Chebyshev's Theorem: The fraction of data values, *x*, within *k* standard deviations of the mean is at least $1-\frac{1}{k^2}$.

Here's why:

Let s be the standard deviation for the data set, and therefore s^2 the variance. Also, let the number of data values, *x*, with $|x-\bar{x}|\geq ks$ be n_1, and let the number of data values, *x*, with $|x-\bar{x}|<ks$ be n_2.

It follows that $n_1+n_2=n$, and that the fraction of values, *x*, within *k* standard deviations of the mean is $\frac{n_2}{n}$, while the fraction of values, *x*, greater than or equal to *k* standard deviations of the mean is $\frac{n_1}{n}$.

So we have that $s^2=\frac{\sum(x-\bar{x})^2}{n-1}=\underbrace{\frac{\sum(x-\bar{x})^2}{n-1}}_{\text{where } |x-\bar{x}|\geq k\sigma}+\underbrace{\frac{\sum(x-\bar{x})^2}{n-1}}_{\text{where } |x-\bar{x}|<k\sigma}$. This means that $s^2\geq\underbrace{\frac{\sum(x-\bar{x})^2}{n-1}}_{\text{where } |x-\bar{x}|\geq k\sigma}\geq\frac{k^2s^2n_1}{n-1}>\frac{k^2s^2n_1}{n}$. So we can conclude that $\frac{k^2s^2n_1}{n}\leq s^2\Rightarrow\frac{n_1}{n}\leq\frac{1}{k^2}$. From this we get that $1-\frac{n_1}{n}\geq 1-\frac{1}{k^2}\Rightarrow\frac{n_2}{n}\geq 1-\frac{1}{k^2}$, or that the fraction of data values, *x*, within *k* standard deviations of the mean is at least $1-\frac{1}{k^2}$.

Example: For the data set

x
1
1
1
1
1
1
20
1
2
1

the mean is 3, and the standard deviation is 5.98. According to Chebyshev's Theorem, at least $1-\frac{1}{2^2}=\frac{3}{4}$ or 75% of the values must be within 2 standard deviations of the mean. The theorem predicts that at least 75% of the values in the data set must fall in the interval from

-8.96 to 14.96. For this data set 9 out of the 10 values fall into this interval, which means that the actual percentage is 90%. Chebyshev's Theorem is correct in that 90% is greater than or equal to 75%, but 90% is a lot more than 75%.