Statistical Distributions

Presented by David John Baker
November 2019

FLATIRON SCHOOL

Statistical Distributions ?

- Distributions are to statistics as data structures are to programming
- Reflect assumptions about the underlying processes
- Allow us to make generalizations about data that we have not yet seen

//

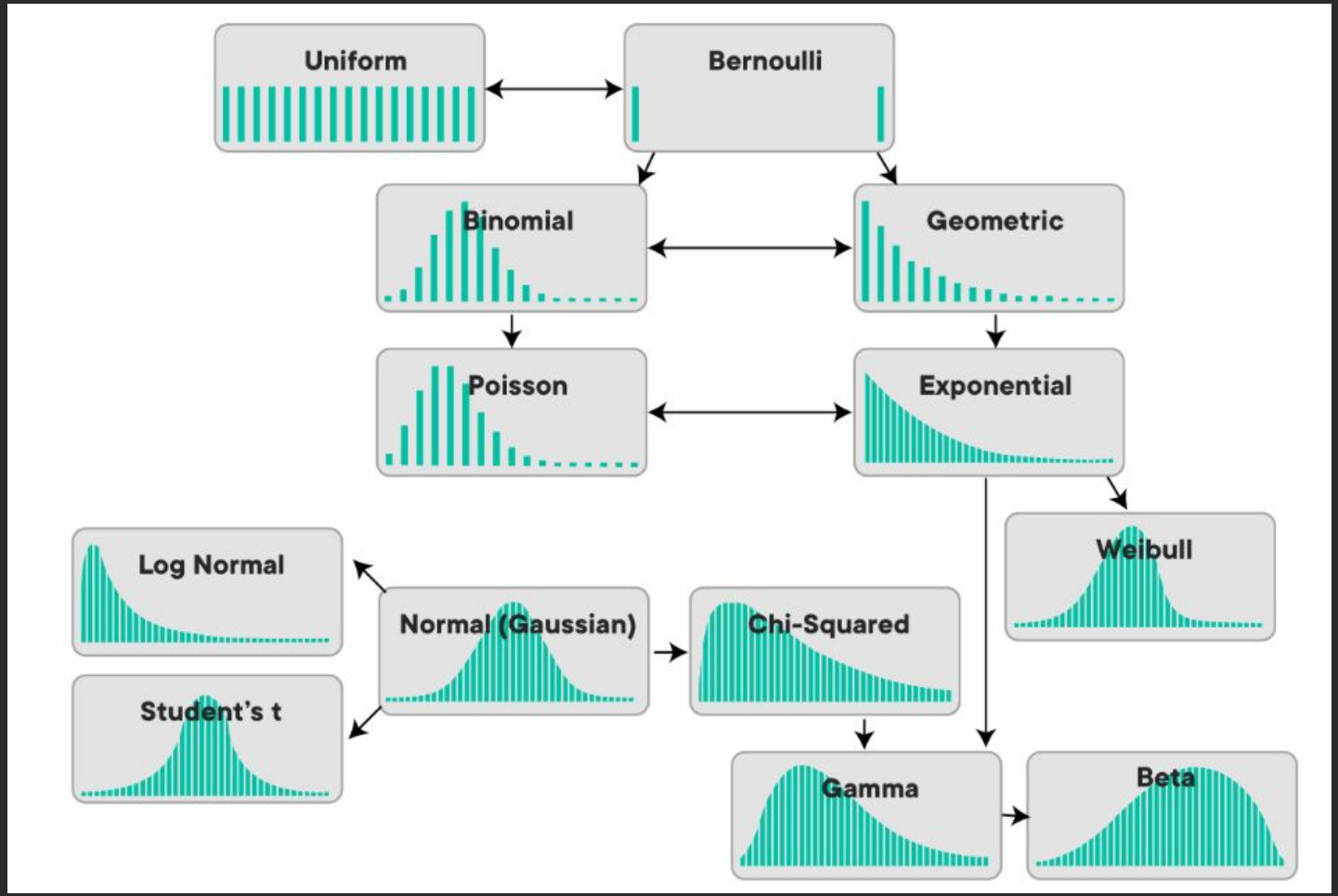
Lesson Goals

- Name and Describe Distributions
- Explain Difference between Continuous and Discrete Distributions
- Know PMF, PDF, CDF
- Describe Normal Distribution + (Standard Normal)
- Give examples of distributions

//

If I measure some quantity, what is the probability that I get one value rather than another?

If I flip a fair coin, what is the probability that I get heads rather than tails? If I measure the height of someone randomly selected from the U.K. population, what is the probability that I get someone who's 5'9"? What is the probability that I get someone who's 7'9"?

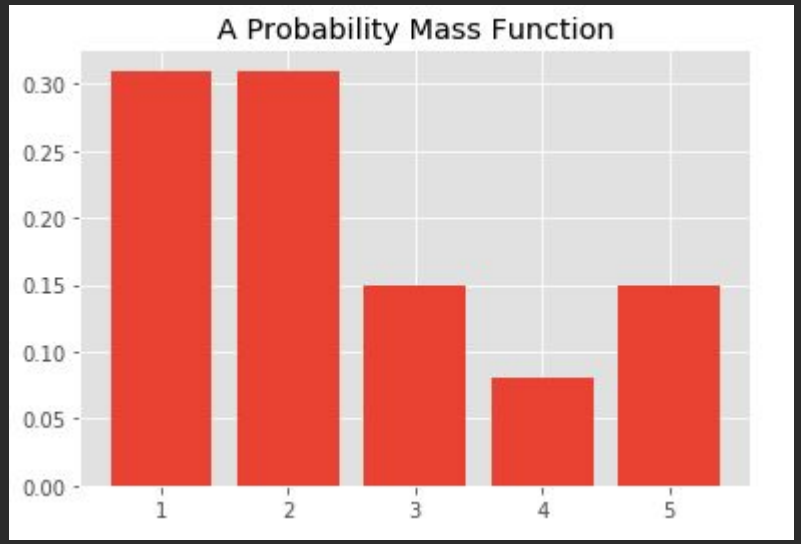

Uniform
Bernoulli
Binomial
Geometric
Poisson
Exponential
Weibull
Log Normal
Normal (Gaussian)
Chi-Squared
Student's t
Gamma
Beta

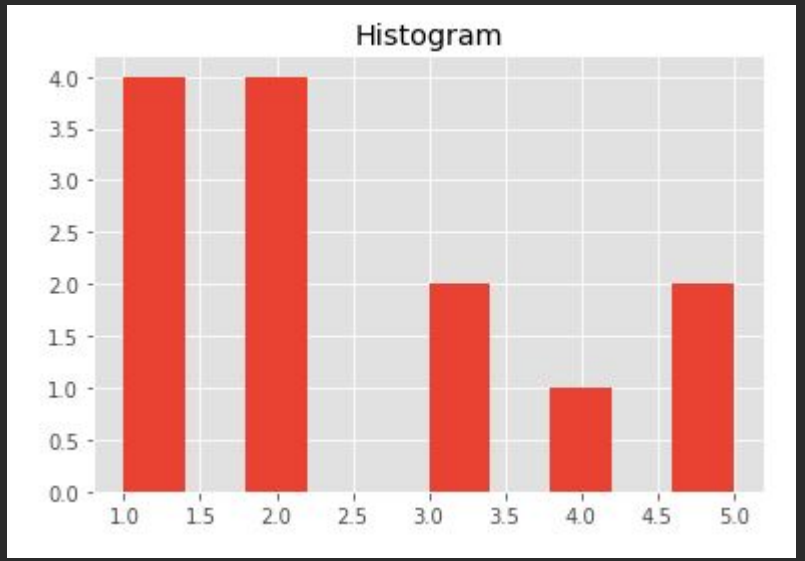
Discrete vs Continuous

- Discrete distributions have finite event space (ex: dice, categories of drawing balls out of an urn, species of animals)
    - Bernoulli
    - Poisson
    - Uniform
- Continuous distributions have infinite event space (ex: heights, time, financial data)
- Normal (Standard Normal)

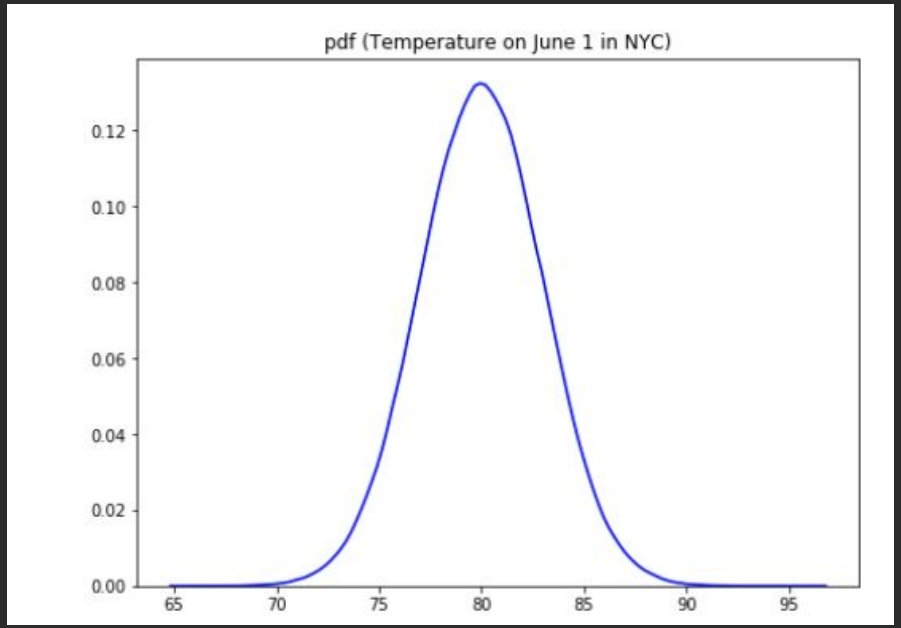
//

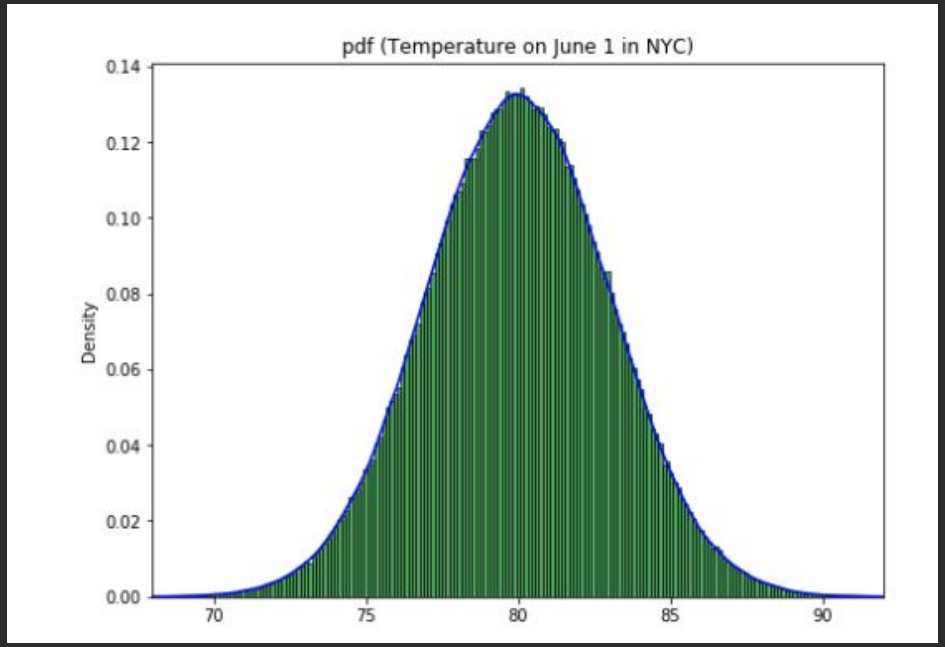
Probability Mass Functions (PMF)
(~A Normalized Barchart~)



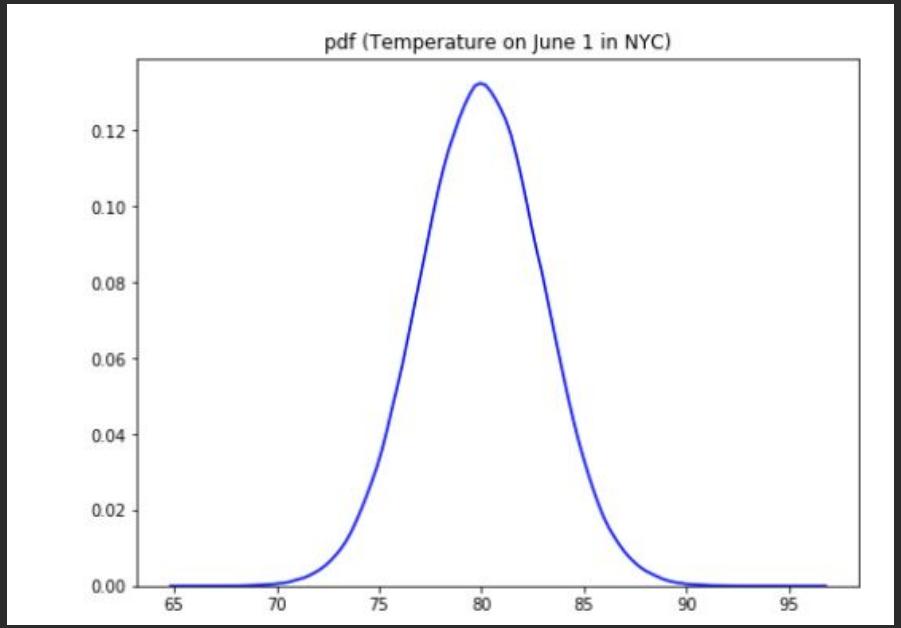


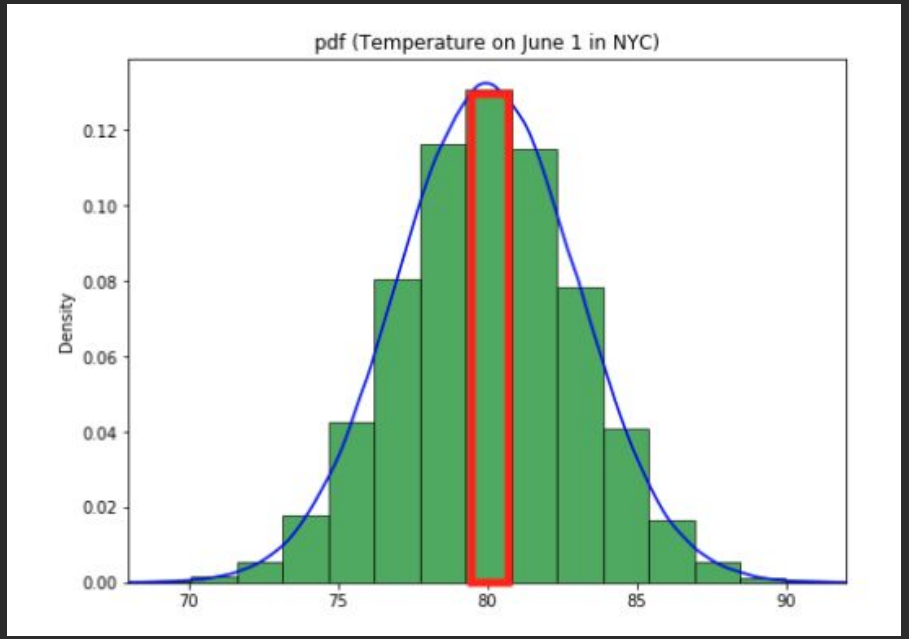
Probability Density Functions (PDF)
(A Normalized Histogram)



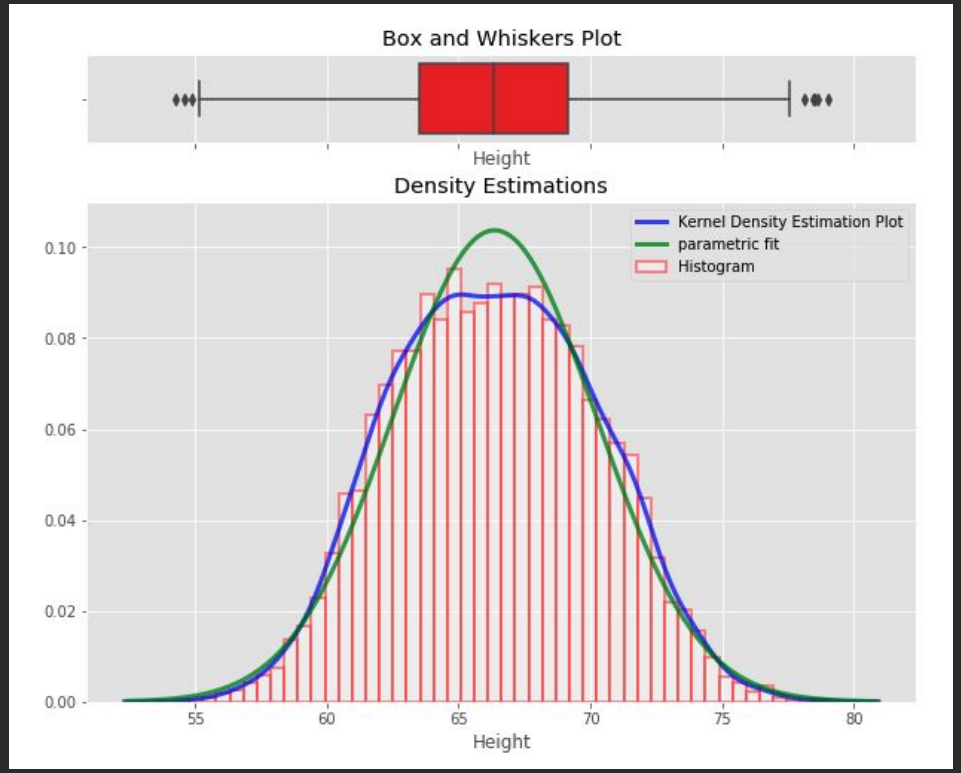


Probability Density Functions (PDF)
(A Normalized Histogram)

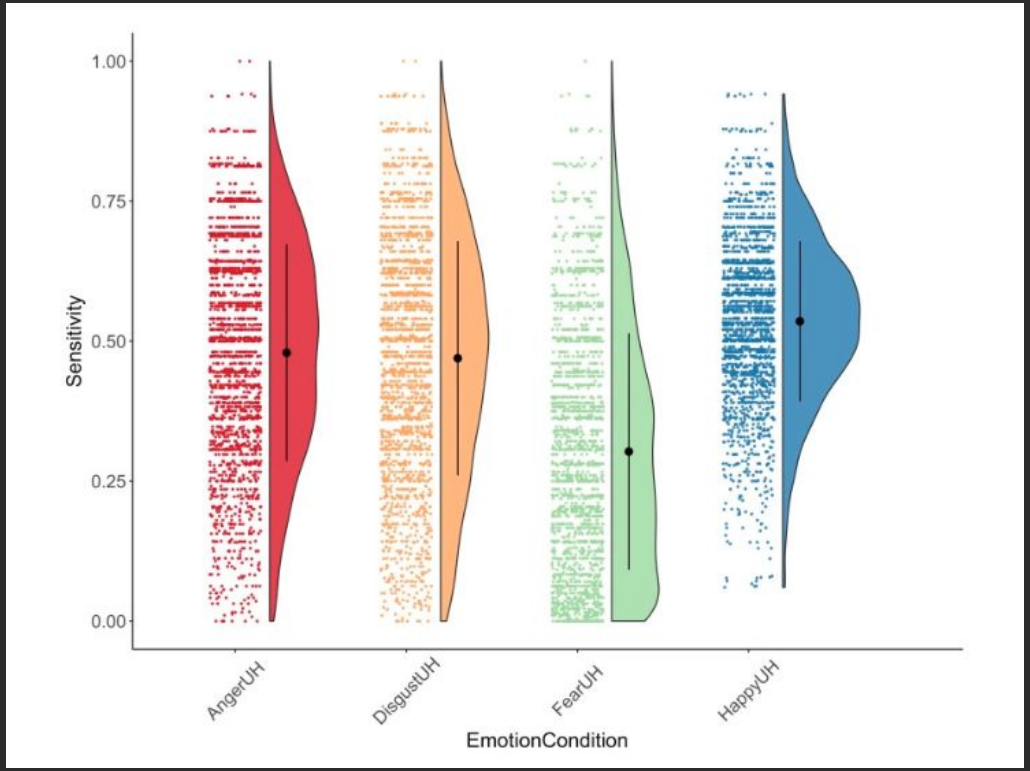




Probability Density Functions (PDF)
(A Normalized Histogram)

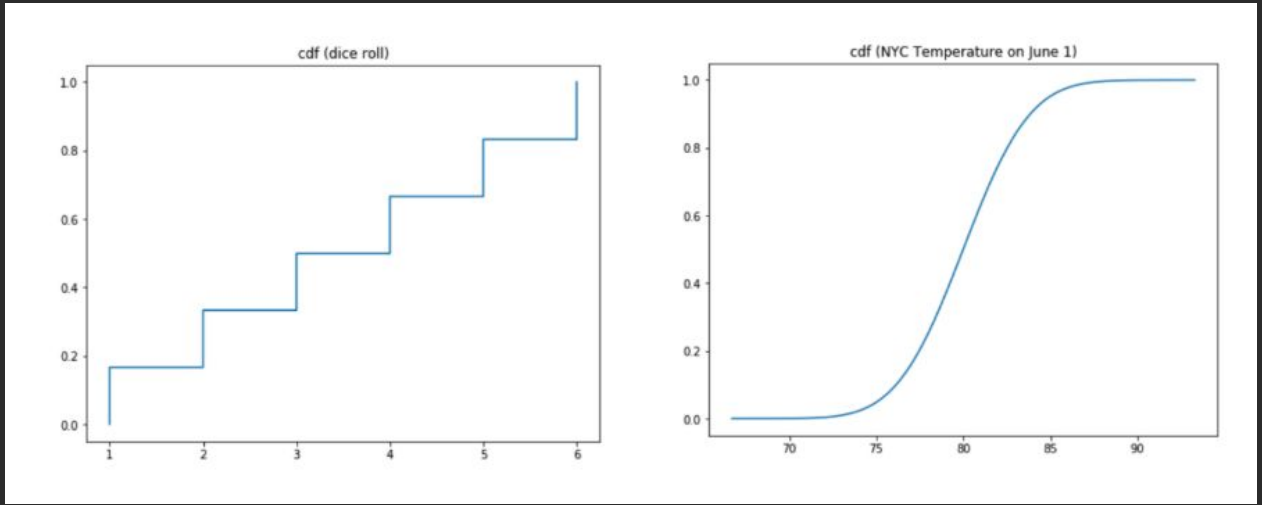


Probability Density Functions (PDF)
(A Normalized Histogram)



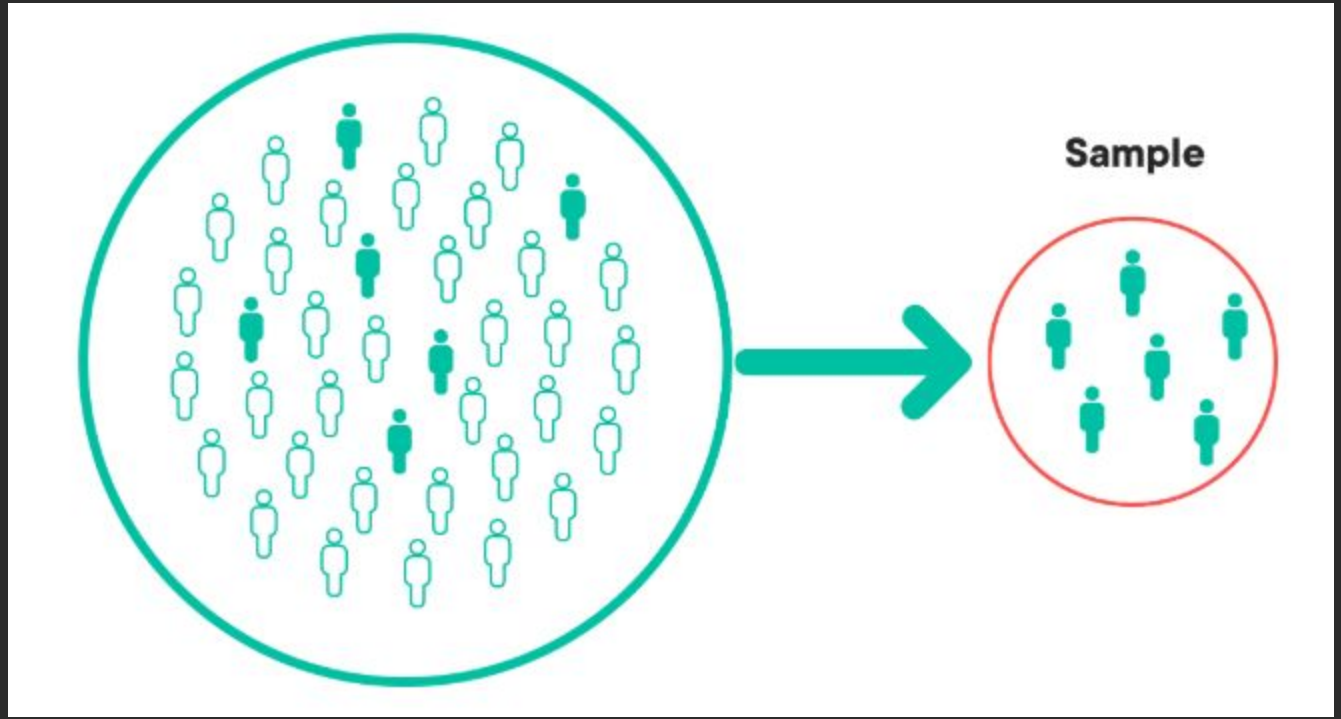
Problems with PMF and PDF?

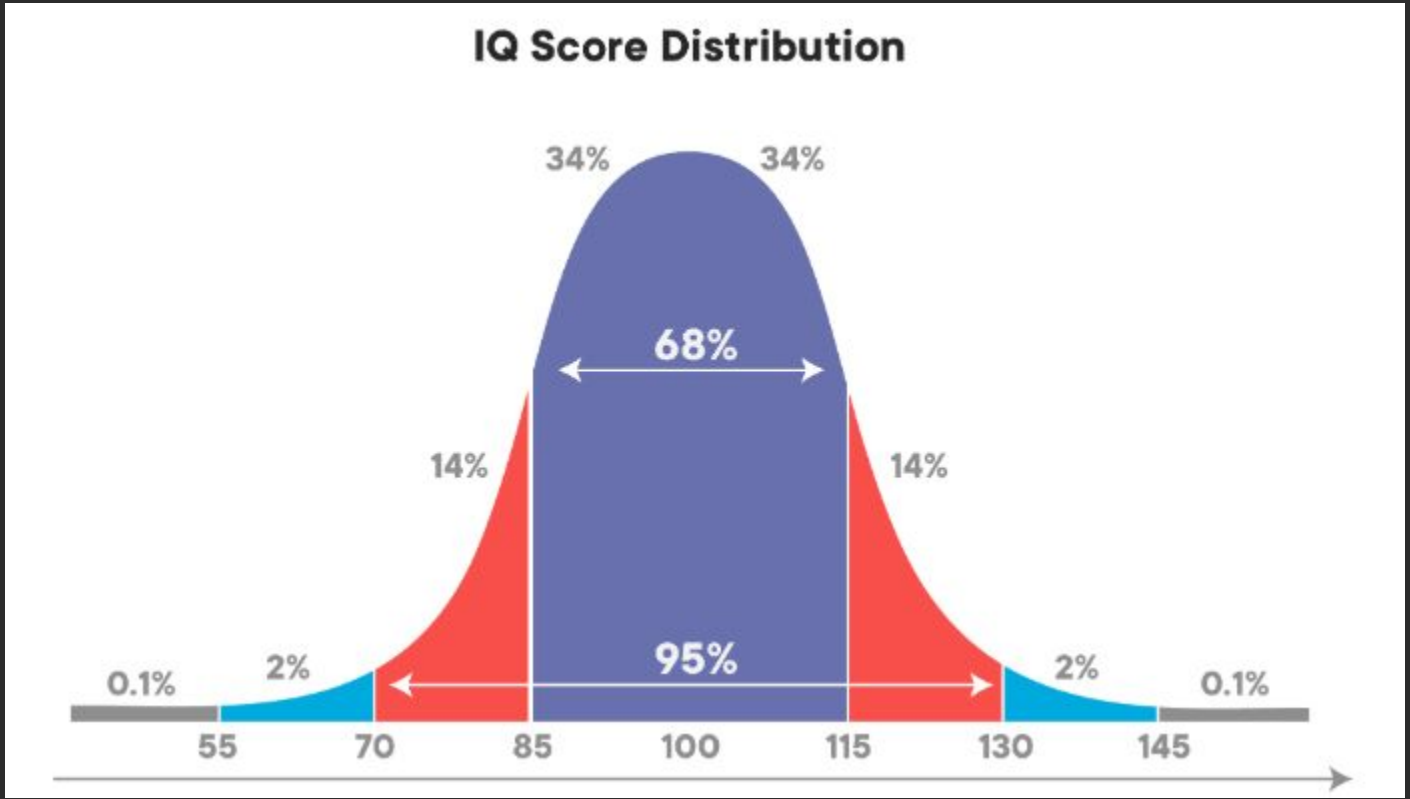
- Interpreting PMF, not to so bad?
- But what about if you wanted to say what is the probability of getting an exact value from a PDF? (a point probability)
- CDF!

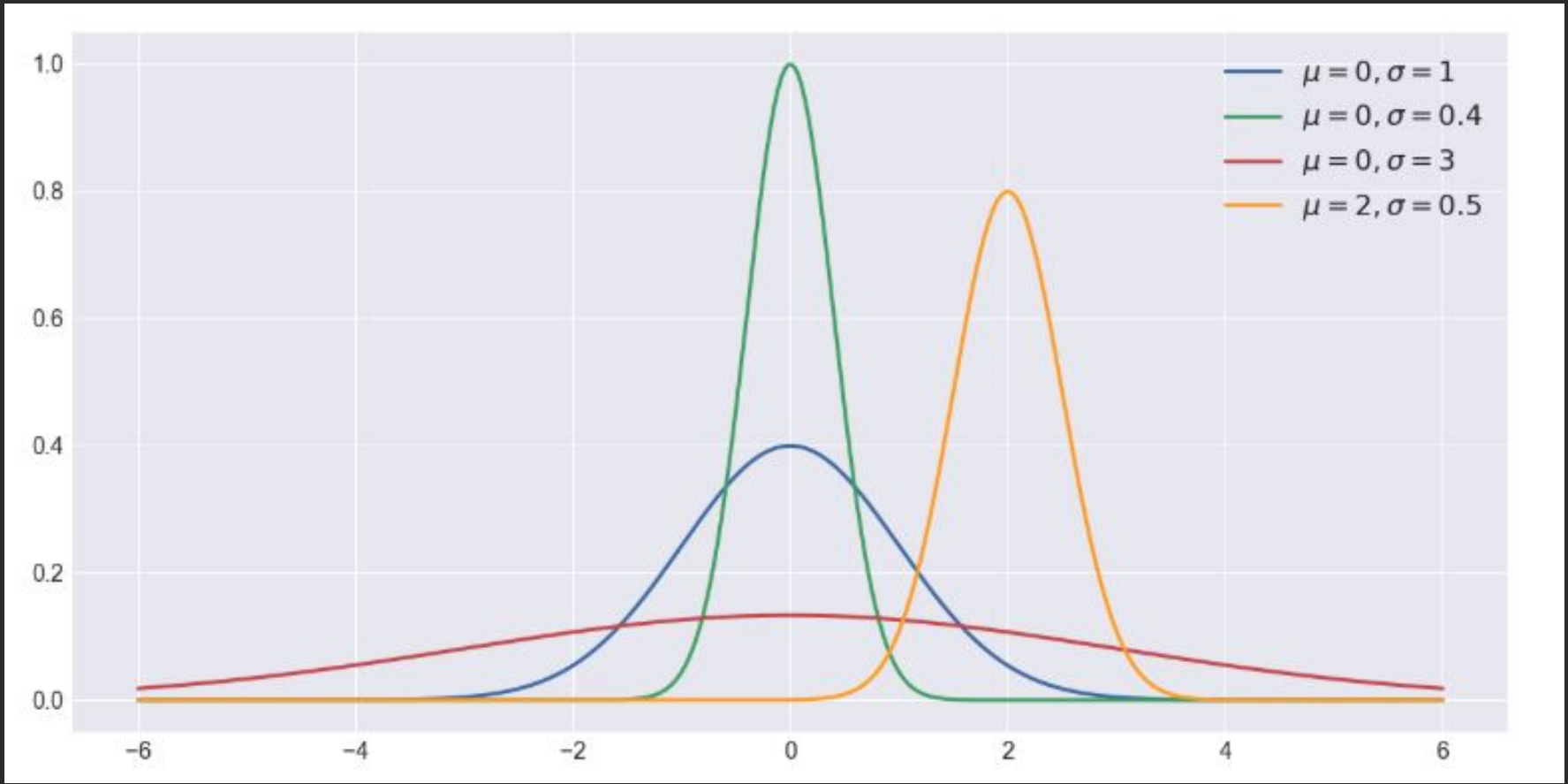


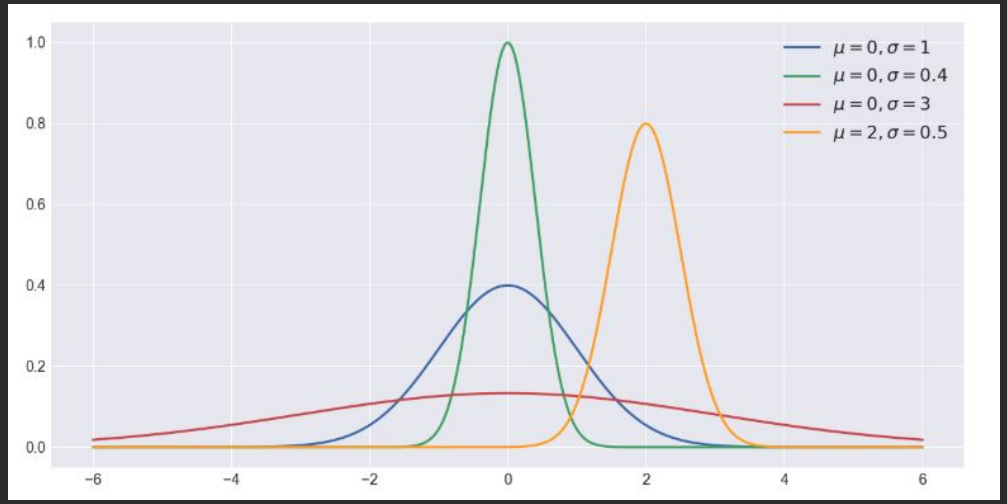
Normal Distributions

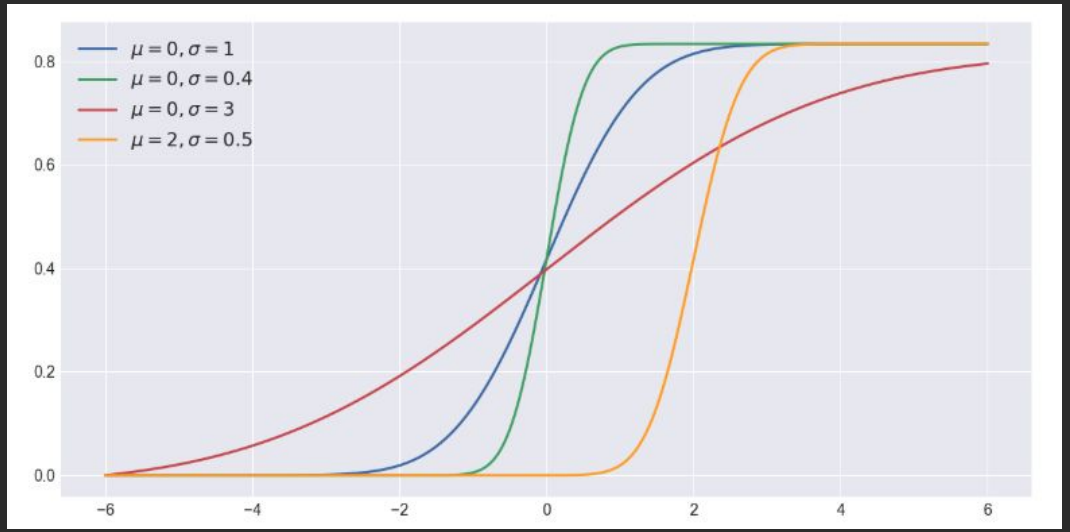
- One of most common distributions found in natural world
- Normal is not a value judgment
- Defined by a mean (mu) and standard deviation (sigma) and general symmetry
- Understanding the properties of the normal distribution will allow us to generalize about data we have not yet seen!
- Reflect on relationship between population, sample, and sampling distribution.

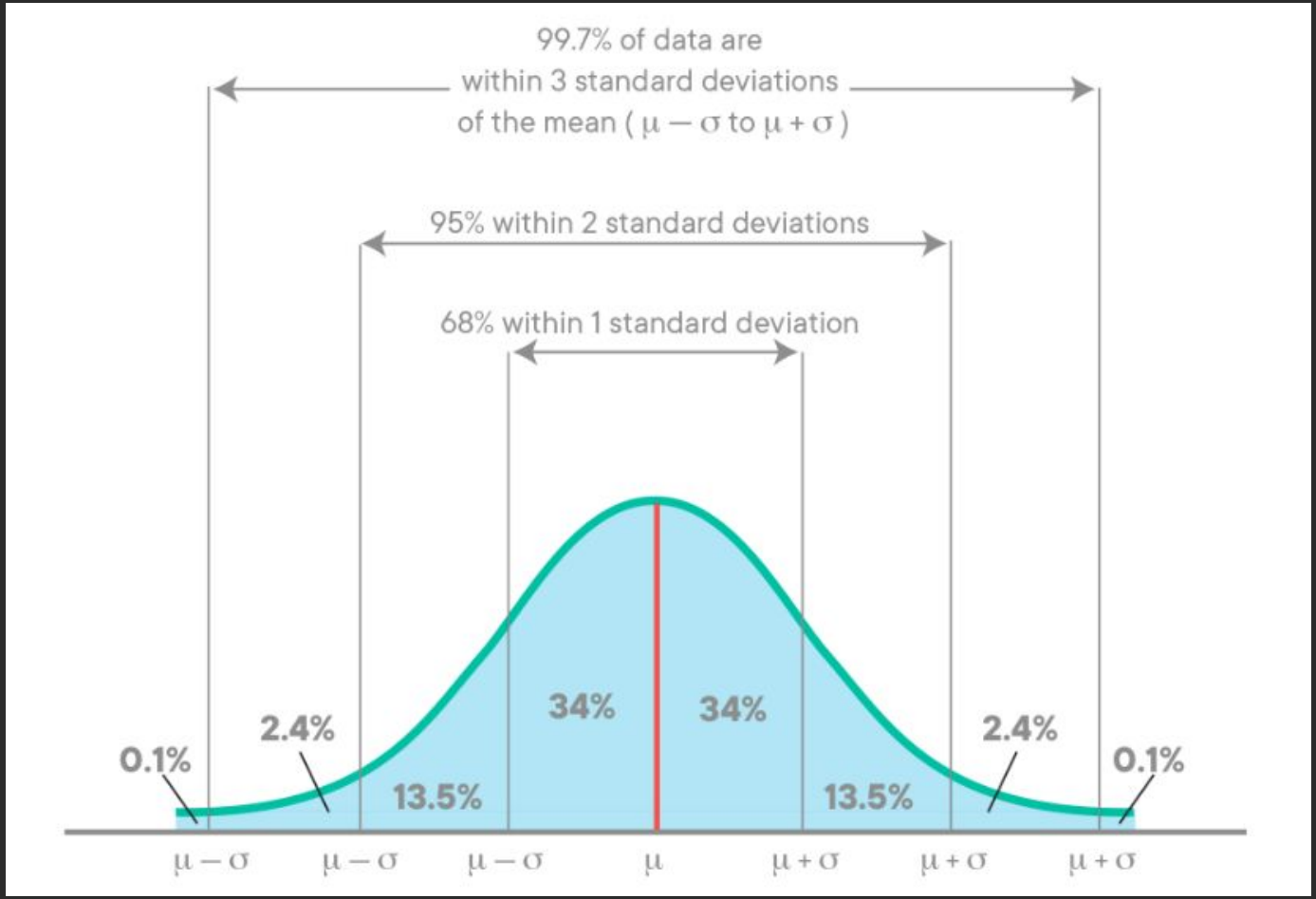

Sample


IQ Score Distribution
34%
34%
68%
14%
14%
95%
0.1%
2%
2%
0.1%
55
70
85
100
115
130
145


1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6
$\mu = 0, \sigma = 1$
$\mu = 0, \sigma = 0.4$
$\mu = 0, \sigma = 3$
$\mu = 2, \sigma = 0.5$


$\mu = 0, \sigma = 1$
$\mu = 0, \sigma = 0.4$
$\mu = 0, \sigma = 3$
$\mu = 2, \sigma = 0.5$
1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6


$\mu = 0, \sigma = 1$
$\mu = 0, \sigma = 0.4$
$\mu = 0, \sigma = 3$
$\mu = 2, \sigma = 0.5$
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6

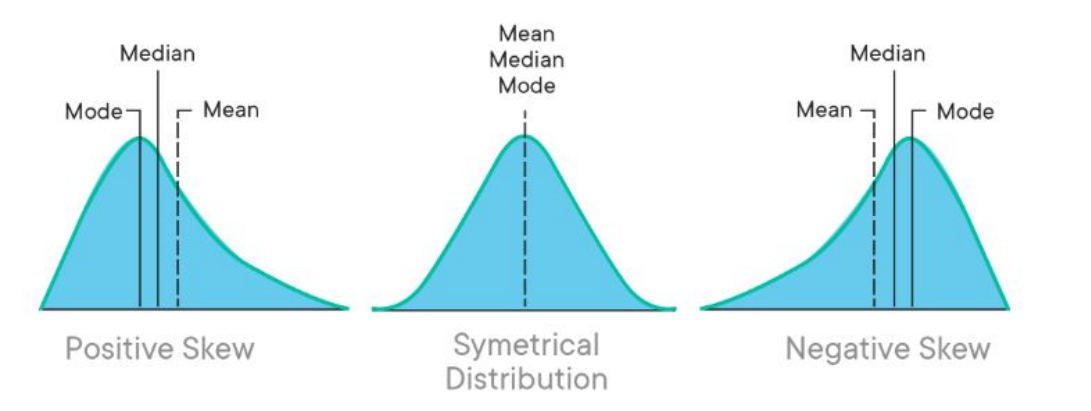

99.7% of data are
within 3 standard deviations
of the mean (μ − σ to μ + σ)
95% within 2 standard deviations
68% within 1 standard deviation
0.1%
2.4%
13.5%
34%
34%
13.5%
2.4%
0.1%
μ − σ
μ − σ
μ − σ
μ
μ + σ
μ + σ
μ + σ

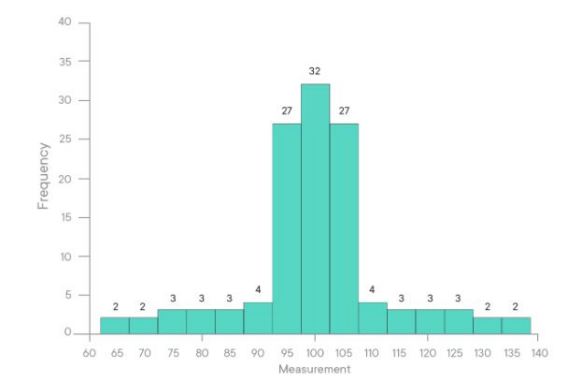
Central Limit Theorem

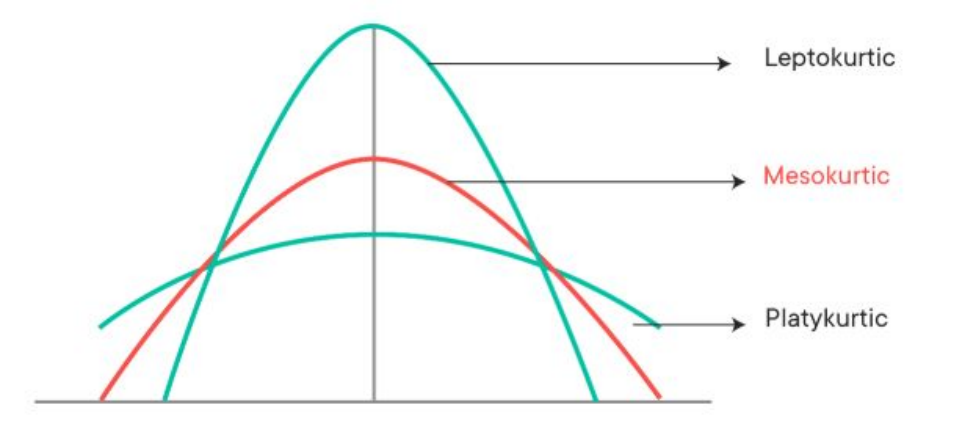
- When you add a large number of independent random variables, irrespective of the original distribution of these variables, their sum tends towards a normal distribution.

//

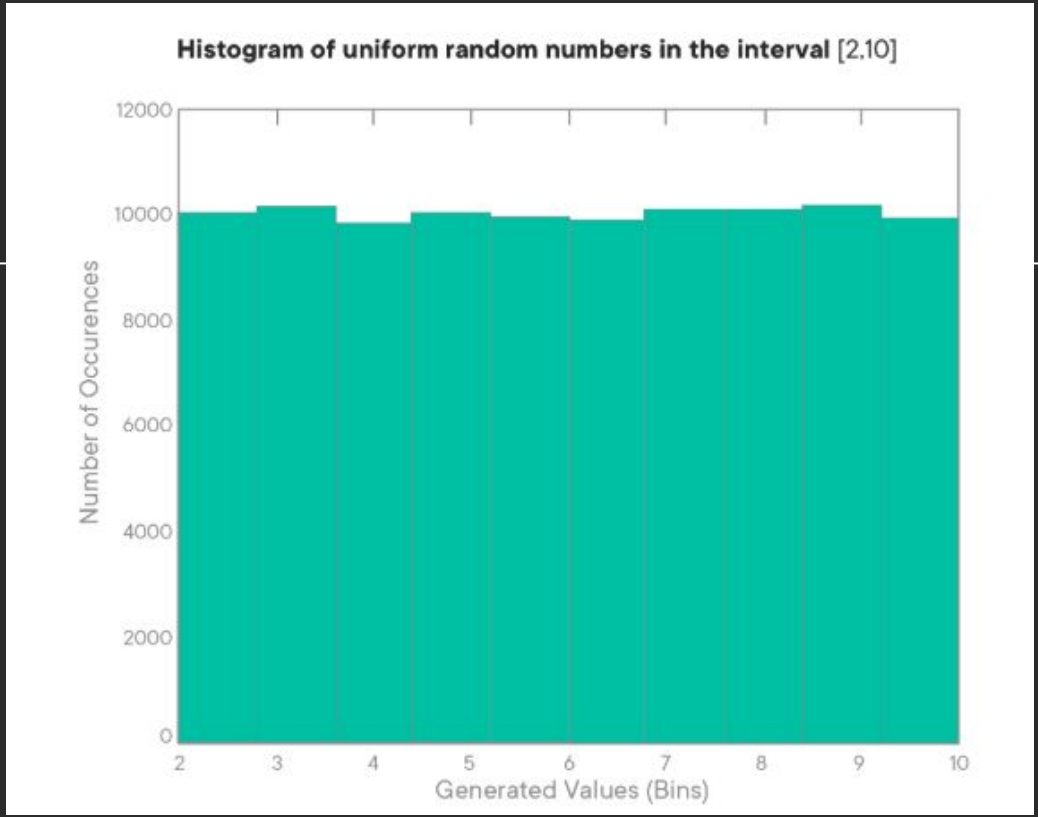
Skew and Kurtosis





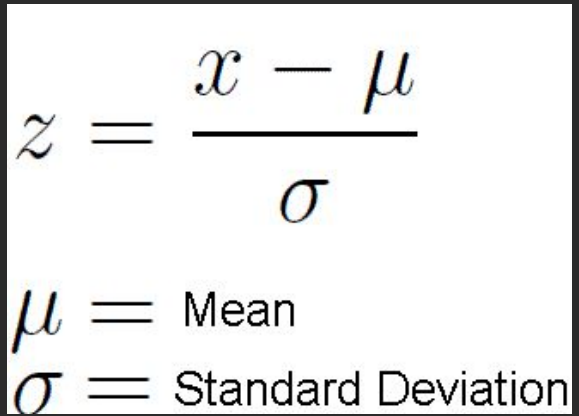


Uniform Distribution



Practice

- In the last portion of class, want to do a class activity
    - Collect heights of all people in class in CM
    - Create small analysis script that is going to import the data
    - First create a plot that plots the raw data
    - Check if that data is normal (visual inspection)
    - Convert all scores to z scores



Checking For Understanding

- Explain discrete vs continuous distributions
- Explain difference between PDF, PMF, CDF
- Explain normal distributions
- Standard Normal Distributions
- Z Scores