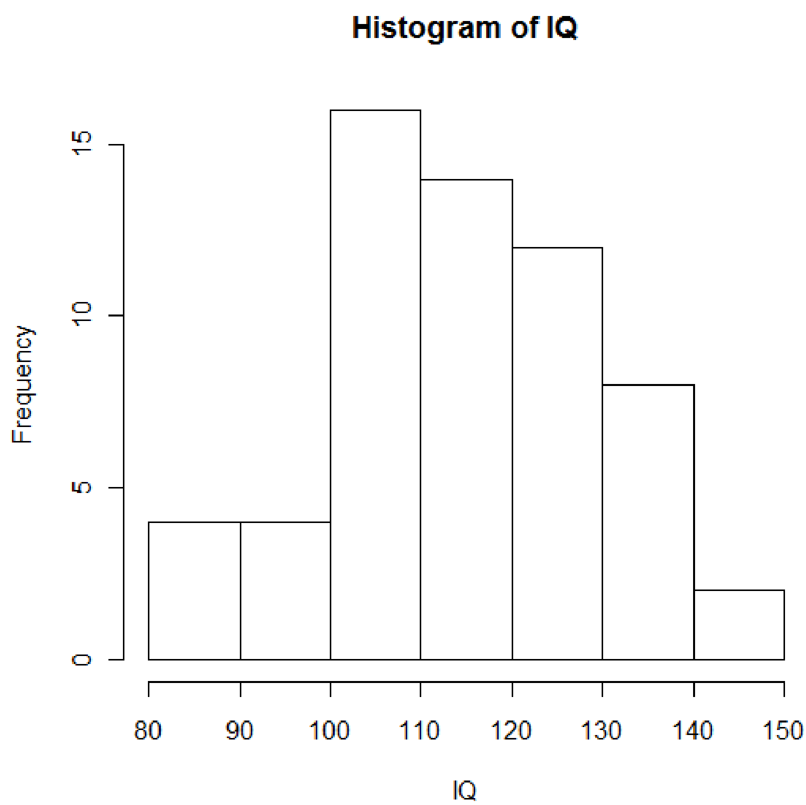
A histogram breaks the range of values of a variable into classes and displays only the count or percent of the observations that fall into each class.

Eg:

The following represent the IQ scores for 60 randomly chosen fifth-grade students:

145	139	126	122	125	130	96	110	118	118
101	142	134	124	112	109	134	113	81	113
123	94	100	136	109	131	117	110	127	124
106	124	115	133	116	102	127	117	109	137
117	90	103	114	139	101	122	105	97	89
102	108	110	128	114	112	114	102	82	101

You can use software (such as Excel or R) to produce a histogram easily:



Sample Mean

To calculate the **mean,** add all values, then divide by the number of individuals.

Example:

Find the mean of the following numbers: 5, 9, 13, 25

Mean = (5 + 9 + 13 + 25) / 4= 13

(Symbol for sample mean: $\bar{x}$)

Median

The **median** is the midpoint of a distribution—the number such that half of the observations are smaller and half are larger.

Eg:

To find the median of a set of numbers, arrange them in ascending order. Let n be the number of observations. If n is odd, then the (n +1)/2 observation is the median.

Example, to find the median of 6, 9, 12, 3, 4, we first arrange the numbers in order: 3, 4, 6, 9, 12.

The median is the (5+1)/2 = 3rd observation (that is, the median is 6).

If n is even, the median is the mean of the two middle observations. Example, to find the median of 6, 9, 12, 3, 4, 5, we arrange the numbers in order: 3, 4, 5, 6, 9, 12. The middle two numbers are 5, 6. The mean of 5 and 6 is 5.5. The median is therefore 5.5.

Quartiles

The **first quartile, Q_1,** is the value in the sample that has 25% of the data at or below it (it is the median of the lower half of the sorted data, excluding *M*).

The **third quartile, Q_3,** is the value in the sample that has 75% of the data at or below it (it is the median of the upper half of the sorted data, excluding *M*).

Example:

	1	1	0.6
	2	2	1.2
	3	3	~~1.6~~ 1.5
	4	4	~~1.9~~ 1.6
	5	5	~~1.5~~ 1.9
	6	6	2.1
$Q_1 = 2.2$	7	7	2.3
	8	1	2.3
	9	2	2.5
	10	3	2.8
	11	4	2.9
	12	5	3.3
Median is 3.4→	13		3.4
	14	1	3.6
	15	2	3.7
	16	3	3.8
	17	4	3.9
	18	5	4.1
$Q_3 = 4.35$	19	6	4.2
	20	7	4.5
	21	1	4.7
	22	2	4.9
	23	3	5.3
	24	4	5.6
	25	5	6.1

Another example: but this time with an even number of observations.

Say you wish to find Q_1, median and Q_3 for the following set of 12 numbers (arranged in ascending order for you):

$Q_3 = 2.85$

0.6 1.2 1.5 1.6 1.9 2.1 2.3 2.5 2.8 2.9 3.3 3.4

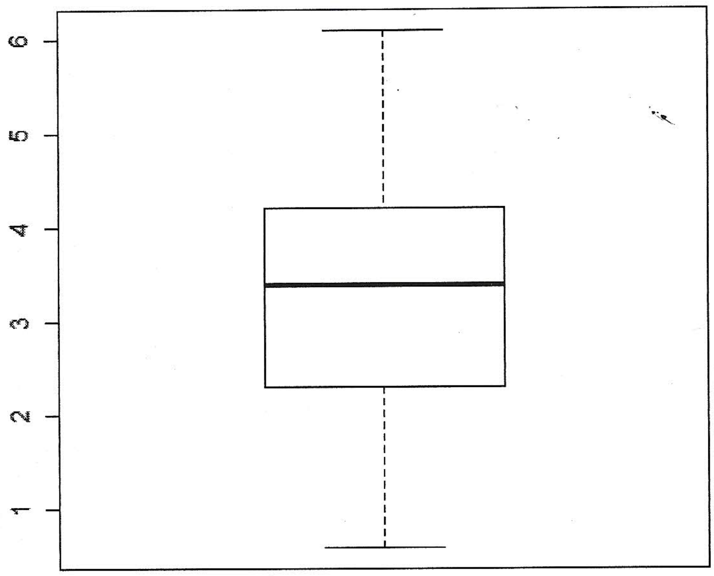
median $= \frac{2.1 + 2.3}{2} = 2.2$

$Q_1 = 1.55$

<u>Interquartile range</u>

IQR = $Q_3 - Q_1$

<u>Boxplot</u>



Maximum

Q_3

median

Q_1

minimum

We call min, Q_1, median, Q_3, max the 5-number summary

Sample variance and standard deviation:

Sample variance:

$$s^2 = \frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x})^2$$

Sample standard deviation is the square root of the sample variance.

$$s = \sqrt{\frac{1}{n-1}\sum_{1}^{n}(x_i - \bar{x})^2}$$

Example:

Learn how to use your calculator to get the sample standard deviation.

Find standard deviation of the following data: 3, 5, 6, 6

[We will learn a shortcut formula in class]

$$\bar{x} = \frac{\sum_{i=1}^{n} x}{n} = \frac{3+5+6+6}{4} = 5$$

$$s^2 = \frac{1}{4-1}\left[(3-5)^2+(5-5)^2+(6-5)^2+(6-5)^2\right]$$

$$= \frac{1}{3}\left[4+0+1+1\right]$$

$$= 2$$

$$\Rightarrow s = \sqrt{2}$$

Does your calculator produce the same answer?

Shortcut formula:

$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$$

For problem above,

$$\sum x = 3+5+6+6 = 20$$

$$\sum x^2 = 3^2+5^2+6^2+6^2 = 106$$

$$s^2 = \frac{1}{3}\left[106 - \frac{20^2}{4}\right]$$

$$= \frac{6}{3} = 2$$