Distributions

The Basics

Definition

The way in which quantities are distributed among a sample space.

Random Variables

- Definition:
  - A variable whose value is subject to variations due to chance
- Types:
  - Discrete Random Variables
  - Continuous Random Variables

Probability Distribution Functions

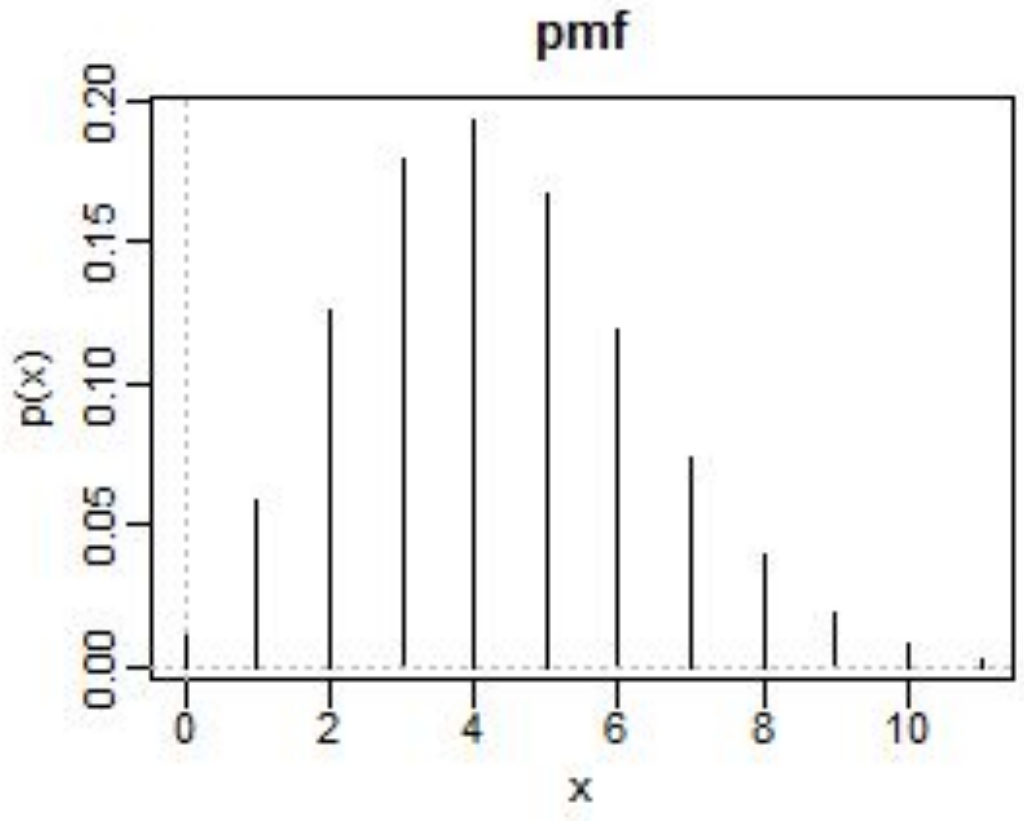
PMF vs PDF

CMF vs CDF

PMF vs. PDF

- PMF:
  - The Probability Mass Function (PMF) is a function that gives the probability that a DISCRETE random variable is exactly equal to some value.
  - Ex: p(x = 3) = 78%

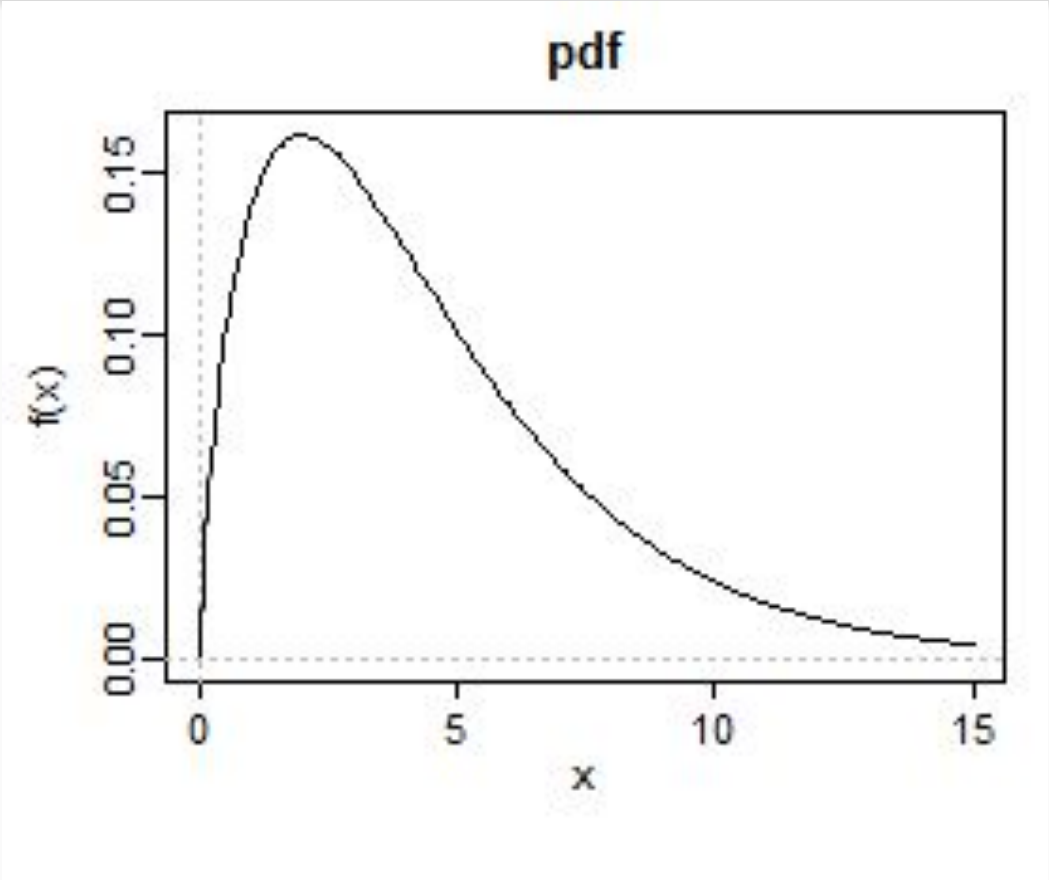
PMF



PMF vs. PDF

- PDF:
  - The Probability Density Function (PDF) is a function that gives the probability that a CONTINUOUS variable lies within a particular range. We cannot say anything about the probability of a variable equaling a single value in the space.
  - In other words: We must refer to probabilities in terms of intervals.
  - Ex: $p(1 < x < 3) = 43\%$

PDF



CMF vs. CDF

- CMF:
  - The Cumulative Mass Function (CMF) Is often represented as a series of probabilities on a table, that add up to 1.
  - Ex: Sum(p(X)) = 1; where X represents {x1, x2, x3, ..., xn}

CMF

CMF table

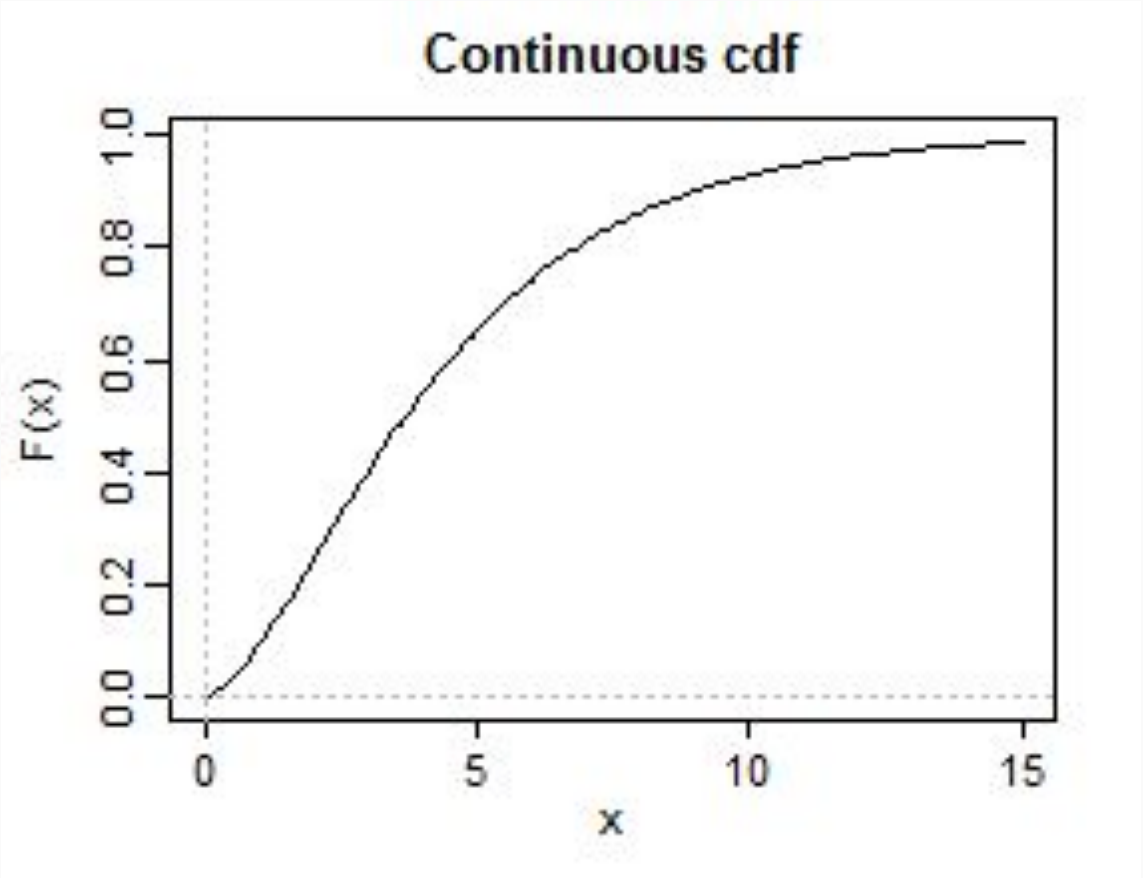
x	1	2	3	4
f(X = x)	0.5	0.1	0.2	0.2

Total (sum)	1

CMF vs. CDF

- CDF:
  - The Cumulative Density Function (CDF) is a probability distribution that is used to find the probability that the variable takes a value less than or equal to x. The probability that the variable is EXACTLY x will ALWAYS be 0.
  - Ex: p(x <= 3) = 72%
  - Ex: p(x = 3) = 0

CDF

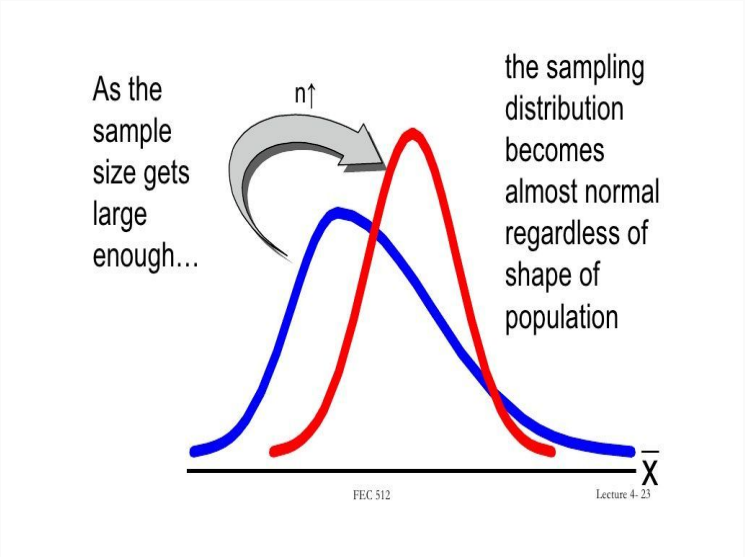


Properties of Distributions

Normal Distributions and more...

Central Limit Theorem

- Given certain conditions, the arithmetic mean of a sufficiently large enough set will be approximately normally distributed, regardless of the underlying distribution.
- This implies that methods that normally work for NORMAL distributions (bell curve), can be applied to many problems involving OTHER types of distributions.



SoftStack Factory 2018

Binomial Distribution and Bernoulli

Bernoulli Trials are a series of binary experiments with each trial either being a success/ failure.

A Binomial distribution is the distribution of the number of successes in a series of n independent bernoulli trials.

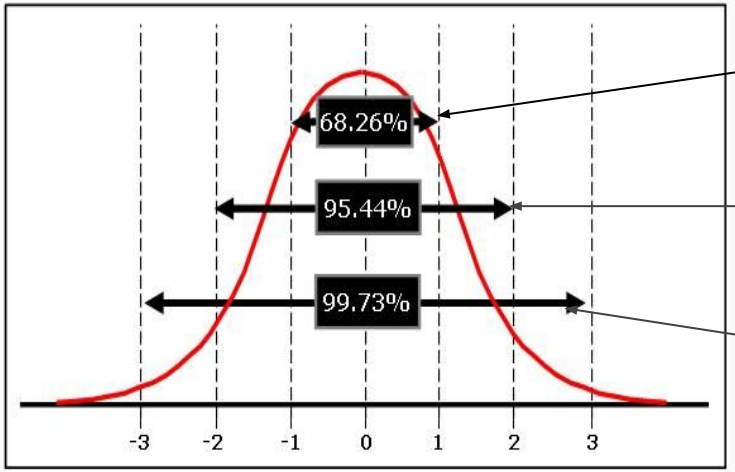
Ex: 100 coin tosses in a series are a Bernoulli Trial.

Ex: The resulting sample space = {H,H,T,T,T,H,T,T,H, … , H} is a Binomial Distribution.

6-Sigma: 68-95-99 Rule

The 6 Sigma rule:

Essentially, if you are able to represent your set as a normal distribution, normalized by standard deviation (Z-score) as such:



68% of all values lie within 1 std dev from the mean.

95% of all values lie within 2 std devs from the mean.

99% of all values lie within 3 std devs from the mean.

SoftStack Factory 2018

Example Graphs

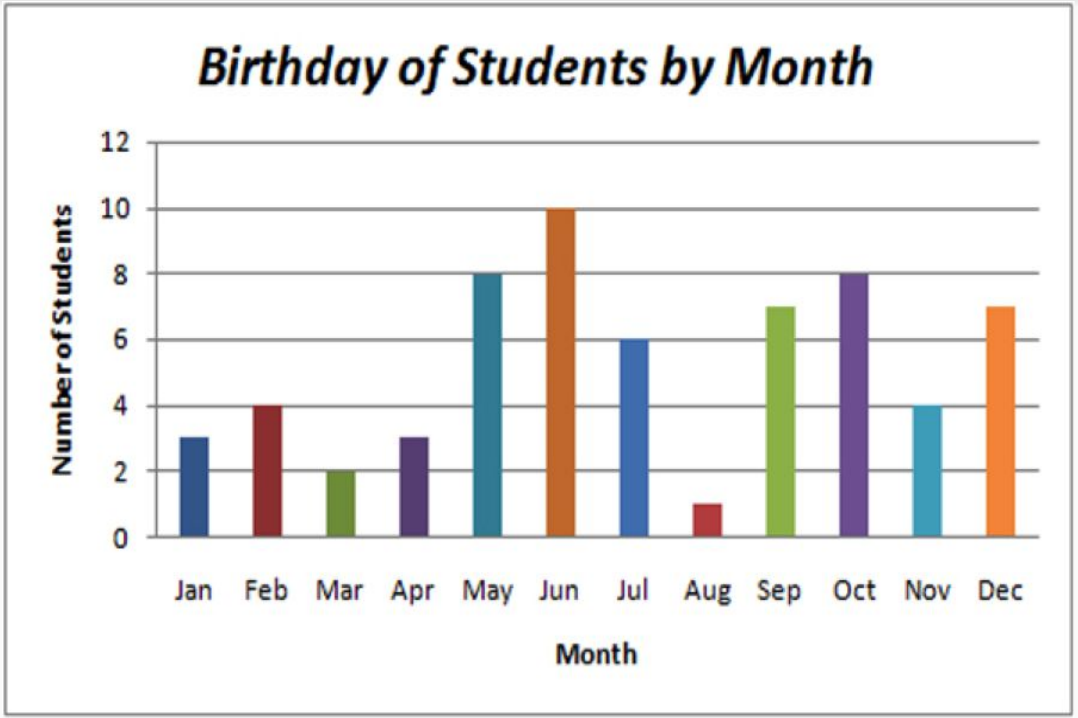
Histograms and Scatter Plots

Histogram

A Histogram is a diagram consisting of rectangles.

Each rectangle’s area is proportional to the frequency of a variable

Each rectangle’s width is equal to the class interval.



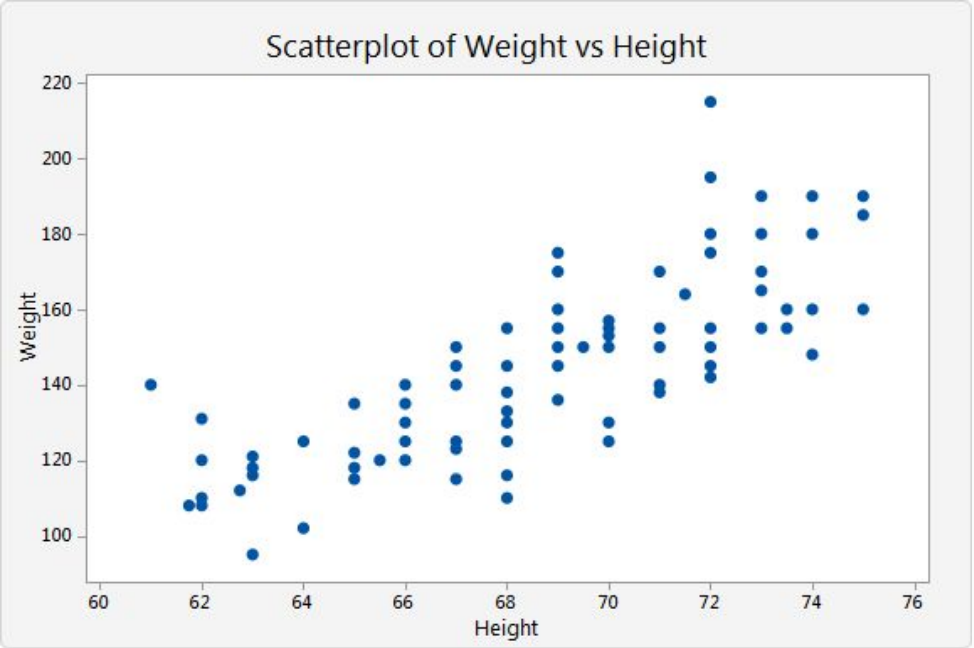
SoftStack Factory 2018

Scatter Plot

A Scatter Plot is a diagram consisting of 2 axes, and several points in an outcome space.

Each point represents a point in the outcome space

Great for finding trends in the data



SoftStack Factory 2018