

## Statistics Handout 2

The **mode** of a data set is the value that occurs most frequently.

The **median** is the central value of an ordered distribution. To find it,

- Order the data from smallest to largest.
- For an *odd* number of data values in the distribution,

$$\text{Median} = \text{Middle data value}$$

- For an even number of data values in the distribution,

$$\text{Median} = \frac{\text{Sum of middle two values}}{2}$$

$$\text{Mean} = \frac{\text{Sum of all the entries}}{\text{Number of entries}}$$

$$\text{Sample mean} = \bar{x} = \sum_n \frac{x}{n}$$

$$\text{Population mean} = \mu = \sum_N \frac{x}{N}$$

The **range** is the difference between the largest and smallest values of a data distribution.

$$\text{Sample standard deviation} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad \text{Sample variance} = s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$\mu = \frac{\sum x}{N} \text{ population mean} \quad \sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}} \text{ population standard deviation}$$

### Results of Chebyshev's theorem

For any set of data:

- At least 75% of the data fall into the interval from  $\mu - 2\sigma$  to  $\mu + 2\sigma$
- At least 88.9% of the data fall into the interval from  $\mu - 3\sigma$  to  $\mu + 3\sigma$
- At least 93.8% of the data fall into the interval from  $\mu - 4\sigma$  to  $\mu + 4\sigma$

## Grouped data

When data are grouped such as in the frequency table or histogram, we can estimate the mean and the standard deviation by using the following formulas. Notice that all data values in a given class are treated as though each of them equals the midpoint  $x$  of the class.

### Sample mean for a frequency distribution

$$\bar{x} = \frac{\sum xf}{n}$$

### Sample standard deviation for a frequency distribution

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

### Computation formula for the standard deviation

$$s = \sqrt{\frac{SS_x}{n - 1}}$$

$$\text{where } SS_x = \frac{\sum x^2 f - (\sum xf)^2}{n}$$

$x$  is the midpoint of a class,

$f$  is the number of entries in that class,

$n$  is the total number of entries in the distribution, and  $n = \sum f$

The summation  $\sum$  is over all classes in the distribution.

