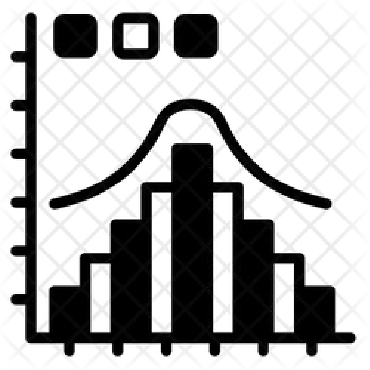
Normal Distribution

Tushar B. Kute,
http://tusharkute.com



Continous Distribution

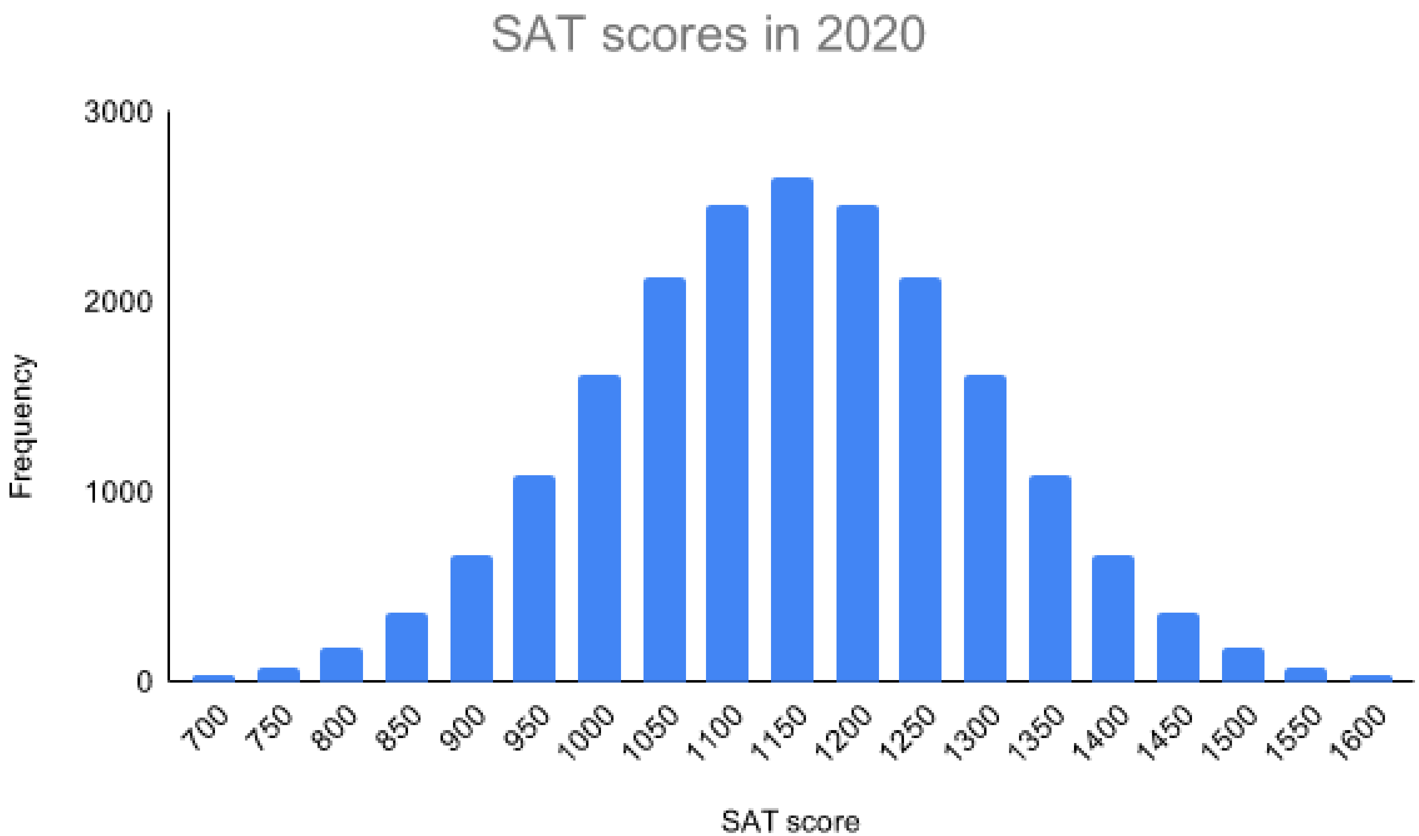
- Probability distributions are either continuous probability distributions or discrete probability distributions.
- A continuous distribution has a range of values that are infinite, and therefore uncountable.
- For example, time is infinite: you could count from 0 seconds to a billion seconds…a trillion seconds…and so on, forever. A discrete distribution has a range of values that are countable.
- For example, the numbers on birthday cards have a possible range from 0 to 122 (122 is the age of Jeanne Calment the oldest person who ever lived).

Normal Distribution

- In a normal distribution, data is symmetrically distributed with no skew.
- When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center.
- Normal distributions are also called Gaussian distributions or bell curves because of their shape.

Ref: https://www.scribbr.com

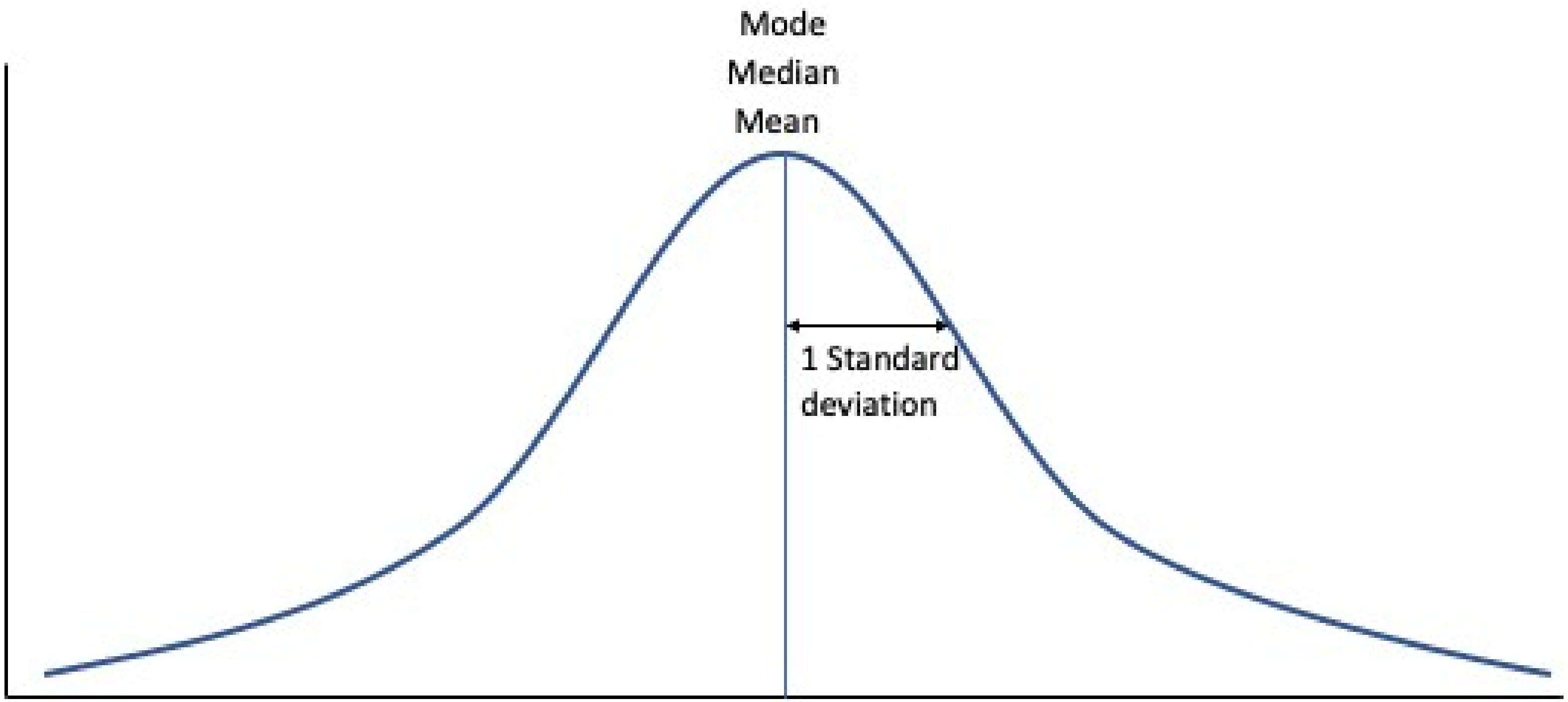
Normal Distribution



Normal Distribution

- Normal distributions have key characteristics that are easy to spot in graphs:
  - The mean, median and mode are exactly the same.
  - The distribution is symmetric about the mean—half the values fall below the mean and half above the mean.
  - The distribution can be described by two values: the mean and the standard deviation.

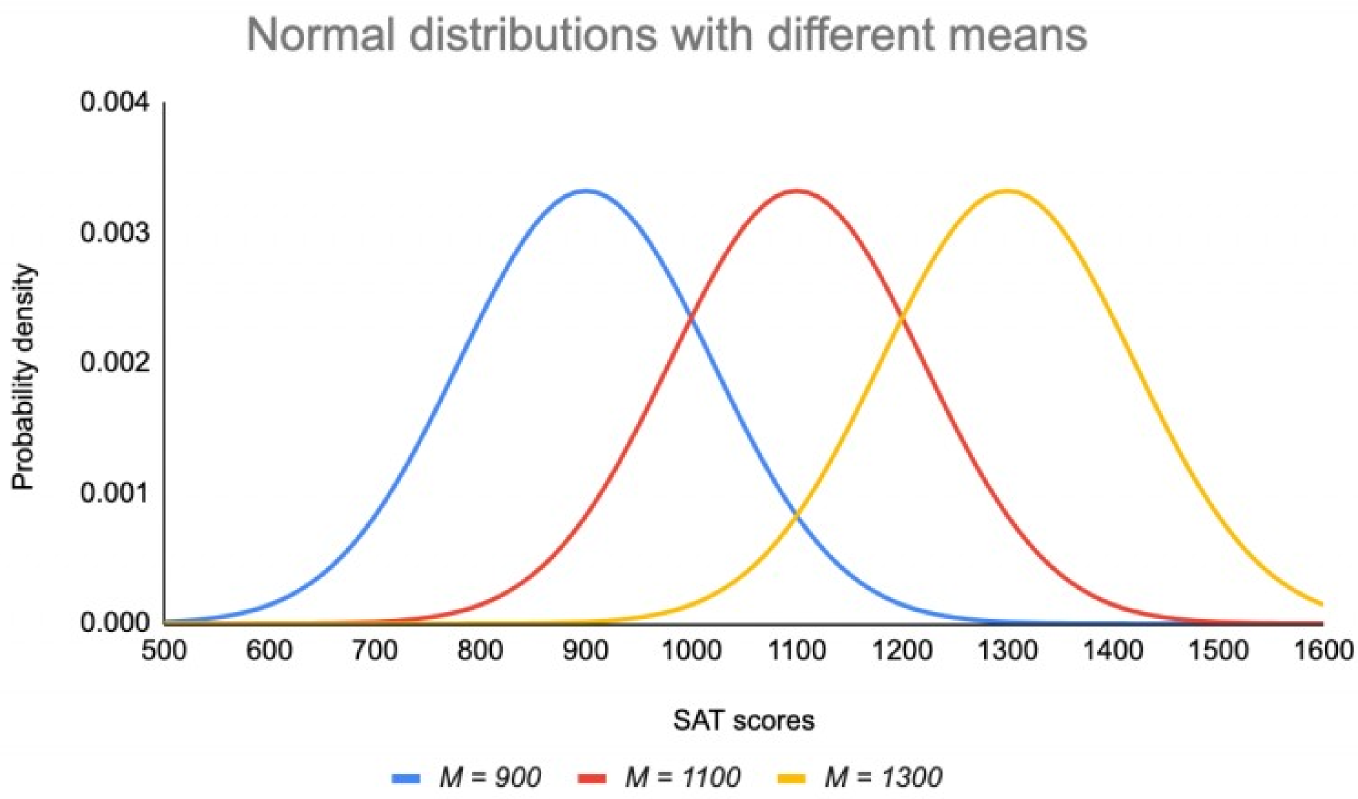
Normal Distribution



Normal Distribution

- The mean is the location parameter while the standard deviation is the scale parameter.
- The mean determines where the peak of the curve is centered.
- Increasing the mean moves the curve right, while decreasing it moves the curve left.

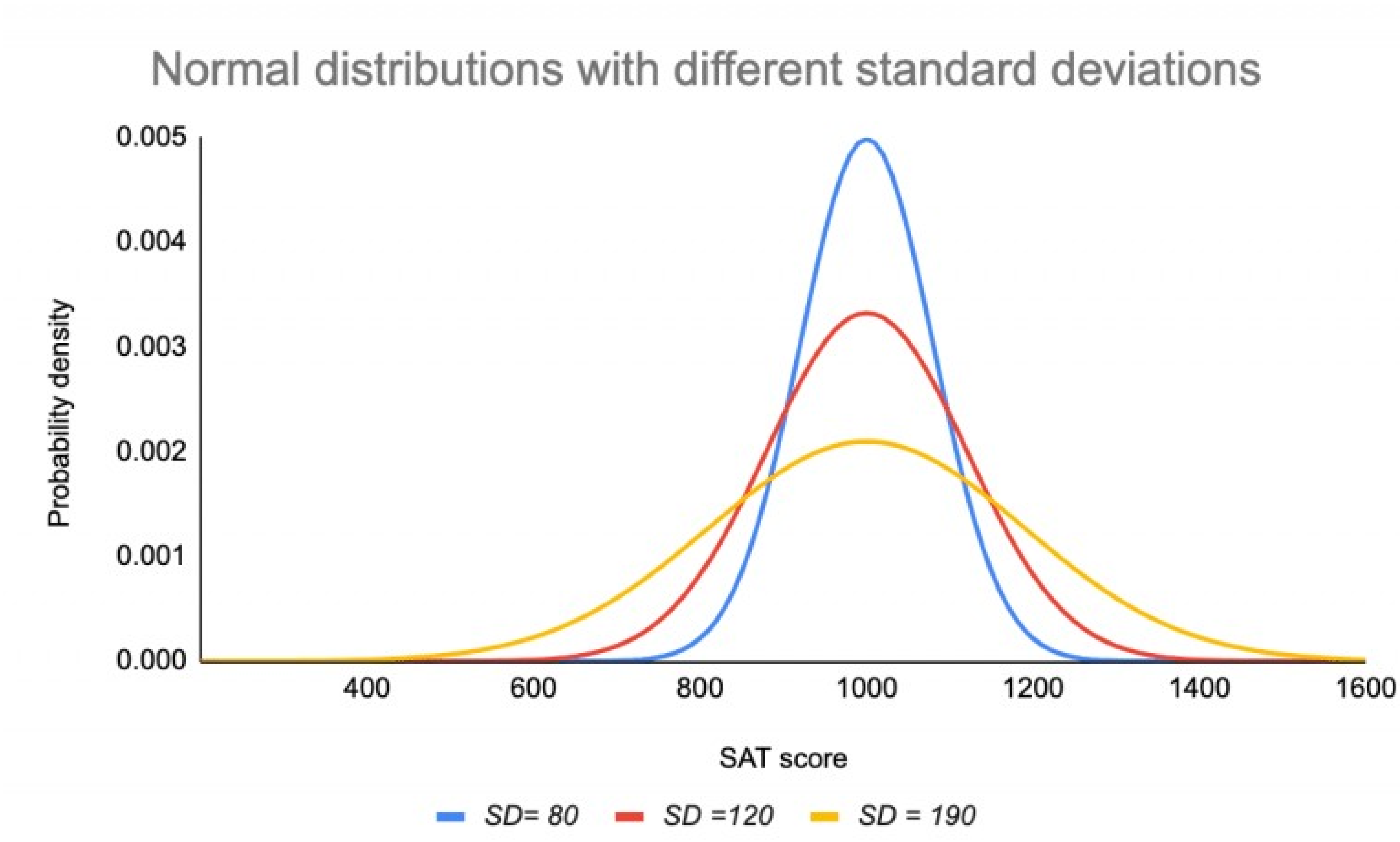
Normal Distribution



Normal Distribution

- The standard deviation stretches or squeezes the curve.
- A small standard deviation results in a narrow curve, while a large standard deviation leads to a wide curve.

Normal Distribution



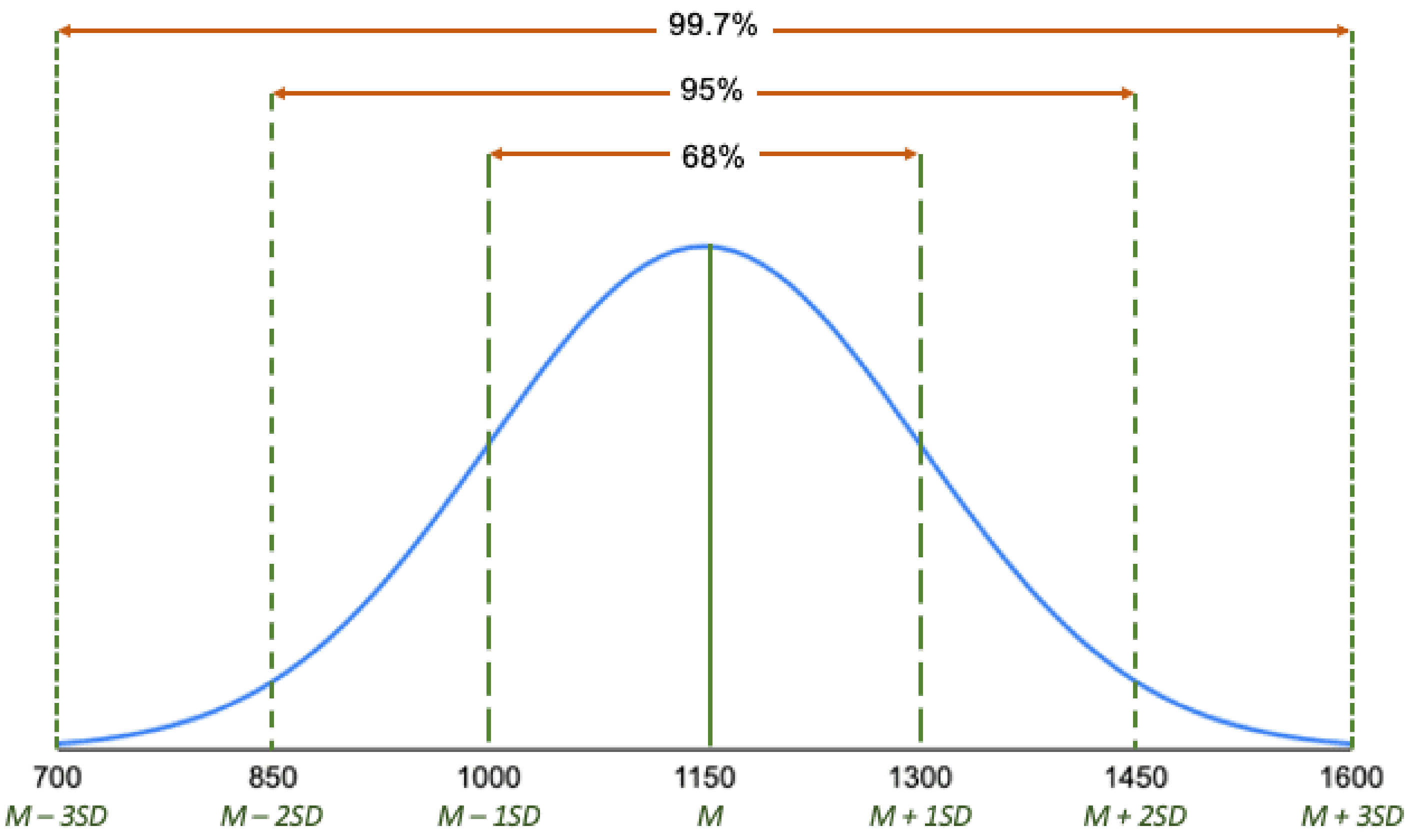
Empirical Rule

- The empirical rule, or the 68-95-99.7 rule, tells you where most of your values lie in a normal distribution:
  - Around 68% of values are within 1 standard deviation from the mean.
  - Around 95% of values are within 2 standard deviations from the mean.
  - Around 99.7% of values are within 3 standard deviations from the mean.

Example:

- Example: Using the empirical rule in a normal distribution
- You collect SAT scores from students in a new test preparation course. The data follows a normal distribution with a mean score (M) of 1150 and a standard deviation (SD) of 150.
- Following the empirical rule:
  - Around 68% of scores are between 1000 and 1300, 1 standard deviation above and below the mean.
  - Around 95% of scores are between 850 and 1450, 2 standard deviations above and below the mean.
  - Around 99.7% of scores are between 700 and 1600, 3 standard deviations above and below the mean.

Example:



99.7%
95%
68%
700
M – 3SD
850
M – 2SD
1000
M – 1SD
1150
M
1300
M + 1SD
1450
M + 2SD
1600
M + 3SD

Empirical Rule

- The empirical rule is a quick way to get an overview of your data and check for any outliers or extreme values that don't follow this pattern.
- If data from small samples do not closely follow this pattern, then other distributions like the t-distribution may be more appropriate.
- Once you identify the distribution of your variable, you can apply appropriate statistical tests.

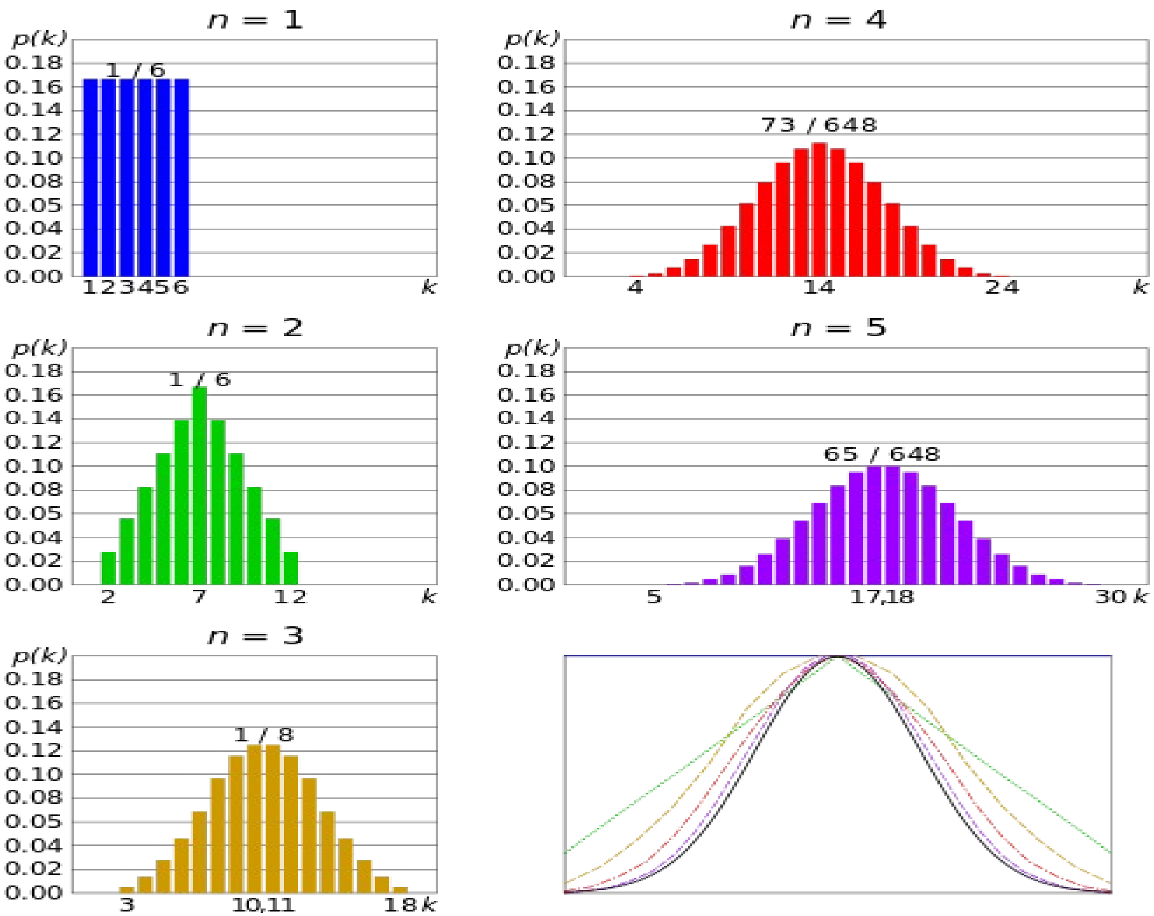
Central Limit Theorem

- The central limit theorem is the basis for how normal distributions work in statistics.
- In research, to get a good idea of a population mean, ideally you'd collect data from multiple random samples within the population.
- A sampling distribution of the mean is the distribution of the means of these different samples.

Central Limit Theorem

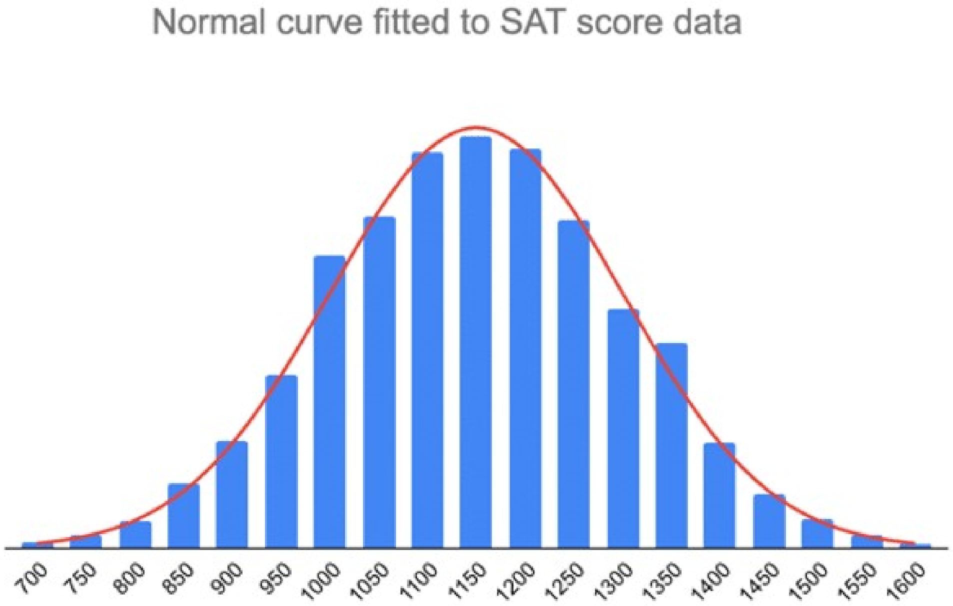
- The central limit theorem shows the following:
  - Law of Large Numbers: As you increase sample size (or the number of samples), then the sample mean will approach the population mean.
  - With multiple large samples, the sampling distribution of the mean is normally distributed, even if your original variable is not normally distributed.

Central limit theorem



Probability Density Function

- Once you have the mean and standard deviation of a normal distribution, you can fit a normal curve to your data using a probability density function.



Probability Density Function

- In a probability density function, the area under the curve tells you probability.
- The normal distribution is a probability distribution, so the total area under the curve is always 1 or 100%.
- The formula for the normal probability density function looks fairly complicated.
- But to use it, you only need to know the population mean and standard deviation.

Probability Density Function

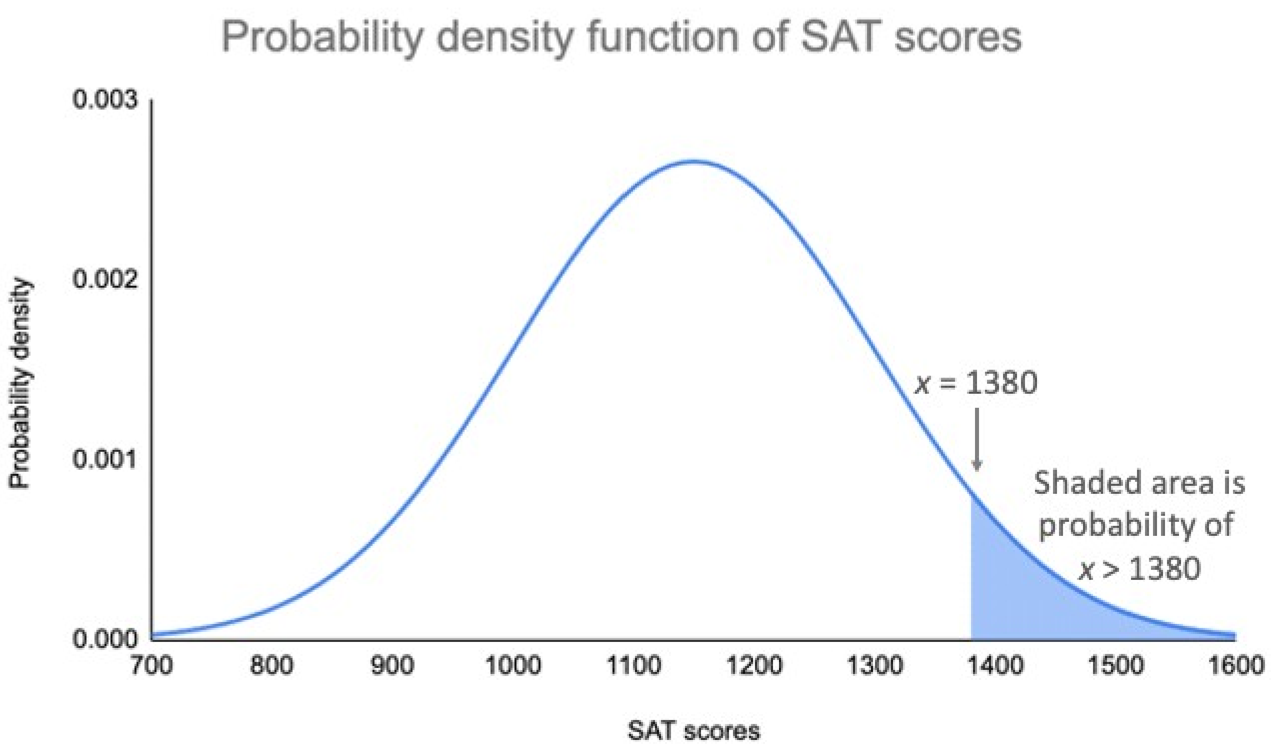
- For any value of x, you can plug in the mean and standard deviation into the formula to find the probability density of the variable taking on that value of x.

Normal Probability Density Formula	Explanation
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	• $f(x)$ = probability • x = value of the variable • μ = mean • σ = standard deviation • σ^2 = variance

Probability Density Function

- Example: Using the probability density function
  - You want to know the probability that SAT scores in your sample exceed 1380.
  - On your graph of the probability density function, the probability is the shaded area under the curve that lies to the right of where your SAT scores equal 1380.

Probability Density Function

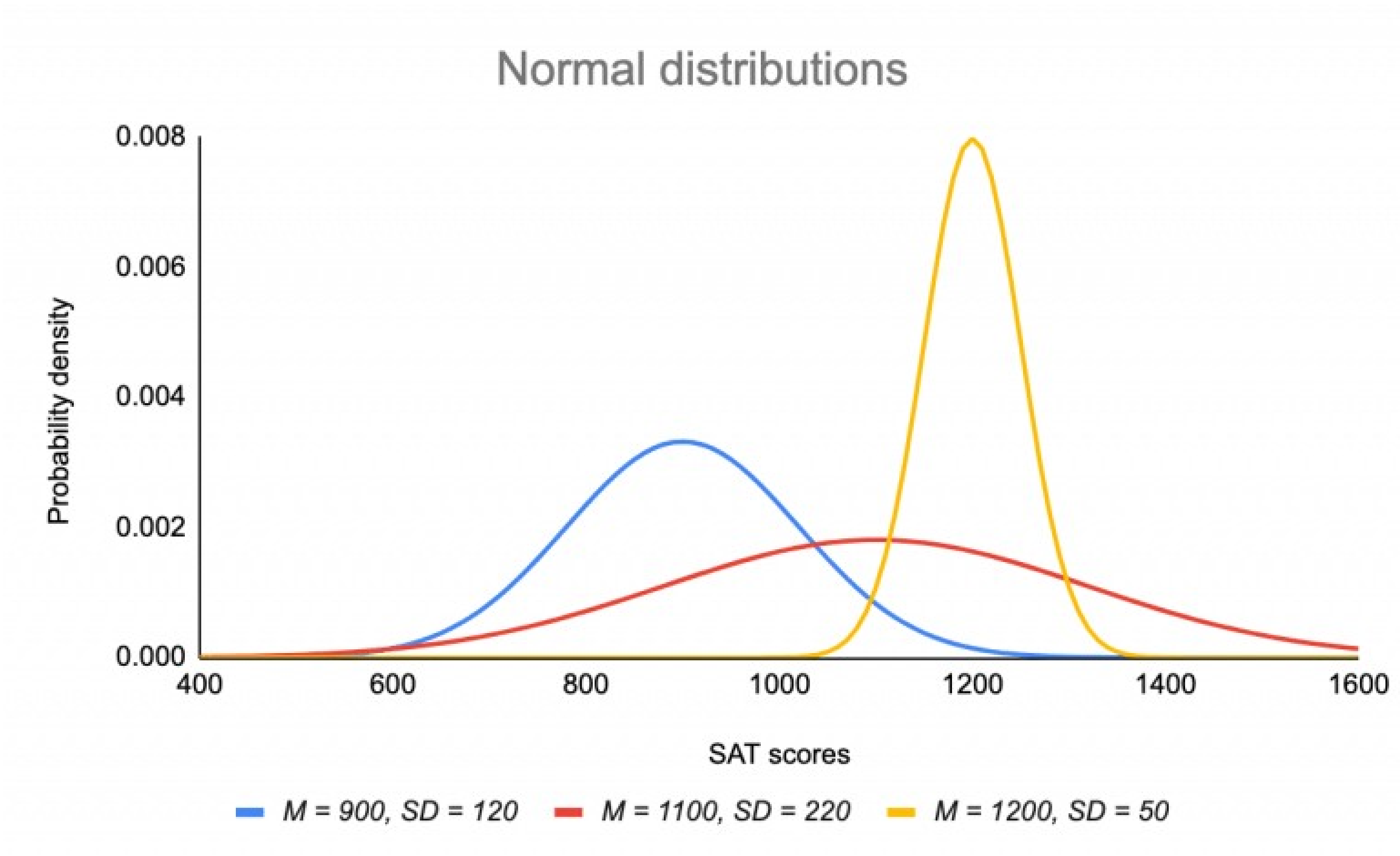


You can find the probability value of this score using the standard normal distribution.

Standard Normal Distribution

- The standard normal distribution, also called the z-distribution, is a special normal distribution where the mean is 0 and the standard deviation is 1.
- Every normal distribution is a version of the standard normal distribution that’s been stretched or squeezed and moved horizontally right or left.

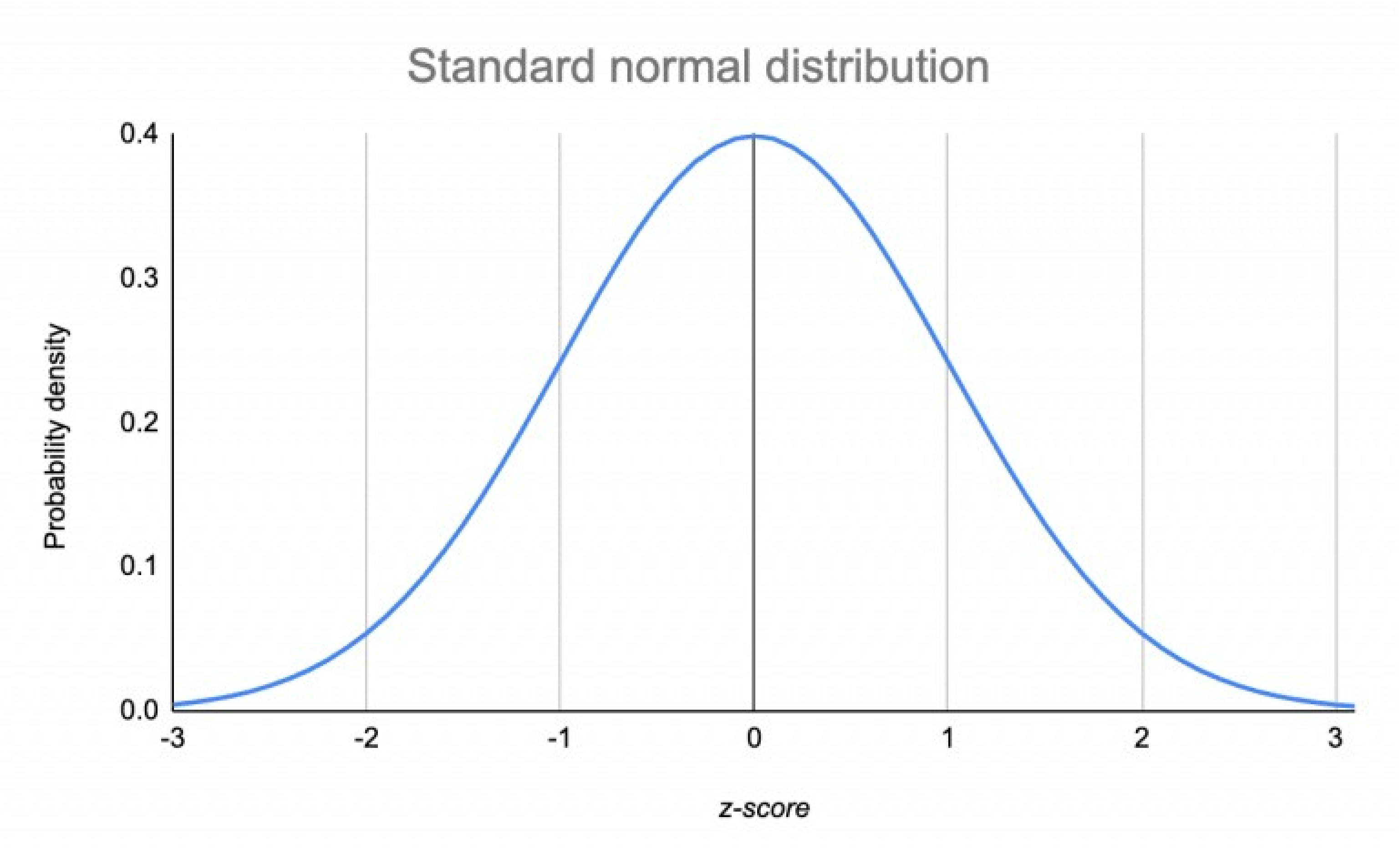
Standard Normal Distribution



Standard Normal Distribution

- While individual observations from normal distributions are referred to as x, they are referred to as z in the z-distribution.
- Every normal distribution can be converted to the standard normal distribution by turning the individual values into z-scores.

Standard Normal Distribution



Standard Normal Distribution

- You only need to know the mean and standard deviation of your distribution to find the z-score of a value.

Z-score Formula	Explanation
$z = \frac{x - \mu}{\sigma}$	• x = individual value • μ = mean • σ = standard deviation

Standard Normal Distribution

- We convert normal distributions into the standard normal distribution for several reasons:
  - To find the probability of observations in a distribution falling above or below a given value.
  - To find the probability that a sample mean significantly differs from a known population mean.
  - To compare scores on different distributions with different means and standard deviations.

Finding probability using the z-distribution

- Each z-score is associated with a probability, or p-value, that tells you the likelihood of values below that z-score occurring.
- If you convert an individual value into a z-score, you can then find the probability of all values up to that value occurring in a normal distribution.

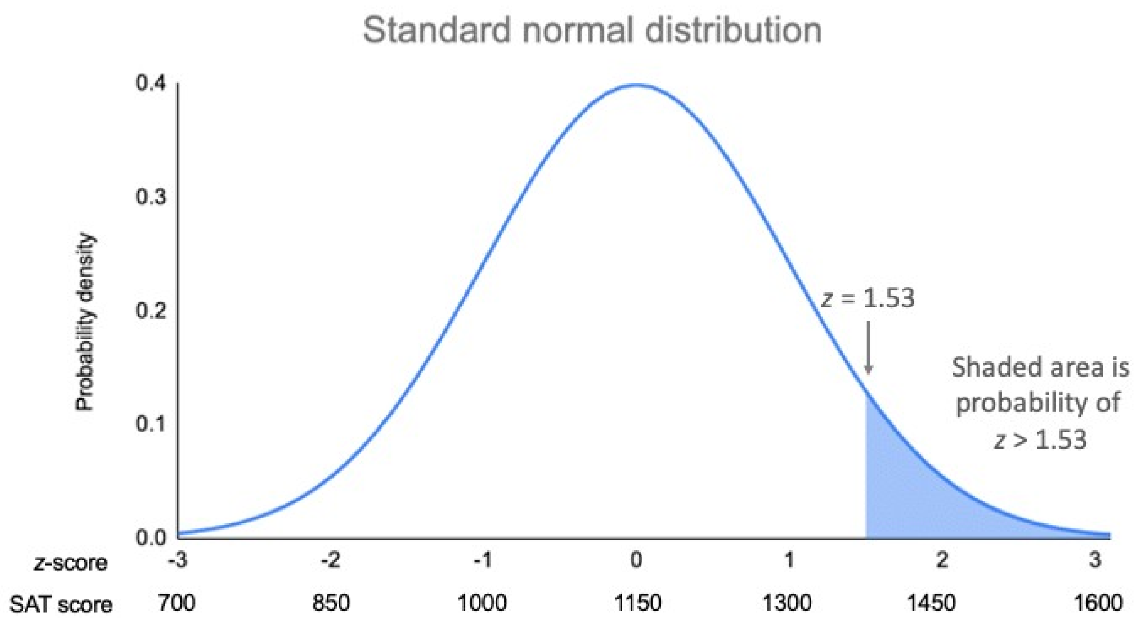
Finding probability using the z-distribution

- Example: Finding probability using the z-distribution
- To find the probability of SAT scores in your sample exceeding 1380, you first find the z-score.
- The mean of our distribution is 1150, and the standard deviation is 150. The z-score tells you how many standard deviations away 1380 is from the mean.

Formula	Calculation
$z = (x - \mu) / \sigma$	$z = (1380 - 1150) / 150$ $z = \mathbf{1.53}$

Finding probability using the z-distribution

- For a z-score of 1.53, the p-value is 0.937. This is the probability of SAT scores being 1380 or less (93.7%), and it’s the area under the curve left of the shaded area.



Finding probability using the z-distribution

- To find the shaded area, you take away 0.937 from 1, which is the total area under the curve.
- Probability of $x>1380 = 1 - 0.937 = 0.063$
- That means it is likely that only 6.3% of SAT scores in your sample exceed 1380.

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com