Formula Sheet for Exam 1 (Chapters 1 and 2)

For a relative frequency table:

Relative Frequency = $\frac{Frequency}{n}$

Class Width $\approx \frac{Range}{number\ of\ classes}$

To Calculate the Mean: $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$

To Calculate the Median:
Arrange the sample data from smallest to largest.

- If n is odd, M is the middle number
- If n is even, M is the mean of the two middle numbers

To Calculate the Variance: $s^2 = \frac{n\sum x^2 - \left(\sum x\right)^2}{n(n-1)}$

To Calculate the Standard Deviation: $s = \sqrt{s^2} = \sqrt{\frac{n\sum x^2 - \left(\sum x\right)^2}{n(n-1)}}$

Theorems Relating to Distributions: the following two Theorems can be used to determine what range of data values we can expect from a given distribution, and it can also be used to determine **what percent** of the data will lie within K standard deviations from the mean:

Chebyshev's Theorem: The proportion of any set of data lying within K standard deviations of the mean is always at least $1-\frac{1}{K^2}$, where K > 1. **Note: $K = \frac{UL-\mu}{\sigma}$

Steps to use Chebyshev's Theorem:
1) Write down your mean, standard deviation, and the given interval (or remaining value, if there is no interval given).
2) Confirm the given interval is symmetric around the mean (or create a symmetric interval around your mean).
3) Get K using the upper limit of your interval.
4) Plug K into: "at least" $\left(1-\frac{1}{K^2}\right)*100\%$
5) The answer above is your solution if you are working to find the percent of data inside an interval, but if you want the amount above or below the given interval, subtract the result from step 4 from 100% and change "at least" to "at most".

Empirical Rule

Steps to using the Empirical Rule:

1. Confirm the problem states the data is mound and symmetric, bell shaped, or normally distributed.
2. Draw a bell curve
3. Put the mean in the center of the curve, and place the interval limits on the curve as well.
4. Then determine how many standard deviations those limits are from the mean.
5. Use the list below to determine the area bounded by the given interval.

   Approximately 68% of the data lies within 1 standard deviation of the mean.
   Approximately 95% of the data lies within 2 standard deviations of the mean.
   Approximately 99.7% of the data lies within 3 standard deviations of the mean.

Z-scores

Z-scores can be used to determine if a given value is unusual, and Z-scores can be used to make relative comparisons between values form different data sets.

Z-scores: $Z = \frac{x-\mu}{\sigma}$