Probability Distributions

Demystifying “Demystifying Probability”

What are key takeaways?

What is likely to be most useful?

What is a probability distribution?

A probability distribution is a table or an equation that links each outcome of a statistical experiment with its probability of occurrence.

What is it good for?

We can model real world data on distributions to make predictions.

We can test for statistical significance.

Discrete Distributions

Uniform

Rolling die
Selecting student at random
Selecting pairs at random

Cumulative Distribution

Discrete Distributions

Bernoulli Experiment/Distribution

RPS, coin flip ($p = .5$)

Free Throws ($p = .905$)

I.I.D.

Discrete Distributions

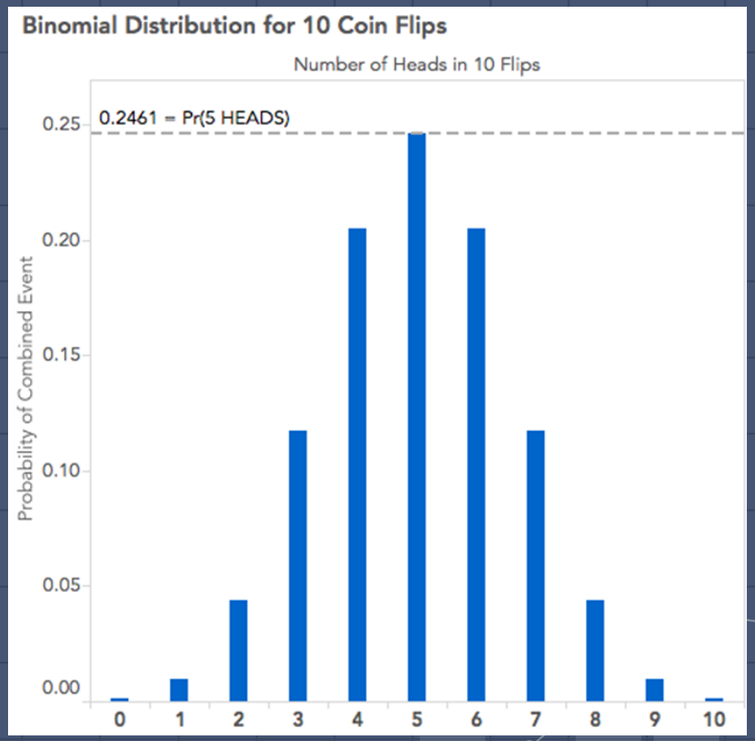
Binomial Distribution

$(p + q)^n$

Probability of k successes in n trials

RPS wins

Free throws made



Discrete vs Continuous

Outcomes are whole numbers

- Binomial
- Poisson
- Geometric
- Hypergeometric
- Negative Binomial
- Zeta (Zipf)

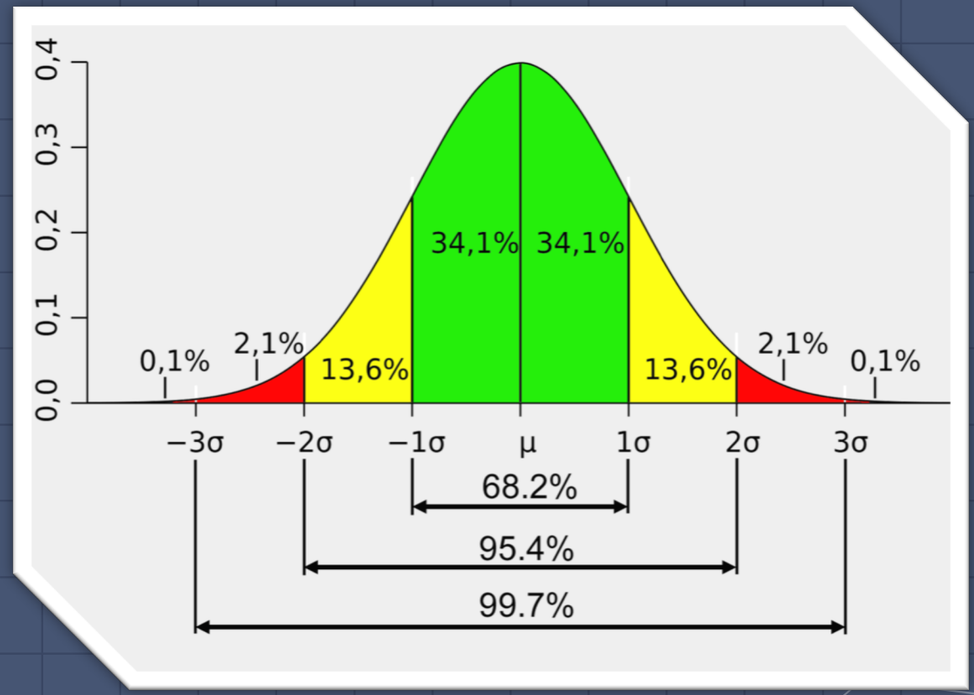
Outcomes can be all Reals

- Normal
- Exponential
- Pareto
- Beta/Gamma
- Student-t
- χ^2 (Chi Squared)

Continuous Distributions

Normal (Gaussian)

Common in nature
Heights/Weights
SAT scores
Sample Means
Error/Residuals SHOULD be

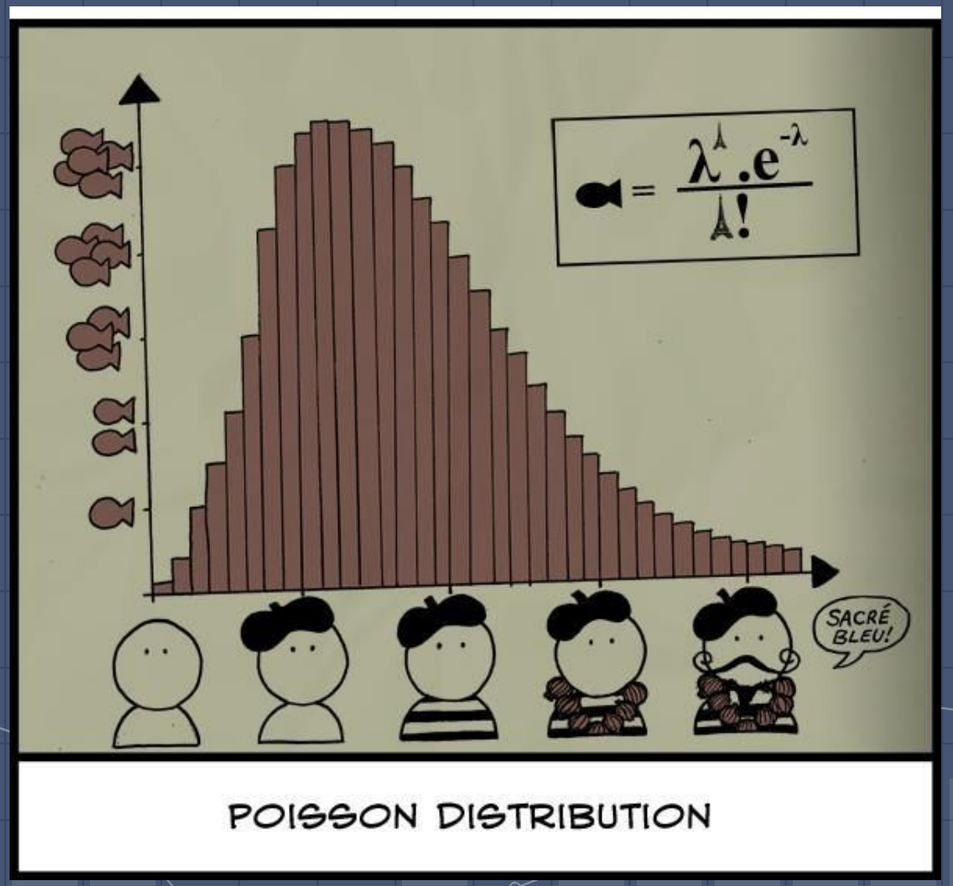


Discrete Distributions

Poisson

Probability of successes in given time interval (λ)

Expected phone calls
Meteor impacts/earthquakes
Mutations in DNA
Insurance claims

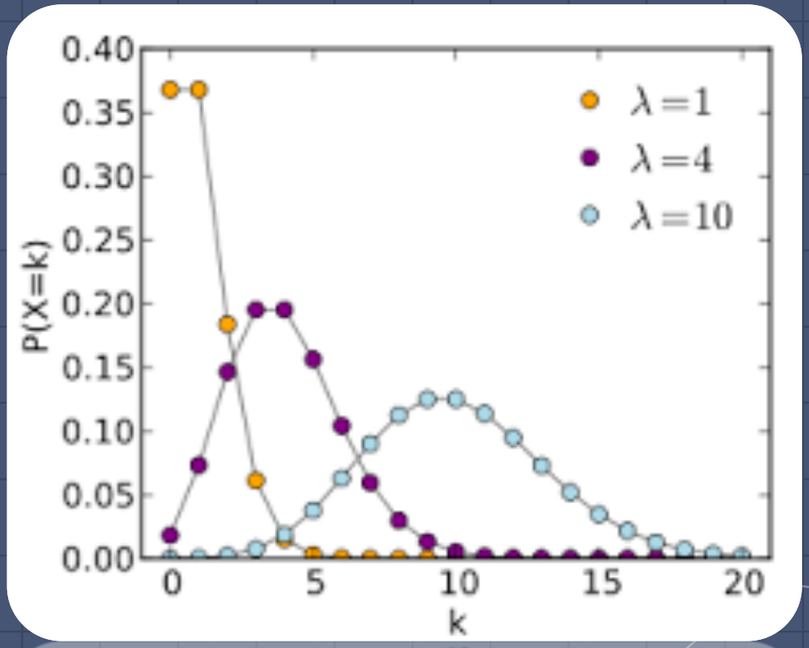


Discrete Distributions

Poisson

Number of expected successes per time unit (λ)

Expected phone calls
Meteor impacts/earthquakes
Mutations in DNA
Insurance claims

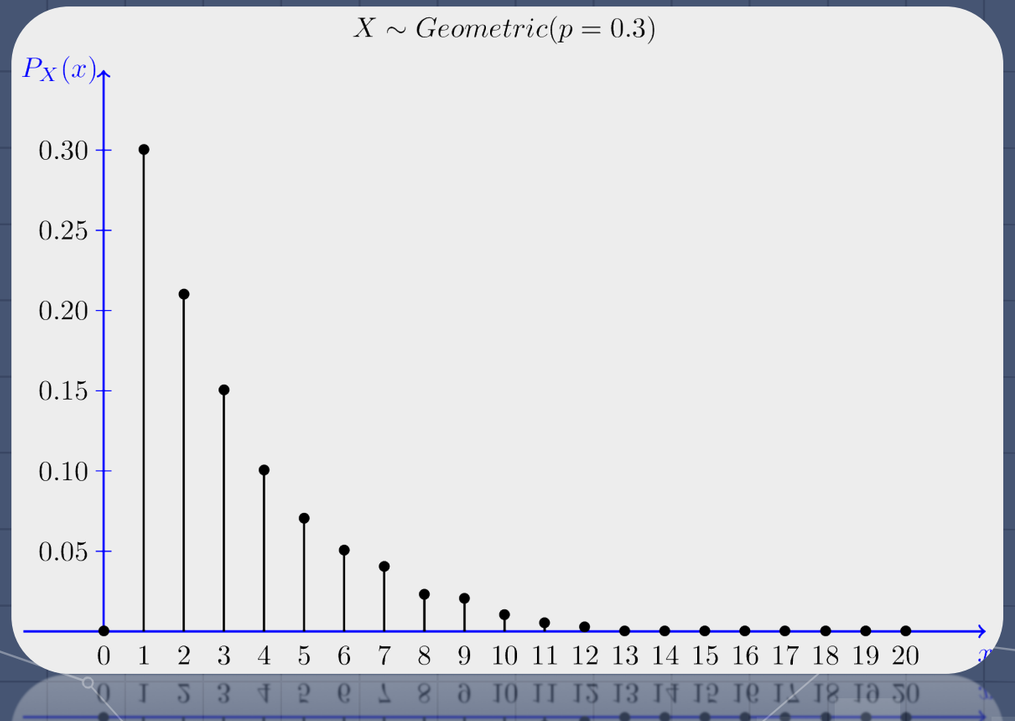


Discrete Distributions

Geometric

Probability of k trials before first success (p)

First boy/girl
First RPS win
First job offer

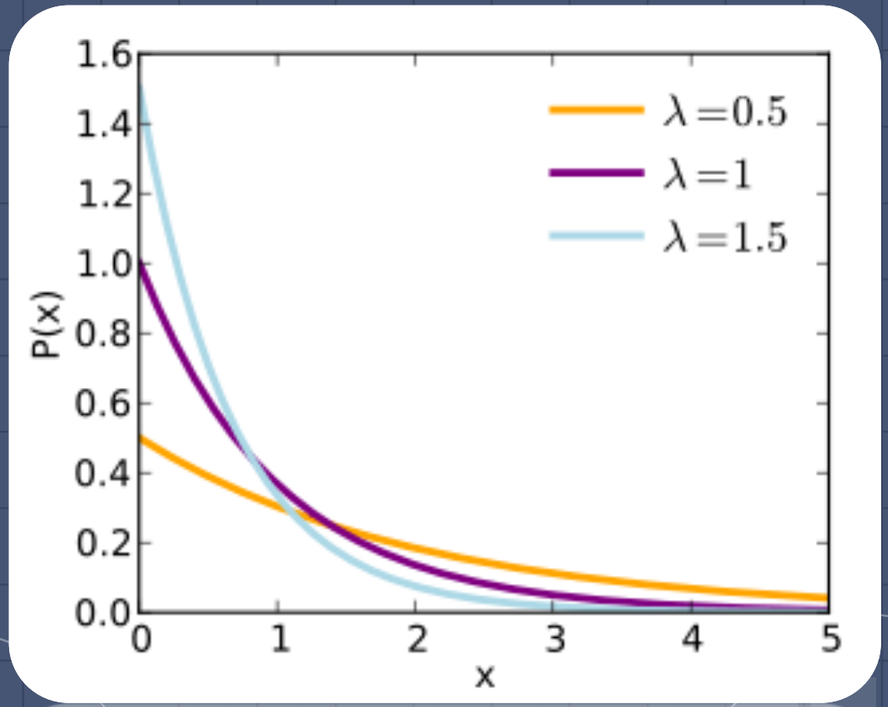


Continuous Distributions

Exponential

Probability of success after some time

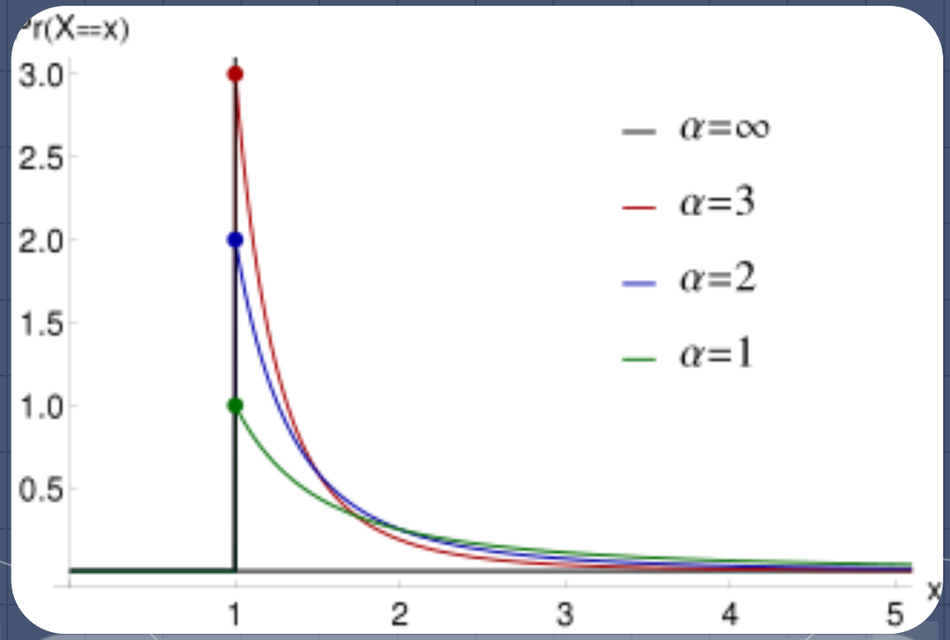
Like geometric for time
Time it takes to be served
Size of raindrop
Time to finish a test



Continuous Distributions

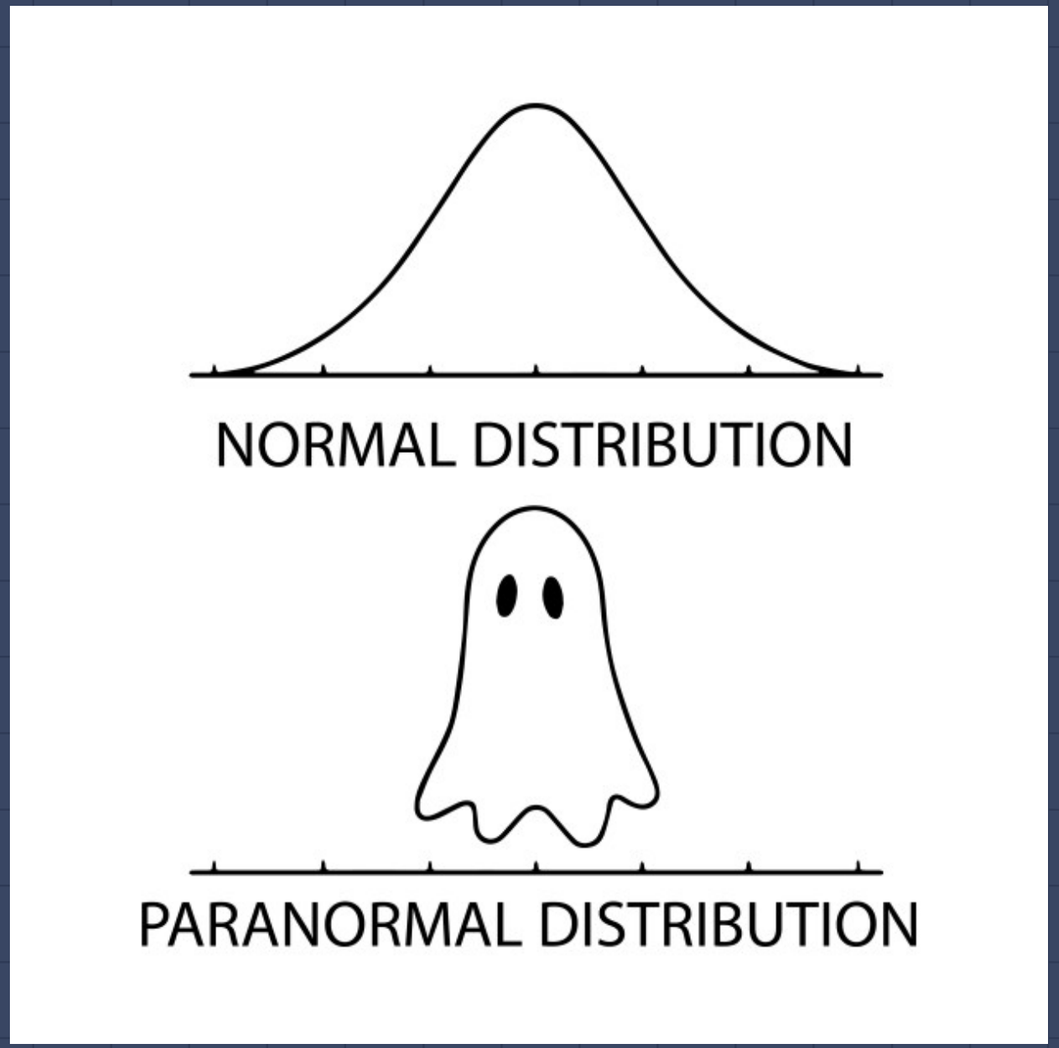
Pareto

80/20 Rule
Income
Housing Price
Rainfall



THANKS!

Any questions?



Works Cited

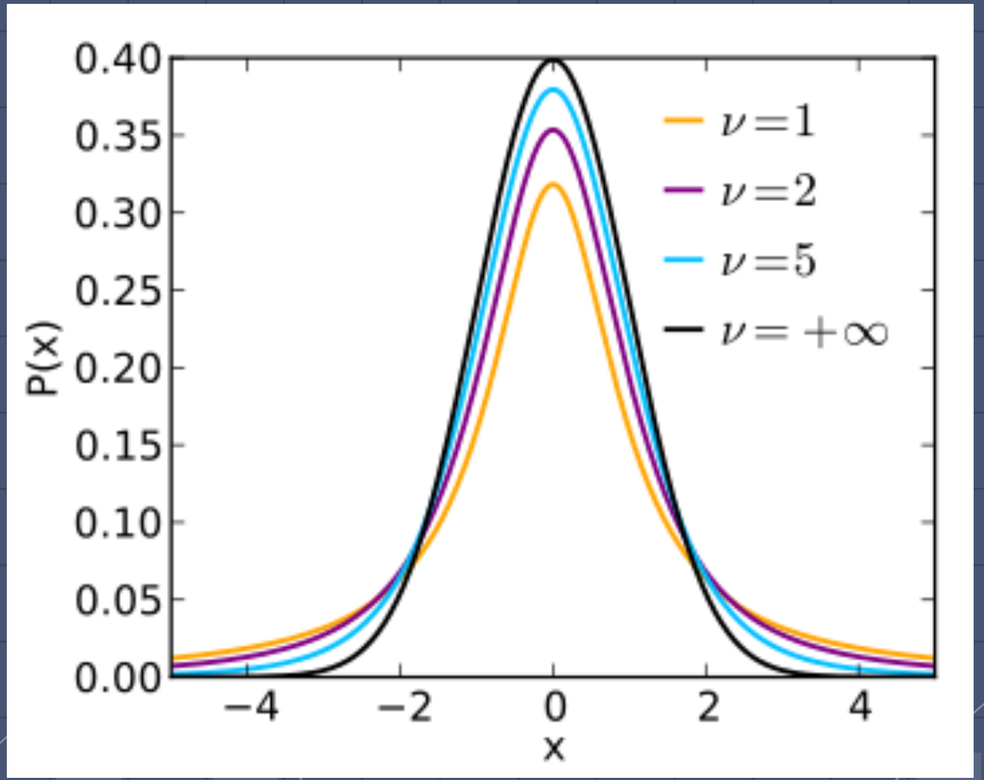
Wikipedia

Appendices

Statistical Significance

Student's t

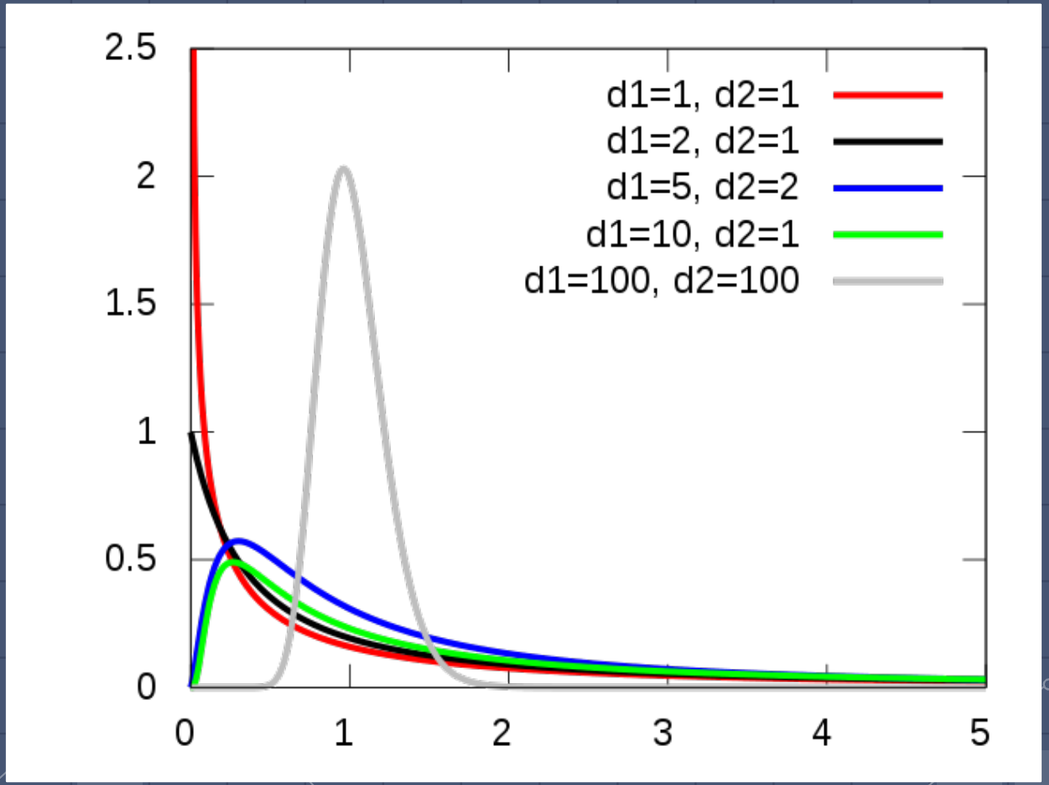
Is a sample mean likely?
No variance, large sample
Regression Fit
Null: slope is zero
p value < .05
coefficient different from 0



Statistical Significance

F

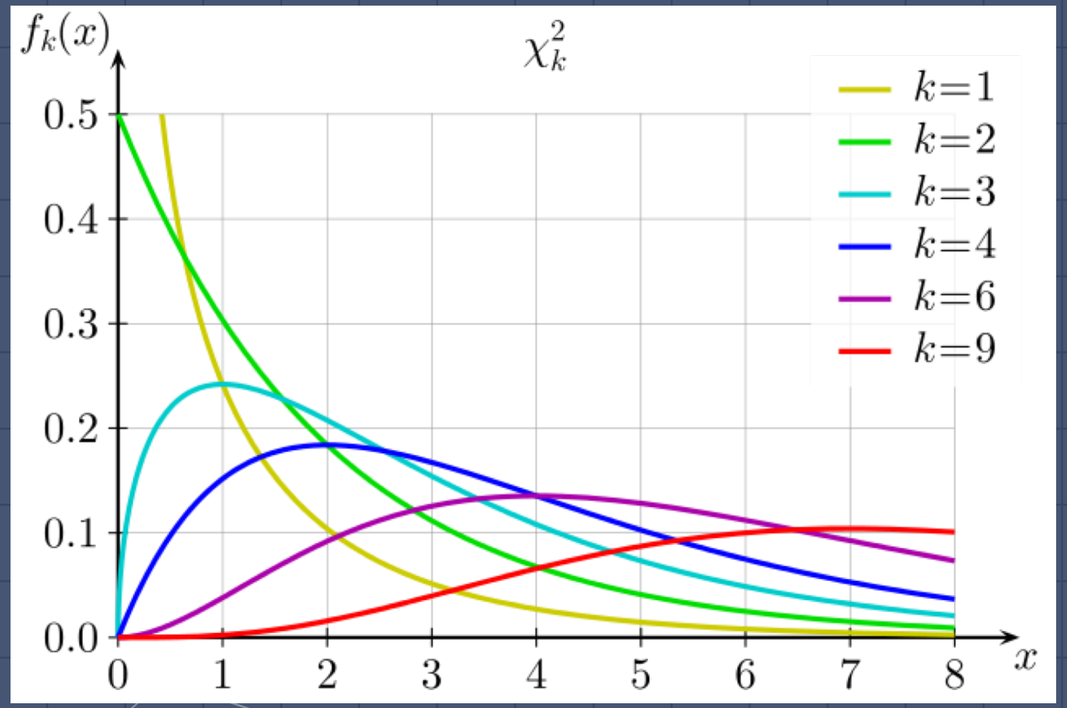
Compare variances
Null: Variances equal
F = 1
In regression:
Null: intercept only model
P < .05, good fit
Proves validity of R^2



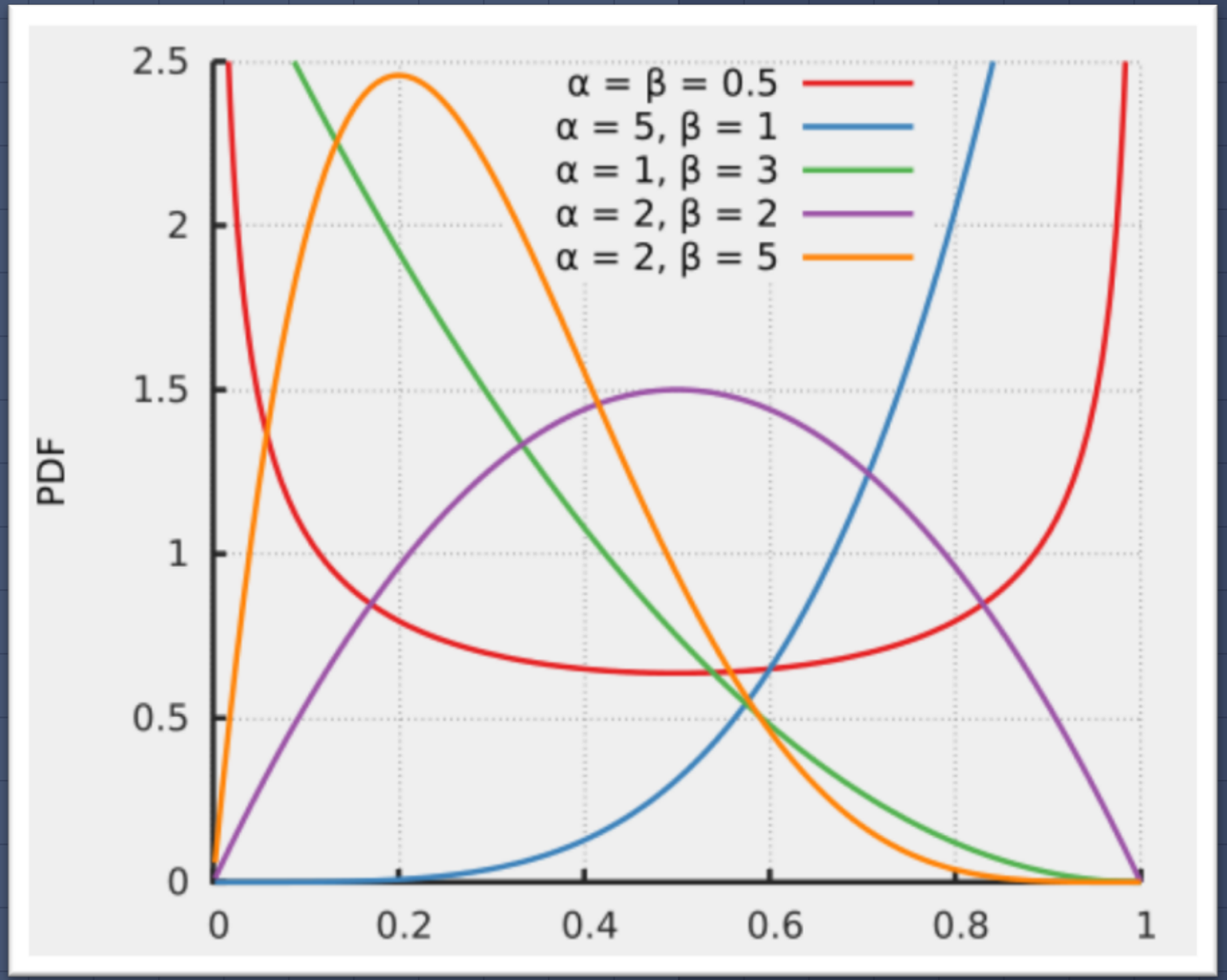
Statistical Significance

χ^2 (**Chi Squared**)

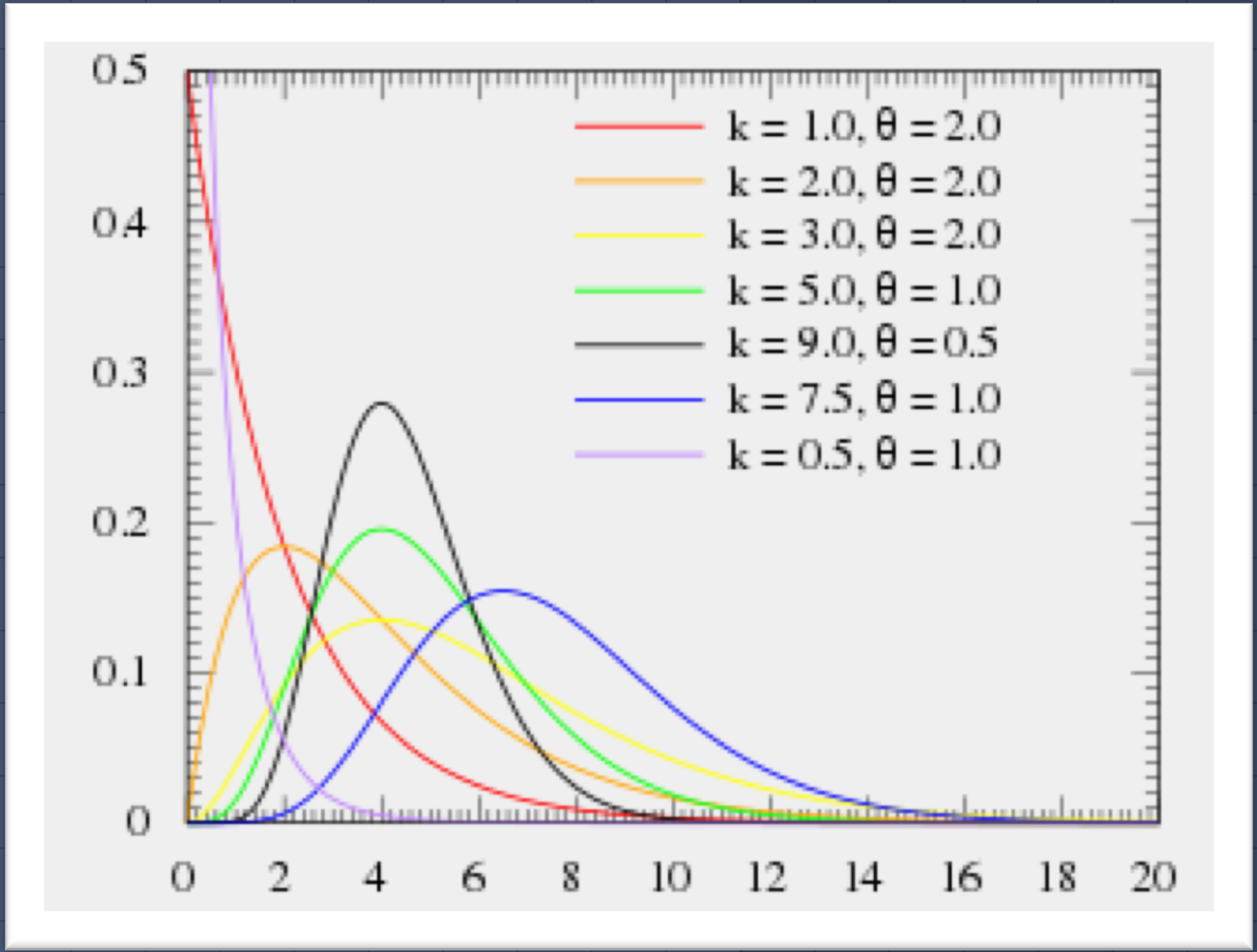
Goodness of Fit
Independence
Homogeneity of Proportions



Beta



Gamma


0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20
k = 1.0, θ = 2.0
k = 2.0, θ = 2.0
k = 3.0, θ = 2.0
k = 5.0, θ = 1.0
k = 9.0, θ = 0.5
k = 7.5, θ = 1.0
k = 0.5, θ = 1.0