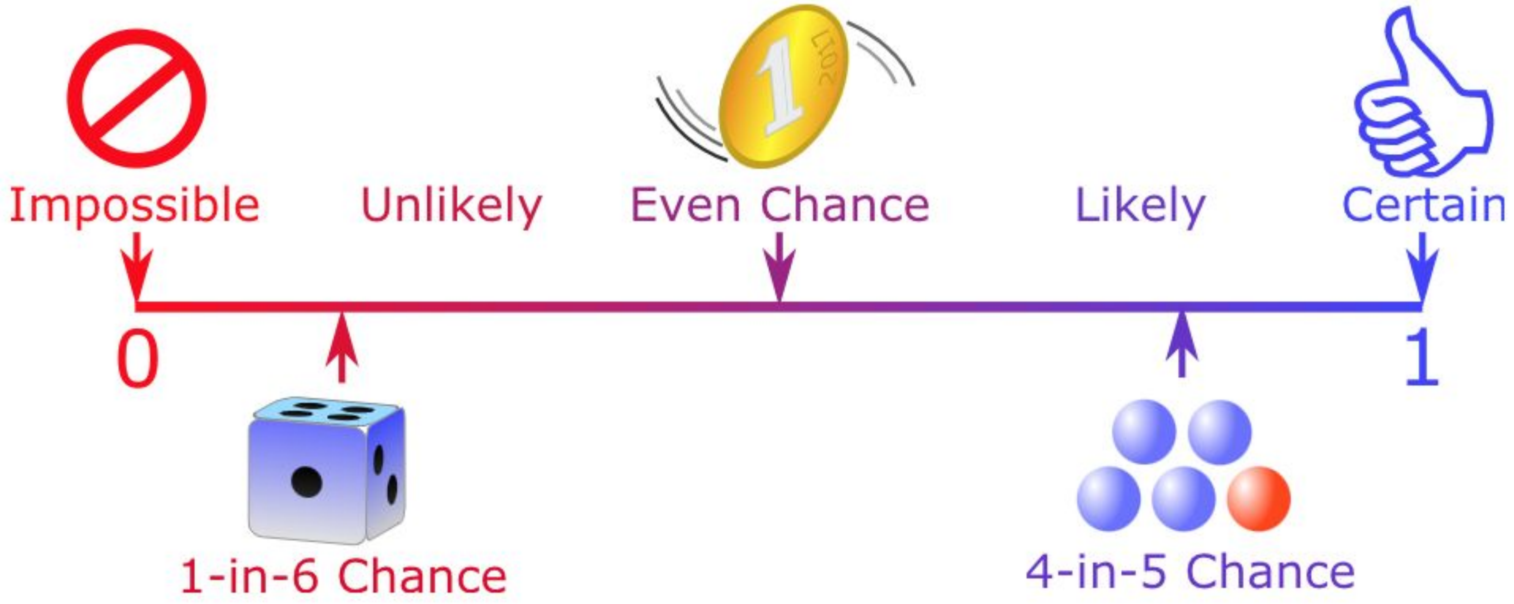
Probability Distributions

- Saurav Poudel

Important Concepts

- Probability
- Probability Distribution
- Important Distributions

Probability



Probability is always between 0 and 1

Important Concepts in Probability

- Event vs Sample Space
- Marginal / Joint / Conditional
- Independence
- Mutually Exclusive
- And / OR

Axioms

$$0 \le P(event) \le 1$$

$$P(\Omega) = 1$$

Random Variable (X)

$$P\,(X = x) = P\,(\text{Dice} = 3)$$

Random Variable

- A random variable is a variable that can take multiple values depending on the outcome of a random event.
- Possible outcomes = Possible values taken
- Random Variable = 2 Types

Random Variable

- If the outcomes are finite (like rolling a dice), discrete random variable.

- If the outcomes are infinite (like drawing a number between 0 and 1), continuous random variable.

Probability Distribution

Why do we think we need Probability Distributions?

Real Life Examples?

Probability Distribution

- It is the ‘description’ of probability of each value that the random variable can take.

- It is a function that gives the probability of each outcome of the random variable.

Notion of function

Notion of $P(X = x)$

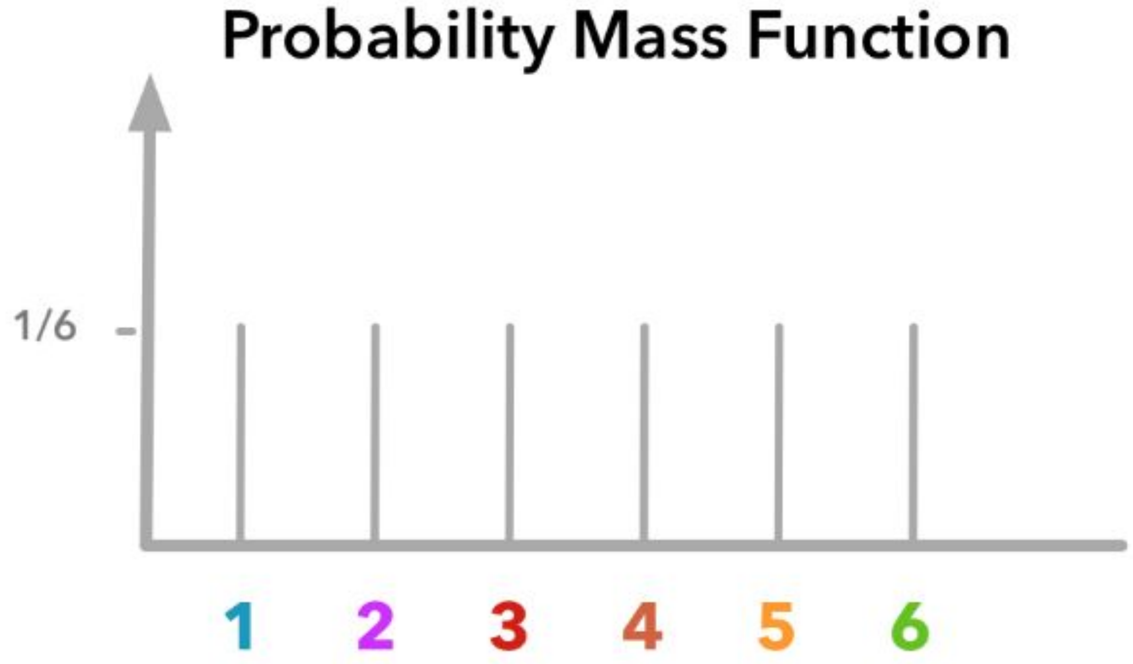
Two types of Probability Distributions

- Probability Mass Function (For Discrete)
- Probability Distribution Function (For Continuous)

Probability Mass Function (PMF)

- PMF is a function that gives $P\ (X = x)$.

- PMF of a dice calculation



Probability mass function of the dice experiment

Properties of PMF

$$\forall x \in \mathrm{x}, 0 \leq P(x) \leq 1$$

$$\sum_{x \in \mathrm{x}} P(x) = 1$$

What is the probability of someone having a weight (technically mass) of 72 kg?

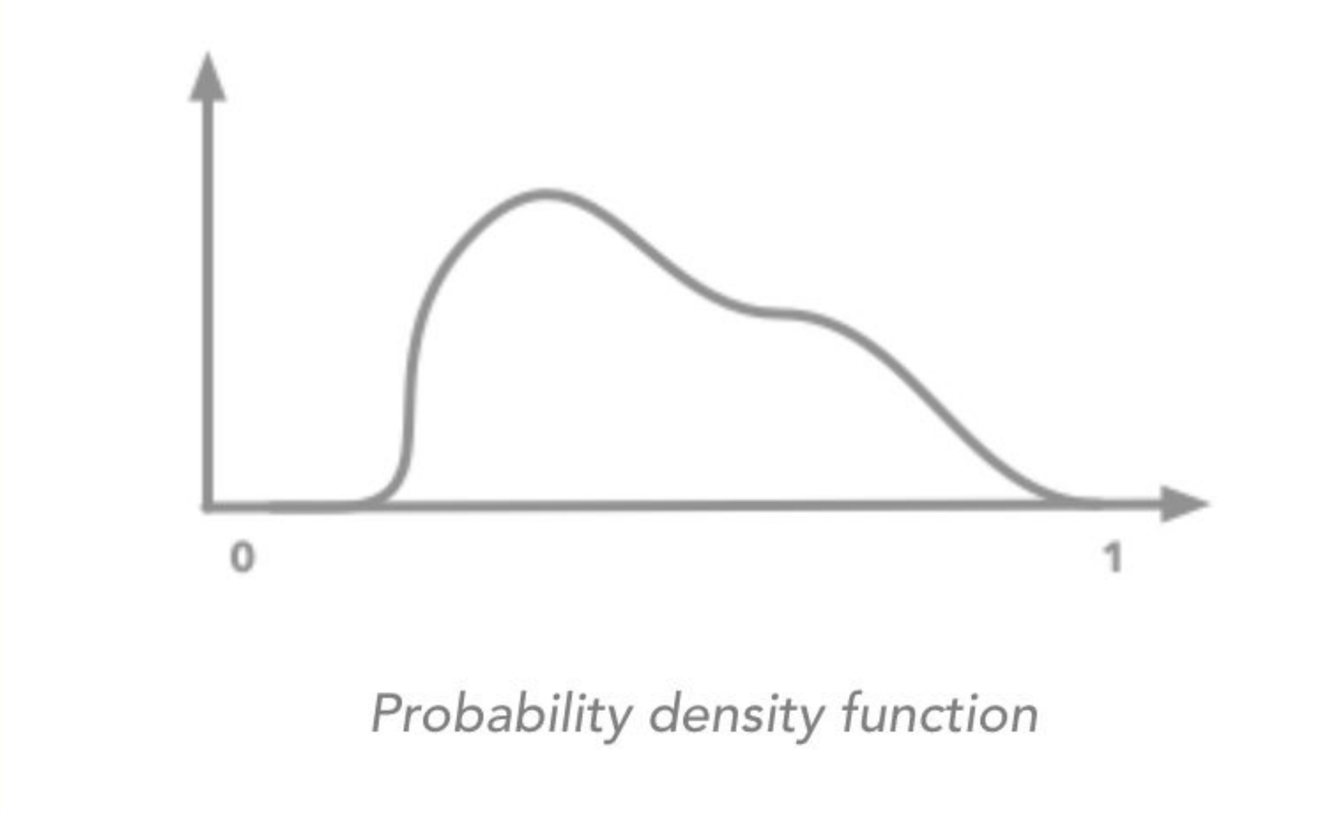
What is the probability of getting 0.25 when randomly drawing values between 0 and 1?

Probability Distribution Function (PDF)

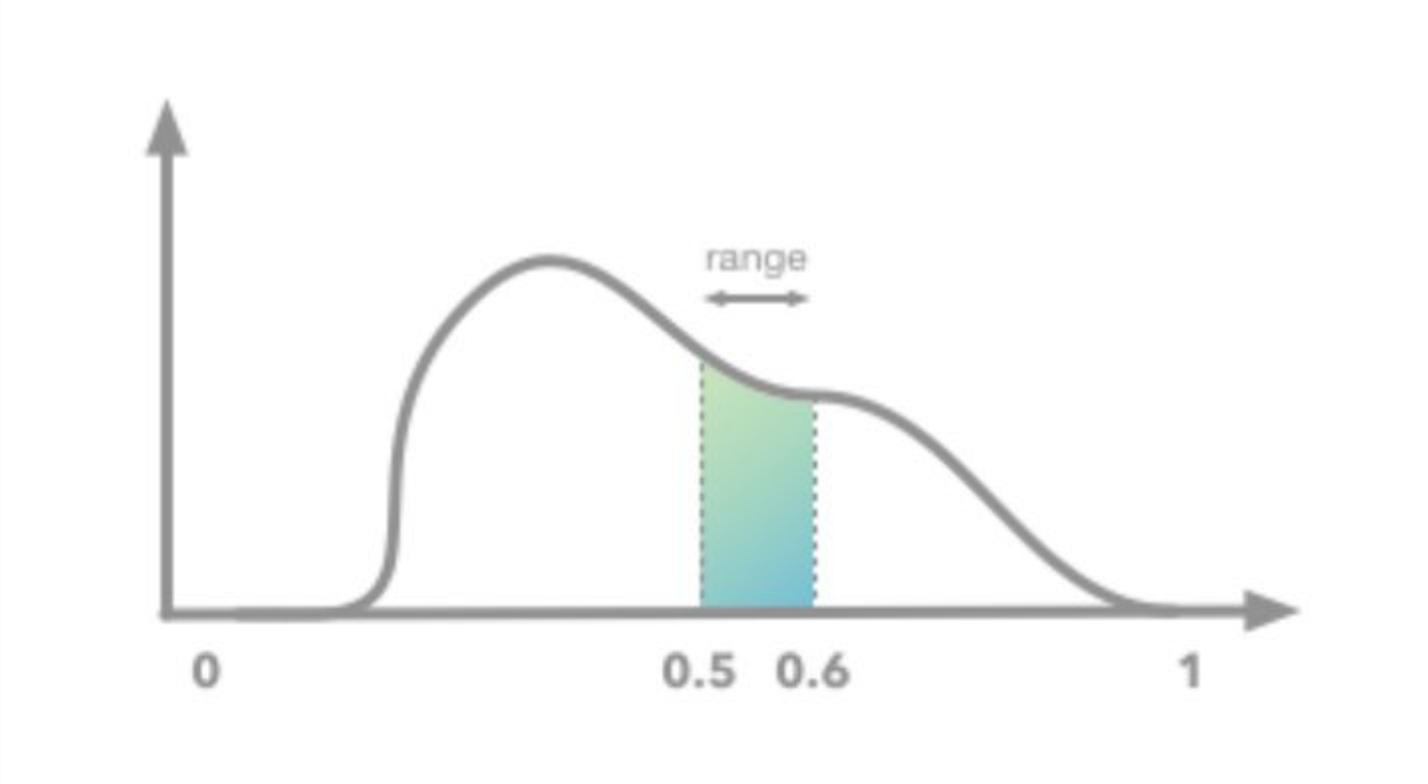
- PDF is a function that gives P(x) for x lying between a and b. (or x less than a)

- We get the probability of a random variable lying within a range.

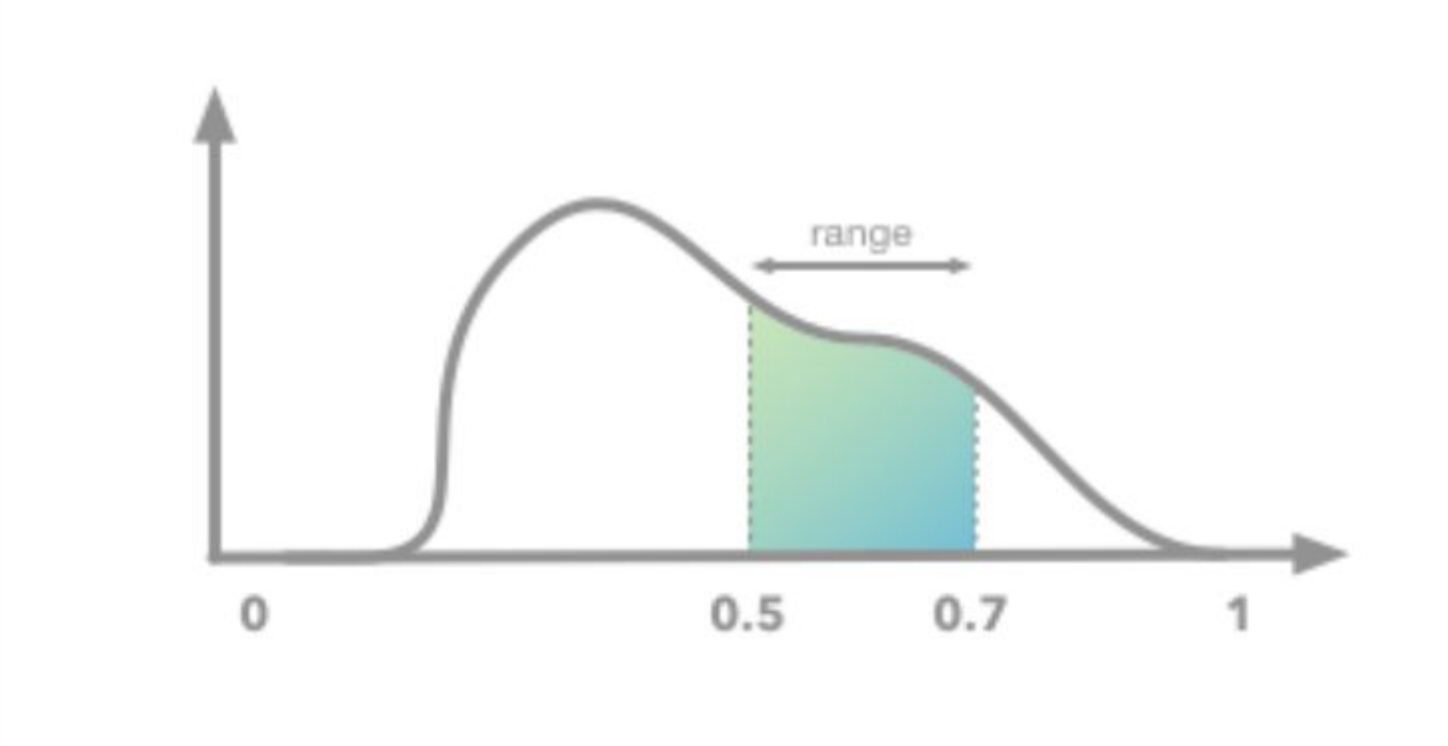
Probability Distribution Function (PDF)



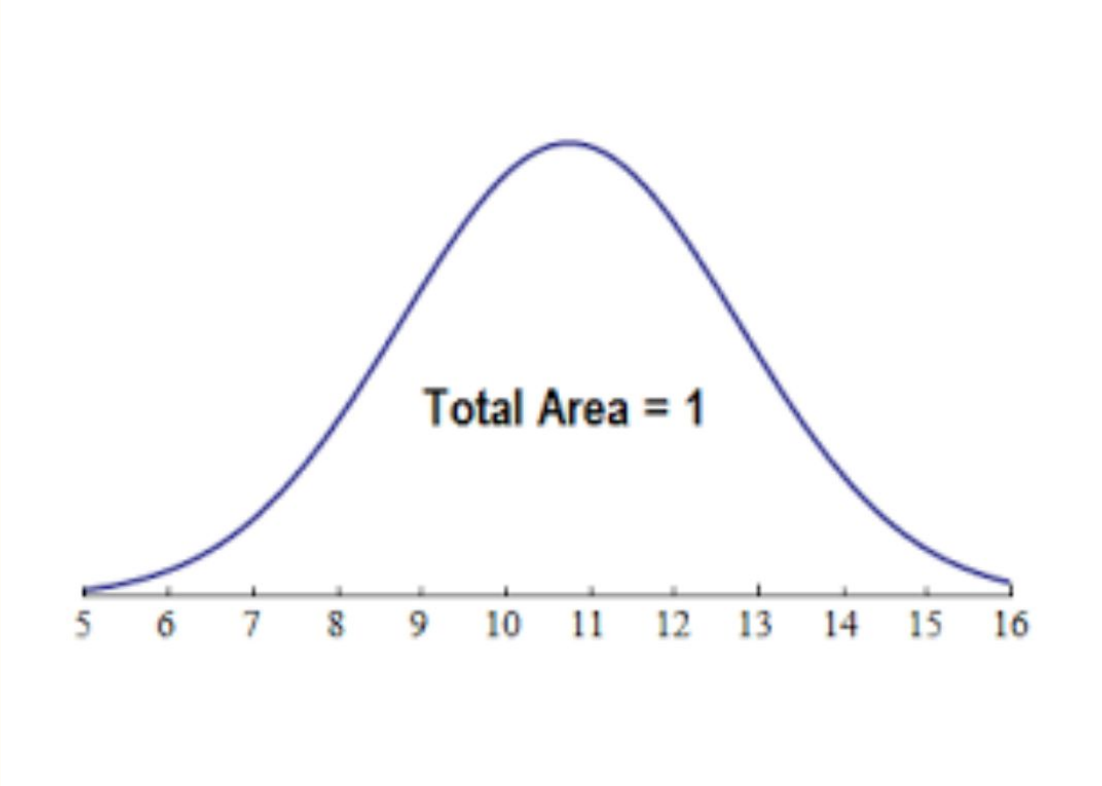
Probability Distribution Function (PDF)



Probability Distribution Function (PDF)



Properties of PDF



Some Important Probability Distributions

Why do we need to know these Probability Distributions?

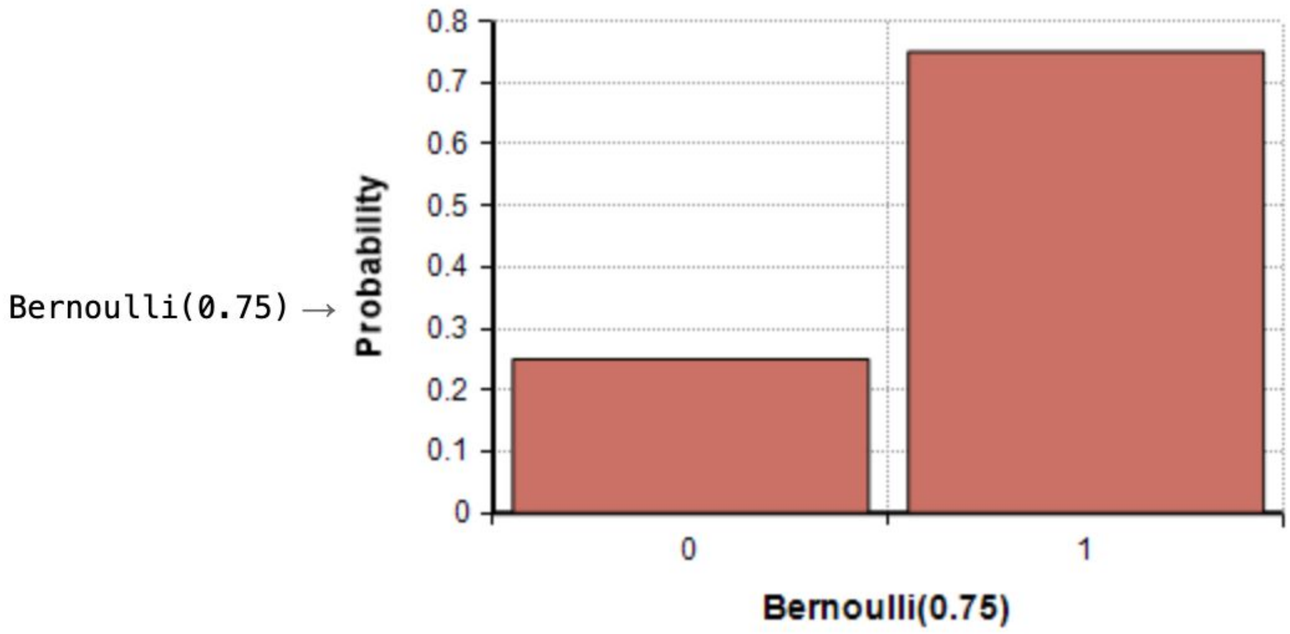
Rather than getting intimidated by the complex formulae, you should ask yourself - where can I use this distribution?

Bernoulli Distribution

- Classic coin toss example.

- Random Variable with two outcomes with probability p and 1 - p.

Bernoulli Distribution



Bernoulli Distribution just gives you the probability of success of 1 event.

What if you want to know the probability of K successes out of N events?

(Let's say the probability of you getting 4 questions out of 10 right, if probability of getting each question right is 0.25)

Binomial Distribution

- Probability Distribution of Success.

- P(Head) in ‘n’ number of coin tosses.

Binomial Distribution

Binomial Distribution Formula

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

where

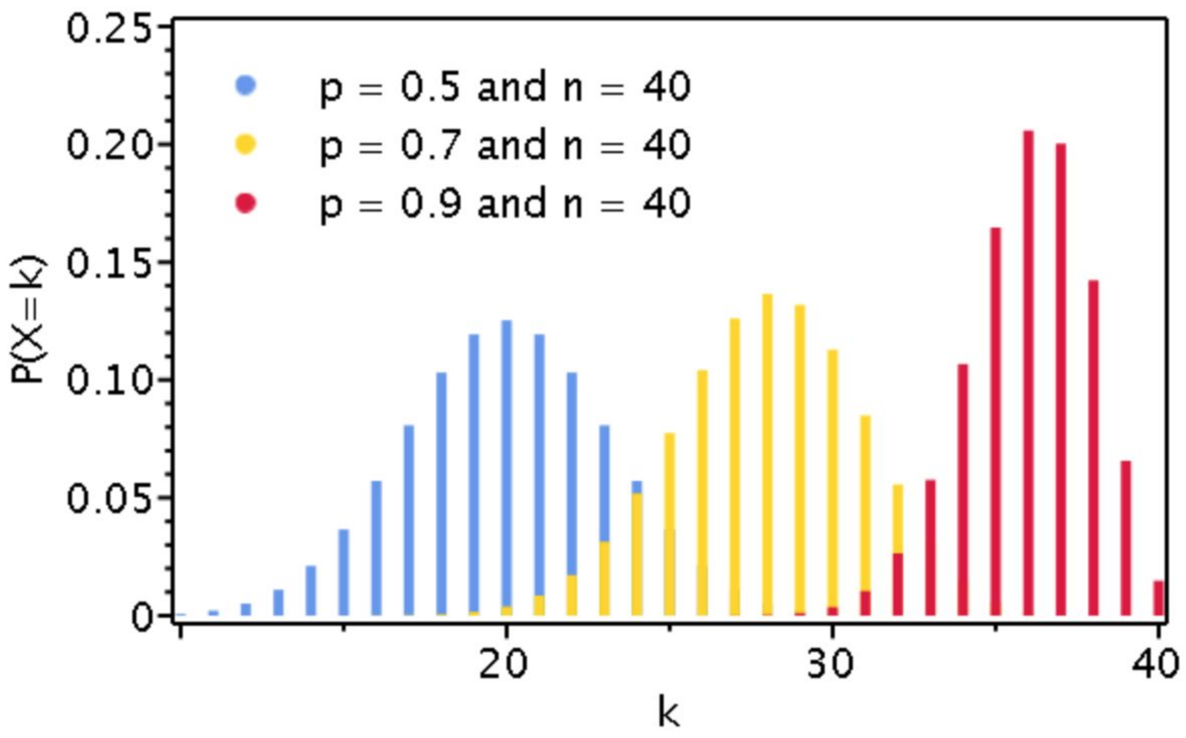
n = the number of trials (or the number being sampled)

x = the number of successes desired

p = probability of getting a success in one trial

$q = 1 - p$ = the probability of getting a failure in one trial

Binomial Distribution



Binomial and Bernoulli Distribution

- Both are discrete distributions.

- Bernoulli is a Binomial Distribution of special case $n = 1$.

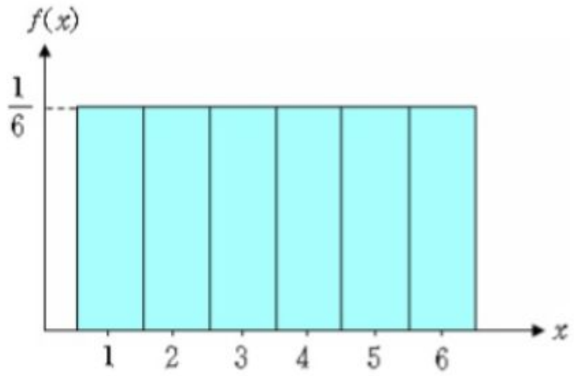
Uniform Distribution

- All the outcomes have equal or uniform probability.

- Classic Dice Example. Or drawing between 0 and 1 with equal probability. (Both discrete and continuous)

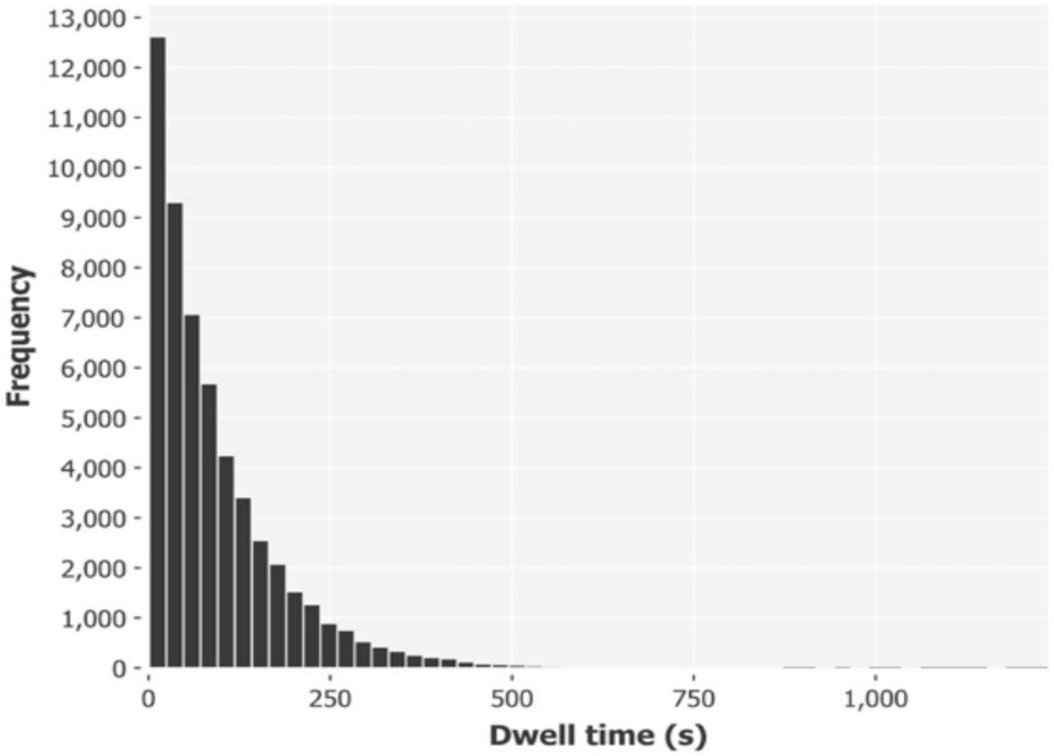
Uniform Distribution

- Consider a case with rolling a fair dice



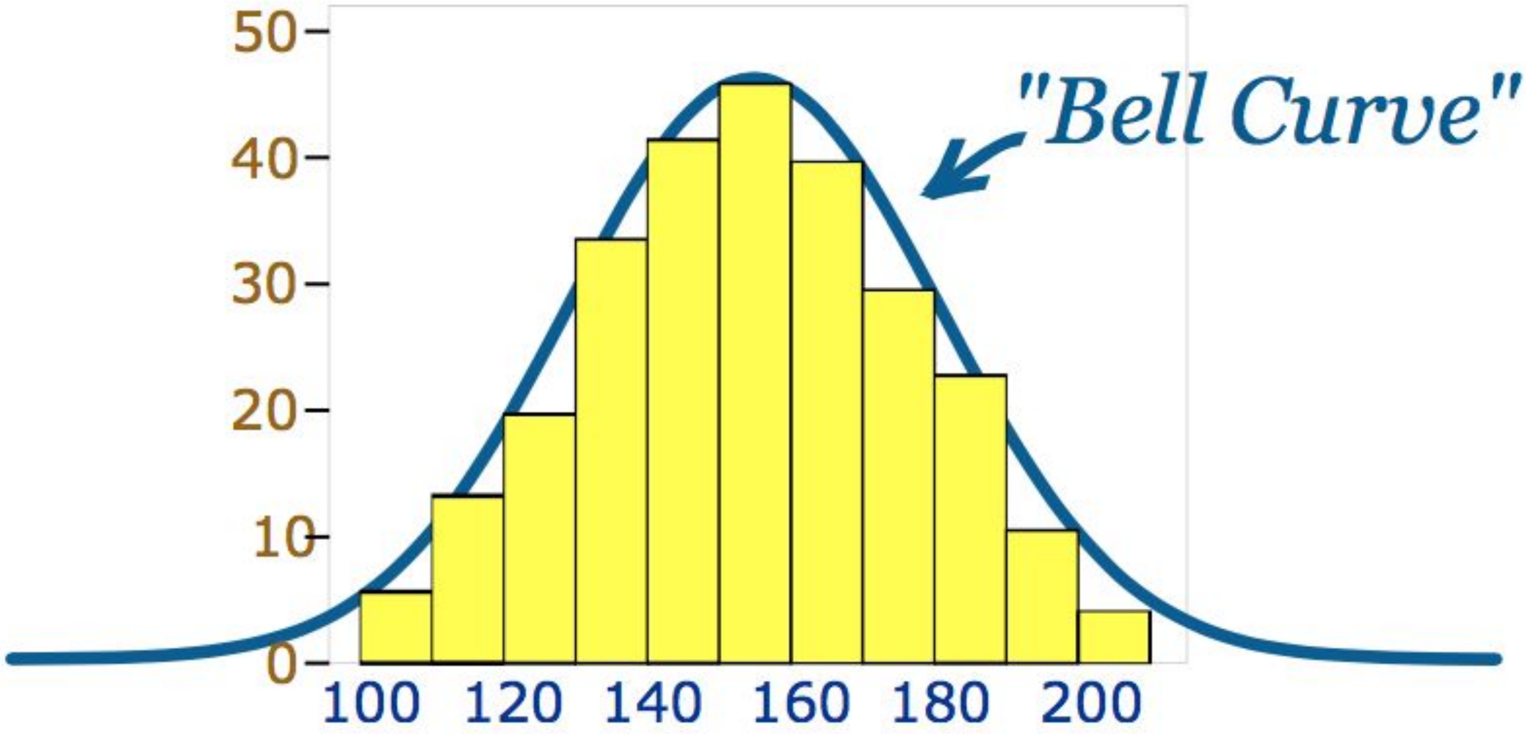
- Each random variable has same probability → Uniform distribution

Exponential Distribution



Now onto the most popular of all distributions!

So normal that we call it ‘Normal Distribution’!



A Normal Distribution

Normal Distribution

- It is a symmetric probability distribution where most of the observations cluster around a center peak.

- Also known as Gaussian Distribution (Bell Curve).

Normal Distribution

Examples around you!

- Height of the people
- Errors in a measurement
- Blood Pressure
- Temperature of a place
- Marks of Students

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

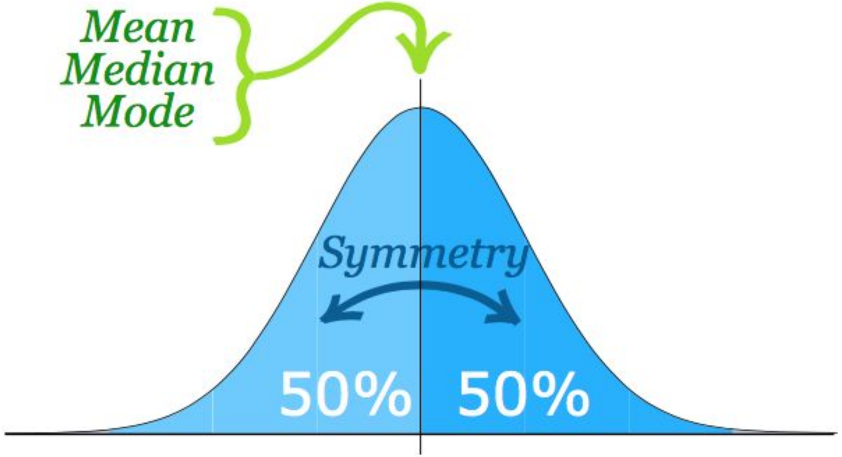
$$\mu = \text{ Mean}$$

$$\sigma = \text{ Standard Deviation}$$

$$\pi \approx 3.14159\cdots$$

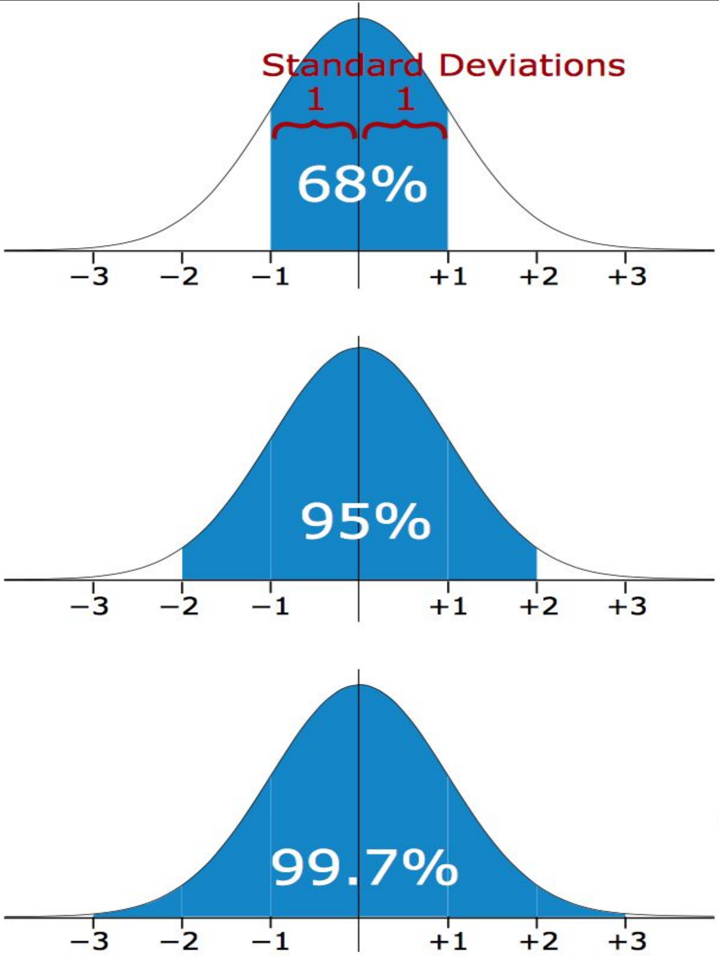
$$e \approx 2.71828\cdots$$

We say the data is "normally distributed":



The **Normal Distribution** has:

- mean = median = mode
- symmetry about the center
- 50% of values less than the mean and 50% greater than the mean



68% of values are within
1 standard deviation of the mean

95% of values are within
2 standard deviations of the mean

99.7% of values are within
3 standard deviations of the mean

Applications of Normal Distribution

The average speed of the vehicles of the sample size 1000 are normally distributed, with mean of 120km/hr and Standard deviation of 10km/hr.

Find the percentage of vehicles with speed less than 100 km/hr.

Why are things Normal?

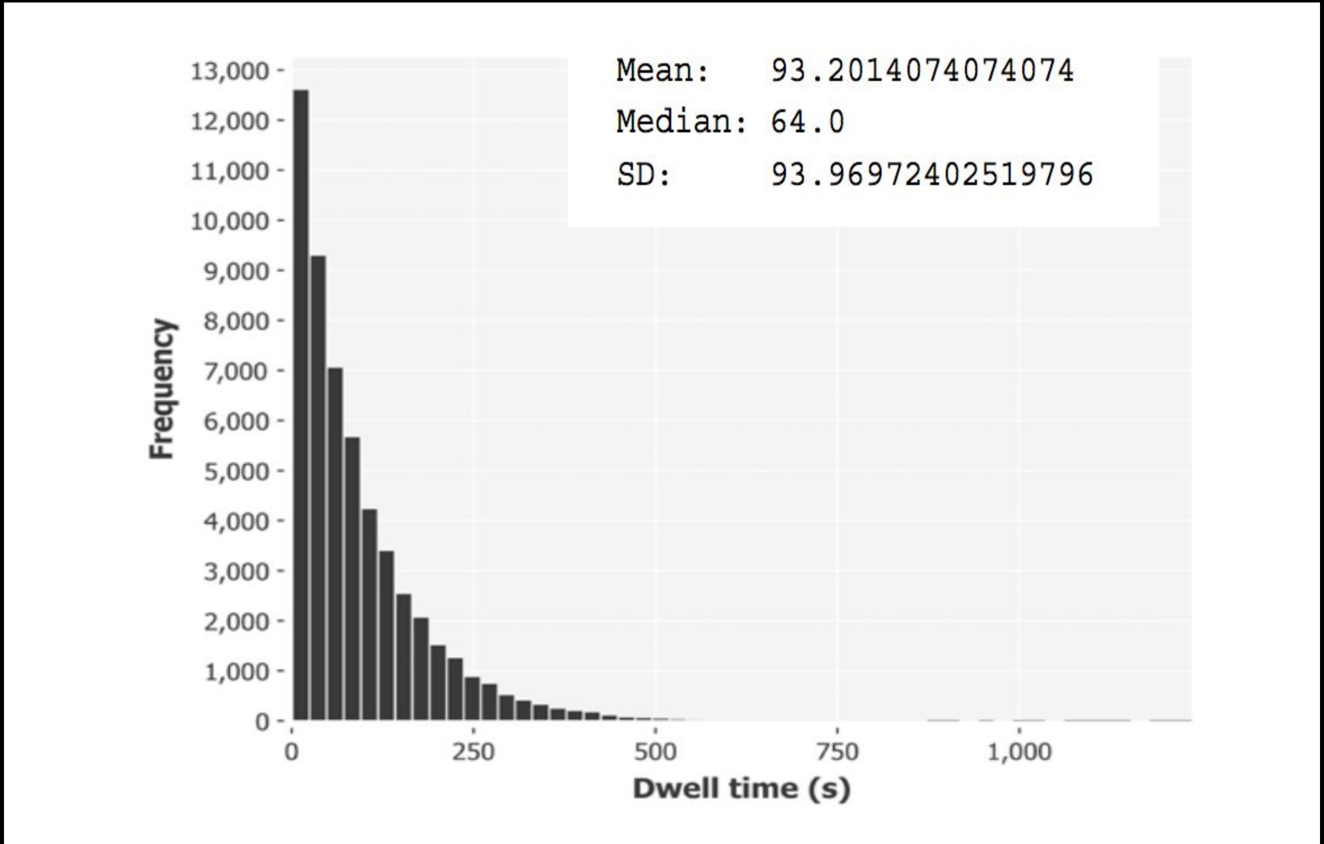
Central Limit Theorem

Textbook Definition:

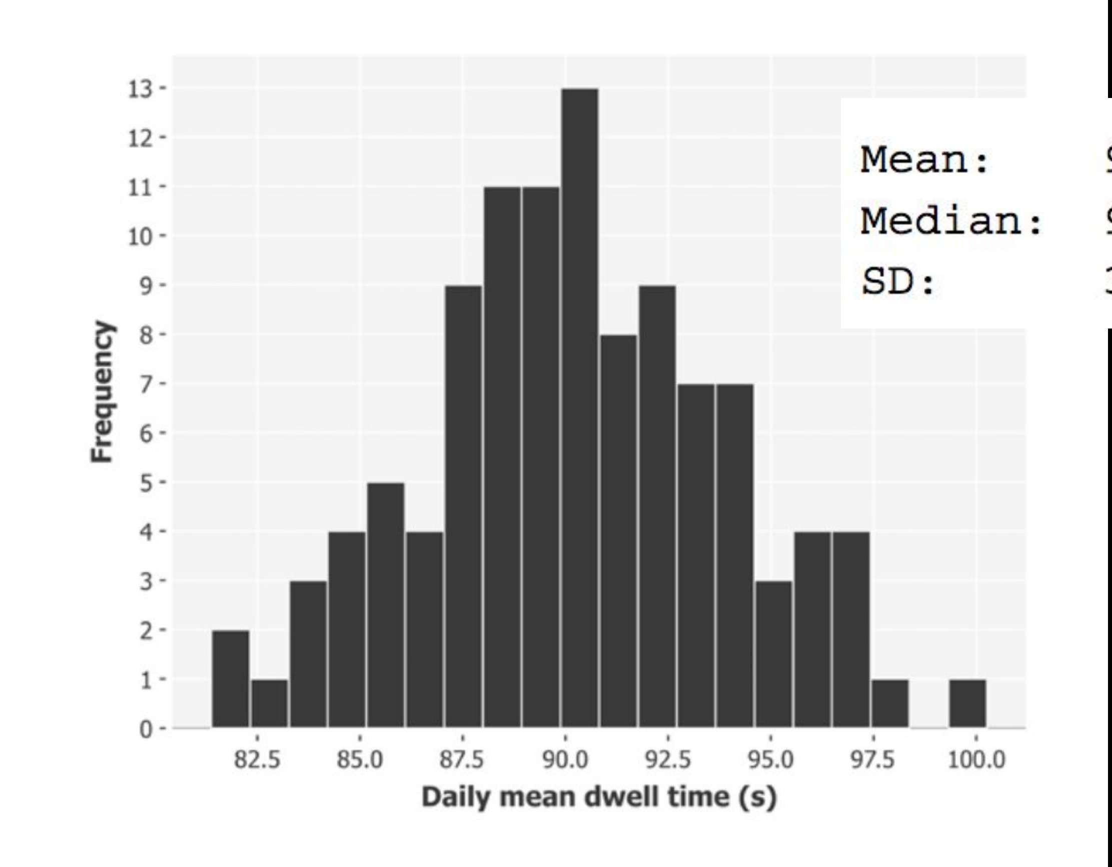
The central limit theorem states that the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, if the sample size is large enough.

Distribution of the mean (or sum) of the samples are Normal, even if the samples come from entirely different distribution.

Dwell time of a website taken over 6 months



Average daily dwell time of the website



Mean:	90.210428650562
Median:	90.13661202185791
SD:	3.722342905320035

Let us check it all out ourselves *(rather than trust some theorem or figures.)*

Why are things in nature Normal?

(I mean, we are not taking any sums or mean in nature right?)

Most of the things in nature are actually the sum or average of multiple effects.

Temperature is the measure of the average heat or thermal energy of the particles in a substance.

Kinetic energy or Pressure of a gas is the average kinetic energy or pressure of all its molecules.

Height of an organism is due to the sum of multiple effects of millions of genes in them.

Application of Standard Deviation!

(And the concept of Standardization)

Standard Scores

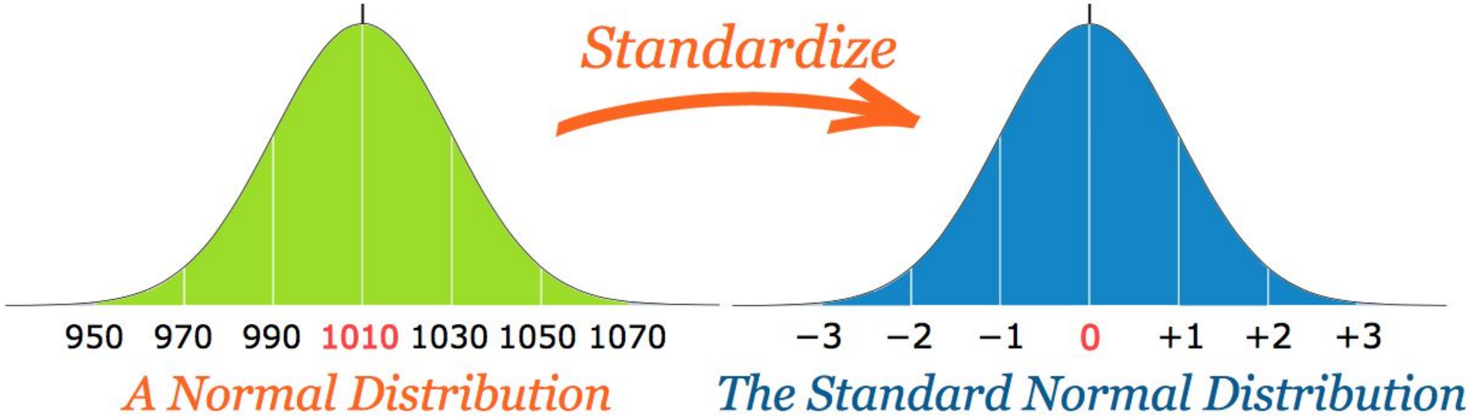
Standardization of Data!

- The number of Standard Deviation away from the Mean.
- Also known as Z scores.
- For example, for a distribution with Mean 10 and SD 2, if your marks is 6, you are 2 SD away from the Mean.

So to convert a value to a Standard Score ("z-score"):

- first subtract the mean,
- then divide by the Standard Deviation

And doing that is called "Standardizing":



Why 'Standardize' data?

Here are the students results (out of 60 points):

20, 15, 26, 32, 18, 28, 35, 14, 26, 22, 17

Most students didn't even get 30 out of 60, and **most will fail**.

The test must have been really hard, so the Prof decides to Standardize all the scores and only fail people 1 standard deviation below the mean.

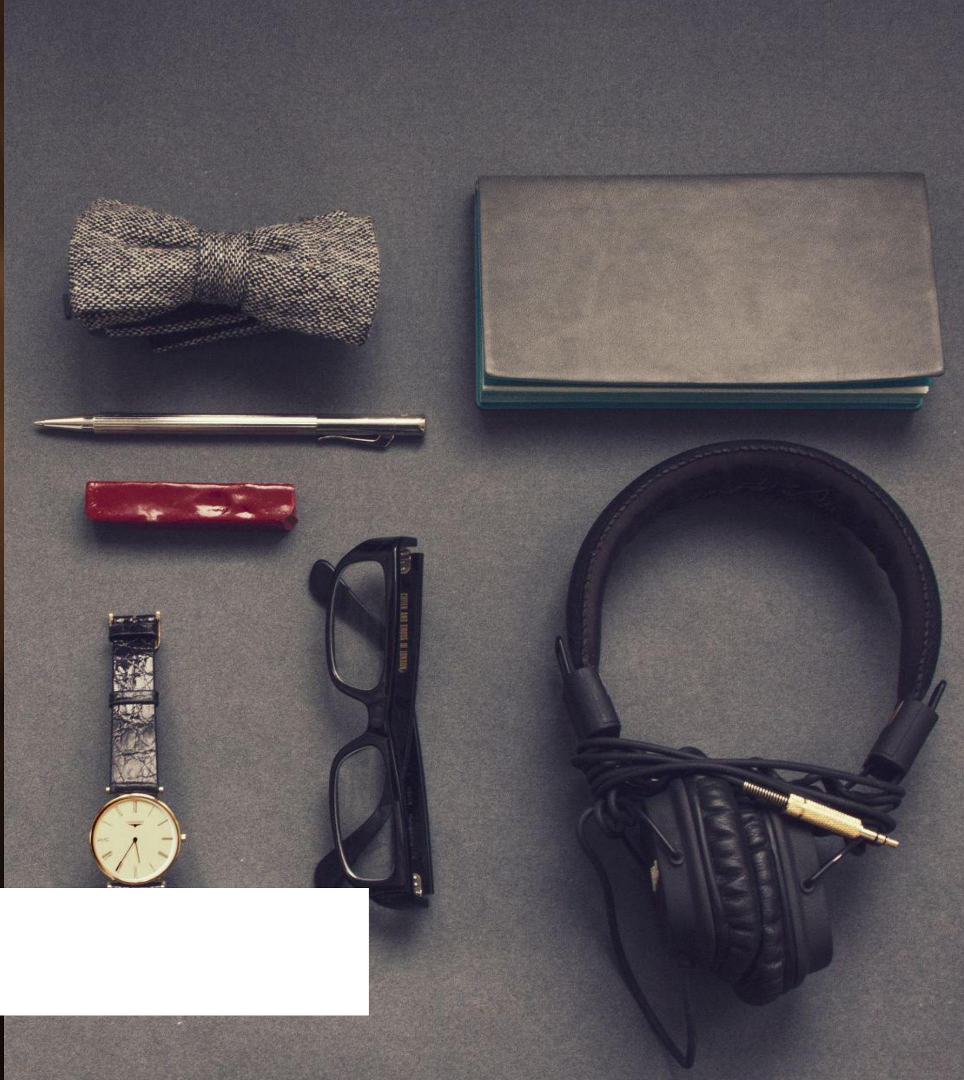
The **Mean is 23**, and the **Standard Deviation is 6.6**, and these are the Standard Scores:

-0.45, -1.21, 0.45, 1.36, -0.76, 0.76, 1.82, -1.36, 0.45, -0.15, -0.91

Now only 2 students will fail (the ones who scored 15 and 14 on the test)

Much fairer!

End Note!



Thank You!