Reminders
Descriptive Statistics, Inferential Statistics
Quantitative, Qualitative Data
(Graphical Methods for Quantitative and Qualitative Data)
Mean, Median, Standard Deviation, Variance

Some things I forgot to mention last time
Effects of Linear Transformation on Mean, Median, Standard Deviation, Variance

Section 5: Using the Mean and Standard Deviation to Describe Data

Chebyshev’s Theorem
Applies to any shape data set
In general, for \( k \geq 1 \), at least \( 1 - \frac{1}{k^2} \) of the data lies in the interval \( \bar{x} - ks \) to \( \bar{x} + ks \).
No useful information about the fraction of data in the interval \( \bar{x} - s \) to \( \bar{x} + s \) since it concludes that at least 0% of the data lies in the interval.
At least \( 1 - \frac{1}{k^2} = \frac{3}{4} \) of the data lies in the interval \( \bar{x} - 2s \) to \( \bar{x} + 2s \)
At least \( 1 - \frac{1}{k^2} = \frac{8}{9} \) of the data lies in the interval \( \bar{x} - 3s \) to \( \bar{x} + 3s \)

Empirical Rule
Applies to data sets that are mound shaped and symmetric
Approximately 68% of the measurements lie in the interval \( \bar{x} - s \) to \( \bar{x} + s \)
Approximately 95% of the measurements lie in the interval \( \bar{x} - 2s \) to \( \bar{x} + 2s \)
Approximately 99.7% of the measurements lie in the interval \( \bar{x} - 3s \) to \( \bar{x} + 3s \)

Show how range/4 is a (somewhat) good first estimate for standard deviation

Section 6: Numerical Measures of Relative Standing

The 100p\(^{th}\) percentile is a number such that about 100p% of the data falls at or below it and about 100(1 - p)% falls at or above it. We denote the number by \( x_p \).
25\(^{th}\) percentile = Lower Quartile (First Quartile) \( x_{0.25} \) or \( Q_L \)
50\(^{th}\) percentile = Median (Second/Middle Quartile) \( x_{0.5} \) or \( Q_M \)
75\(^{th}\) percentile = Upper Quartile (Third Quartile) \( x_{0.75} \) or \( Q_U \)

Give on of those "how to interpret a percentile" problems: If you scored such and such on the SAT

The sample \( z \)-score for a measurement is \( z = \frac{x - \bar{x}}{s} \).
The population \( z \)-score for a measurement is \( z = \frac{x - \mu}{\sigma} \).
Interpretation of \( z \)-Scores for Mound-Shaped Distributions of Data:
Approximately 68% of the measurements will have a \( z \)-score between -1 and 1.
Approximately 95% of the measurements will have a \( z \)-score between -2 and 2.
Approximately 99.7% of the measurements will have a \( z \)-score between -3 and 3.