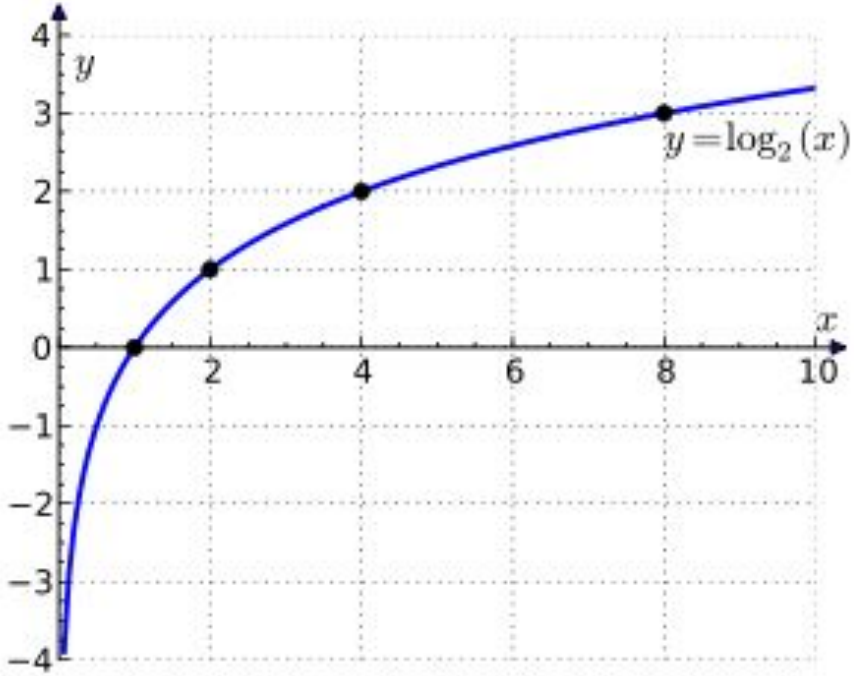
All the Stuff You Need To Know

Chelsea Parlett-Pelleriti

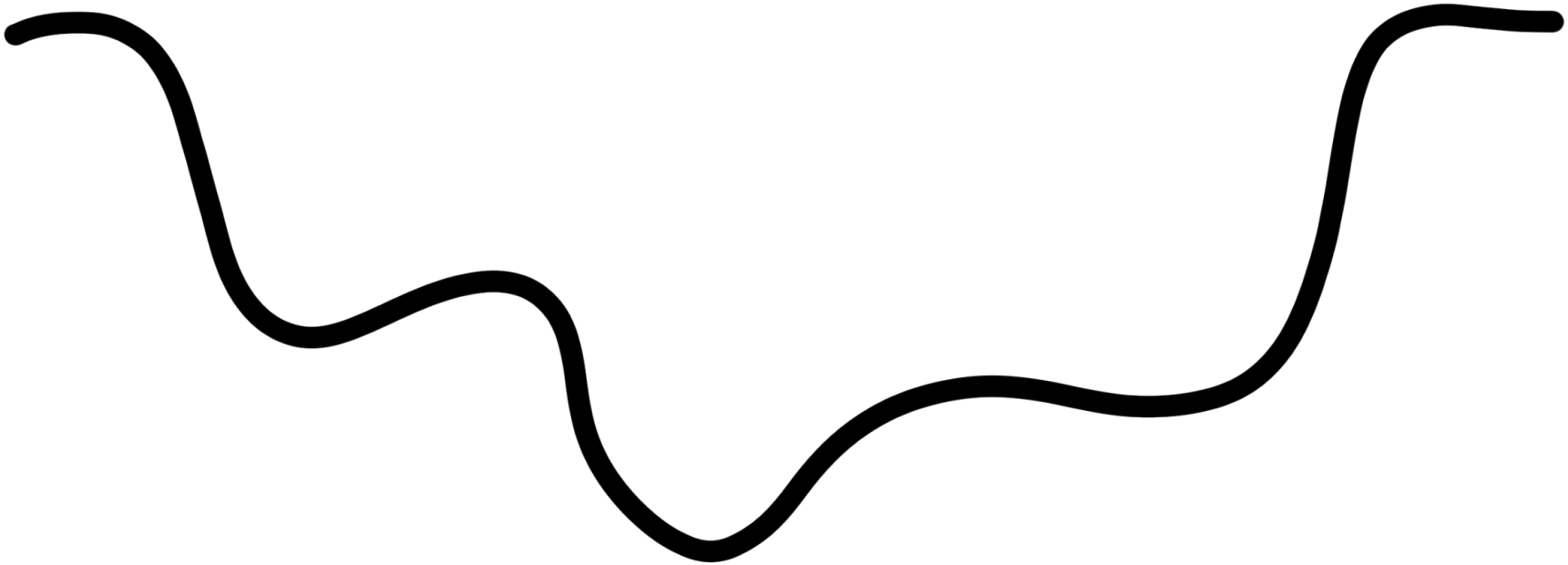

Welcome!
Everything is fine.

Logarithms

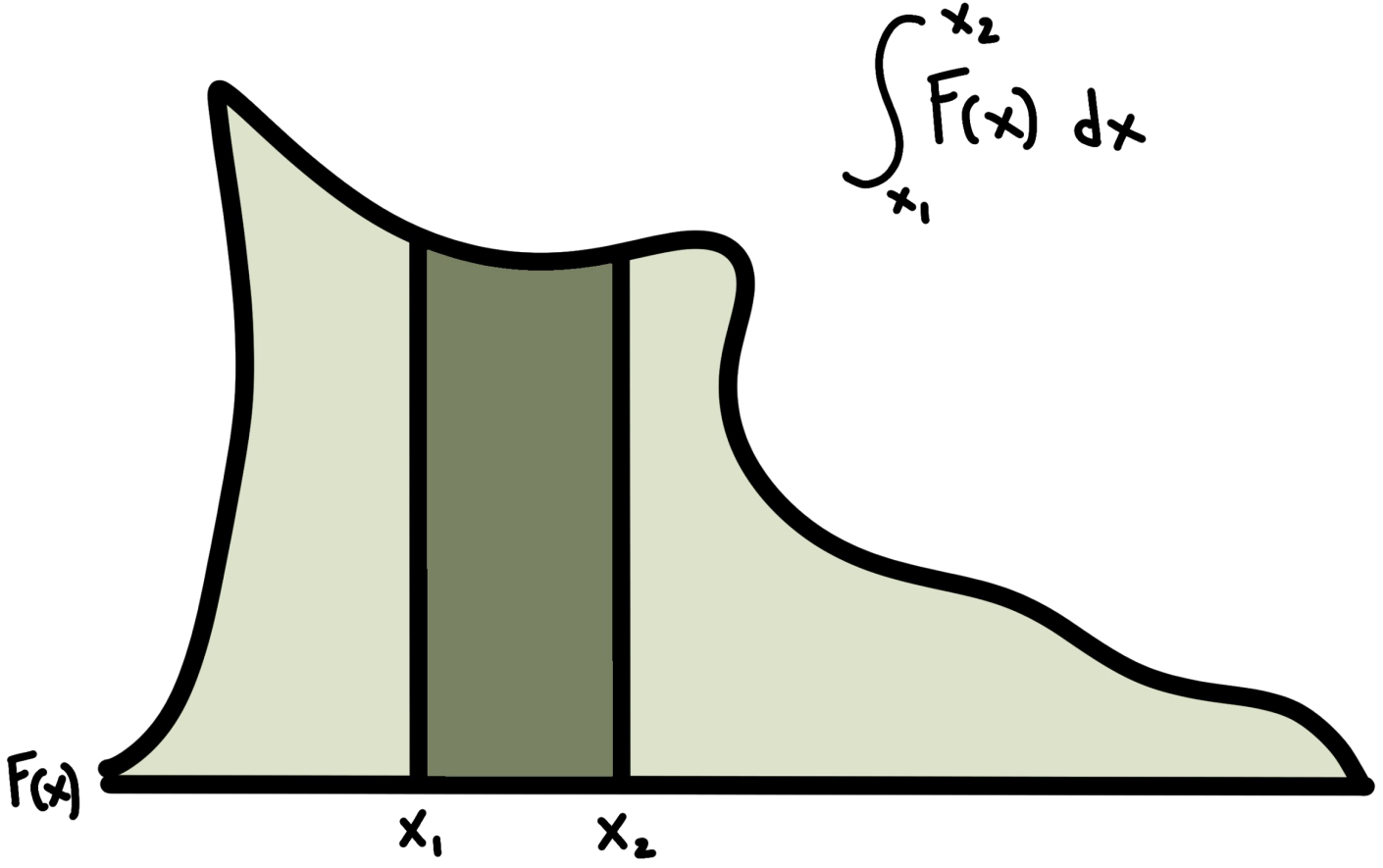
Log rules:



Derivatives

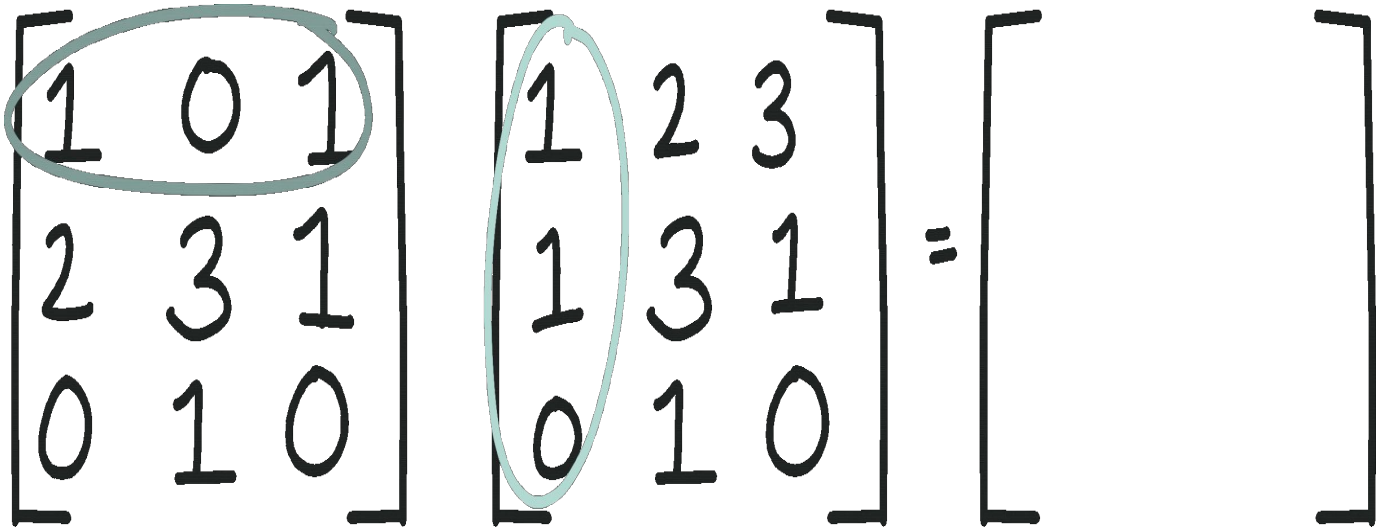


Integrals



Matrices and Vectors

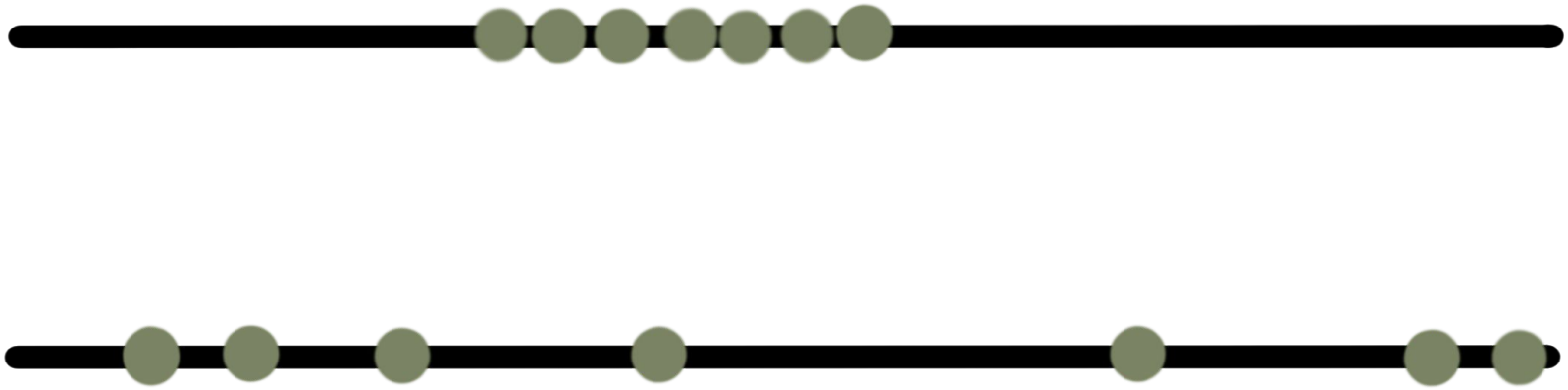
- Data as a Matrix/Vector (it’s just an excel spreadsheet)
- Matrix Algebra



Variance and Covariance

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

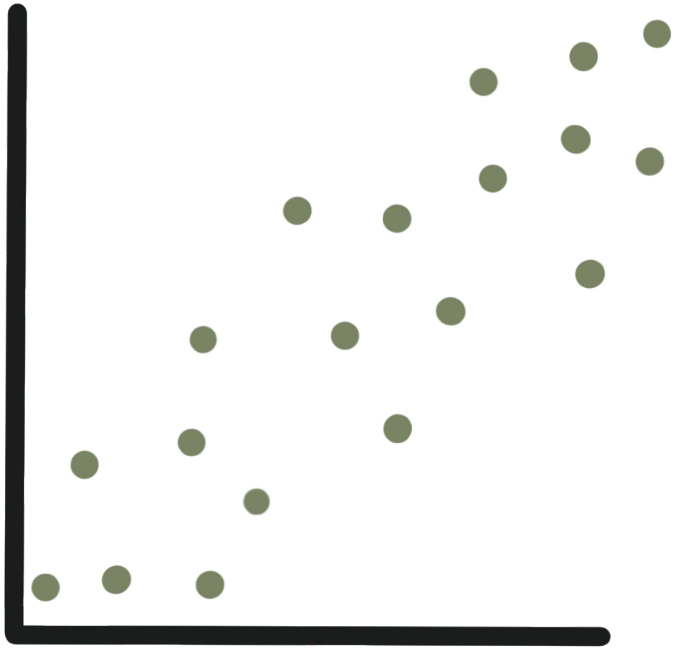
Which has higher variance?

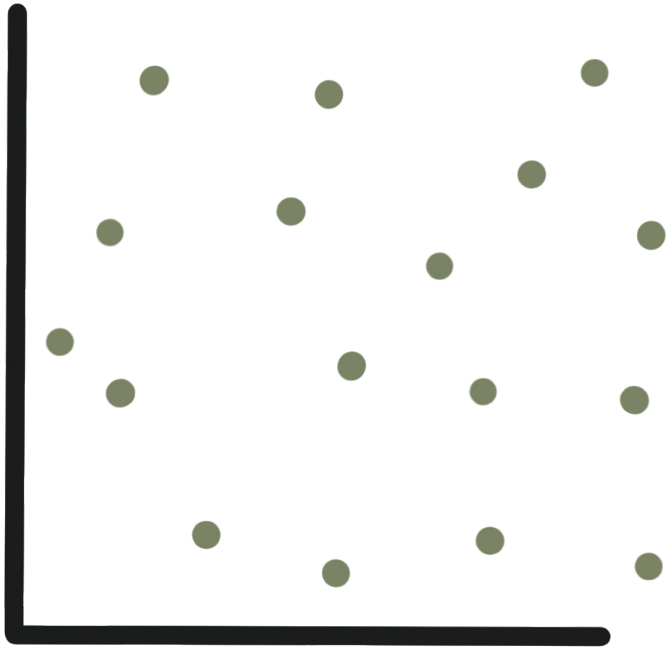


Variance and Covariance

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

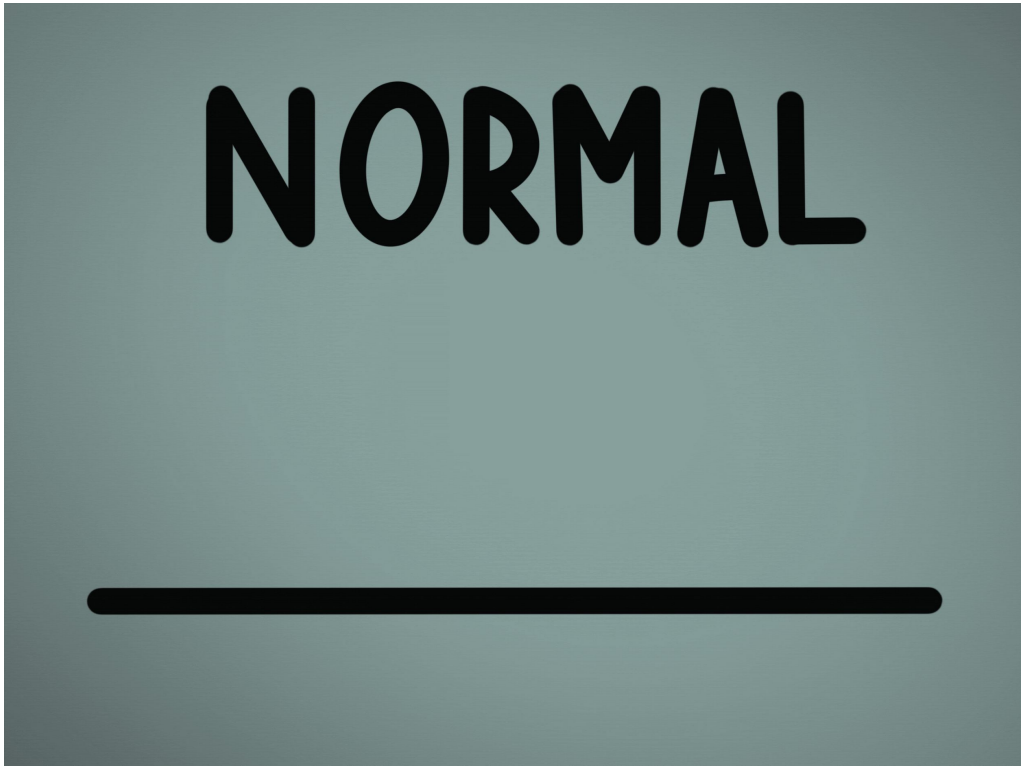
Which has higher covariance?





Normal Distribution

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



Random Variables

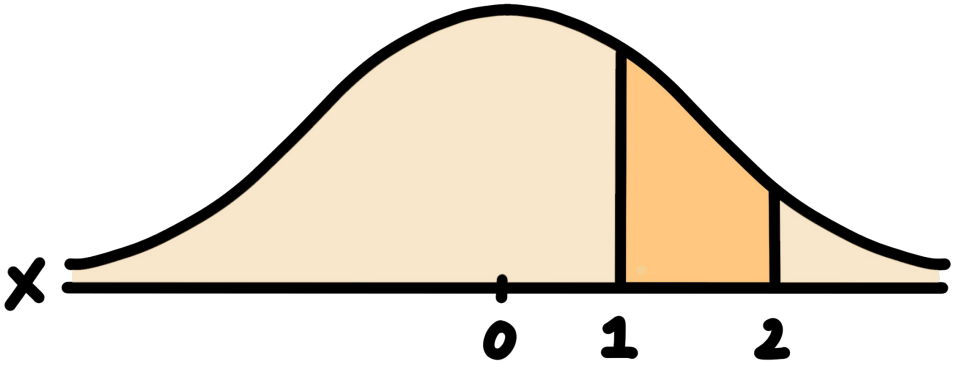
- A variable that depends on some random process.
- Coin flip, Height, Jelly Bean Flavor.

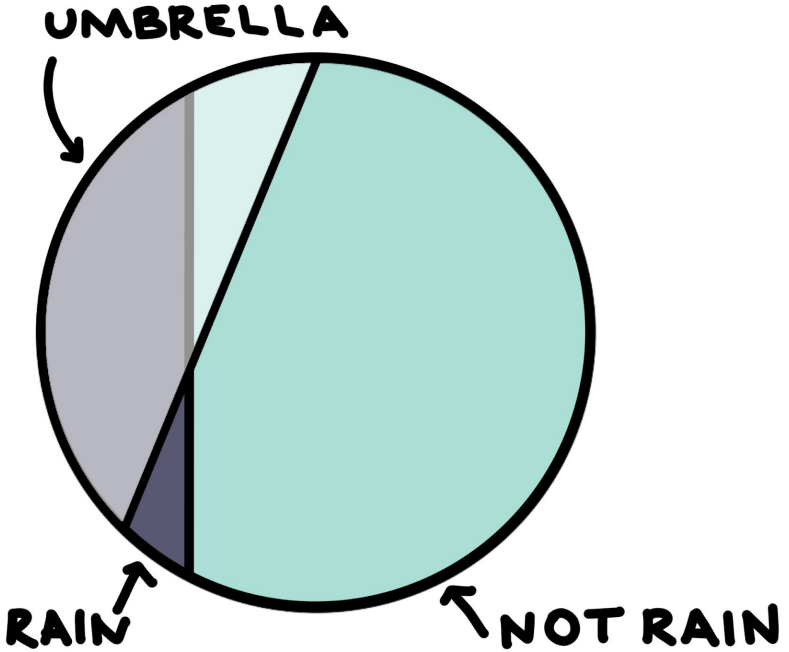
Data Types

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

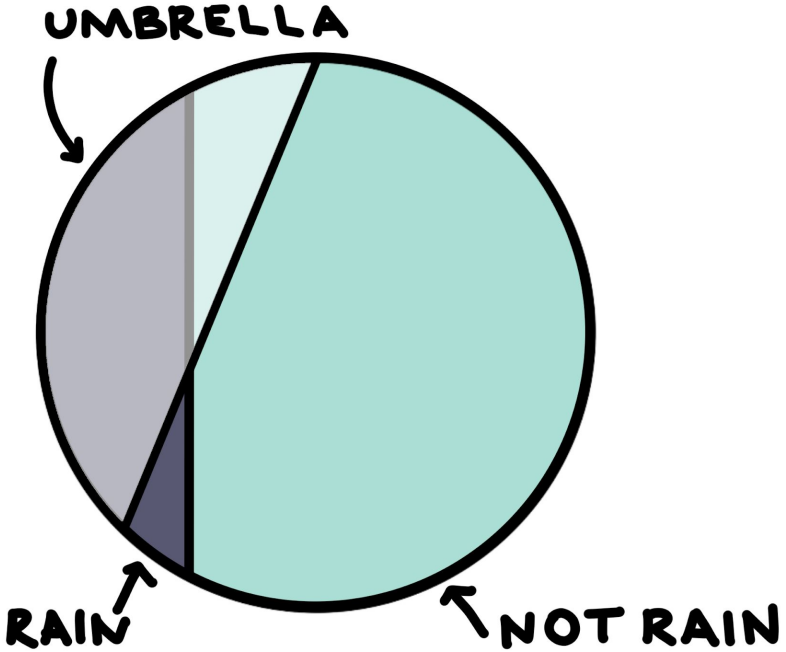
Probability

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





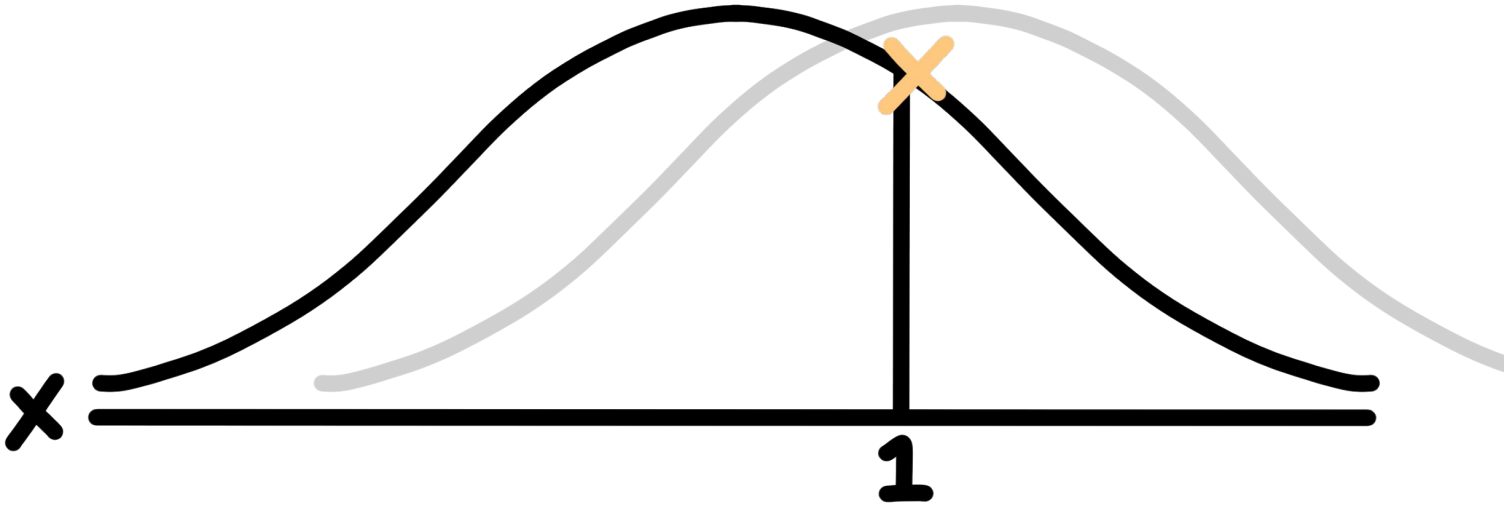
Conditional Probability



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
- Dictonary of Lists
- CSV's

Creating and Accessing a DataFrame

- **Cheatsheet**
- Head
- Indexing
- Mean
- Max
- Min