Mean, Variance, and Standard Deviation

Mean Value:

	Population	Sample
List of values	$\mu = \frac{\sum x}{N}$	$\bar{x} = \frac{\sum x}{n}$
Frequency tables	$\mu = \frac{\sum x f}{\sum f}$	$\bar{x} = \frac{\sum x f}{\sum f}$

Standard Deviation:

The *variance* is **defined** by the formulas:

	Population	Sample
List of values	$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$	$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$
Frequency tables	$\sigma^2 = \frac{\sum (x-\mu)^2 f}{\sum f}$	$\sigma^2 = \frac{\sum (x-\mu)^2 f}{\sum f - 1}$

We also have the following easier to **compute** formulas:

	Population	Sample
List of values	$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}$	$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$
Frequency tables	$\sigma^2 = \frac{\sum x^2 f - \frac{(\sum x f)^2}{\sum f}}{\sum f}$	$s^2 = \frac{\sum x^2 f - \frac{(\sum x f)^2}{\sum f}}{\sum f - 1}$

The **standard deviation** is the square root of the variance:

$$\sigma = \sqrt{\sigma^2} \qquad s = \sqrt{s^2}$$

Discrete Random variables

The mean or expected value

$$\mu = E(X) = \sum x\, p(x)$$

The variance and the standard deviation

The definition of the variance

$$\sigma^2 = \sum (x-\mu)^2\, p(x)$$

The formula that's easier to compute

$$\sigma^2 = \left(\sum x^2 p(x)\right) - \mu^2$$

The standard deviation is the square root of the variance:

$$\sigma = \sqrt{\sigma^2}$$

Probability Formulas

Union (<u>or</u>) formula:

$$p(A \textbf{ or } B) = p(A) + p(B) - p(A \textbf{ and } B)$$

If A, B are *mutually exclusive*, then

$$p(A \textbf{ or } B) = p(A) + p(B)$$

Intersection (<u>and</u>) formula:

$$p(A \textbf{ and } B) = p(A) \cdot p(B \textbf{ given } A)$$

If A, B are *independent*, then

$$p(A \textbf{ and } B) = p(A) \cdot p(B)$$

Complement (<u>not</u>) formula:

$$p(\textbf{not } A) = 1 - p(A)$$

Events A, B are *independent* if:

$$p(A \textbf{ and } B) = p(A)p(b)$$

$$p(A \textbf{ given } B) = p(A)$$

$$p(B \textbf{ given } A) = p(B)$$

If these probabilities are not equal then they are *not* independent.

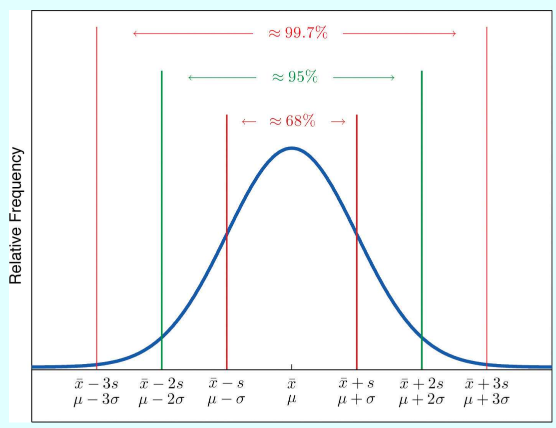
z–scores, Empirical Rule, and Chebyshev's theorem

Population	**Sample**
$z = \dfrac{x-\mu}{\sigma}$	$z = \dfrac{x-\bar{x}}{s}$
$x = \bar{x} + z\,s$	$x = \bar{x} + z\,\sigma$

Empirical Rule

For approximately <u>bell shaped</u> data sets. we have *that approximately*

68%	of z–scores is between	-1 and 1
95%	of z–scores is between	-2 and 2
99.7%	of z–scores is between	-3 and 3



The Empirical Rule.

Chebyshev's Theorem

For any data set, a proportion of *at least* $\dfrac{k^2-1}{k}$ of z–scores is between $-k$ and k.

For $k = 2$ and $k = 3$ we have *at least*:

75%	of z–scores is between	-2 and 2
88.89%	of z–scores is between	-3 and 3