



All the Stuff You Need To Know

Chelsea Parlett-Pelleriti

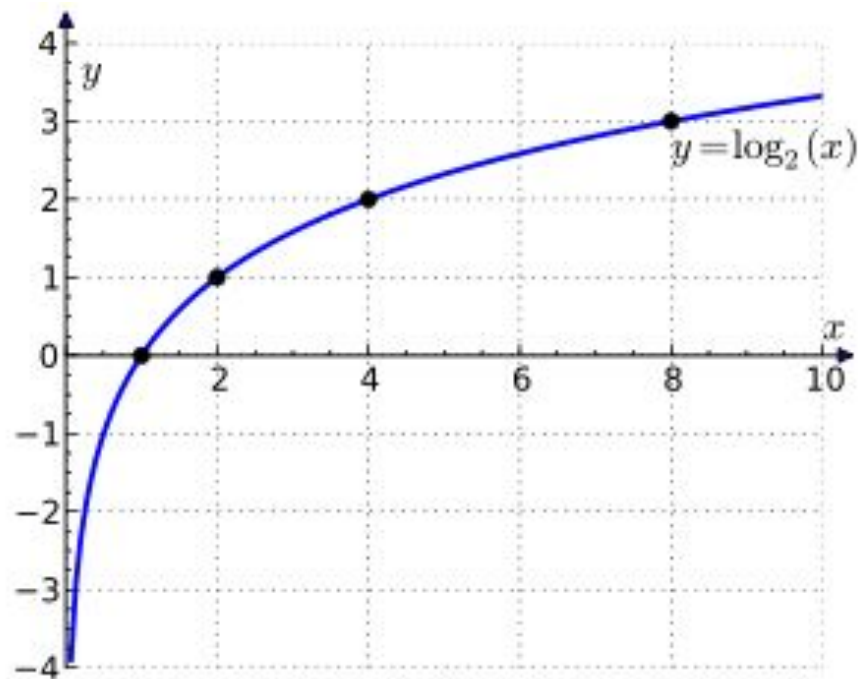


Welcome!
Everything is fine.

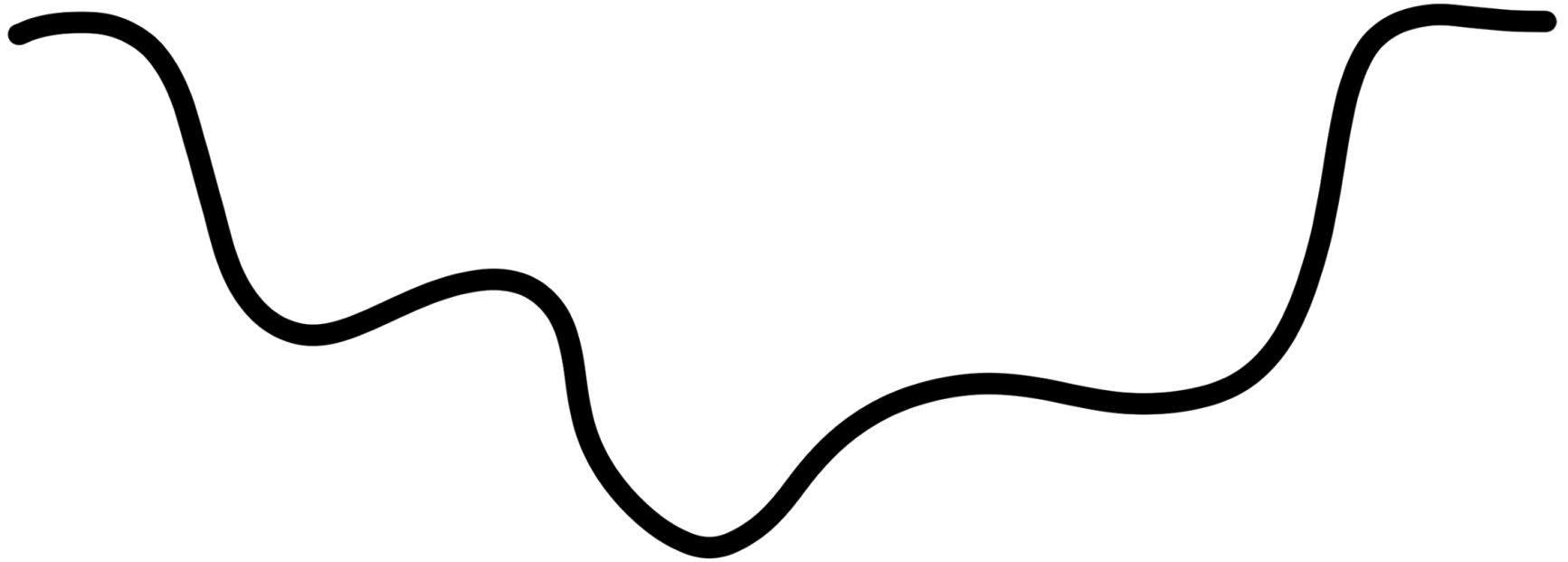


Logarithms

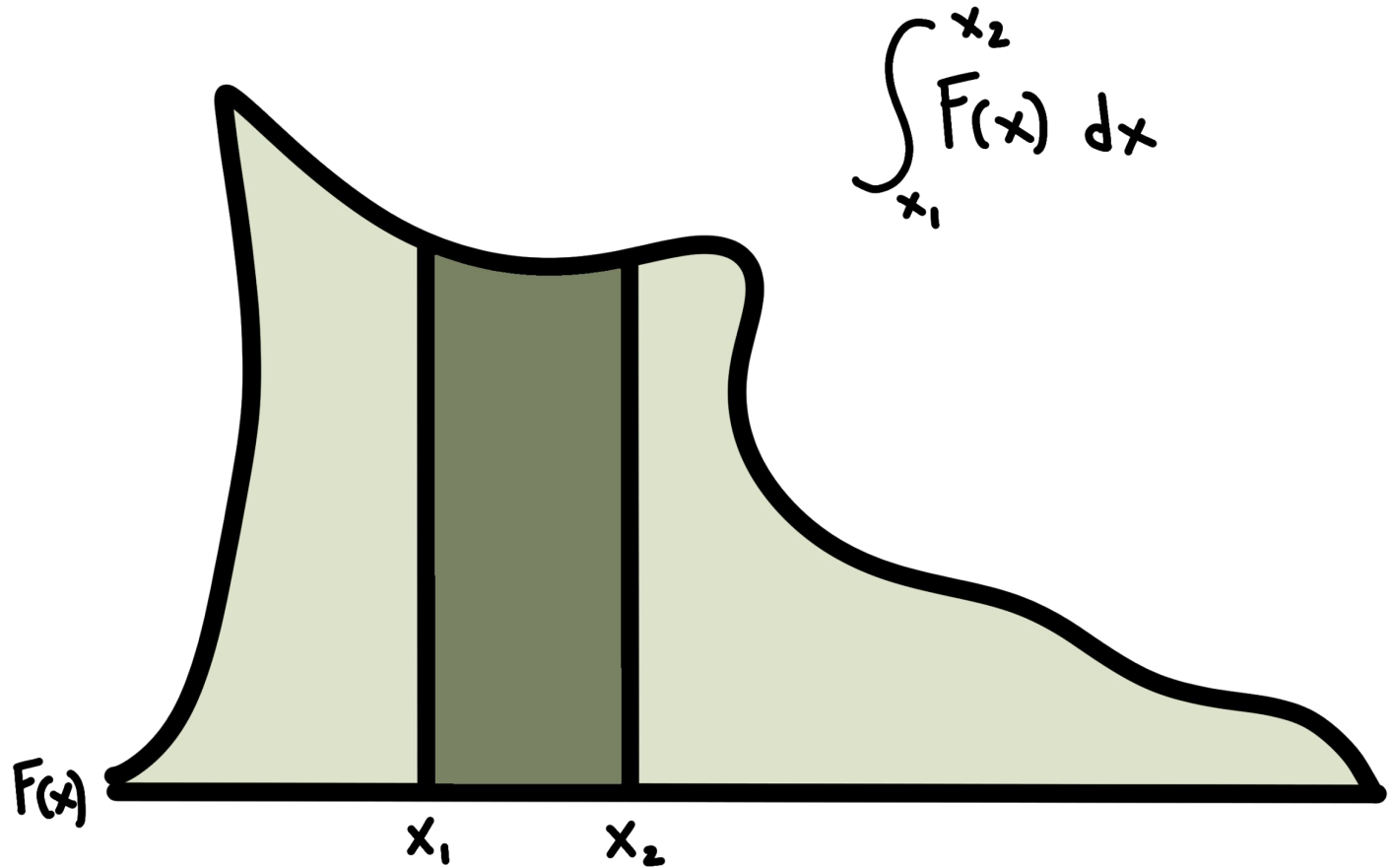
Log rules:



Derivatives

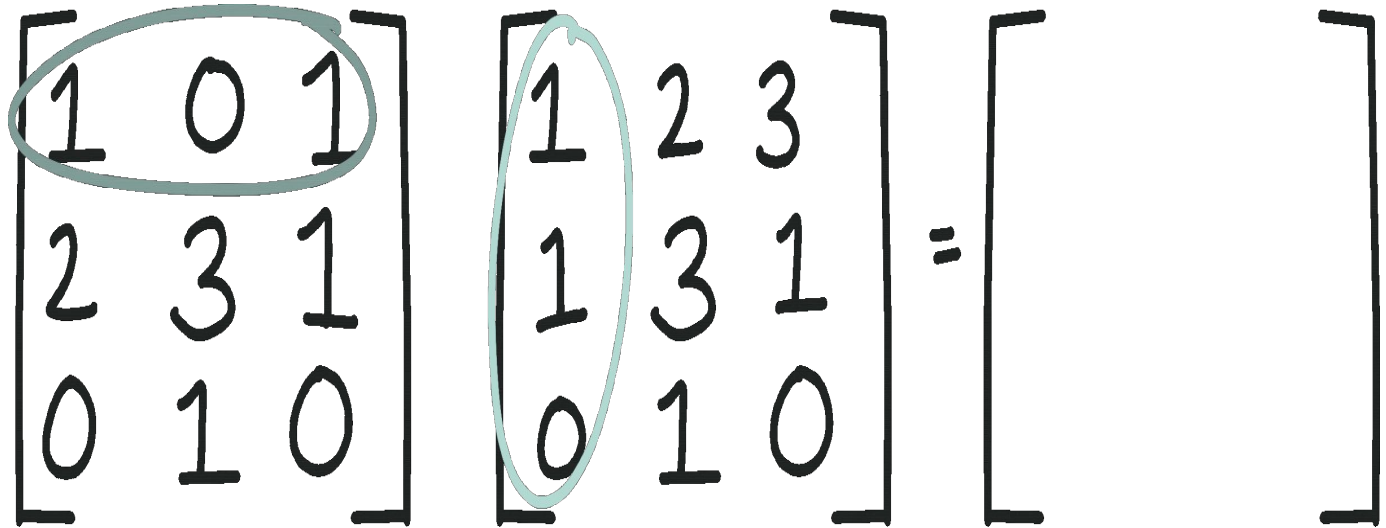


Integrals



Matrices and Vectors

- Data as a Matrix/Vector (it's just an excel spreadsheet)
- Matrix Algebra



A hand-drawn equation representing matrix multiplication. The first matrix is a 3x3 matrix with elements $\begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. The second matrix is a 3x3 matrix with elements $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. The third part of the equation is an equals sign followed by an empty 3x3 matrix bracket $\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$. A green oval highlights the first row of the first matrix (1, 0, 1). A blue oval highlights the first column of the second matrix (1, 1, 0).

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

Variance and Covariance

Which has higher variance?

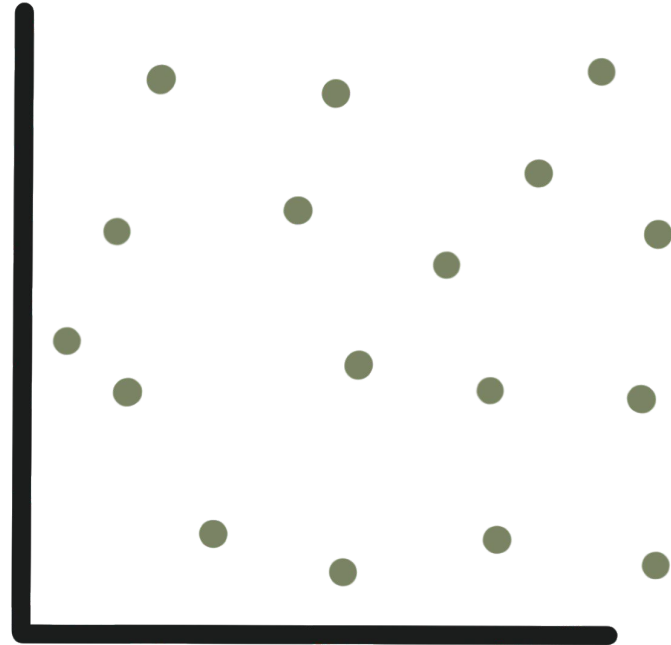
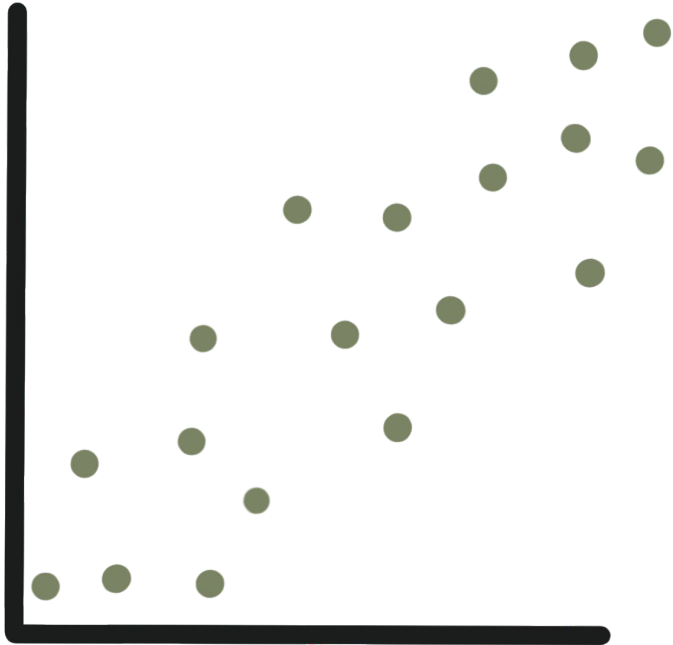
$$\frac{\sum (x_i - \mu)^2}{N}$$



Variance and Covariance

Which has higher covariance?

$$\frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N}$$



Normal Distribution

- Symmetric, Unimodal
- “Bell Curve”
- 68-95-99.7 rule
- CLT



NORMAL

Random Variables

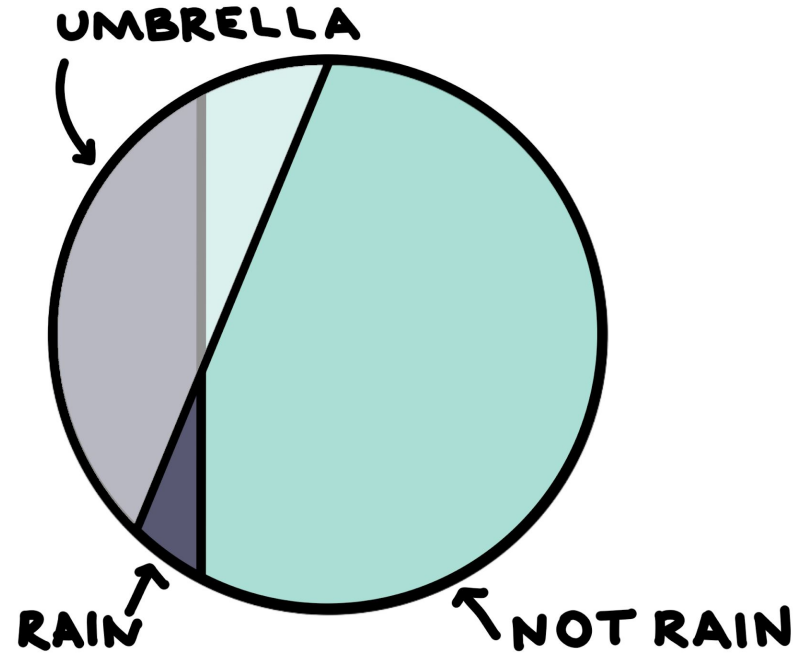
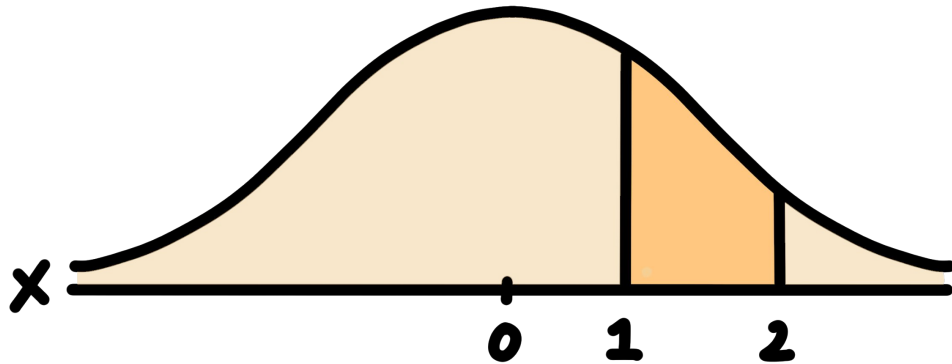
- A variable that depends on some random process.
- Coin flip, Height, Jelly Bean Flavor.

Data Types

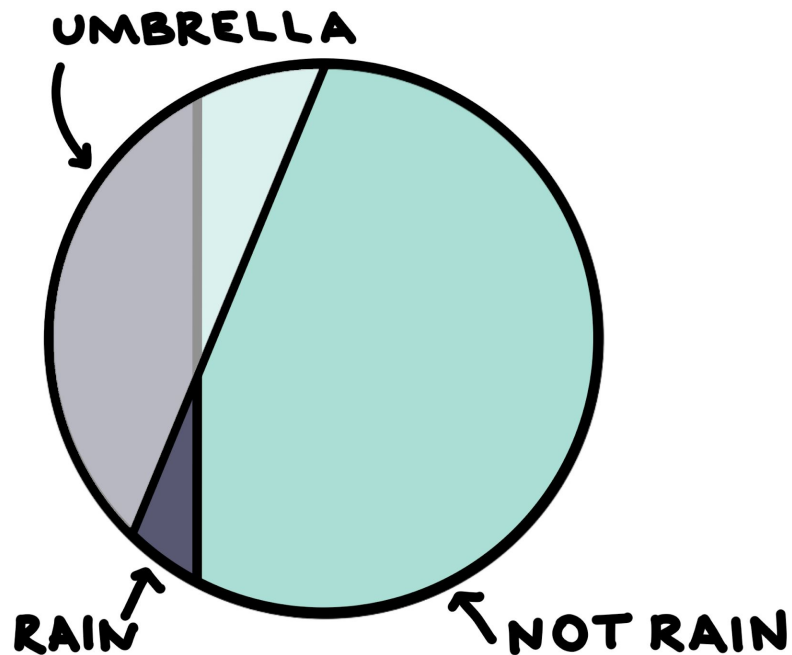
- Continuous
- Categorical
 - Nominal
 - Dummy
 - Ordinal
 - Interval
- Boolean
- Text

Probability

$$P(1 > x > 2 \mid \mu = 0, \text{sd} = 1)$$



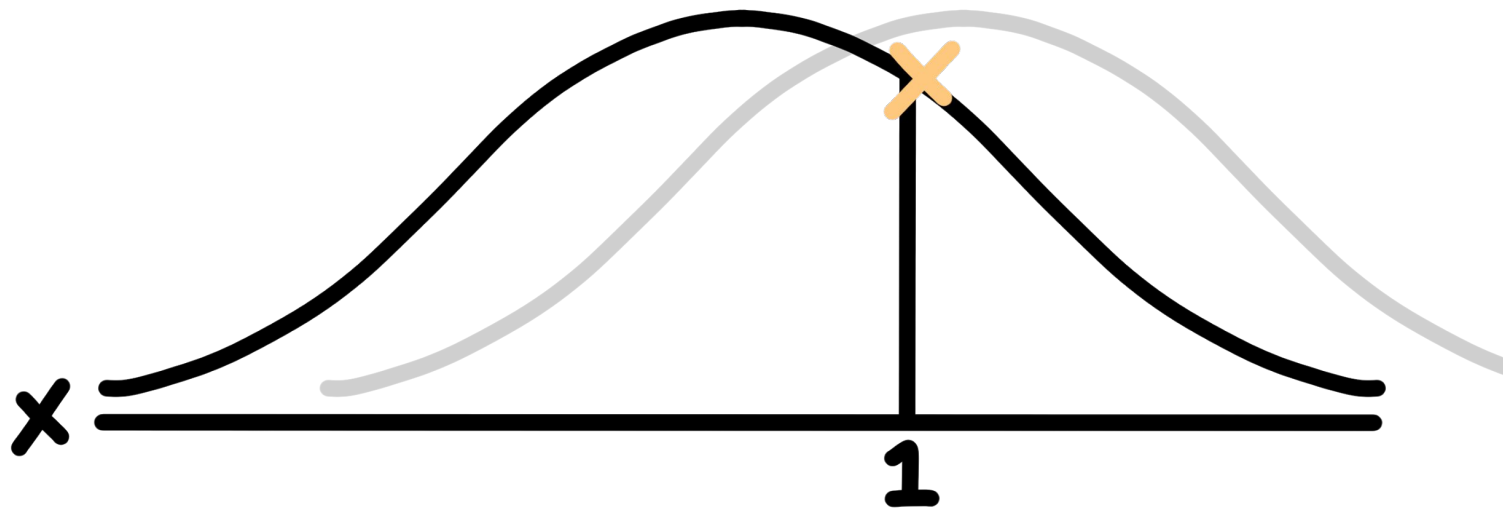
Conditional Probability



Odds

Likelihood

$$P(u=0, sd=1 \mid x=1)$$



Inference vs. Prediction

DOWNLOADS Check

- Python
- Numpy
- Pandas
- Sklearn
- Keras
- Plotnine
- Jupyter Notebooks

Python and sklearn

- `.transform()`
- `.fit()`
- `.predict()`



What is a Pandas DataFrame?

- List of Lists
- Dictionary of Lists
- CSV's

Creating and Accessing a DataFrame

- [Cheatsheet](#)
- Head
- Indexing
- Mean
- Max
- Min