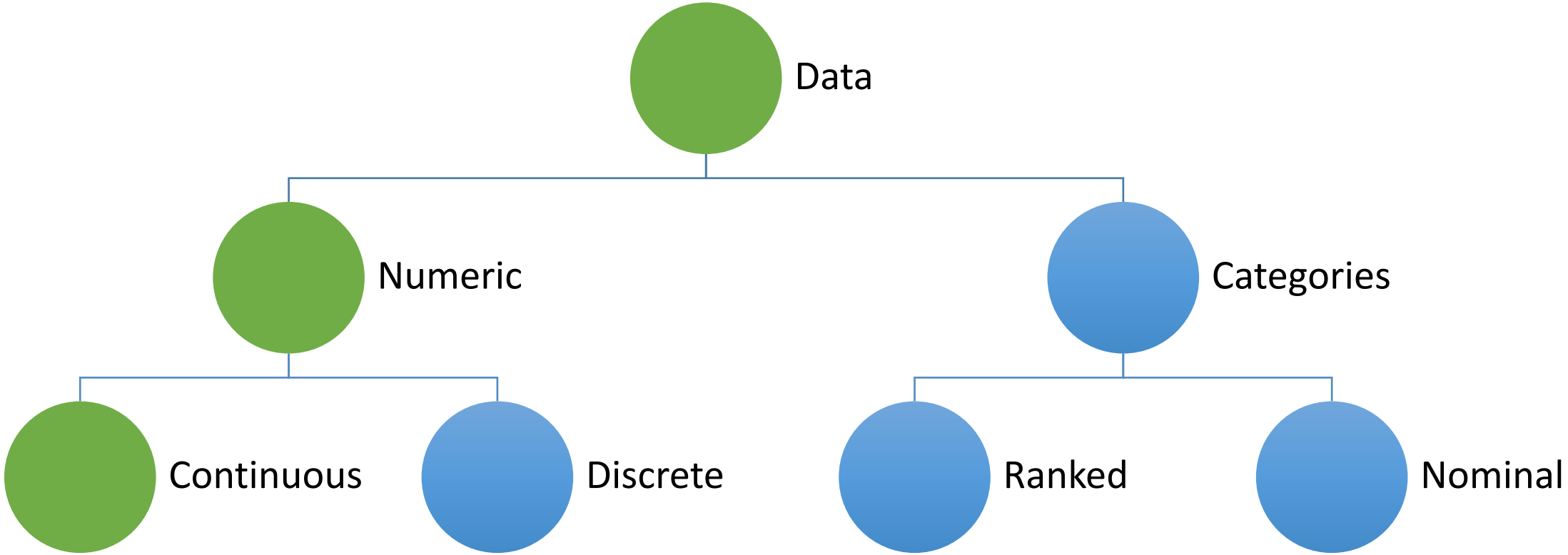
Statistics with Spa  ows

Lecture 3

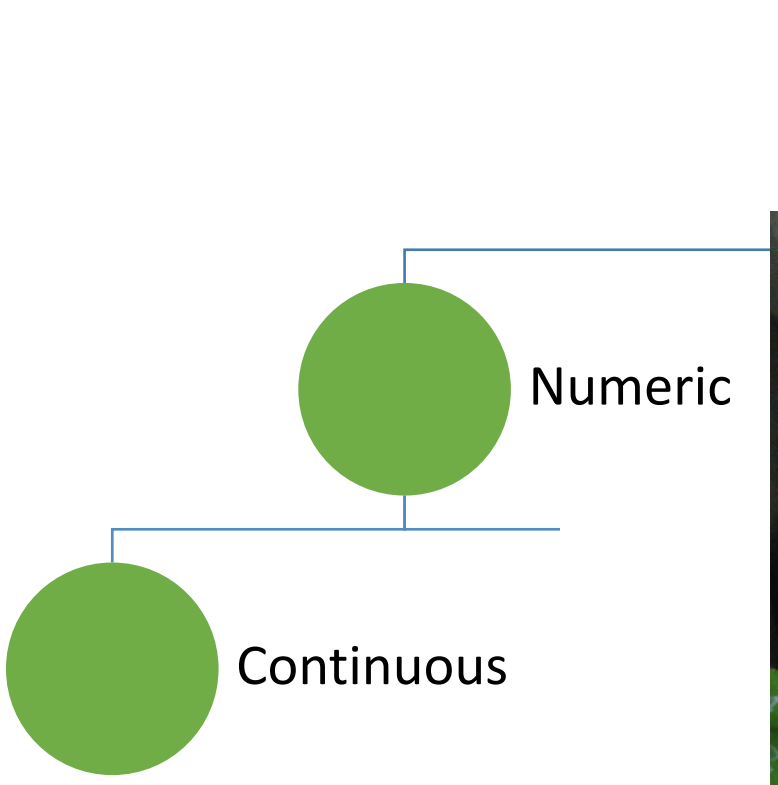
Julia Schroeder

Julia.schroeder@imperial.ac.uk

Data types

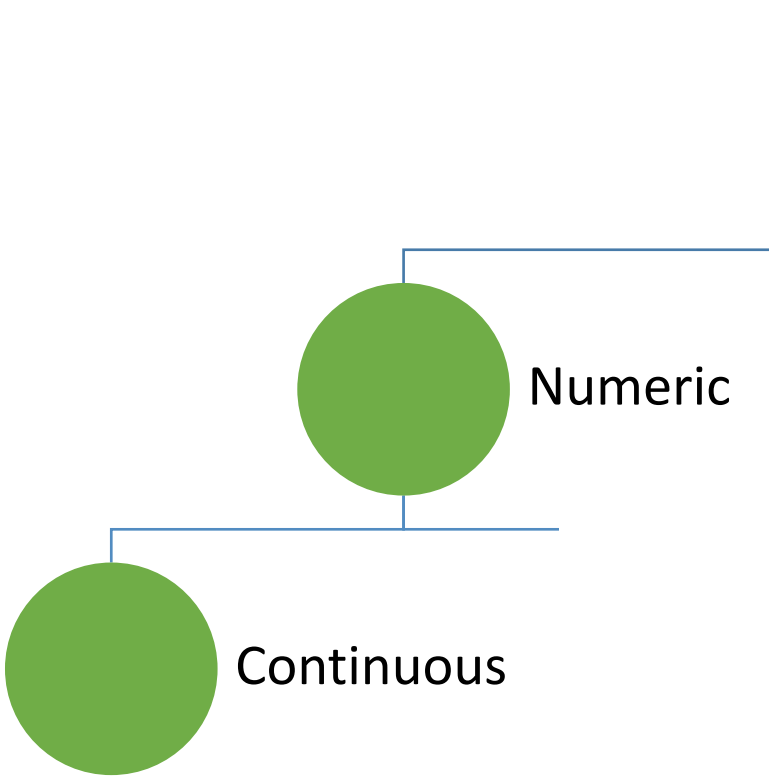

Data
Numeric
Categories
Continuous
Discrete
Ranked
Nominal

Data types

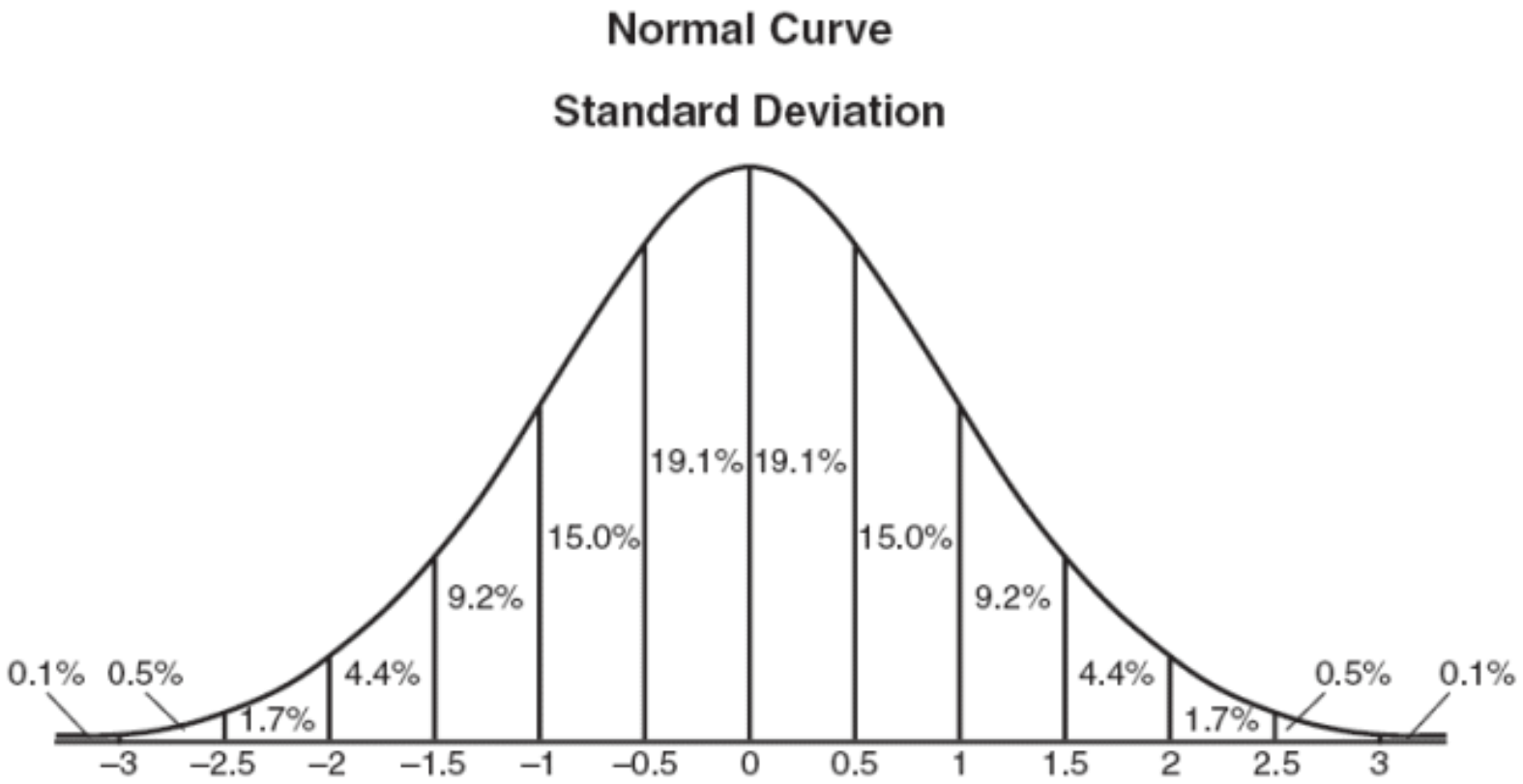


0.3, -5.985,
189.652589,
ect

Data types

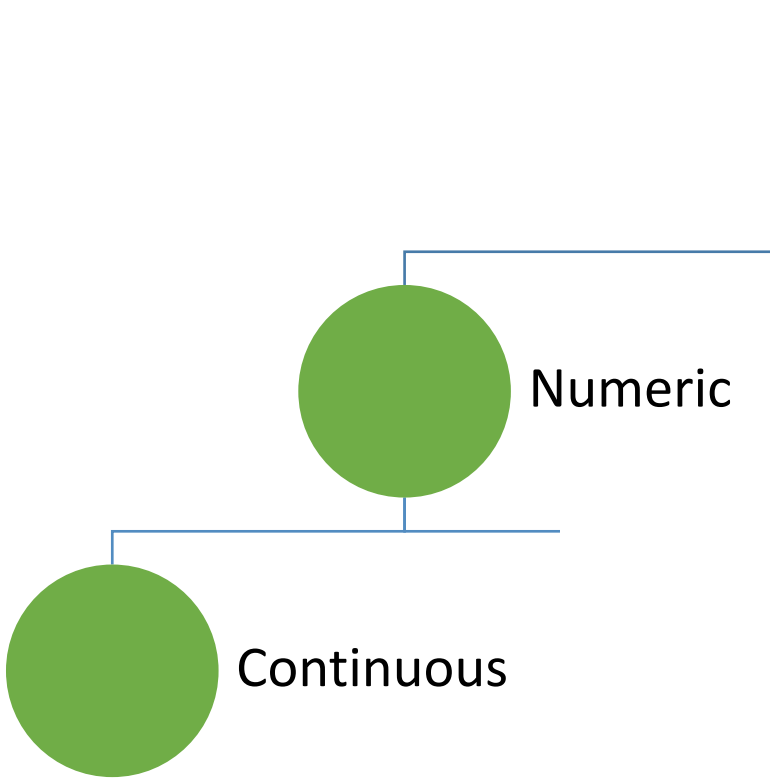


Continuous values
Sampled from normal distribution

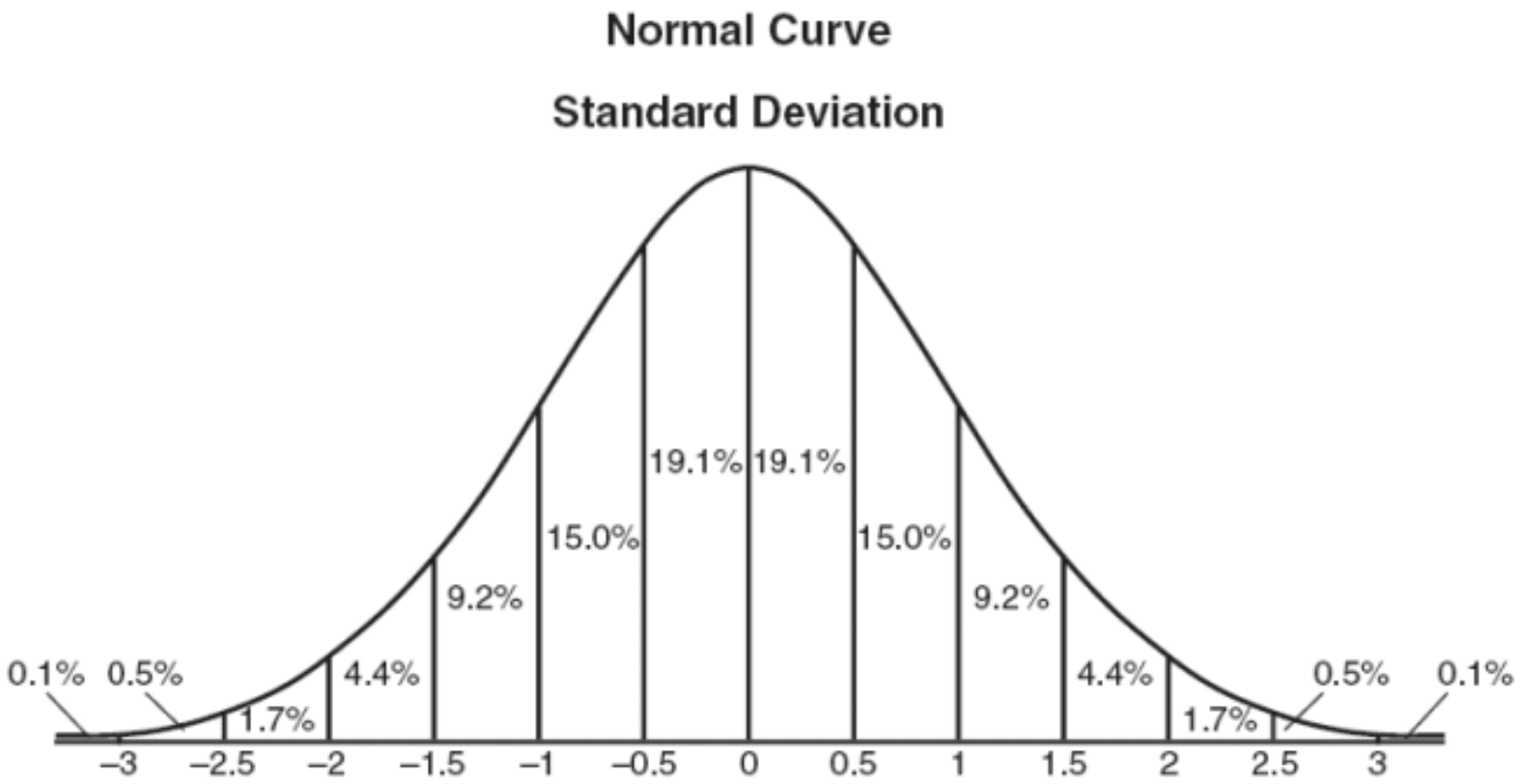


Data types

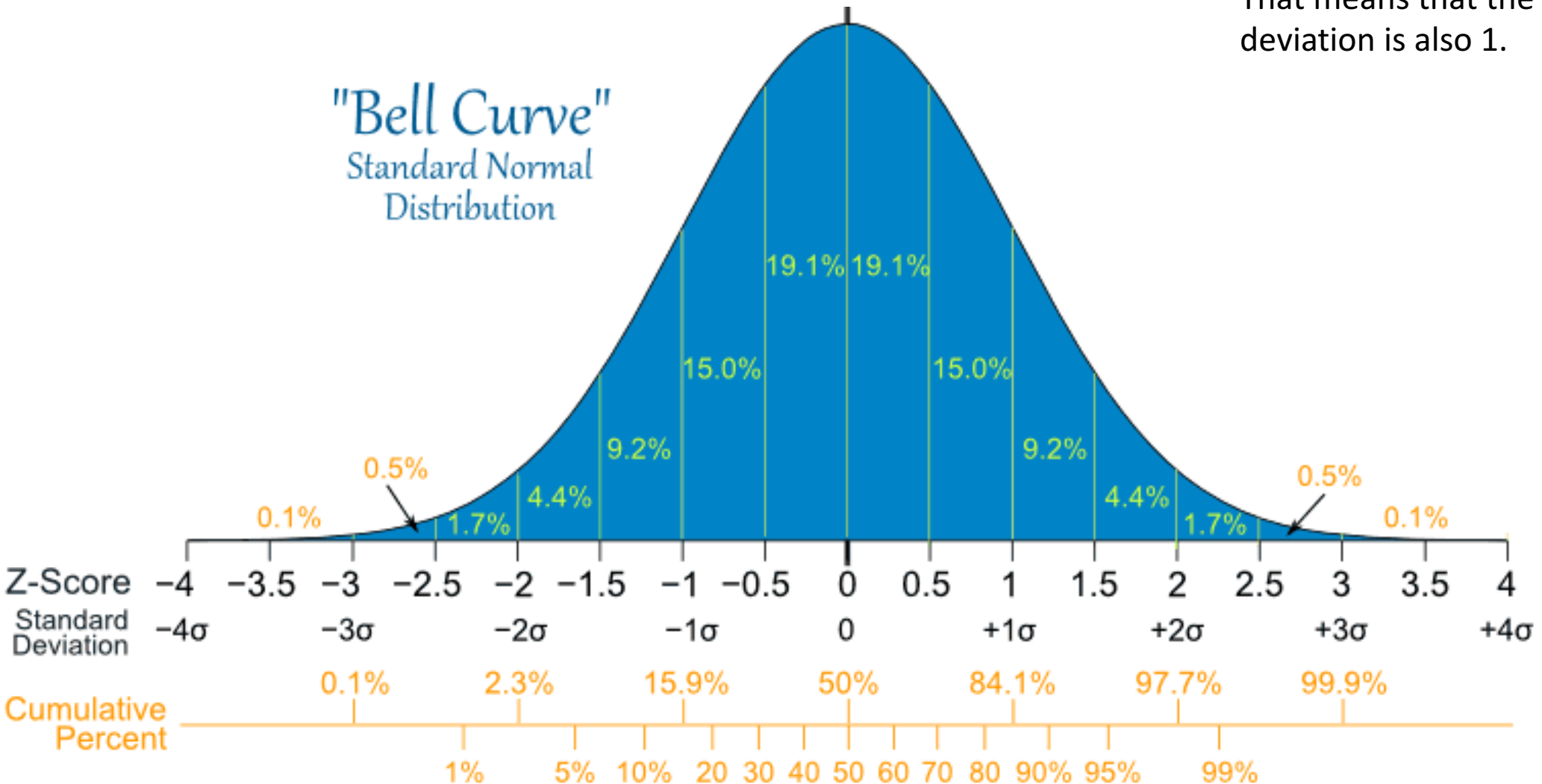
$$p(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \; e^{\frac{-(X-\mu)^2}{2\sigma^2}}$$



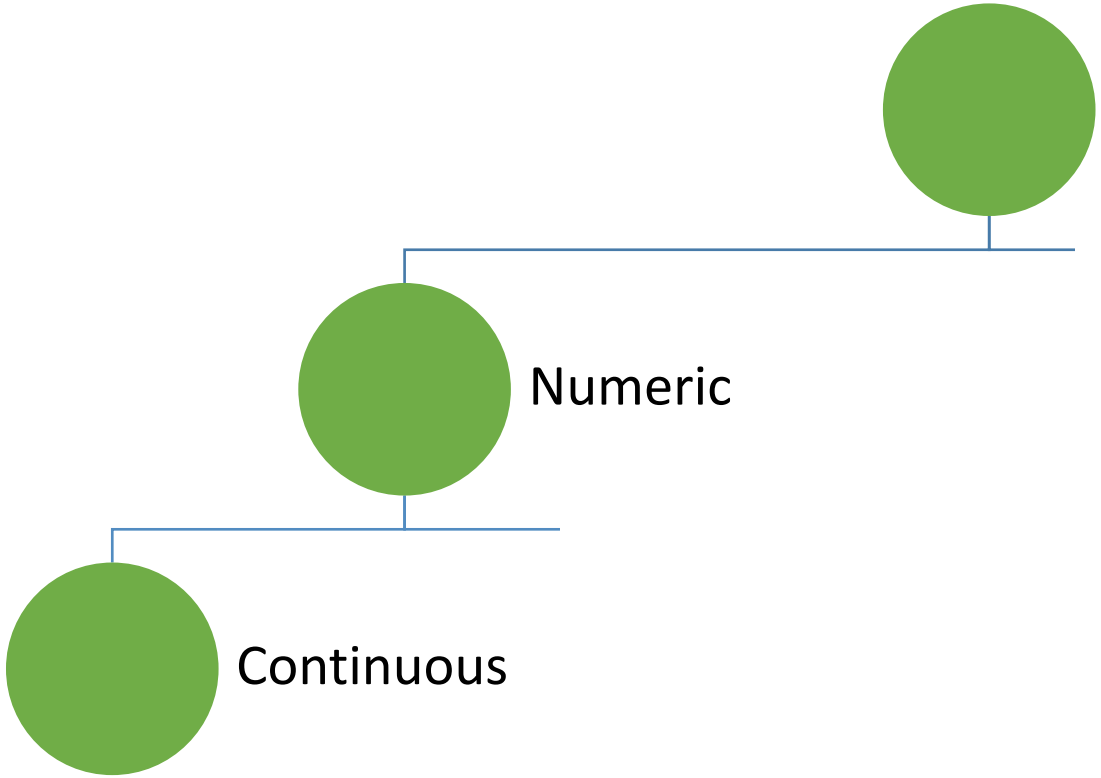
Continuous values
Sampled from normal distribution
Probability distribution defined by mean and variance!

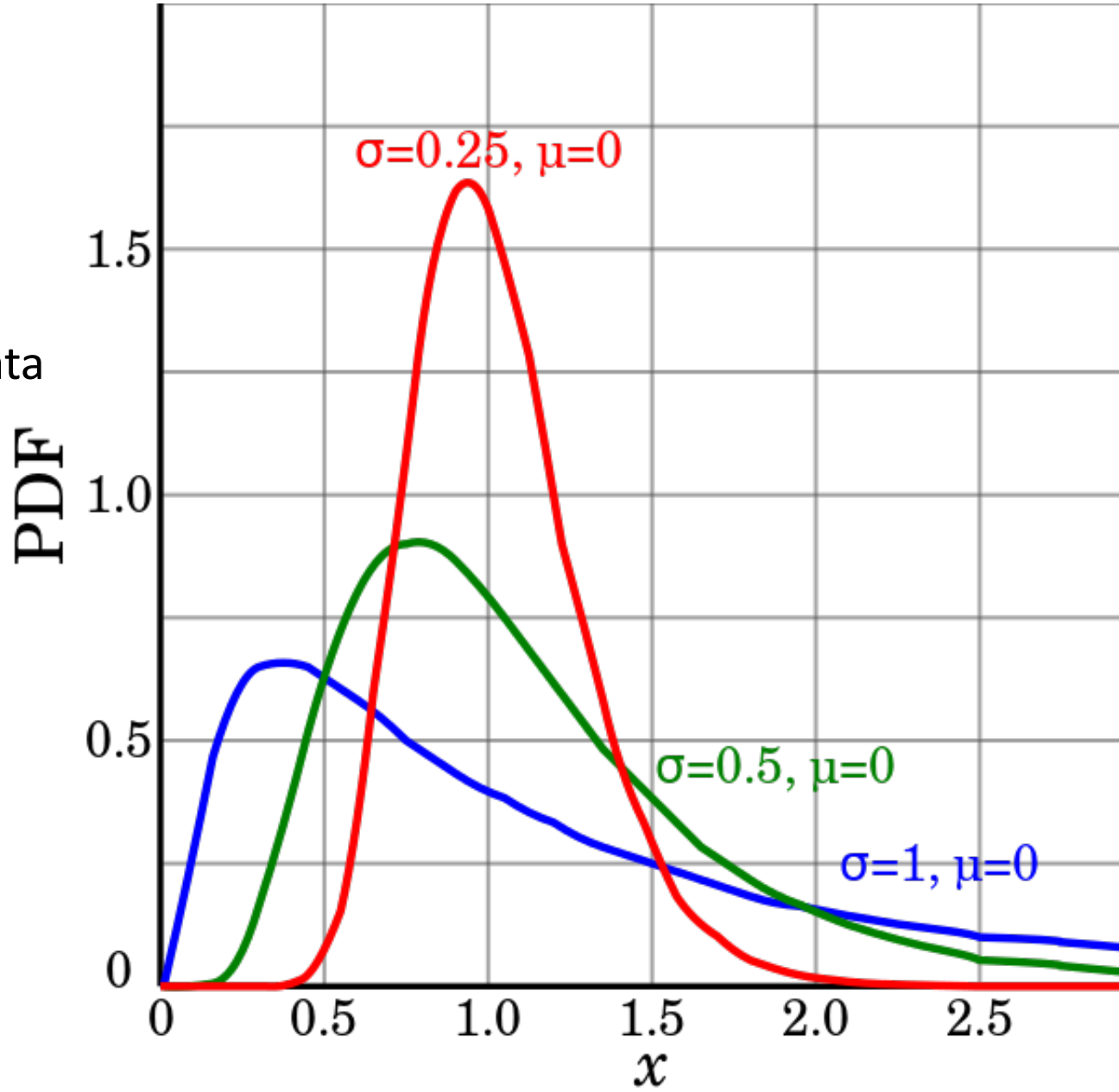


Z-distribution is a normal distribution with a mean of 0, and a variance of 1.
That means that the standard deviation is also 1.



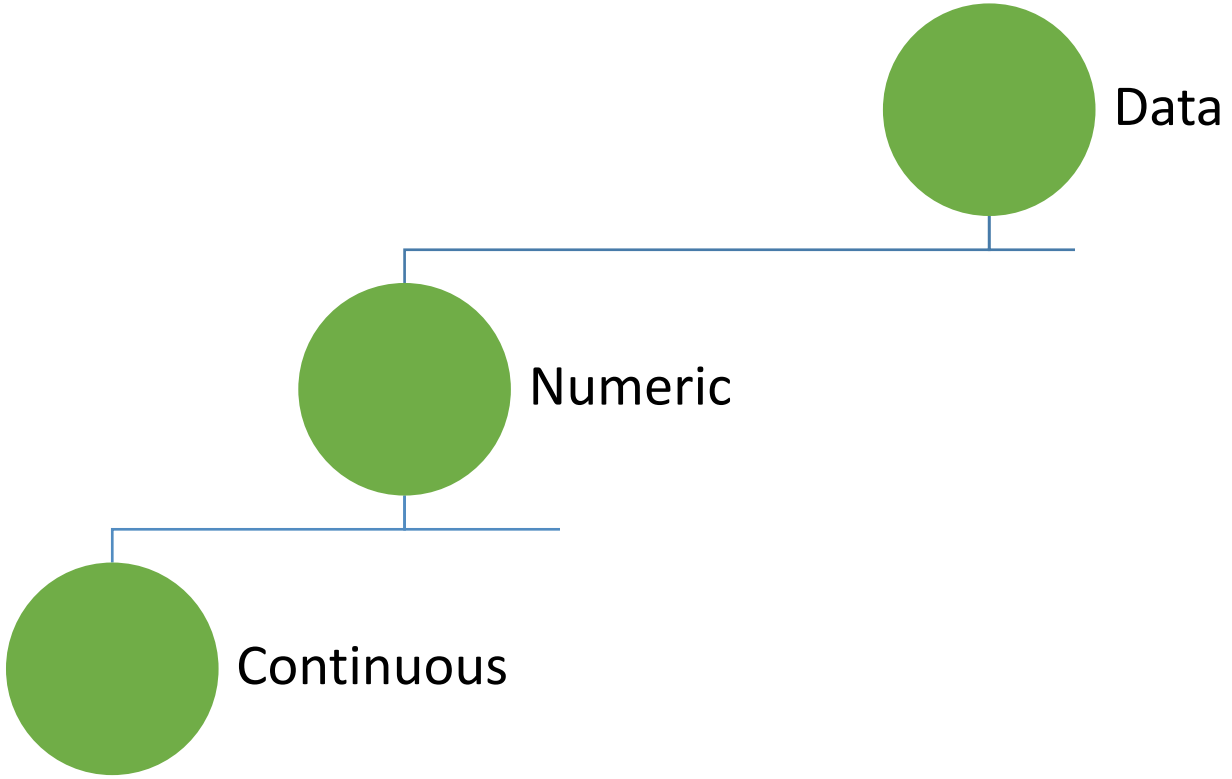
Data types

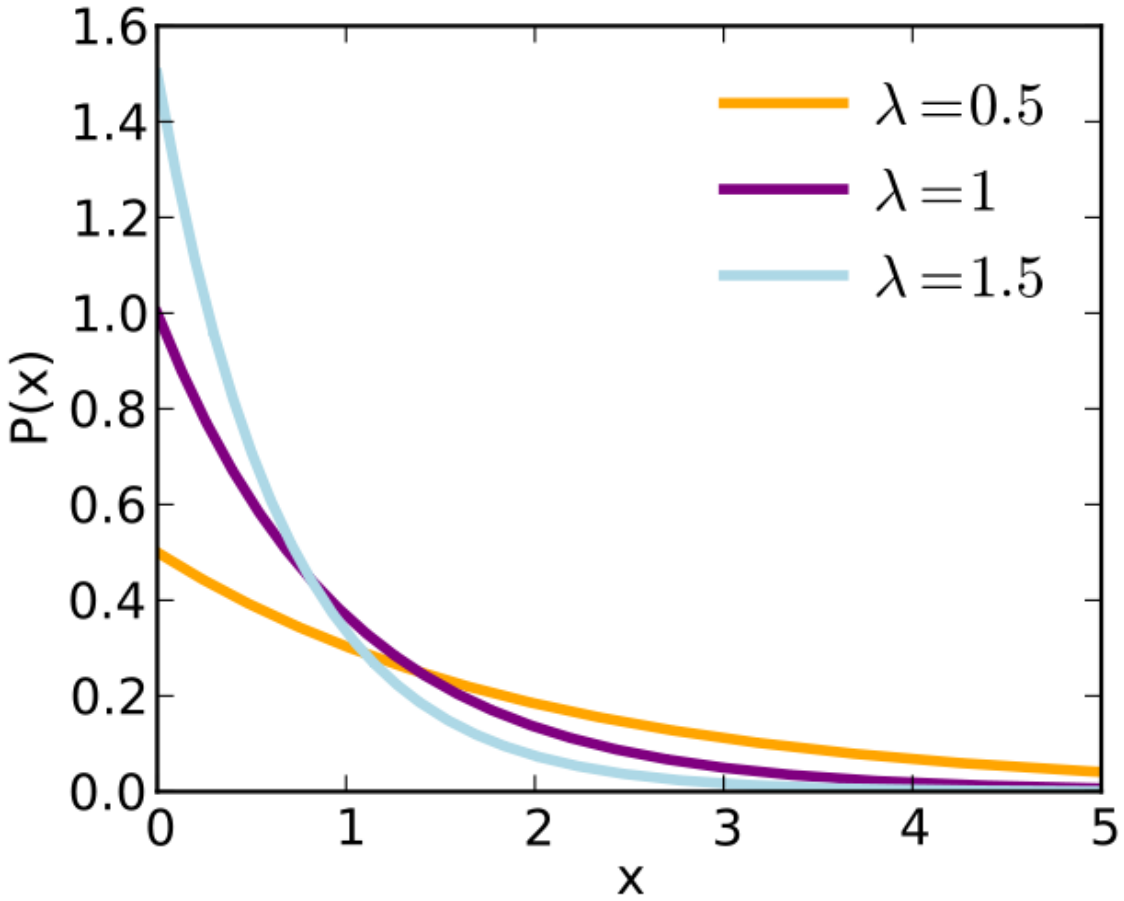




Continuous values
Log-normal
-> measurements that cannot be zero
→ Positive relationship between mean and variance

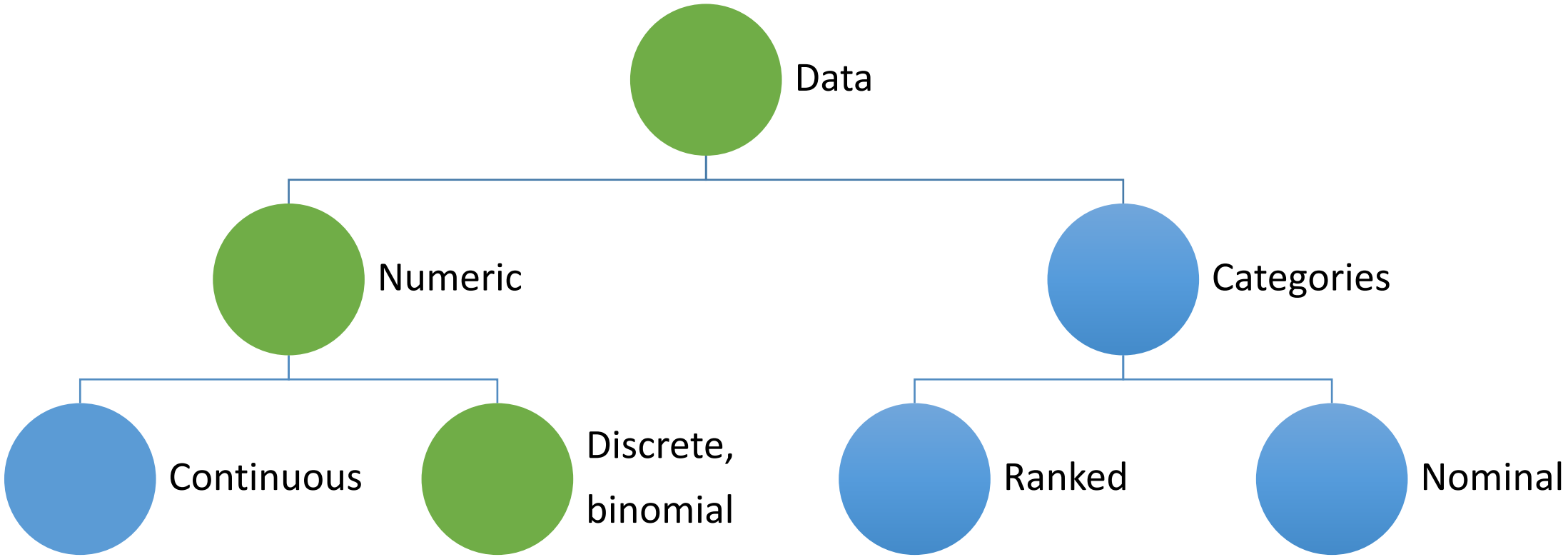
Data types





Continuous values
Exponential distribution
e.g. time interval (from diagnosis to death)

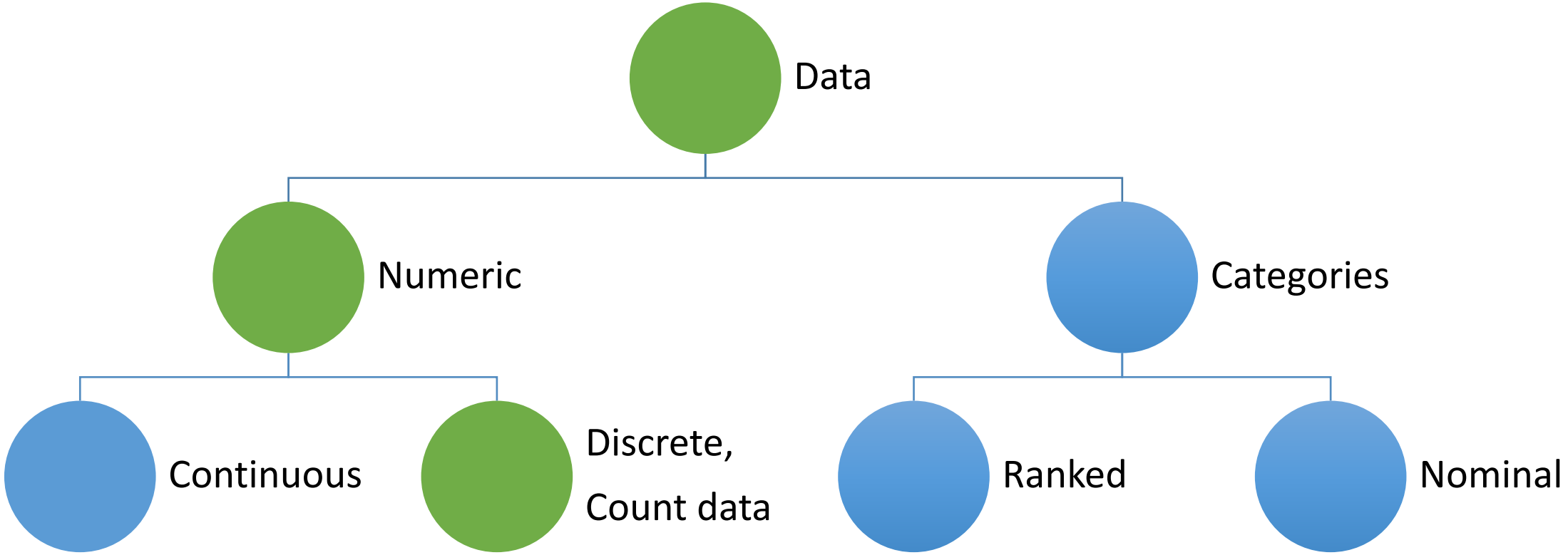
Data types


Data
Numeric
Categories
Continuous
Discrete, binomial
Ranked
Nominal
0,0,1,0,1,1,1,0
Female, male, male, female
Absent, absent, present, absent
Heads, heads, heads, tails, heads

Multinomial data

More than two possible outcomes

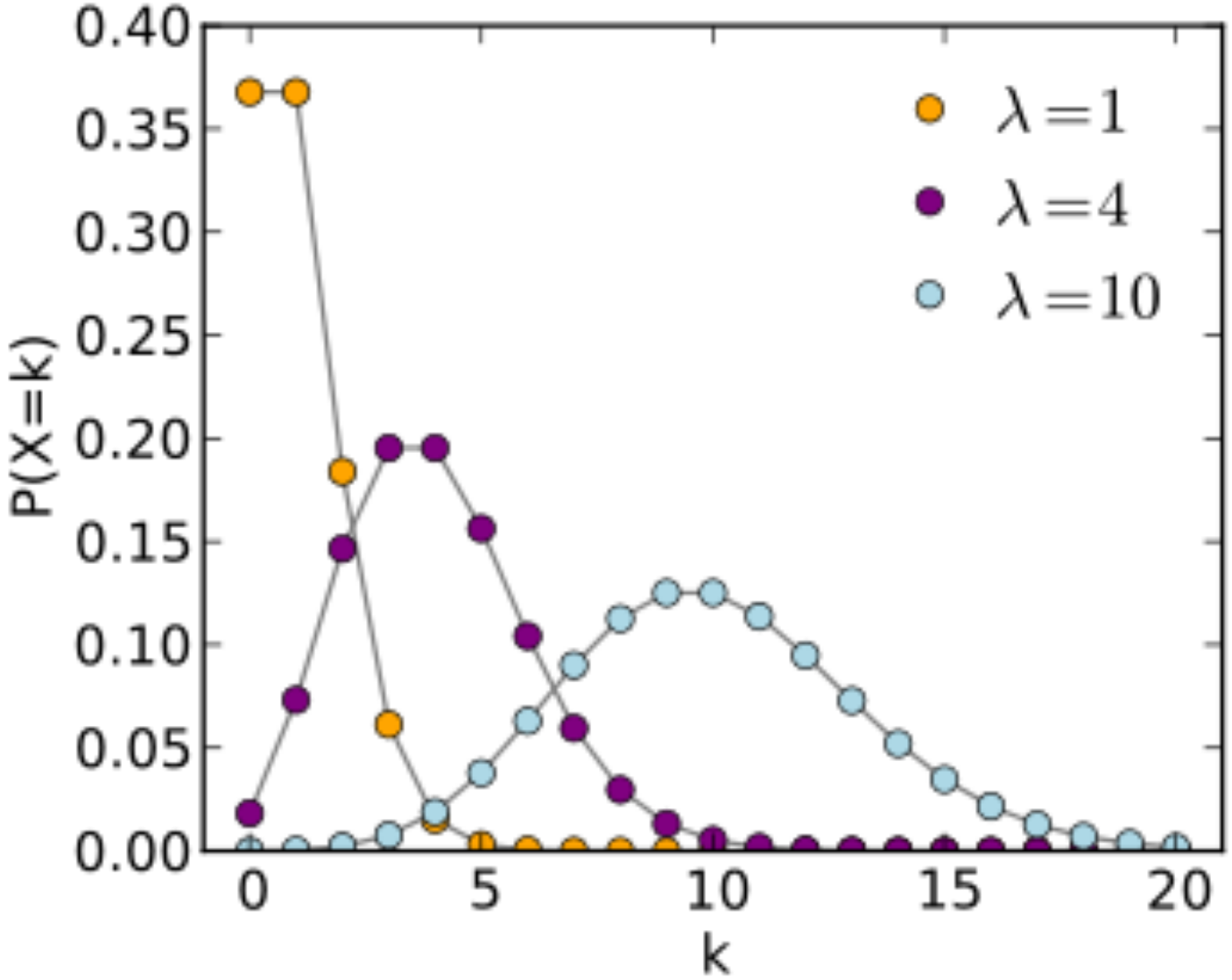
Data types


Data
Numeric
Categories
Continuous
Discrete,
Count data
Ranked
Nominal
1,2,3,3,4,5,6,8
10,2,10,30,50

Count data – Poisson distribution

Counting
How many offspring has a bird?
Species number in a forest
…

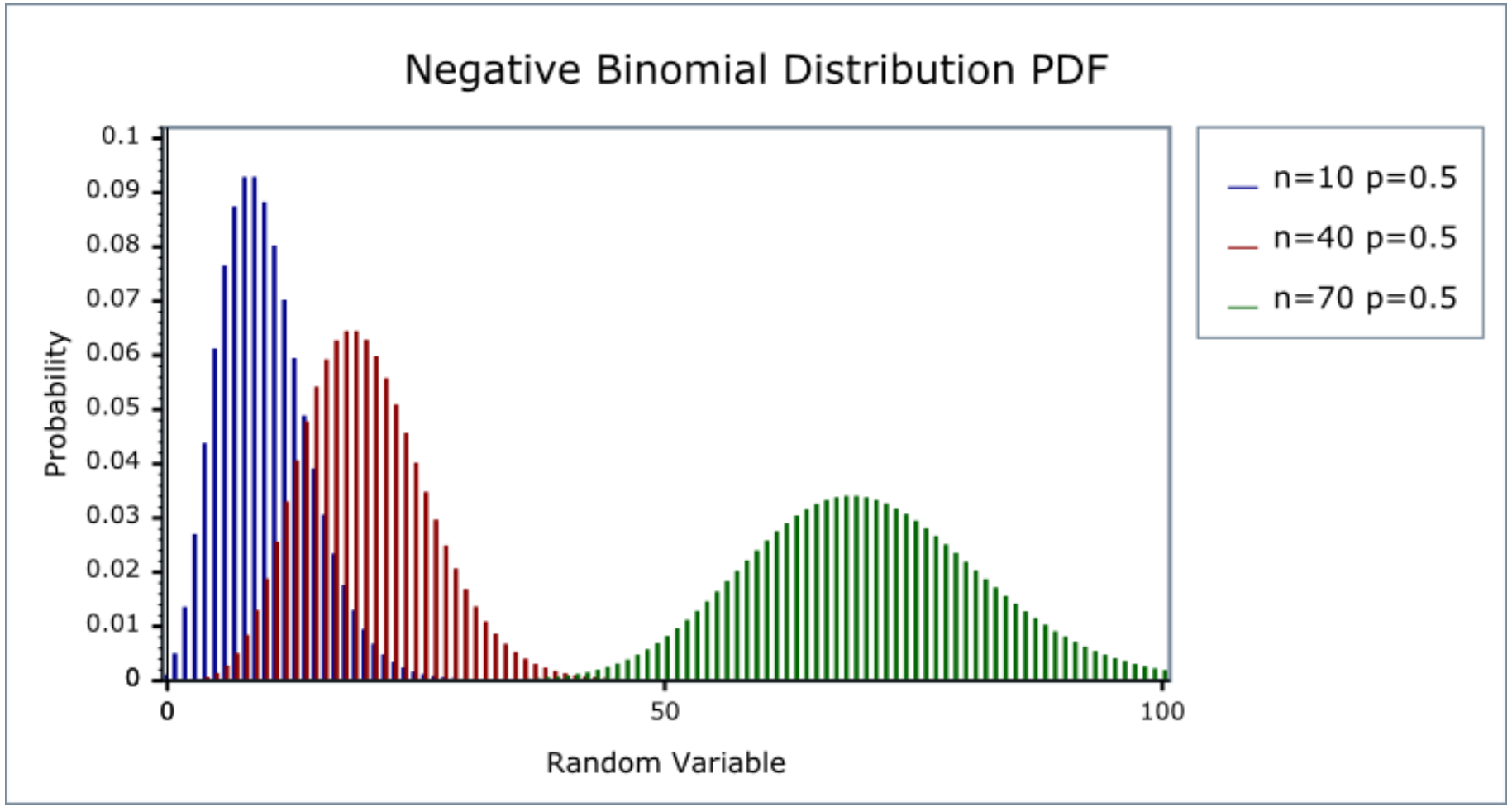
Mean = variance!



Count data – negative binomial

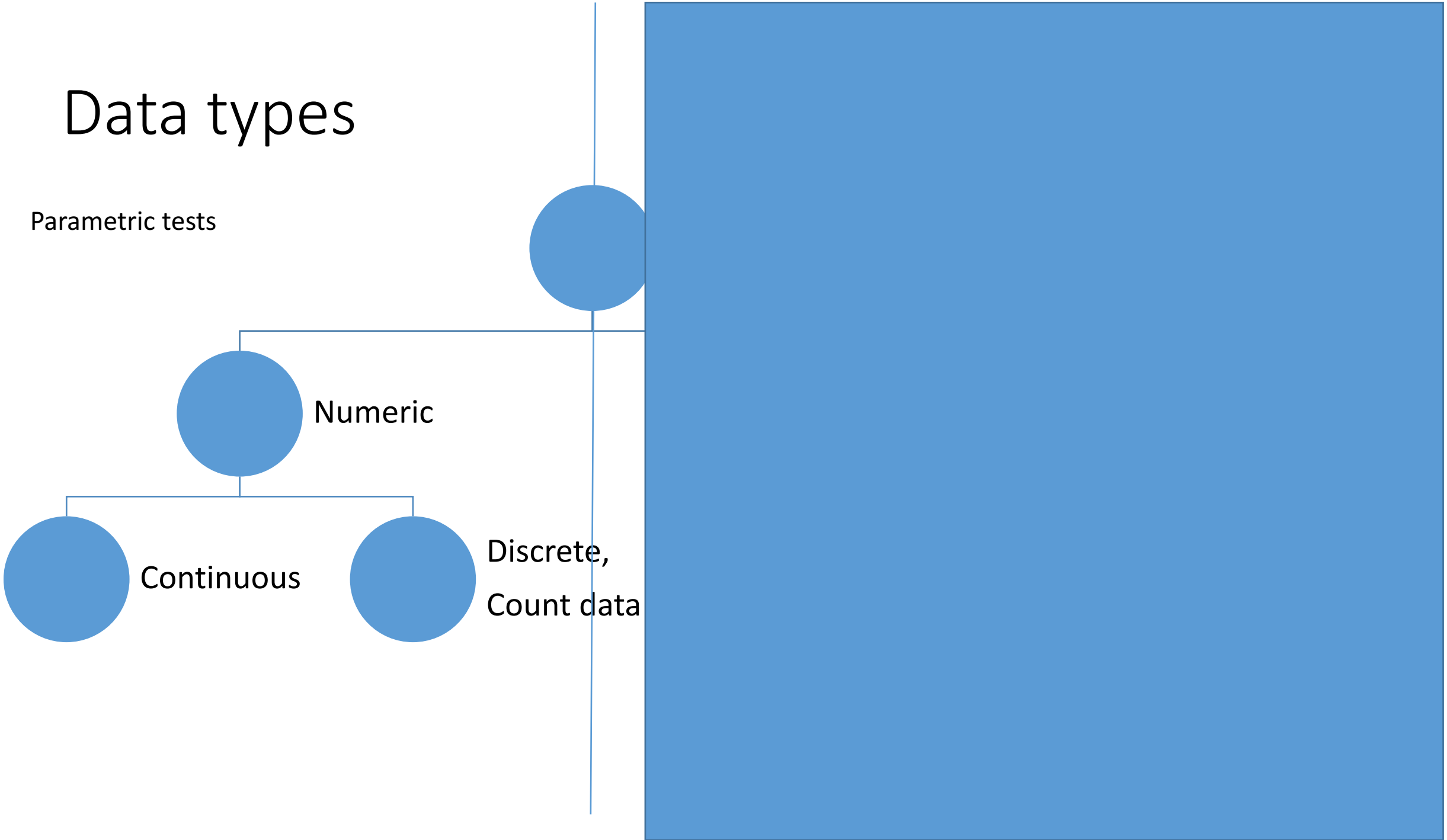
Mean != variance

No need for independence

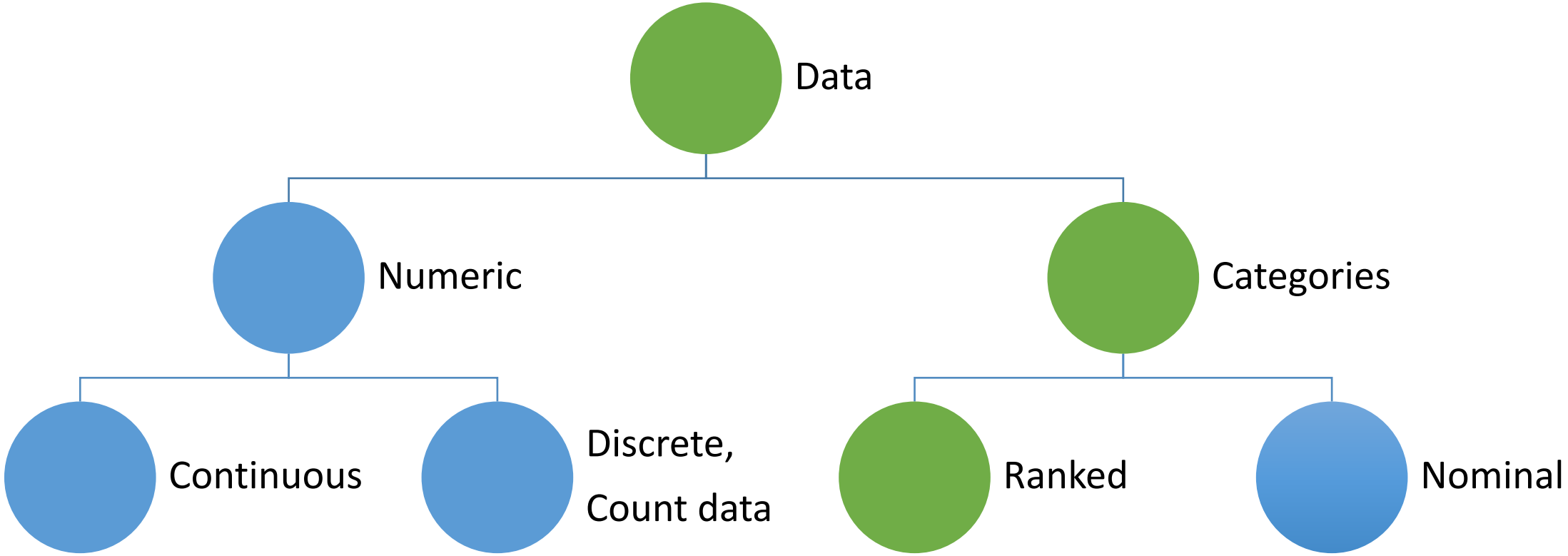


Data types

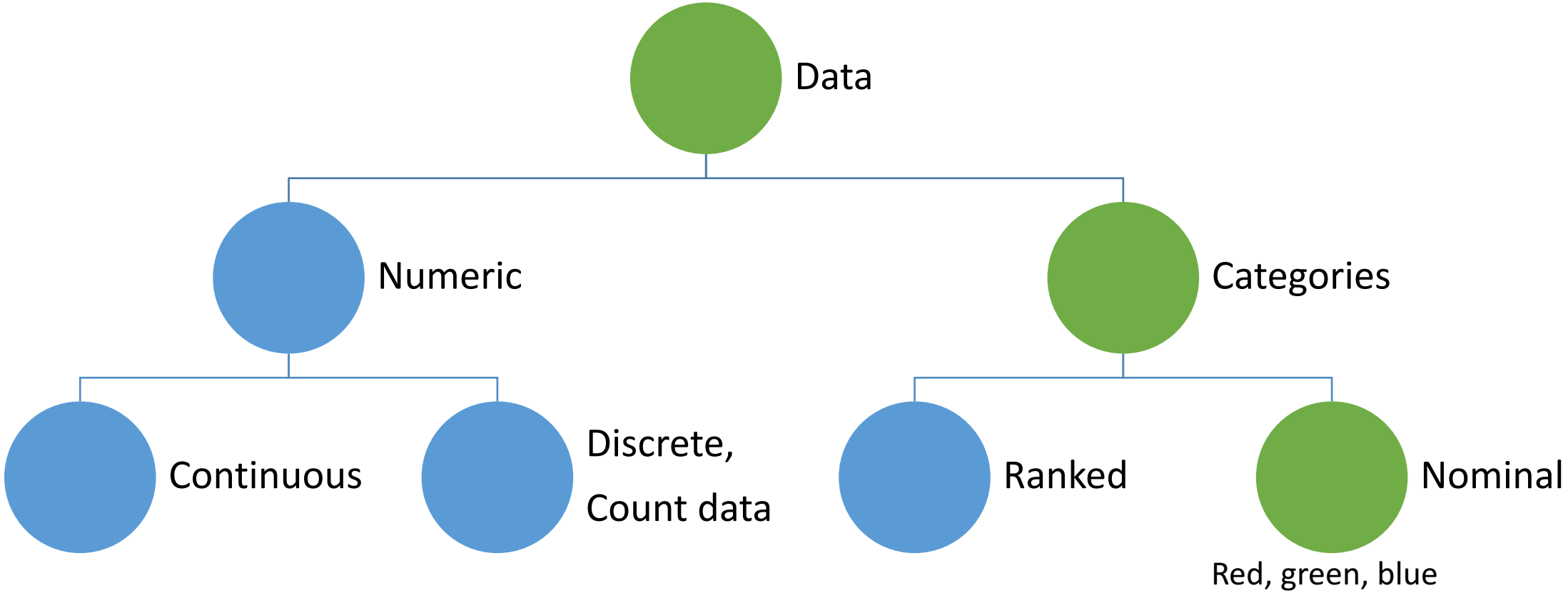
Parametric tests



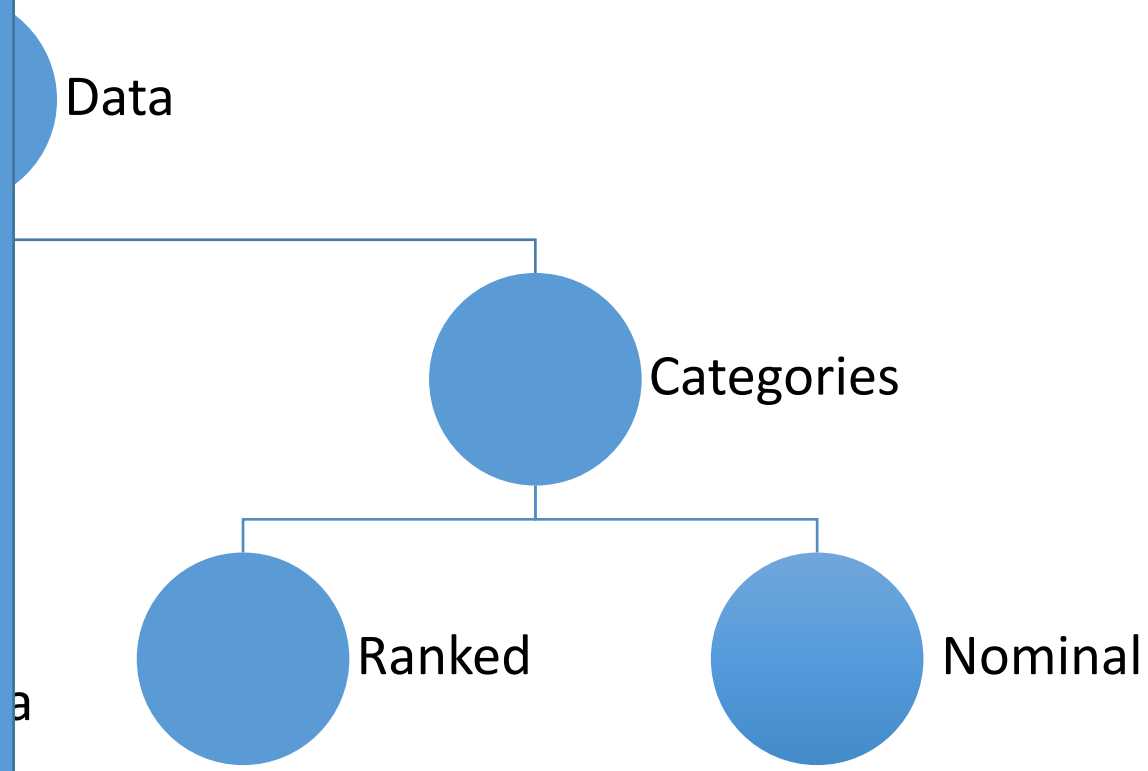
Data types


Data
Numeric
Categories
Continuous
Discrete,
Count data
Ranked
Nominal
Young, middle-age, old
Cold, warmer, hot

Data types


Data
Numeric
Categories
Continuous
Discrete,
Count data
Ranked
Nominal
Red, green, blue

Non-parametric tests



Check-in – learning goals!

- **Description**
  - Descriptive statistics of a distribution
  - Variance, standard deviation describe distribution

- **Different data types**
  - Non-parametric data – can be problematic

- **Parametric data**: continuous and discrete
  - Data can follow different distributions

Exercise – DO IT NOW (no handout)!

- Determine the type of each variable in the sparrow dataset
- Look at their distributions
- Find r functions to draw the probability and densitty distributions for
- Gaussian
- Poisson
- Binomial
- Random

Exercise – DO IT NOW (no handout)!

- Determine the type of each variable in the sparrow dataset
- Tell us all about these variables

Exercise – DO IT NOW (no handout)!

- Determine the type of each variable in the sparrow dataset
- Tell us all about these variables

- How to draw a Poisson distribution?

Exercise – DO IT NOW (no handout)!

- Determine the type of each variable in the sparrow dataset
- Tell us all about these variables

- How to draw a Poisson distribution?
- How to draw a binomial distribution?

Exercise – DO IT NOW (no handout)!

- Determine the type of each variable in the sparrow dataset
- Tell us all about these variables

- How to draw a Poisson distribution?
- How to draw a binomial distribution?
- Gaussian?

Exercise – DO IT NOW (no handout)!

- Determine the type of each variable in the sparrow dataset
- Tell us all about these variables

- How to draw a Poisson distribution?
- How to draw a binomial distribution?
- Gaussian?
- Random?

Exercise – DO IT NOW (no handout)!

- Determine the type of each variable in the sparrow dataset
- Tell us all about these variables

- How to draw a Poisson distribution?
- How to draw a binomial distribution?
- Gaussian?
- Random?
- What is the difference between a density plot and a probability plot?