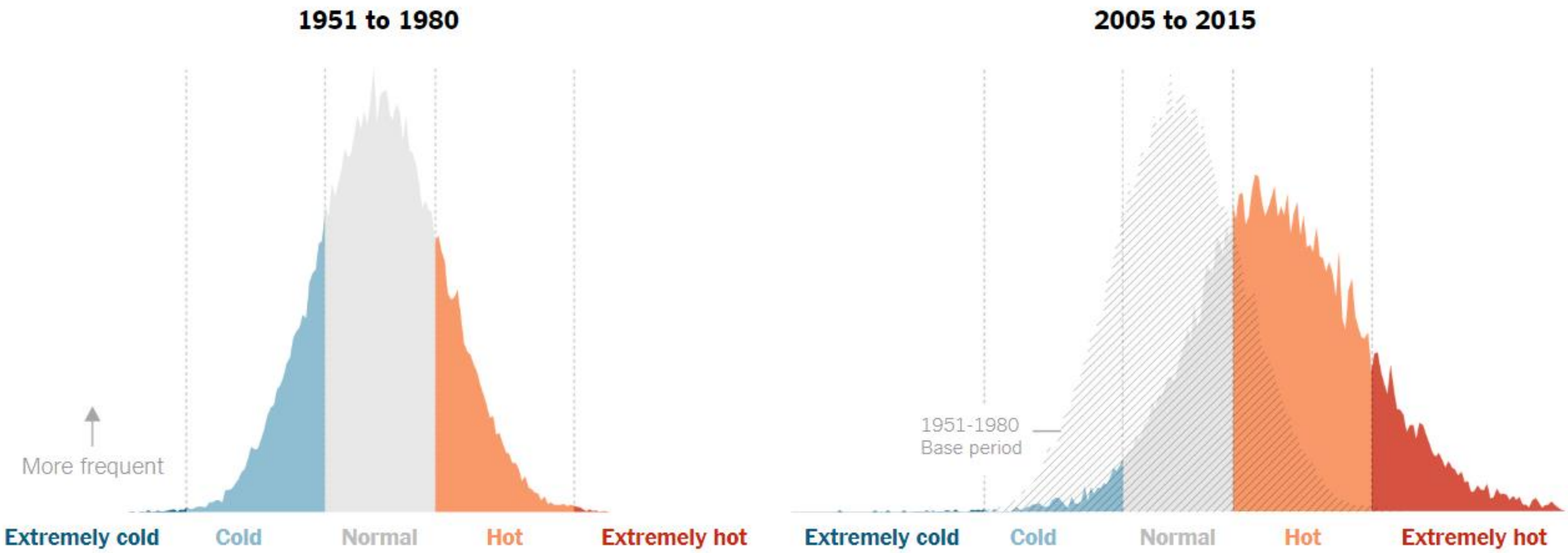
Probability and Mathematical Statistics in Data Science

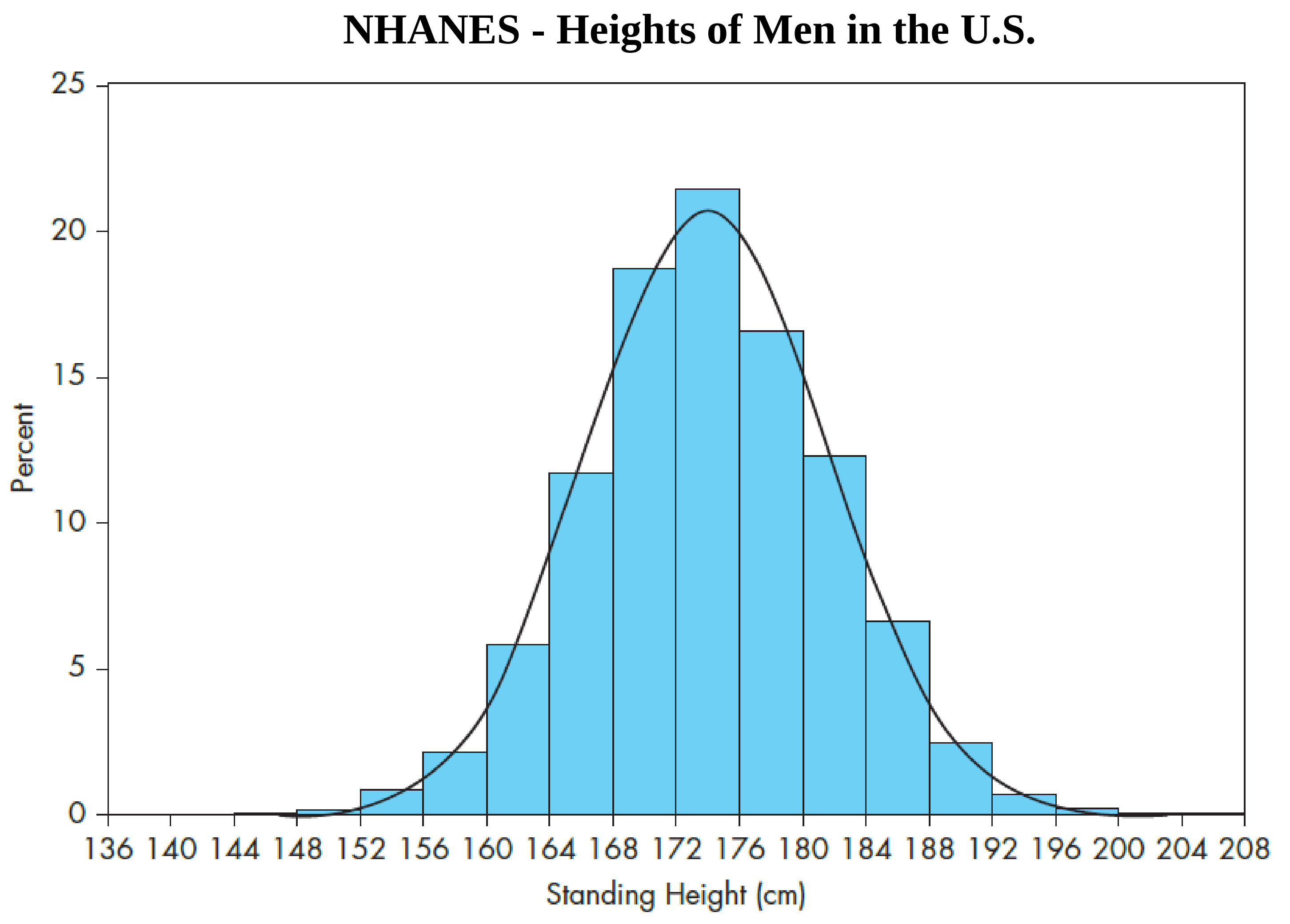
Lecture 21: Section 8.2: Standard Normal Curve

It’s Not Your Imagination. Summers Are Getting Hotter.



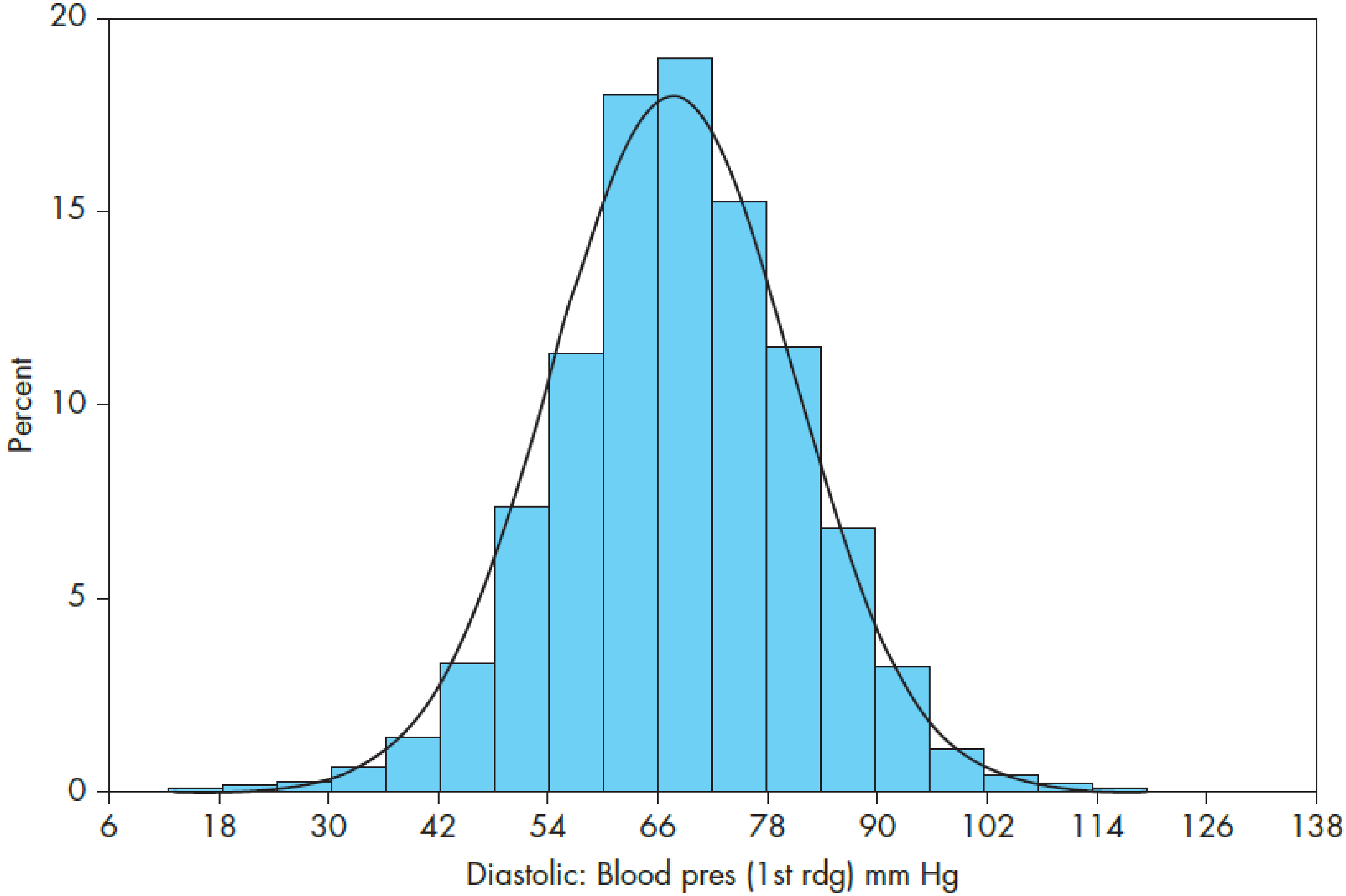
https://www.nytimes.com/interactive/2017/07/28/climate/more-frequent-extreme-summer-heat.html

Visualizing Variation in Quantitative Data



Source Data: https://www.cdc.gov

The Shape of Natural Variation: The Normal Distribution



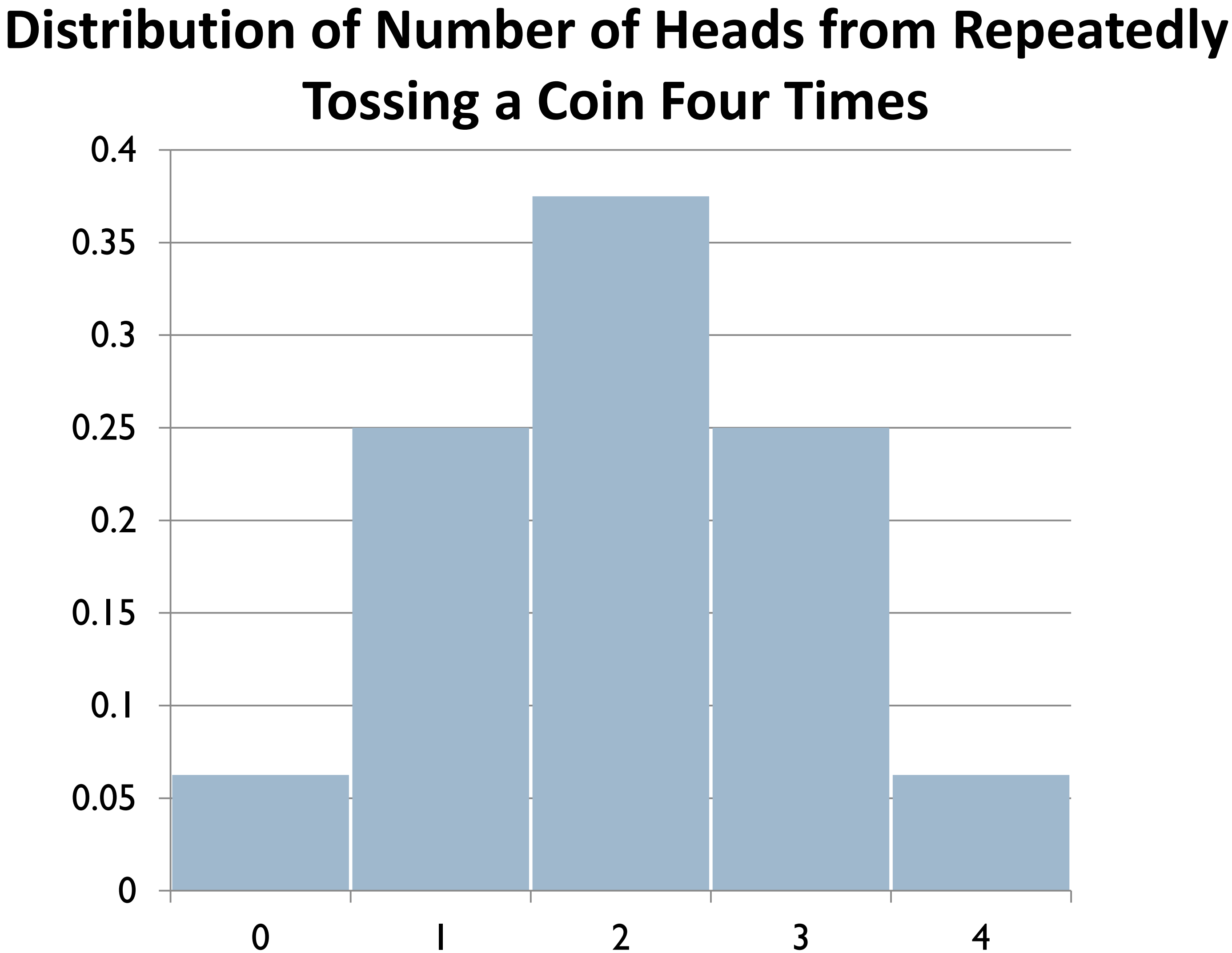
2011-2012 NHANES – Diastolic Blood Pressure

Order in Random Processes

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

—Sir Francis Galton

Order in Random Processes



Distribution of Men's Heights

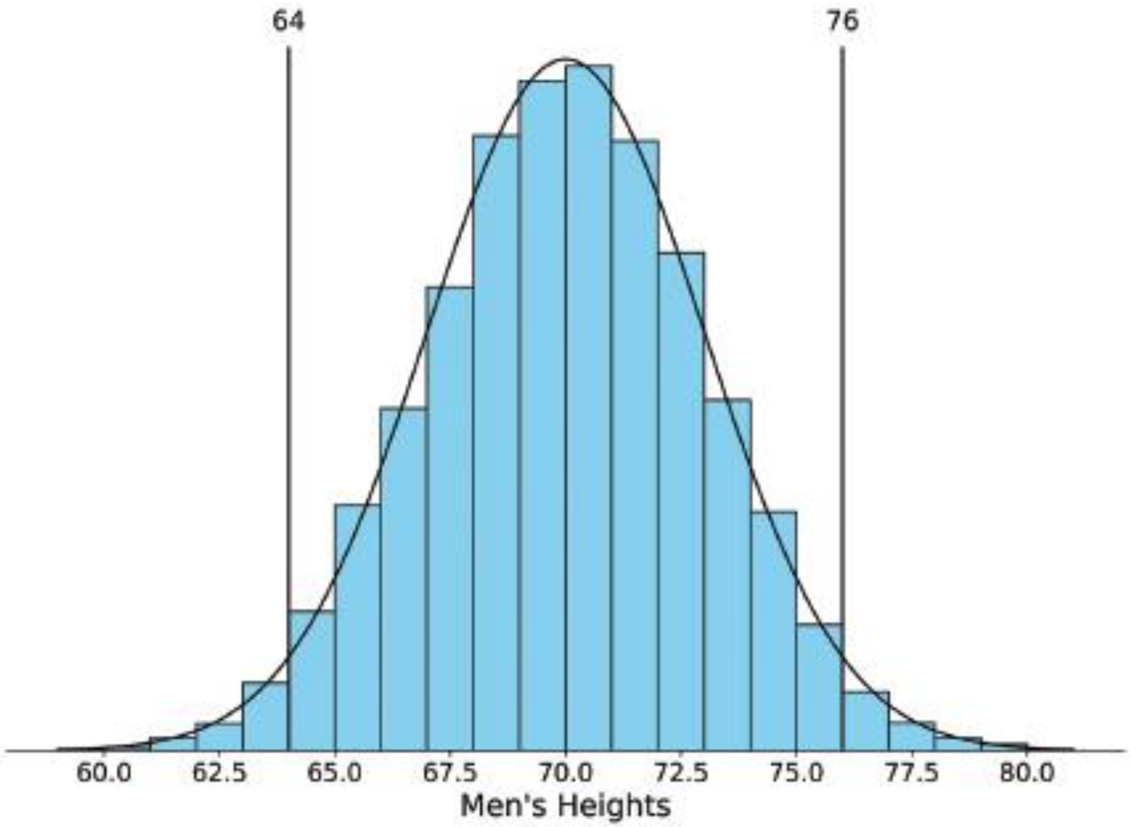


Figure 6.2: Population Distribution of Men's Heights in the United States

Sampling Distribution of Men's Heights

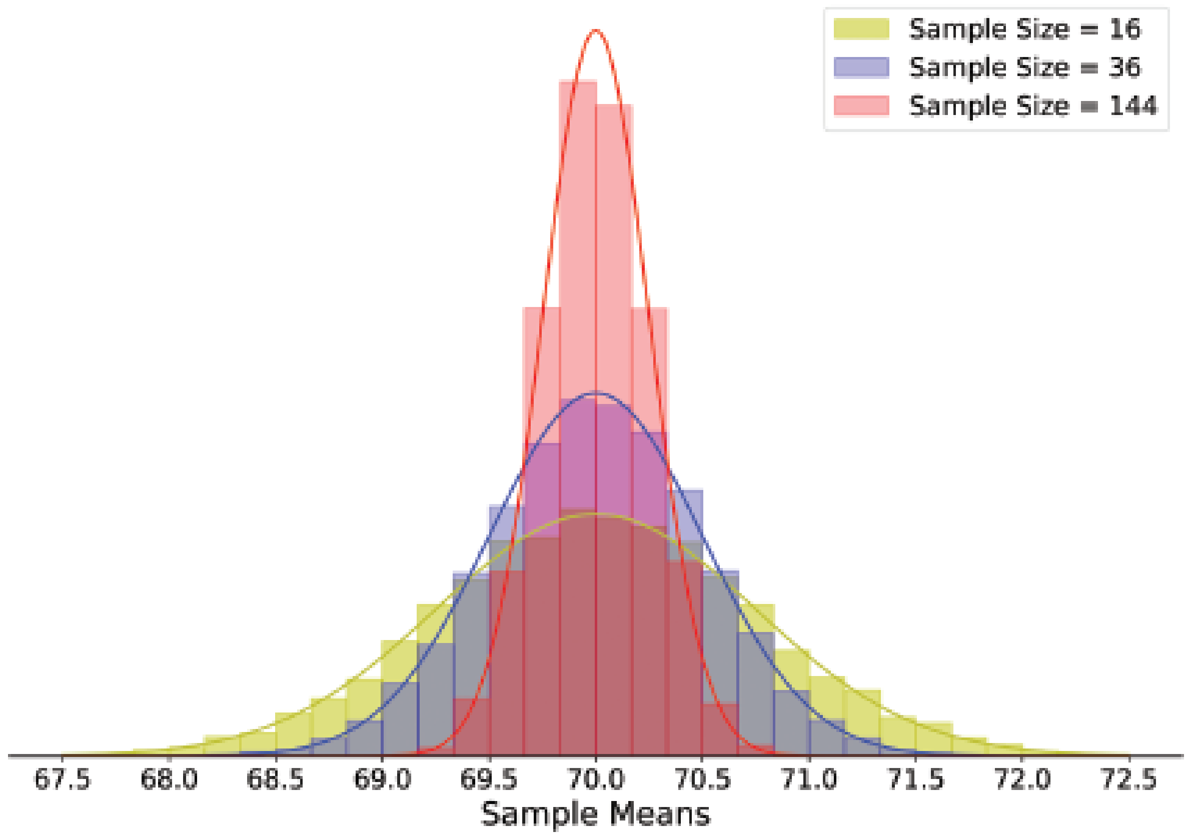


Figure 6.5: Simulation of 10,000 Sample Mean Heights—Sample Sizes Equal to 16, 36 and 144 Men

The Normal Distribution

- It's probably the most important distribution in the world!

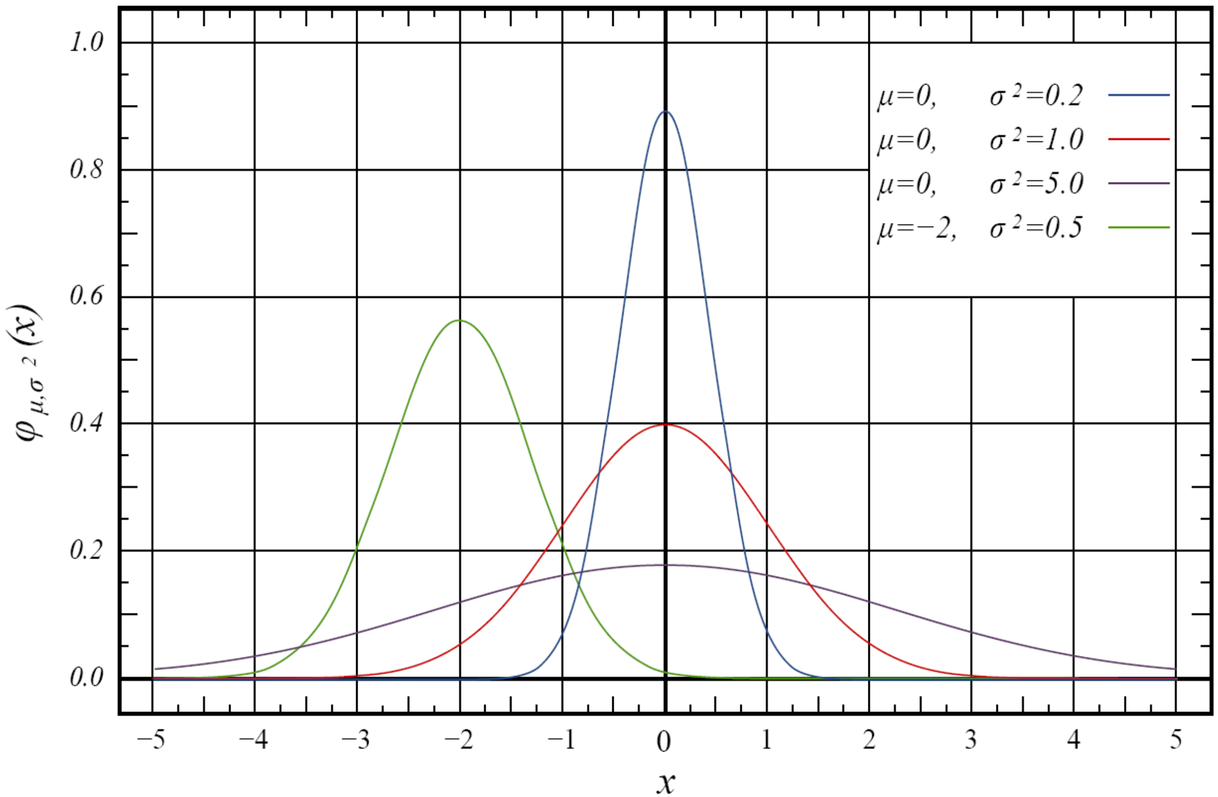
- Many numerical populations have distributions that can be fit very closely by an appropriate normal curve. (people's height/weight; testing scores; etc.)

- A continuous random variable is said to have a normal (Gaussian) distribution with parameters μ and σ, where $-\infty<\mu<\infty$, and $0<\sigma$, if the pdf of X is

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty$$

The Normal PDF

- Normal distribution is a bell-shaped, single peaked and symmetric distribution.



Parameters

- Clearly $f(x; \mu, \sigma) \geq 0$, but a somewhat complicated calculus argument must be used to verify that

$$\int_{-\infty}^{\infty} f(x; \mu, \sigma) dx = 1.$$

- Parameter μ, stands for the expected value of the normal distribution.

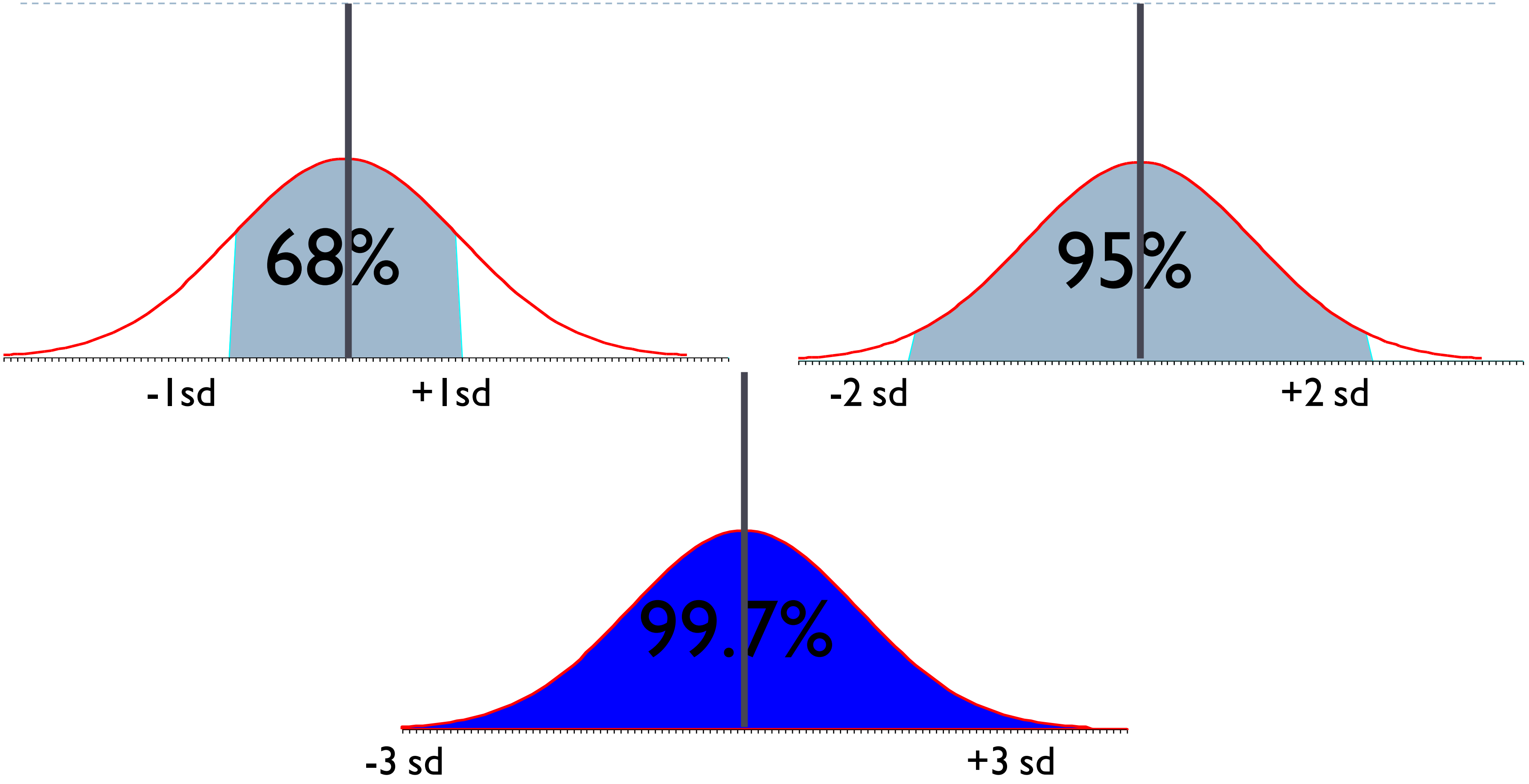
- Parameter σ, stands for the standard deviation of the normal distribution.

Characteristics of a Normal Distribution

- **Symmetric and Bell-Shaped**

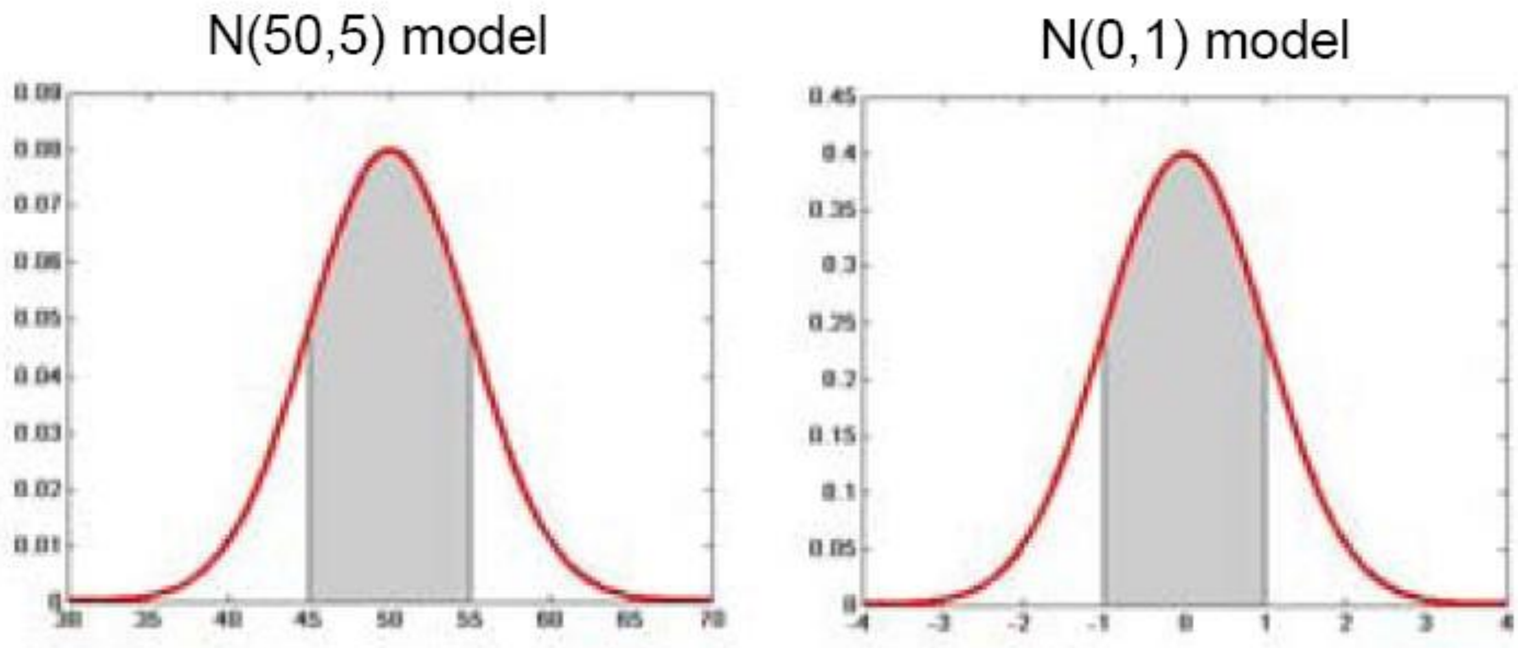
- **Empirical Rule**
- For any **normal curve**, approximately …
  - **68%** of the values fall within **1 standard deviation** of the mean in either direction
  - **95%** of the values fall within **2 standard deviations** of the mean in either direction
  - **99.7%** of the values fall within **3 standard deviations** of the mean in either direction

Empirical Rule for Any Normal Curve



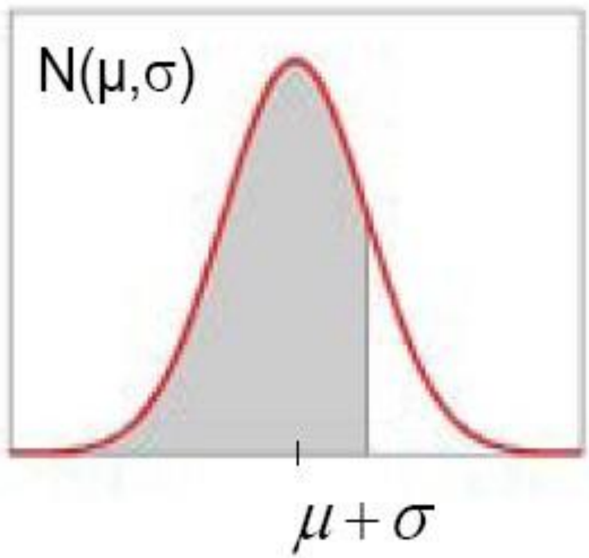
Basic Properties

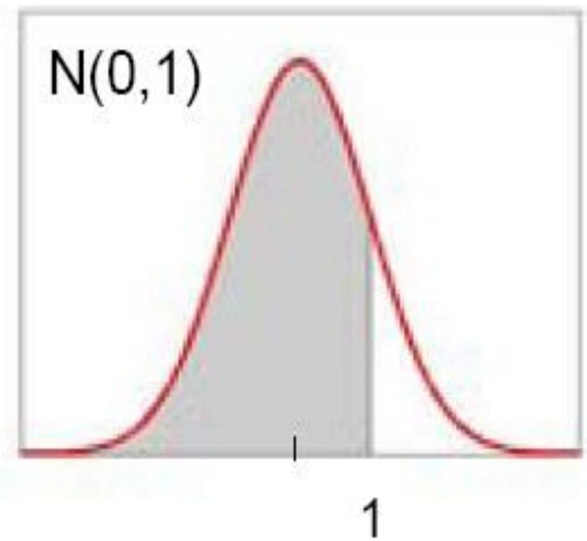
- All normal models have the same shape and the same area within *x* standard deviations of its mean.



Area between (45,55) = Area between (-1,1)

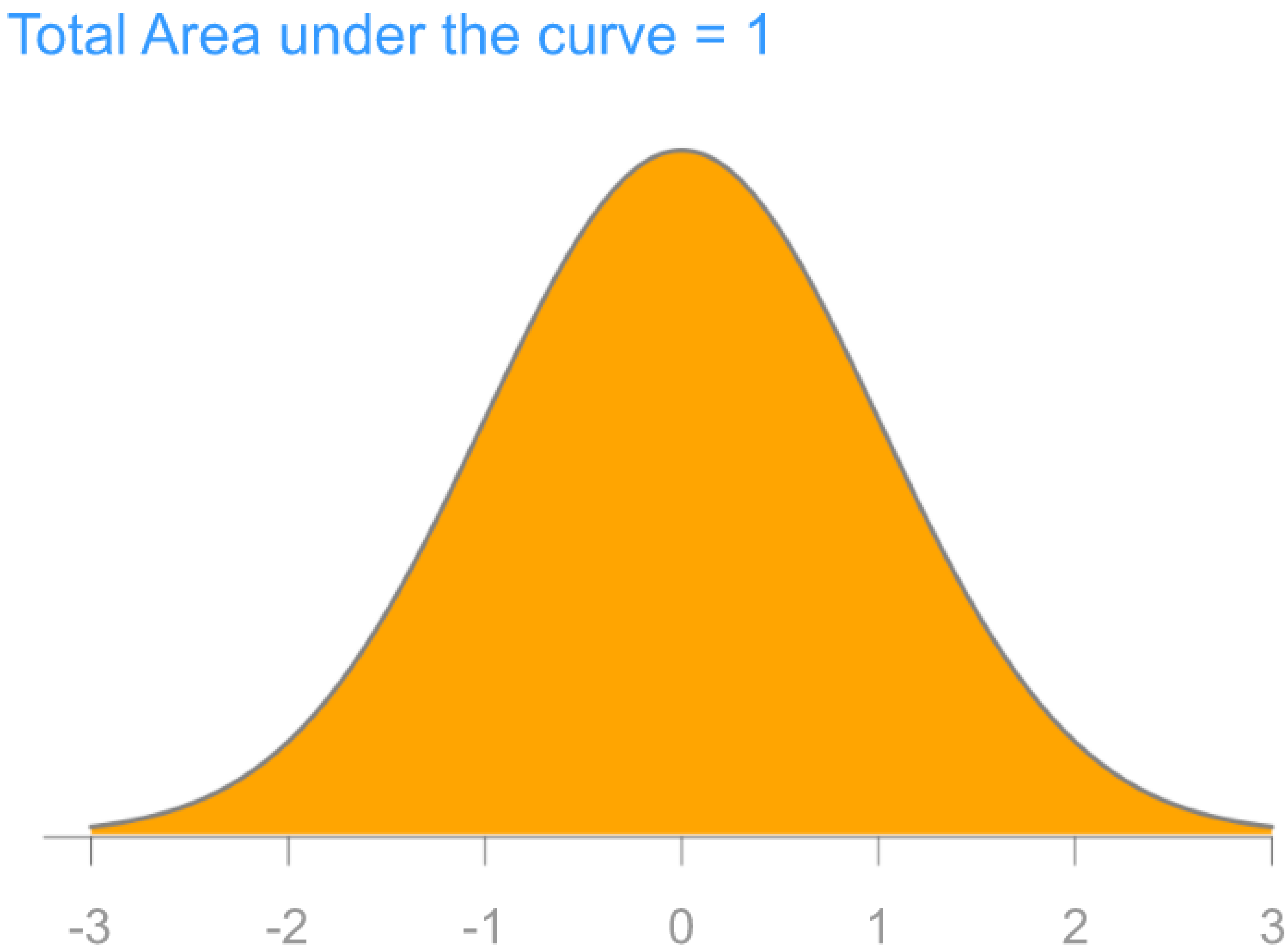
Key Result





$$area\{y < \mu + \sigma\} = area\{z < 1\}$$

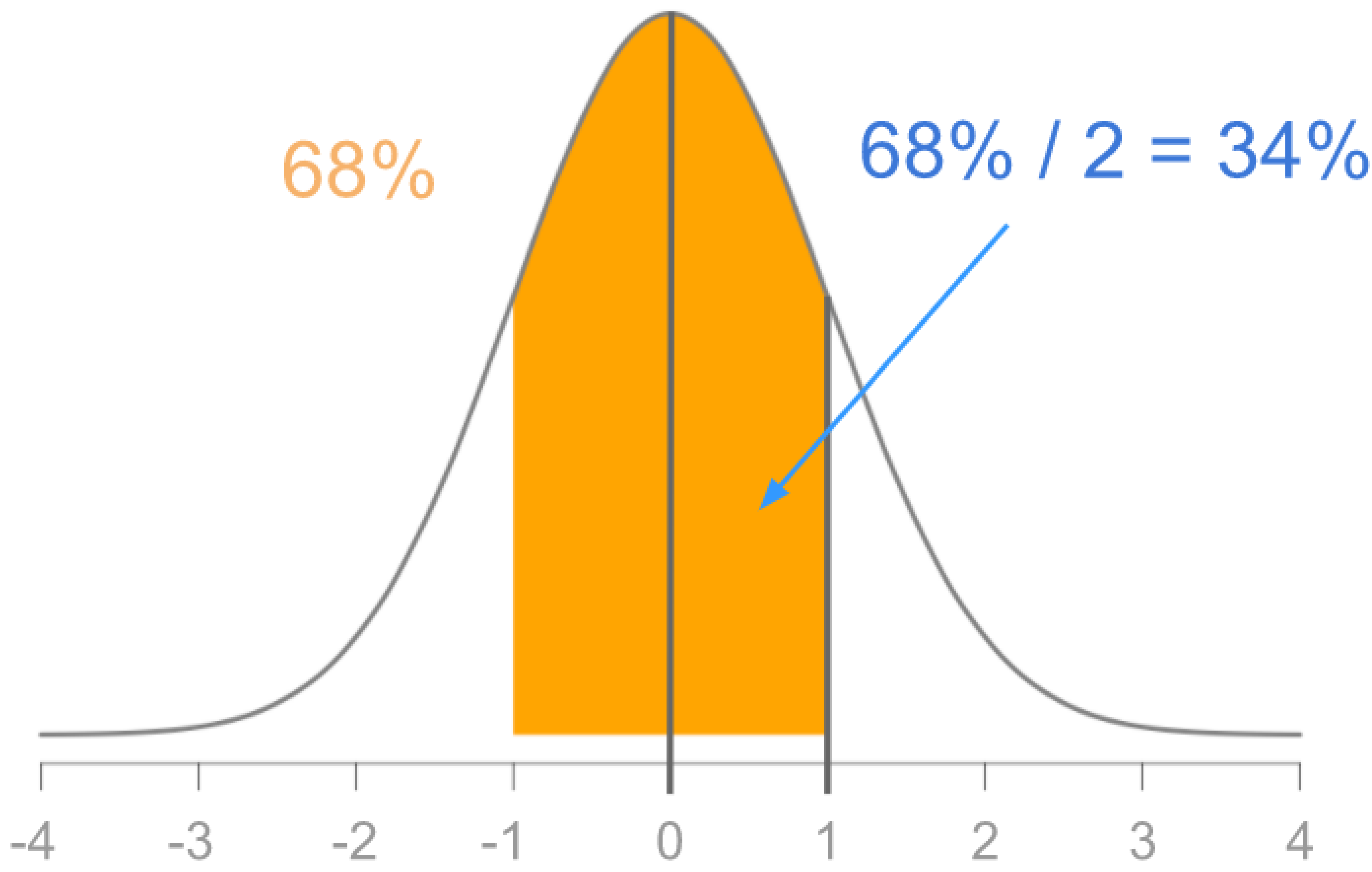
About the Standard Normal Curve



http://www.gastonsanchez.com/stat2/

About the Standard Normal Curve

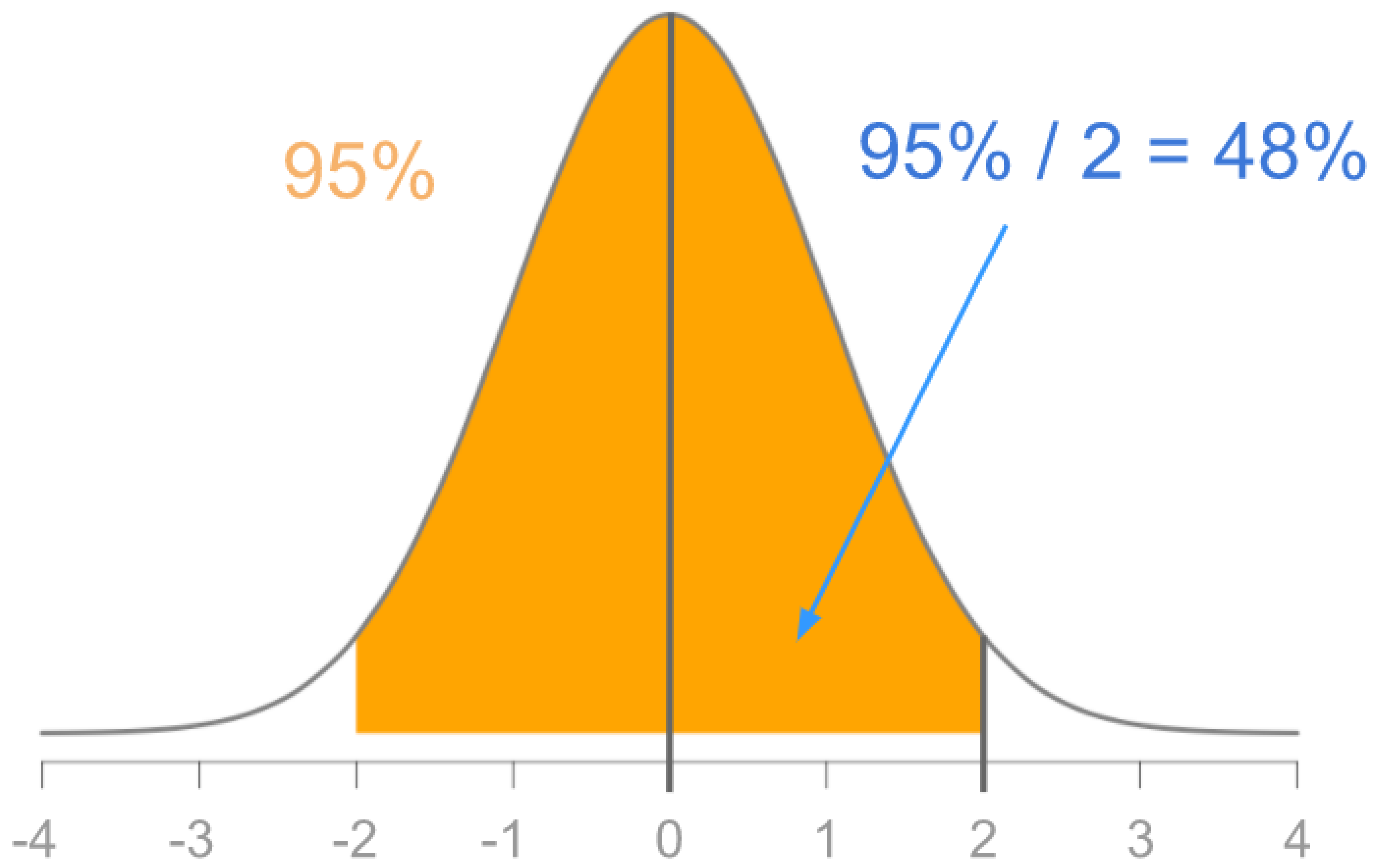
68% of area within 1 unit from 0



http://www.gastonsanchez.com/stat2/

About the Standard Normal Curve

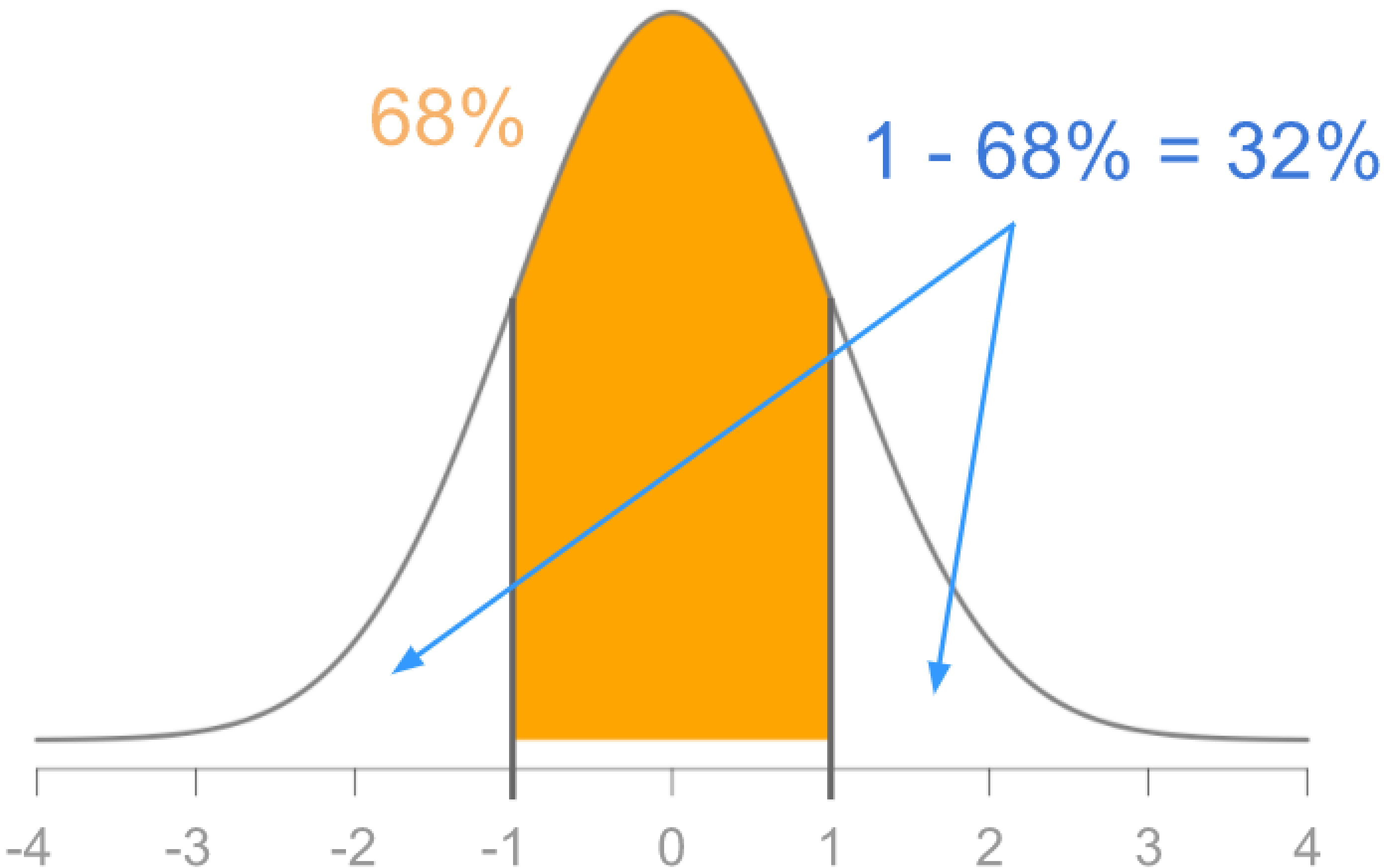
95% of area within 2 units from 0



http://www.gastonsanchez.com/stat2/

About the Standard Normal Curve

68% of area within 1 unit from 0



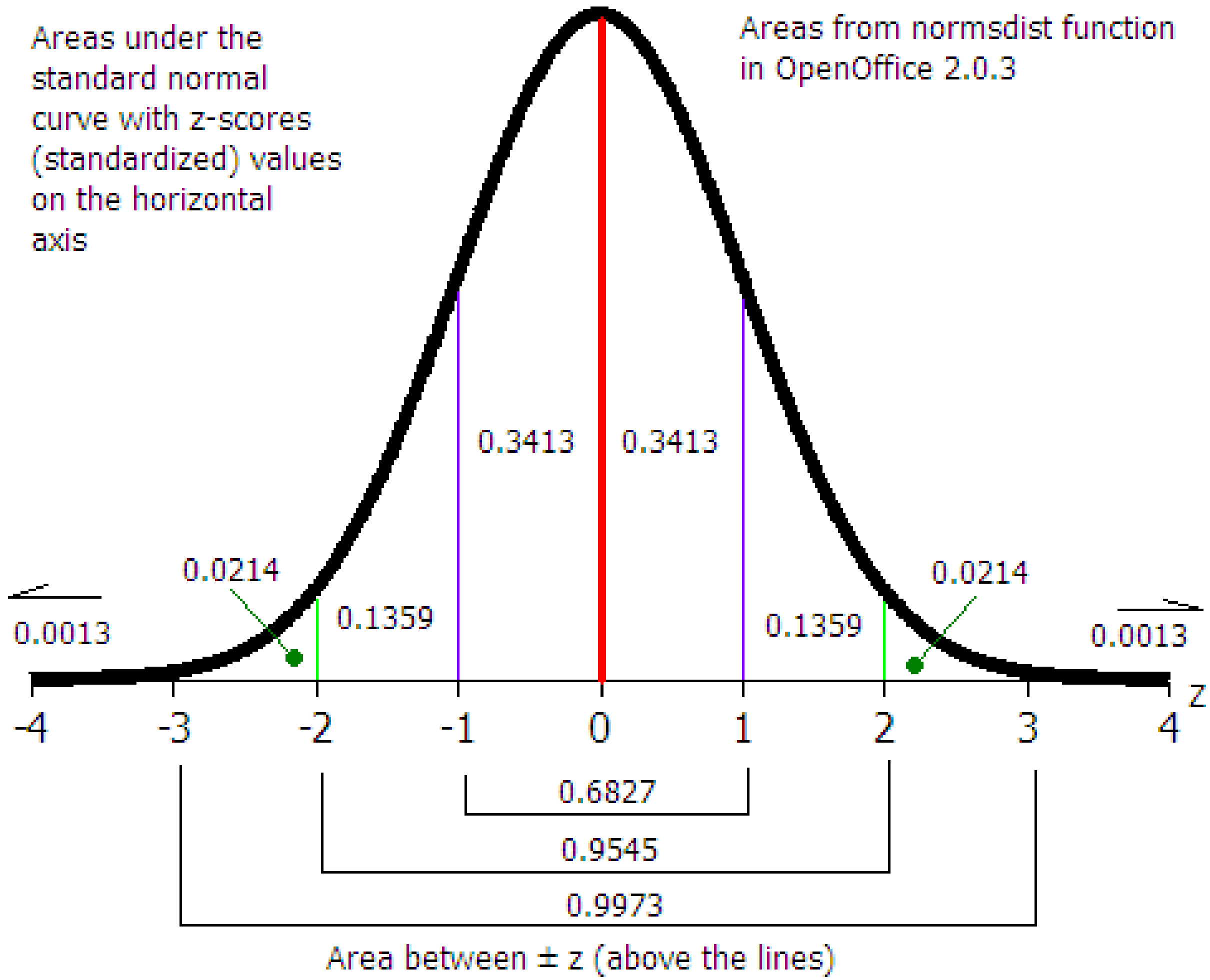
http://www.gastonsanchez.com/stat2/

The Standard Normal Curve

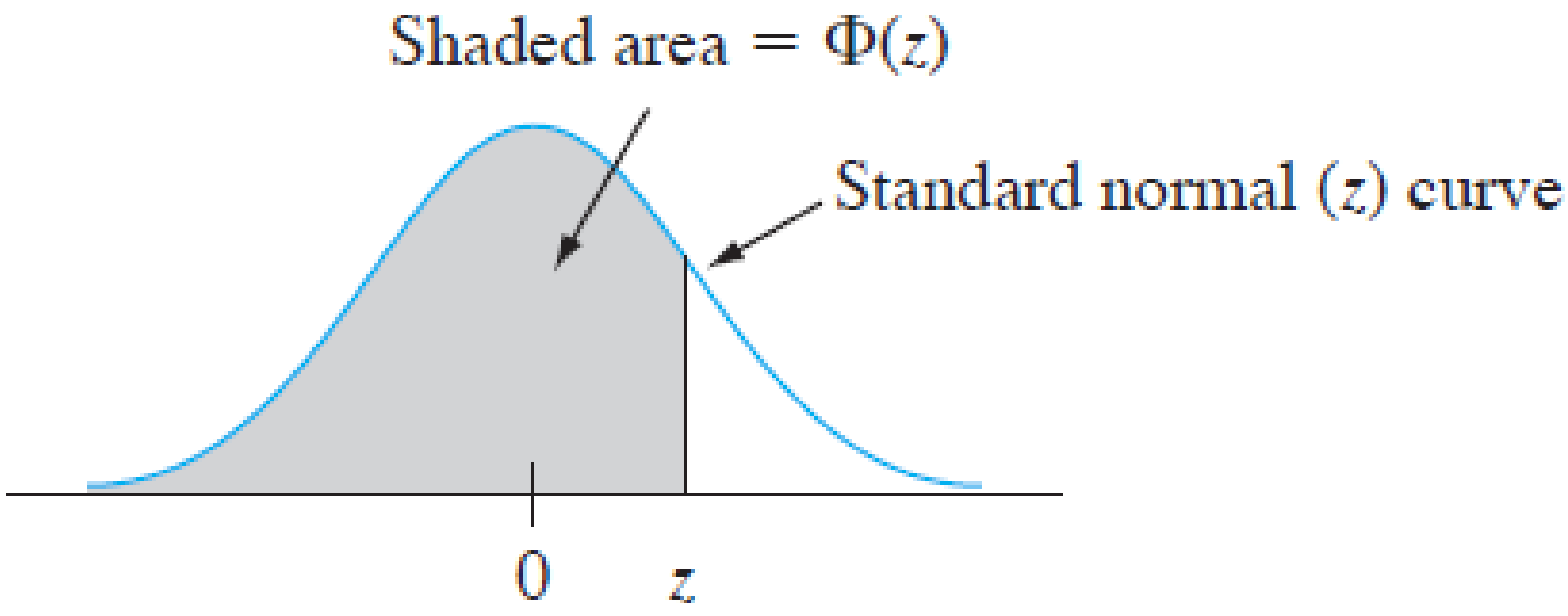
1. The standard normal curve has a mean of 0 and standard deviation of 1.

2. We convert our data values to the their corresponding values on the standard normal curve.

3. This enables us to get exact probabilities of certain events.



The Standard Normal Curve



Standard normal cdf, symmetries, percentiles

Example: Find the area

Q1.(a) to the right of 1.25

(b) between -0.3 and 0.9

(c) outside -1.5 and 1.5.

Q2. (a) Find z such that $\Phi(z) = 0.95$

(b) Find z so that the area in the middle is 0.95 ($\Phi(z) - \Phi(-z) = 0.95$)