MAT123 Spring 2025 Notes

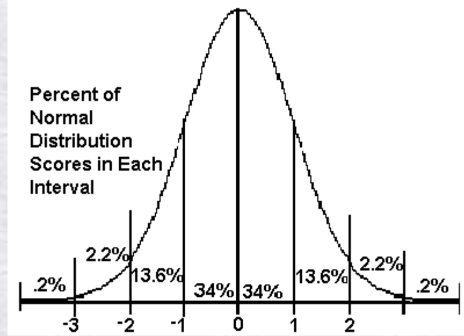
Mon, 2/3, 2.5, 2.6, 2.7

2.5 Measures of the Center of the Data
 a) Distinguish notations for population and sample statistics
 b) Discuss Law of Large Numbers
2.6 Skewness and the Mean, Median, and Mode
 a) Discuss symmetry and skewness in data
2.7 Measures of the Spread of the Data
 a) Use software to calculate and interpret measures of dispersion (Lab 2)
 b) Calculate z-scores for data
 c) Apply the Empirical Rule and Chebyshevâ€™s Inequality to data
3

Comments ...

Center & Spread
Hi,
From a middle school introduction to statistics, a student is coached on the computation of an average, mean, mode, and median as a measures of the center.
Now consider an average as "a one number representation" of the entire sample or population.
In middle school, range was the measure of how the set of numbers spreads out -- the difference between the highest and lowest -- a difference of the most extreme data points.
In more recent years, middle school statistics focused more on discrete data representation -- quartiles, inner quartile range.
Now consider spread in more specific ways. Here's a pdf of this "page" or view the video of "Center & Spread".
Go with confidence. Stay safe, A^2


Agnes (Asquared) Azzolino



TI-83 Excel Sketchpad formulas

New Words, New Symbols

- CAPITOL SIGMA -- symbol $\sum$ -- a symbol meaning "add up the terms." So, the formula for mean says, "Add up all the numbers, then divide by the number of numbers."
  Lower case letters are used for statistical symbols -- in English for a sample, in Greek for a population.
  The Greek capital letter sigma is used indicate a sum or to "add them up." The Greek lower case sigma is the symbol for a population standard deviation, a spread.
  The lower case mu is used for a population mean, a center. The English x bar is for the mean, a center of a sample.

$$\bar{x} = \frac{\sum x}{n}$$

Statistics Symbols

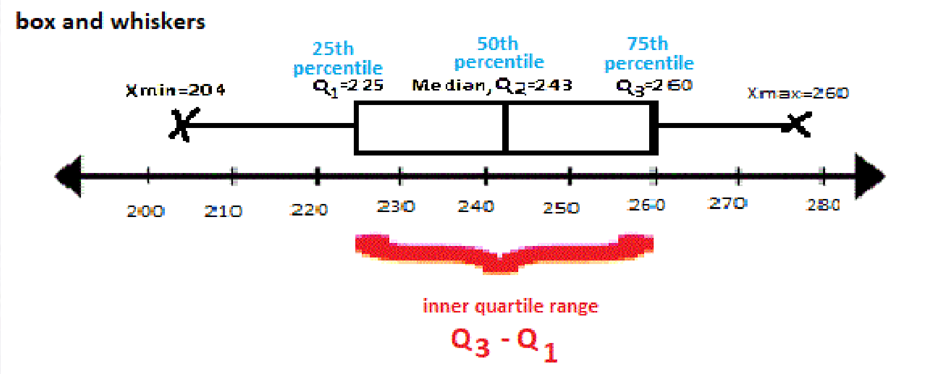
	mean	standard deviation
sample	$\bar{x}$	s
population	μ	σ

	Mean	Variance	Standard Deviation
For a Sample	$\text{mean} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
For a Population	$\text{mean} = \mu = \frac{\sum_{i=1}^{n} x_i}{n}$	$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$	$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$

- **CENTER** - "a one number representation" of the entire sample:
  AVERAGE -- any of the following three sample statistics
  MEAN -- symbol: $\bar{x}$, read as "x bar" - arithmetic average.
  MODE -- most frequent score or data point
  MEDIAN -- symbol: $\hat{x}$, read as "x hat" --median, Q_2, the 50th percentile, the middle data point when the data is ordered from lowest to highest

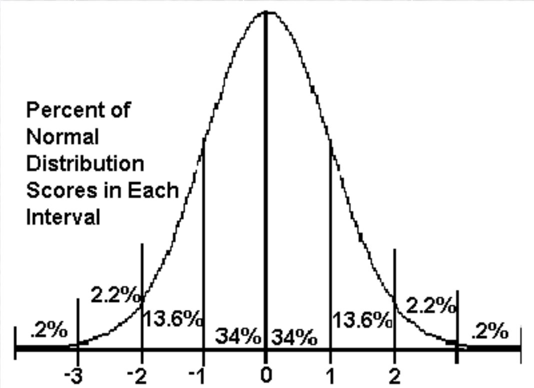
- **SPREAD** - "how the data spreads out
  RANGE - the spread of the data from the highest data point to the lowest data point, $x_{max} - x_{min}$

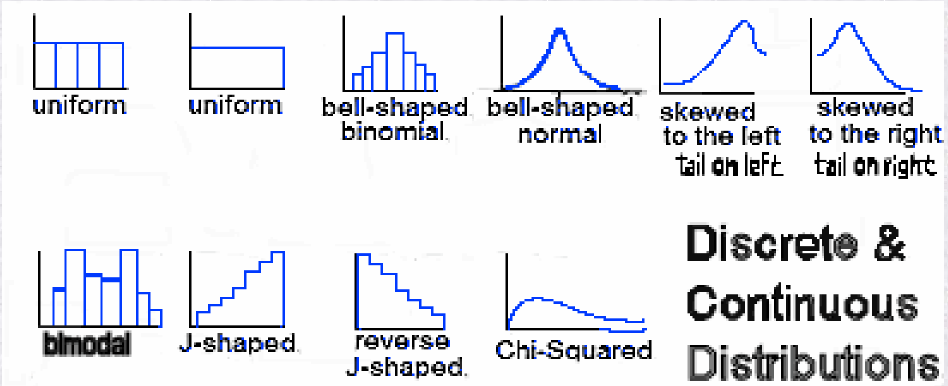
[INNER-QUARTILE RANGE](#) - the spread of the middle 50% of the data, the difference between the 3rd quartile and the 1st quartile, Q_3 - Q_1, where Q_3 is the 75th percentile and Q_1 is the 25th percentile.



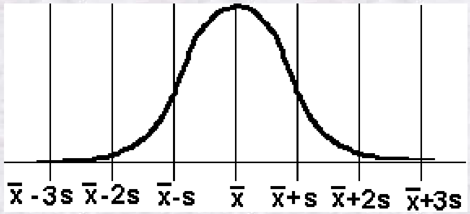
[VARIANCE](#) - the square of the standard deviation.
STANDARD DEVIATION - the average spread of the data computed in the standard way.

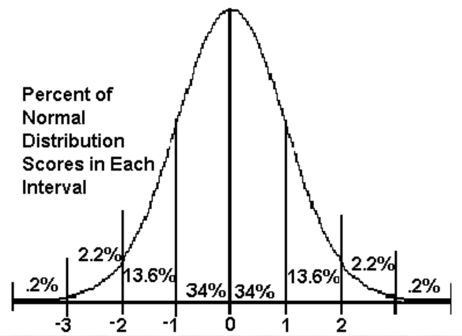
- [SYMMETRIC](#) - having a left to right (bilateral) symmetry: the mean or median is in the middle and the tails are the same length and "fattness."
- DISTRIBUTION - how the data is spread out, what shape the frequency distribution or histogram has.





- NORMAL DISTRIBUTIONS - or Gaussian distribution, a continuous probability distribution (so the area under the curve equals 1), where the mean, mode, median are all the same, so the data gathers about a center making a symmetric bell-shaped curve. Many data points -- heights of people, lengths of fish, errors measurements, standardized test scores have normal distributions.

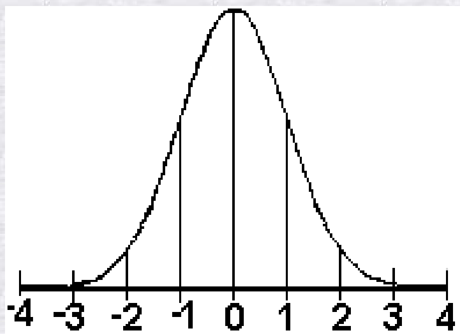


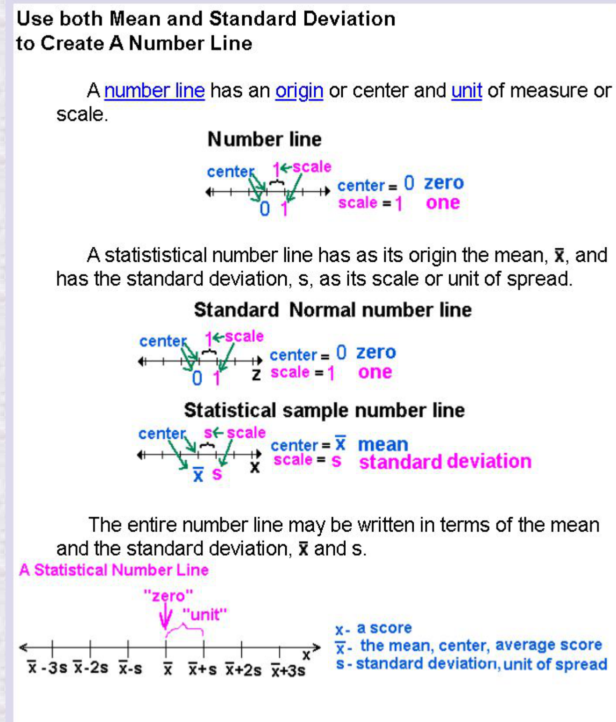


$$f(x,\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

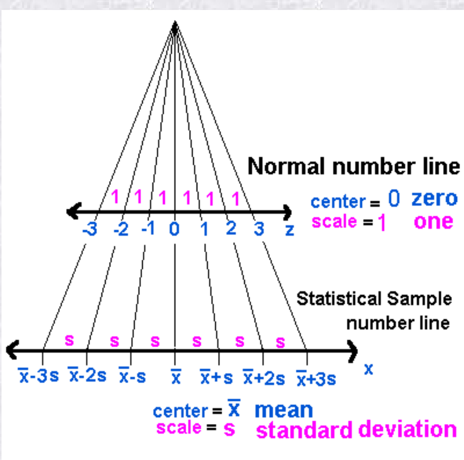
[stat123sp25.A.xls](#) - basic 1-variable & 2-variable statistics, scatterplot, regression
[sheet 6b of ReimannSums.gsp](#) -- cummulative probability distribution

- STANDARD NORMAL DISTRIBUTION - a normal distribution having a mean of 0 and a standard deviation of 1.
  It is very useful in computing, and looking up, probabilities, comparing samples and populations, and analysis and hypothesis testing.





Two very useful formulas make it easy to translate x scores into z scores or z into x. Below the calculators do the work for you. Use the percent images above and mental computation or the calculators to answer the questions below the calculators.



Compute x or z
Complete the computation by entering the values and pressing the buttons.
Enter negative two as "-2."

$z = \frac{x - \bar{x}}{s}$ = [2.5] - [3] / [1] so, [z is] [ ]

$x = \bar{x} + zs$

x = [3] +([1])([2.5])

[x =] [ ]

- CHEBYCHEV'S RULES -- for any distribution, the percent of scores within k standard deviations of the mean, k > 0, is $1/k^2$

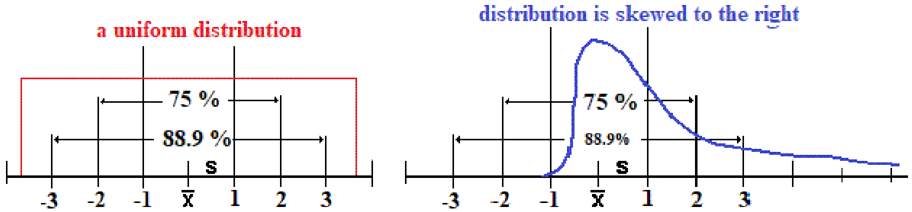
Chebychev's Rule - weaker, used for non-bell-shaped distributions

for $k > 1$ the percents of scores are at least $1 - \frac{1}{k^2}$

$k = 2$ 75 %

$k = 3$ 88.9 %

for example:



Chebyshev's Rule & Empirical Rule

___% of the data is within ...

For ANY data set, no matter what the distribution of the data is:

- At least 75% of the data is within two standard deviations of the mean.
- At least 89% of the data is within three standard deviations of the mean.
- At least 95% of the data is within 4.5 standard deviations of the mean.
- This is known as Chebyshev's Rule.

For data having a distribution that is BELL-SHAPED and SYMMETRIC:

- Approximately 68% of the data is within one standard deviation of the mean.
- Approximately 95% of the data is within two standard deviations of the mean.
- More than 99% of the data is within three standard deviations of the mean.
- This is known as the Empirical Rule.
- It is important to note that this rule only applies when the shape of the distribution of the data is bell-shaped and symmetric. We will learn more about this when studying the "Normal" or "Gaussian" probability distribution in later chapters.

Introductory Statistics

SENIOR CONTRIBUTING AUTHORS
BARBARA ILLOWSKY, DE ANZA COLLEGE
SUSAN DEAN, DE ANZA COLLEGE

openstax

OpenStax
Rice University
6100 Main Street MS-375
Houston, Texas 77005

pg. 120

"Download for free at https://openstax.org/details/books/introductory-statistics."