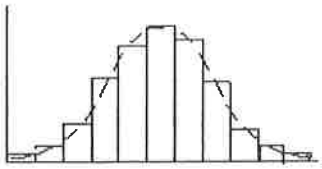
## Histograms and the Normal Distribution

A histogram that has symmetric bell-shape is also called a Normal Distribution
See how a curve drawn through the bars makes a bell shape?



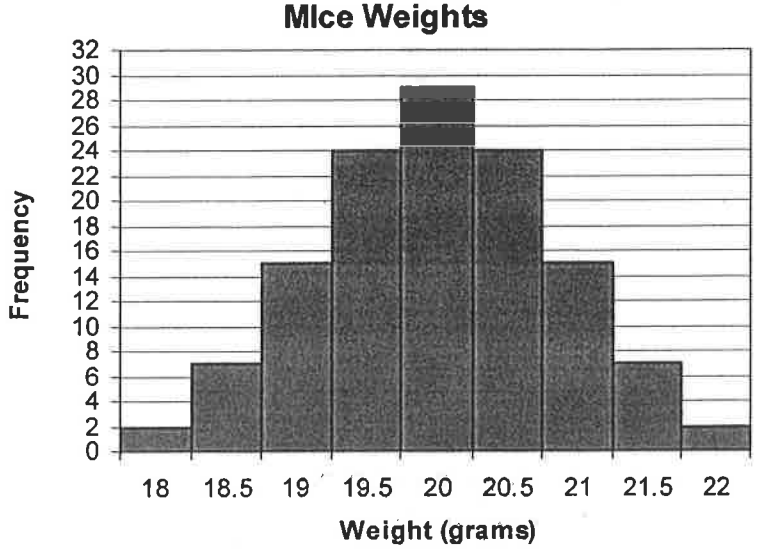
If data is "normally distributed" then we can use the symmetry and the shape of the graph to help us answer many other questions.

1. The highest bar on any normal curve is the mean, median, and mode of the data set!
2. We will use the symmetry to find what percent of data fall in certain intervals of data

Mean, Median, and Mode

Lab "A" measured the weights of 125 mice in grams. The histogram shows the distribution.

| Weights | Frequency |
|---|---|
| 18.0-18.5 | 4 |
| 18.5-19.0 | 8 |
| 19.0-19.5 | 12 |
| 19.5-20.0 | 16 |
| 20.0-20.5 | 20 |
| 20.5-21.0 | 16 |
| 21.0-21.5 | 12 |
| 21.5-22.0 | 8 |
| 22.0-22.5 | 4 |



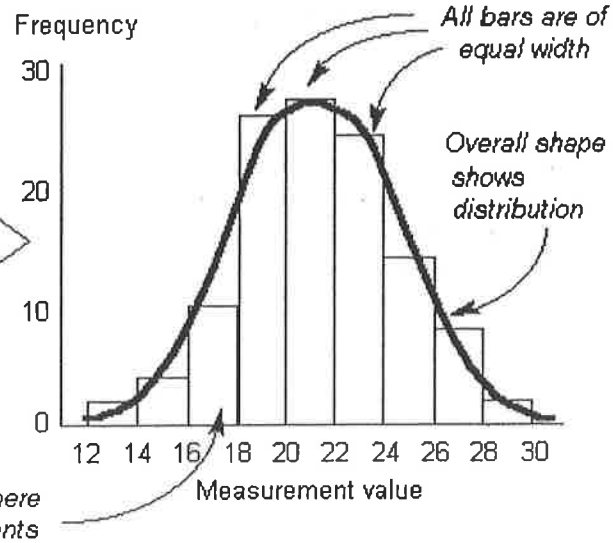

The weights appear to be "normally distributed" since most of the mice weigh about twenty grams, and fewer and fewer mice weight a lot more or a lot less. In this lab the weights have mean = 20 grams, median = 20 grams, and the mode = 20 grams

The shapes of the bars do not have to make a perfect bell-shape to be considered "Normal". Lab "B's" mice are shown below - it is approximately normally distributed.

Set of measurements

| | | |
|---|---|---|
| 21.2 | 20.5 | 19.9 |
| 23.1 | 21.4 | 18.1 |
| 14.2 | 23.9 | 12.2 |
| 25.1 | 29.2 | 26.6 |
| 16.1 | 23.1 | 23.8 |
| 23.6 | 22.5 | 22.0 |
| 20.7 | 21.1 | 15.6 |
| 19.5 | 20.4 | 21.1 |
| | etc. | |

Grouped into bars



**Theorem** **The Empirical Rule**

For a distribution that is symmetrical and normally distributed:

1. Approximately 68.2% of all data values will lie within one standard deviation on either side of the mean.
2. Approximately 95.4% of all data values will lie within two standard deviations on either side of the mean.
3. Approximately 99.7% of all data values will lie within three standard deviations on either side of the mean.

These percentages are based on the amount of **area under the normal curve** between the locations specified in the empirical rule. In theory, the curve extends infinitely in both directions, and the total area under the whole curve is 1 (100%). Because the curve is **symmetrical**, these percentages will be split evenly on either side of the mean. Figure 8-15 shows these areas and values.

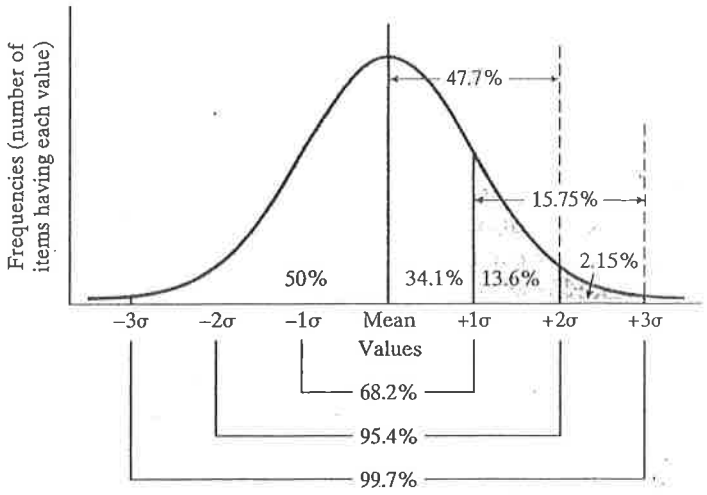



**FIGURE 8-15**
Standard normal curve.

**Definition** ***z*-Score (Standard Score)**

A *z*-score is the number of standard deviations ($s$) that a particular piece of data ($x$) is from the mean average ($\bar{x}$) for the set of data. If a *z*-score is positive, then the piece of data in question is above the mean average, and if the *z*-score is negative, then it is below the mean average. A *z*-score is calculated as follows:

$$z = \frac{x - \bar{x}}{s}$$

(See Figure 8-16.)

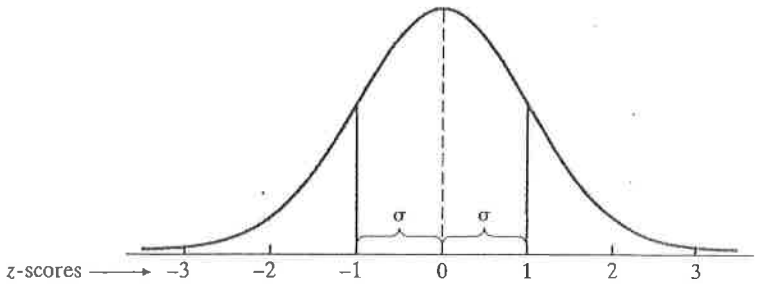



**FIGURE 8-16**
Normal curve and *z*-scores.