Statistics & Probabilities

What we will cover today:

0️⃣ Intro
1️⃣ Descriptive Statistics
2️⃣ Summary Statistics
3️⃣ Probabilities
4️⃣ Random Variables
5️⃣ Central Limit Theorem

Let's start by importing some useful packages

In [ ]:
```
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns

import math
import scipy
import scipy.stats as stats
```

0️⃣ Introduction

Quoting Wikipedia (https://en.wikipedia.org/wiki/Statistics):

> Statistics is the discipline that concerns the collection, organization, displaying, analysis, interpretation and presentation of **data**.

Statistics allows to **summarize data** in a small number of indicators

Combined with probability, statisticians can **draw conclusions** 💪 based on **samples** 👌

Statistical Work

- **Data Analysis**: gathering, displaying and summarizing data
- **Probability**: laws of chance, in or out of the casino
- **Inference**: drawing statistical conclusions from specific data, using probability

1️⃣ Descriptive Statistics

- How can we discover **underlying patterns** in a heap of numbers?
- How can we **represent** data in useful ways?
- How can we **summarize** the data?

First step: **gather** some data 🕵️

Experiment: ask students in a University class to give their **weight** (in pounds).

Male (57)

```
140 145 160 190 155 165 150 190 195 138 160 155 153 145 170 175 175 170 18
0
135 170 157 130 185 190 155 170 155 215 150 145 155 155 150 155 150 180 16
0
135 160 130 155 150 148 155 150 140 180 190 145 150 164 140 142 136 123 15
5
```

Female (35)

```
140 120 130 138 121 116 125 145 150 112 125 130 120 130 131 120 118 125 13
5
125 118 122 115 102 115 150 110 116 108  95 125 133 110 150 108
```

We can convert this raw data into a **DataFrame**:

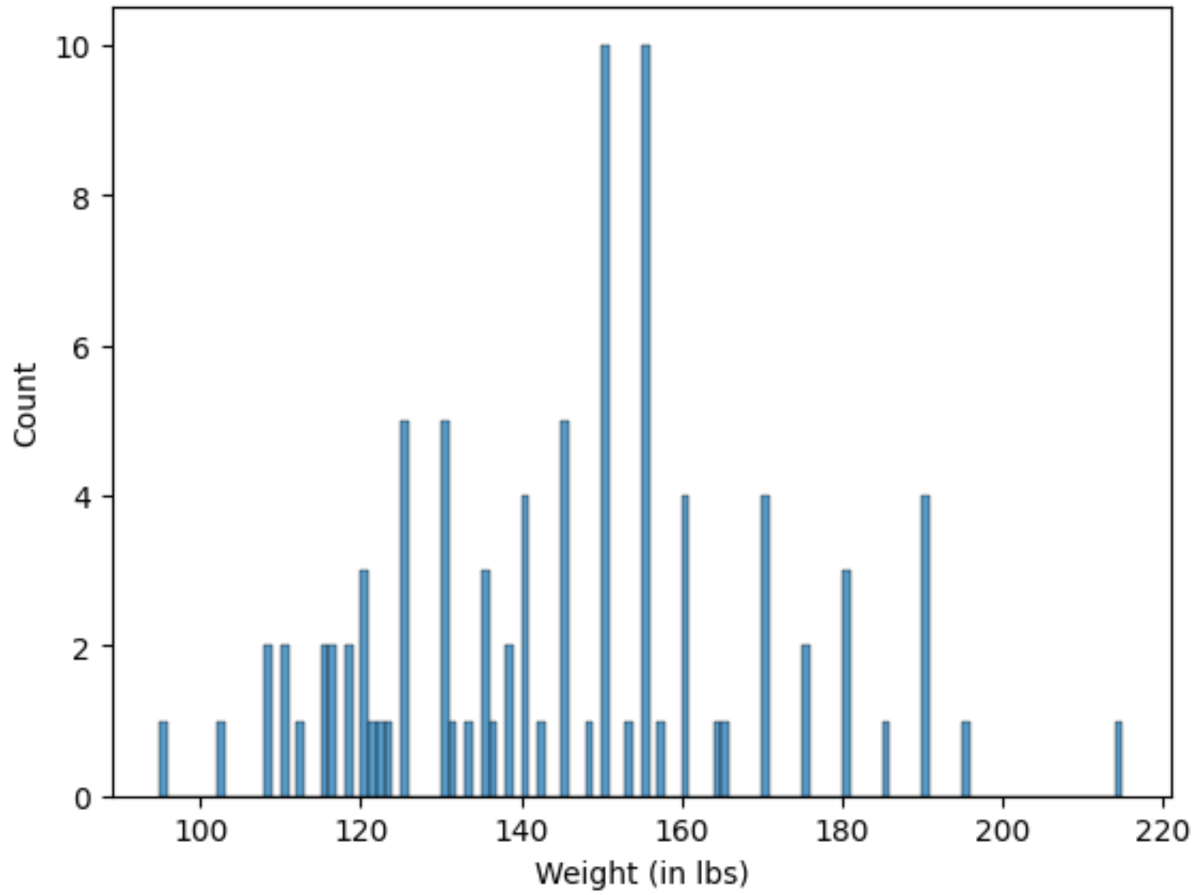
In [ ]:
```
male_df = pd.DataFrame([140, 145, 160, 190, 155, 165, 150, 190, 195, 138, 160, 155, 153, 145, 170, 175, 175, 170, 180, 135, 170, 157, 130, 185, 190, 155, 170, 155, 215, 150, 145, 155, 155, 150, 155, 150, 180, 160, 135, 160, 130, 155, 150, 148, 155, 150, 140, 180, 190, 145, 150, 164, 140, 142, 136, 123, 155],
    columns=['weight'])
male_df['sex'] = 'male'
female_df = pd.DataFrame([140, 120, 130, 138, 121, 116, 125, 145, 150, 112, 125, 130, 120, 130, 131, 120, 118, 125, 135, 125, 118, 122, 115, 102, 115, 150, 110, 116, 108, 95, 125, 133, 110, 150, 108],
    columns=['weight'])
female_df['sex'] = 'female'

weights_df = pd.concat([male_df, female_df], ignore_index=True)
weights_df.sample(5)
```

Out[ ]:

	weight	sex
49	145	male
39	160	male
25	155	male
75	135	female
33	150	male

We can now **plot** the data. For every number between `95` and `215`, plot a bar chart counting the number of people for a given weight.

In [ ]:
```
ax = sns.histplot(weights_df["weight"], bins=weights_df["weight"].max() - weights_df["weight"].min())
ax.set_xlabel("Weight (in lbs)")
plt.show()
```



🤔 What's the name of this graph?

Histogram

A histogram is a representation of the **distribution** of **numerical** data.

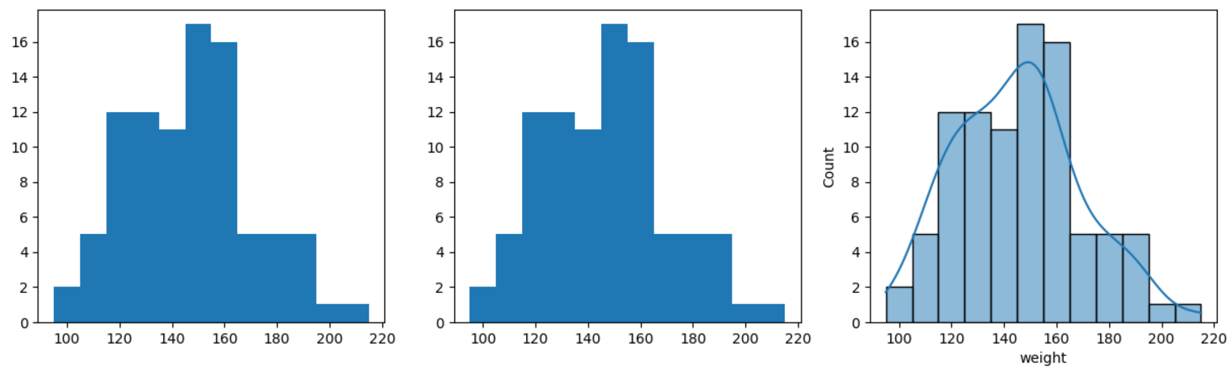
It is an **estimate of the probability distribution** of a continuous variable.

⚠️ Histogram (continuous variable) ≠ Bar chart (categorical or discrete variable)

Histogram Bins

Instead of drawing one bar per integer in $[95, 215]$, we can create **12** bins and count weights falling into these intervals.

In [ ]:
```
f, (ax1, ax2, ax3) = plt.subplots(1, 3, figsize=(15, 4))
ax1.hist(weights_df["weight"], bins=[95, 105, 115, 125, 135, 145, 155,
165, 175, 185, 195, 205, 215])
ax2.hist(weights_df["weight"], bins=12)
sns.histplot(weights_df["weight"], bins=12, ax=ax3, kde=True)
plt.show()
```

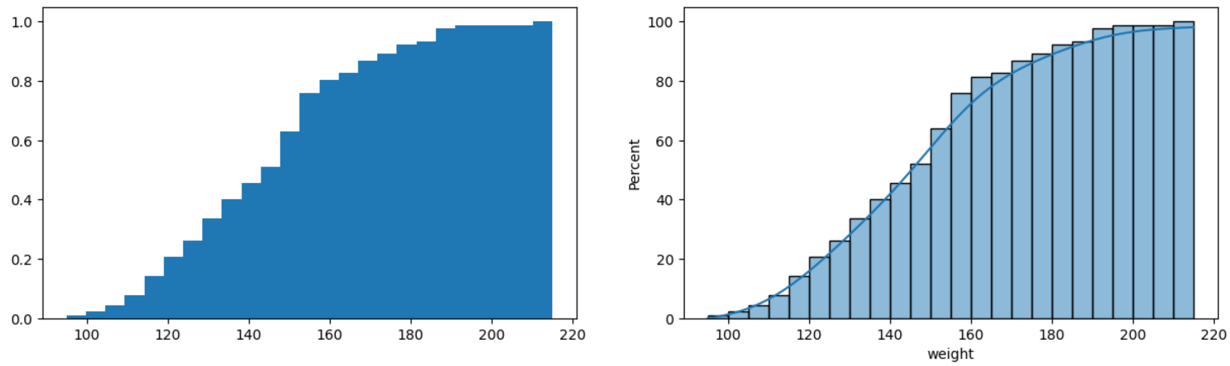


Cumulative plots

Alternatively, we can plot the count of weights *inferior* to a certain value.

Instead of the *counts*, we can also plot the *density* (sums up to 1) or *percentage* (sums up to 100).

In [ ]:
```
f, (ax1, ax2) = plt.subplots(1, 2, figsize=(15, 4))
ax1.hist(weights_df["weight"], bins=25, cumulative=True, density=True)
sns.histplot(weights_df["weight"], binwidth=5, ax=ax2, kde=True,
             cumulative=True, stat="percent");
```

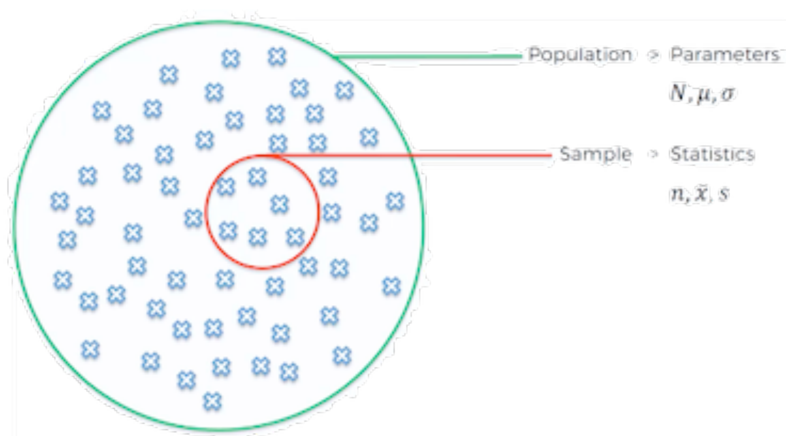


2️⃣ Summary statistics

🎯 Goal: **Summarize** and provide information about the data in a few measures.

Typical summary statistics include measures of:

- Location / central tendency (e.g. **mean**)
- Statistical Dispersion / spread (e.g. **variance**)
- Shape of the distribution (e.g. **skewness** & kurtosis)
- Linear **correlation** of two variables X and Y



The mean of a **population** of N elements is defined by:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

The mean of a **sample** of the population $x_1, x_2, \ldots, x_n$ ($n < N$) is defined by:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Median

The median is the value **separating** the higher half from the lower half of a data sample

Odd number of values:

1 3 3 **6** 7 8 9

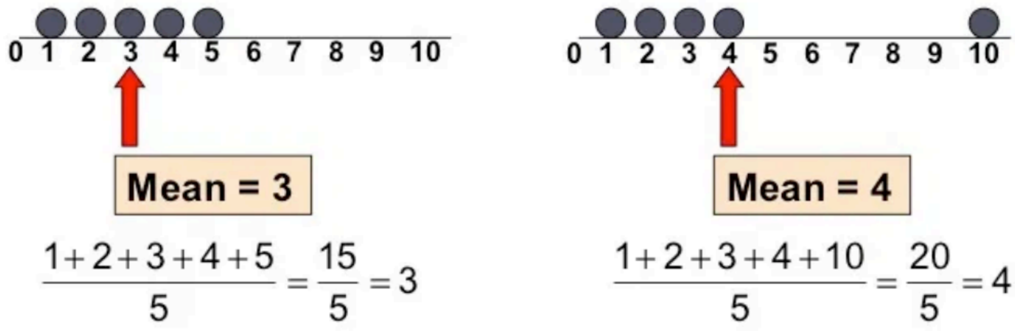
Even number of values

1 2 3 **4** **5** 6 8 9

Then the median is **4.5**

Mean vs Median

Median is robust against outliers.



Mode

The mode is the value that appears **most often**

One mode

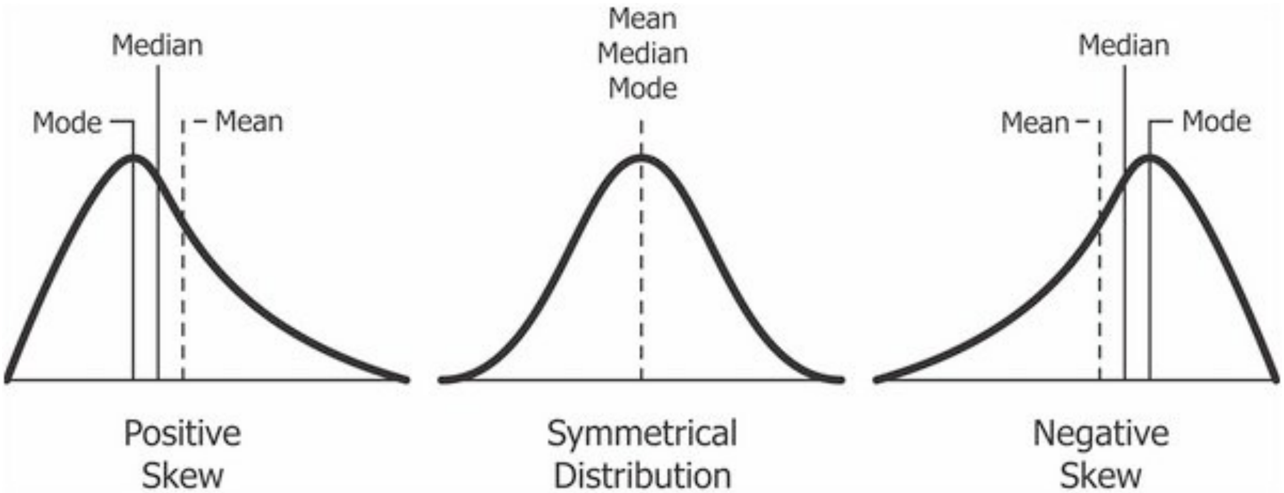
1 3 6 6 6 6 7 7 12 12 17

The mode is **6** and it is unique.

Bimodal dataset

1 1 2 4 4

There are two modes: **1** and **4**



👉 [Skewness (https://en.wikipedia.org/wiki/Skewness)](https://en.wikipedia.org/wiki/Skewness) (*[Asymétrie (https://fr.wikipedia.org/wiki/Asym%C3%A9trie_(statistiques))](https://fr.wikipedia.org/wiki/Asym%C3%A9trie_(statistiques))* 🇫🇷)

Statistical dispersion

Dispersion (also called variability, scatter, or spread) is the extent to which a distribution is stretched or squeezed.

Examples:

- **Variance** σ^2
- **Standard deviation** σ (*Écart-Type* 🇫🇷)
- Interquatile Range IQR
- [etc. (https://en.wikipedia.org/wiki/Statistical_dispersion)](https://en.wikipedia.org/wiki/Statistical_dispersion)

The **variance** of a population of N elements is:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

The **standard deviation** of a **population** of N elements is the square root of the variance:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Based on a **sample** of a population, a commonly used estimator of σ is the *sample standard deviation*:

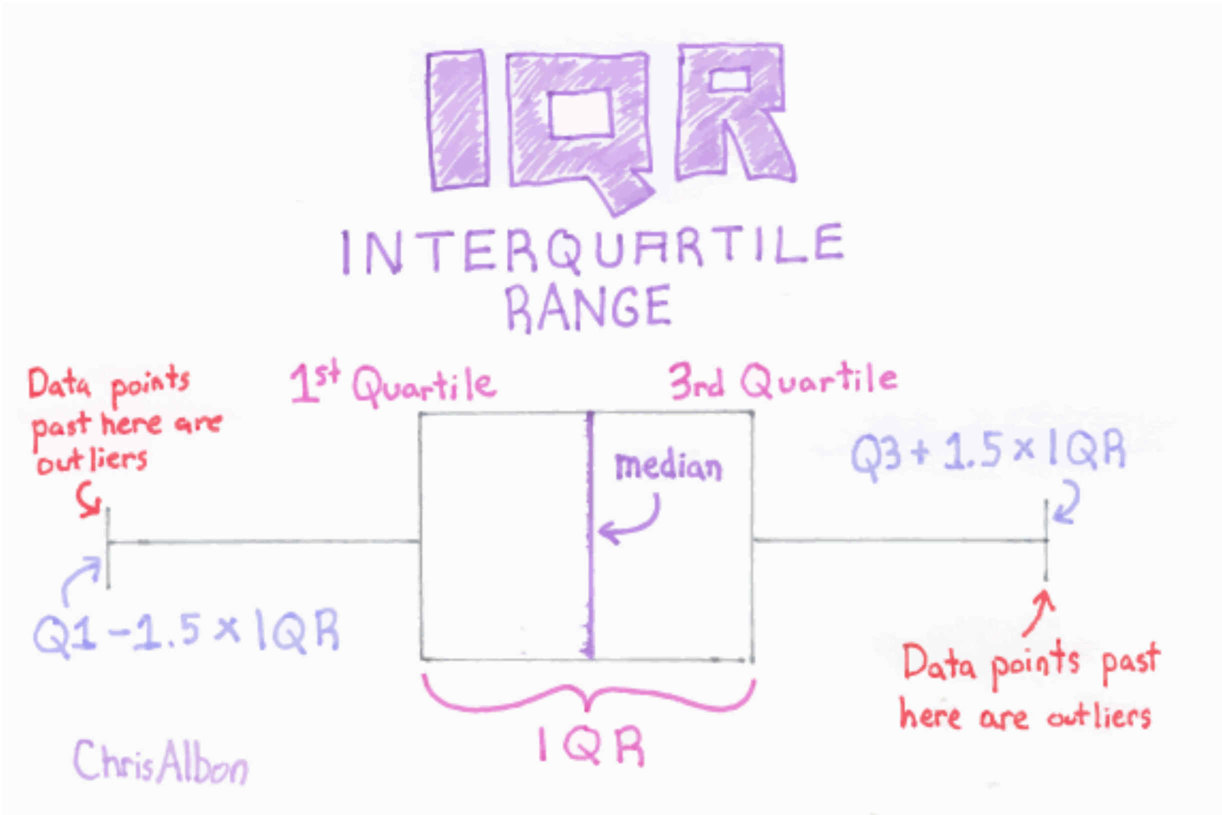
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

🤔 $\frac{1}{n}$ would give an underestimate of the true population variance (Bessel's correction (https://en.wikipedia.org/wiki/Bessel%27s_correction))

Interquartile range (IQR)

The difference between upper and lower quartiles: $\text{IQR} = Q_3 - Q_1$

💡 It can be used to identify outliers in the data set: they are defined as observations that fall below Q_1 - 1.5 IQR or above Q_3 + 1.5 IQR. 📈 IQR is very useful for **boxplots**!



❓ [*But my whiskers are not symmetrical!* (https://stackoverflow.com/questions/51694935/seaborns-boxplot-whiskers-meaning)](https://stackoverflow.com/questions/51694935/seaborns-boxplot-whiskers-meaning)

Summary Statistics

Summary statistics of a sample are the five following numbers:

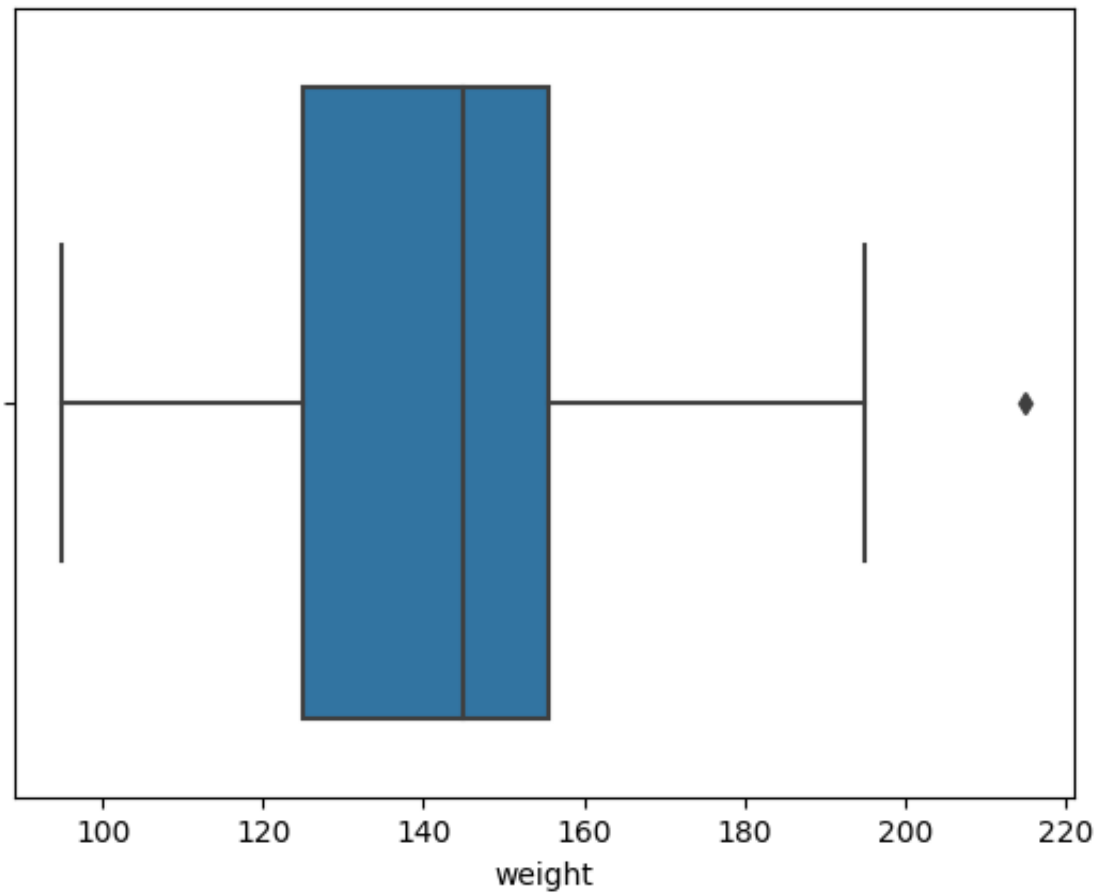
- min
- lower quartile (25%)
- median (50%)
- upper quartile (75%)
- max

➕ a boxplot

In [ ]:
```
weights_df['weight'].describe()
```

Out[ ]:
```
count     92.000000
mean     145.152174
std       23.739398
min       95.000000
25%      125.000000
50%      145.000000
75%      155.500000
max      215.000000
Name: weight, dtype: float64
```

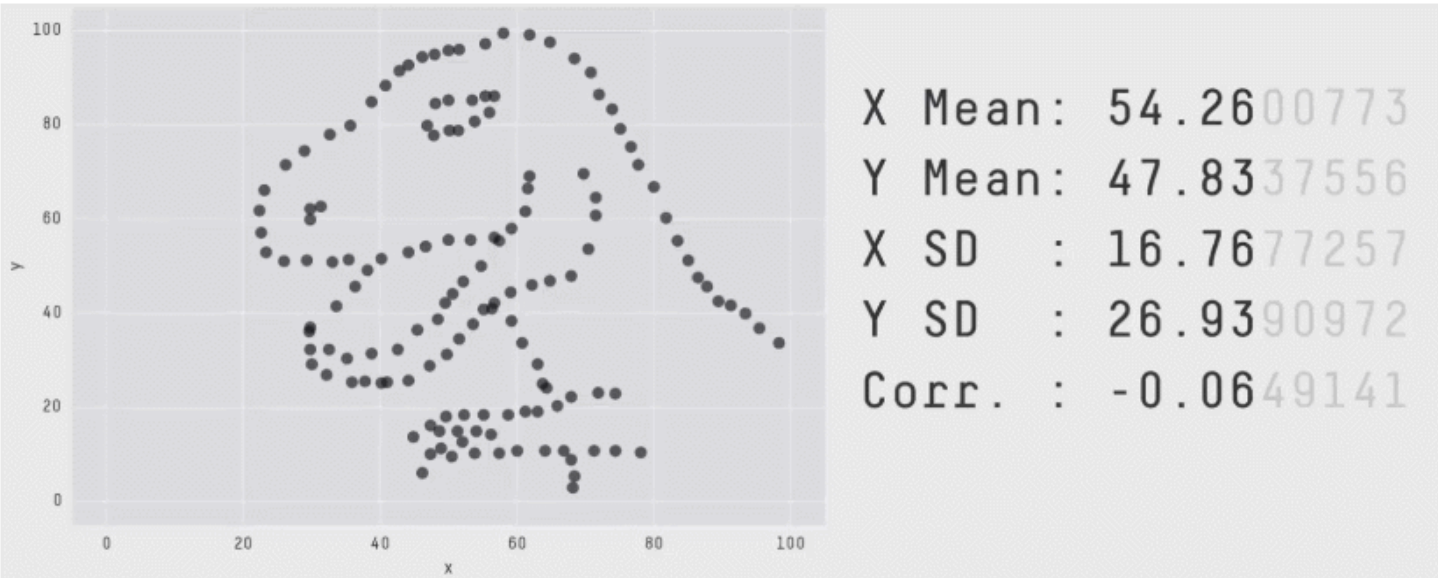
In [ ]:
```
sns.boxplot(x=weights_df['weight'])
plt.show()
```



Summary statistics are important, but...

⚠️ Beware of the Datasaurus (https://dl.acm.org/doi/10.1145/3025453.3025912) 🦕

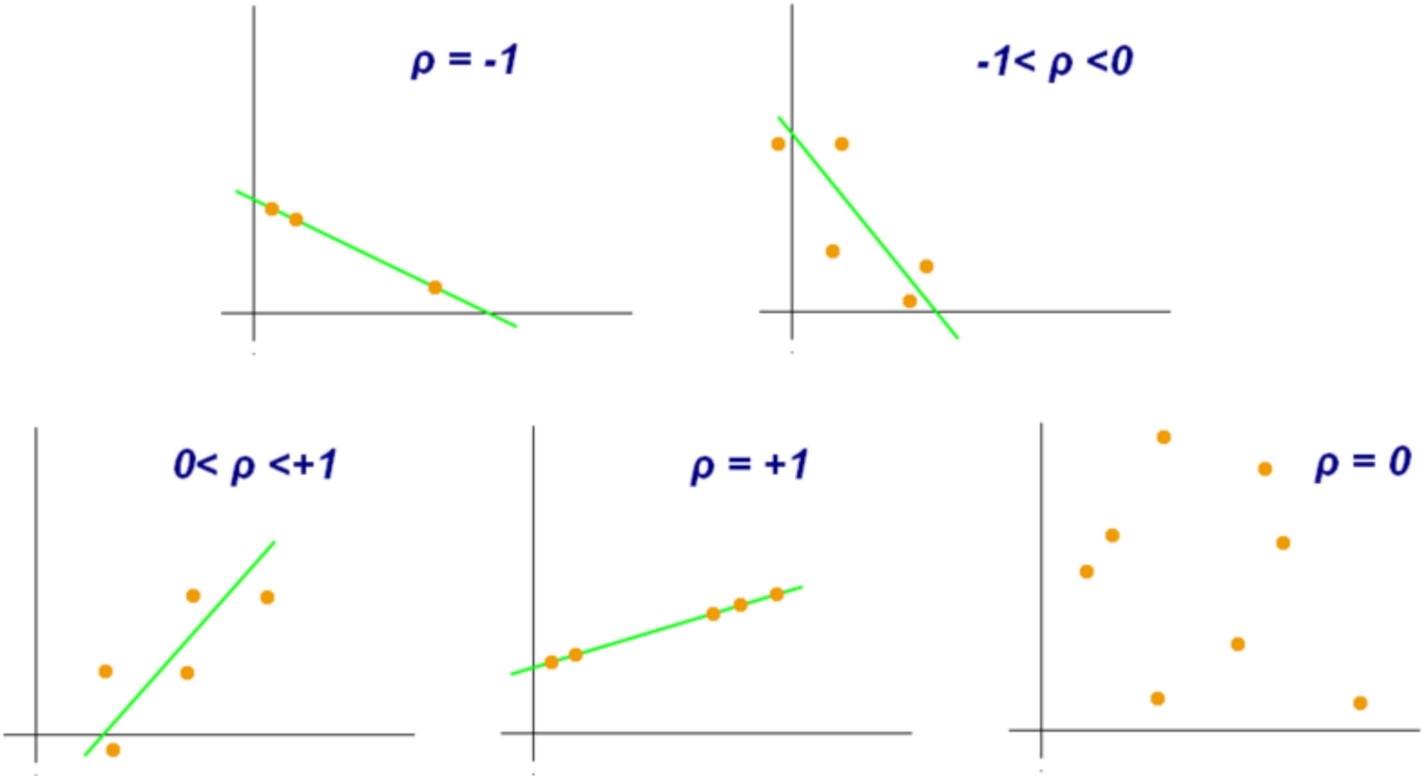
Summary statistics are not enough, you need to conduct Exploratory data analysis (https://en.wikipedia.org/wiki/Exploratory_data_analysis) too!

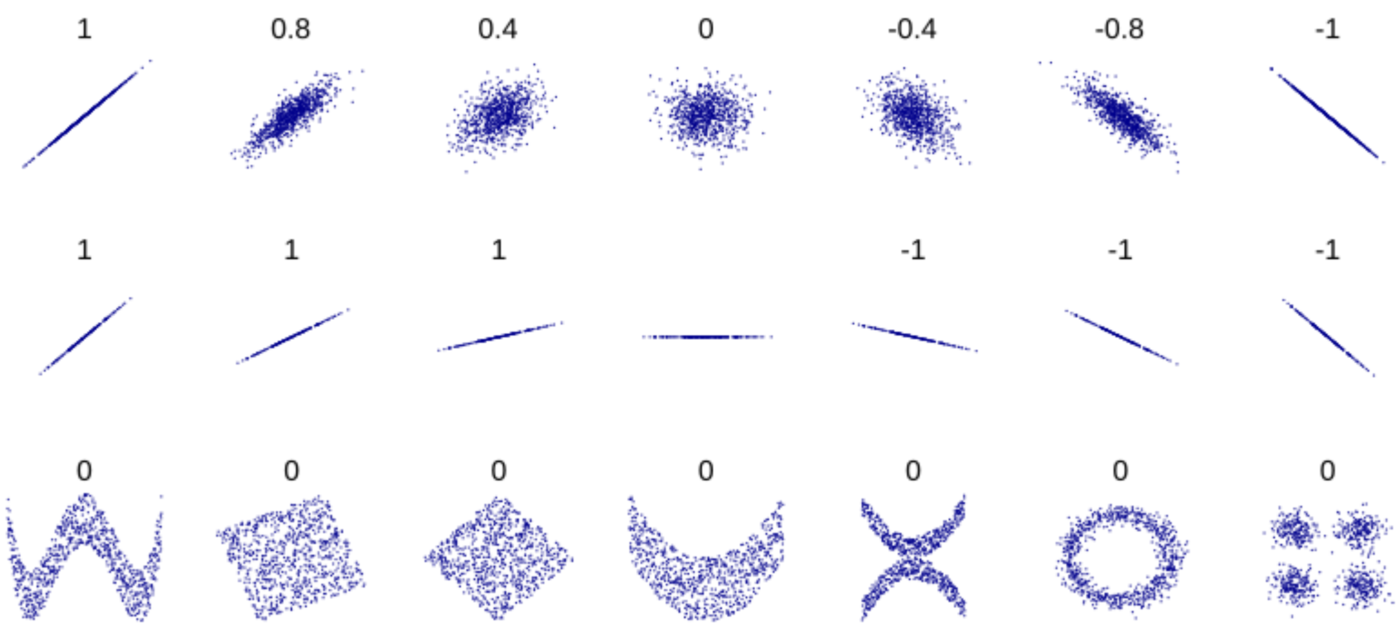


Correlation between 2 variables

The linear correlation between X and Y is also called the Pearson's coefficient (https://en.wikipedia.org/wiki/Pearson_correlation_coefficient)

$$r = Corr(X, Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}$$


$\rho = -1$
$-1 < \rho < 0$
$0 < \rho < +1$
$\rho = +1$
$\rho = 0$


1 0.8 0.4 0 -0.4 -0.8 -1
1 1 1 -1 -1 -1
0 0 0 0 0 0 0

Correlation vs. Independence?

- (X,Y) independent $\Rightarrow$ Corr(X,Y) = 0
- Corr(X,Y) = 0 $\nRightarrow$ (X,Y) independent (as r only captures **linear** dependence - cf Wikipedia (https://en.wikipedia.org/wiki/Correlation_and_dependence#Correlation_and_independence))

3️⃣ Probabilities

Sets

Probability theory uses the language of sets. A set is a collection of some items (elements).

Example:
$A=\{\clubsuit, \diamondsuit\}$

You can perform operations (https://www.probabilitycourse.com/chapter1/1_2_2_set_operations.php) on sets and visualize them with Venn diagrams:

- Union
- Intersection
- Complement
- Substraction
- Partition

Random experiment

- A random experiment is a process by which we observe something uncertain
- After the experiment, the result of the random experiment is known: it is the (**outcome**)
- The set of all possible outcomes is called the **sample space** $S, \Omega \text{ or } U$ (*Univers* 🇫🇷)

Examples:

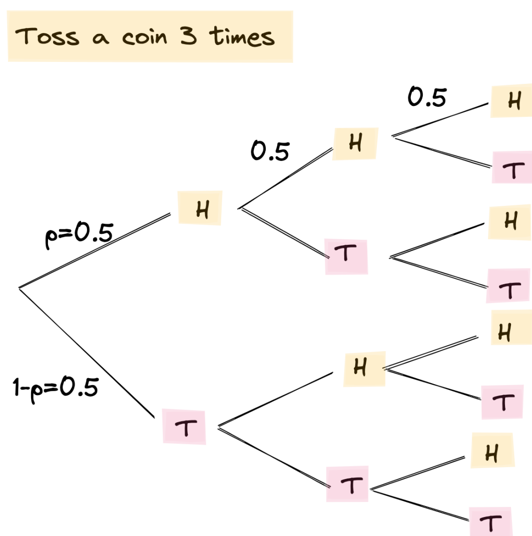
- Toss a coin. $S=H,T$
- Roll a die. $S=1,2,3,4,5,6$
- Observe the number of goals in a soccer match. $S=0,1,2,3,\ldots$

When we repeat a random experiment several times, we call each one of them a **trial** (épreuve 🇫🇷).

Sample space S is defined based on **how you define** the random experiment...

❓ If the experiment is *Toss a coin **three** times*, what's the sample space?

$$S=\{ (H,H,H),(H,H,T),(H,T,H),(T,H,H),(H,T,T),(T,H,T),(T,T,H),(T,T,T) \}$$



☝️ The goal is to assign a **probability** to events, defined as **subsets** of a sample space S.

Probability

We assign a probability measure $P(A)$ to an event A.

This is a value between 0 and 1 that shows how likely the event is.

Union & Intersection

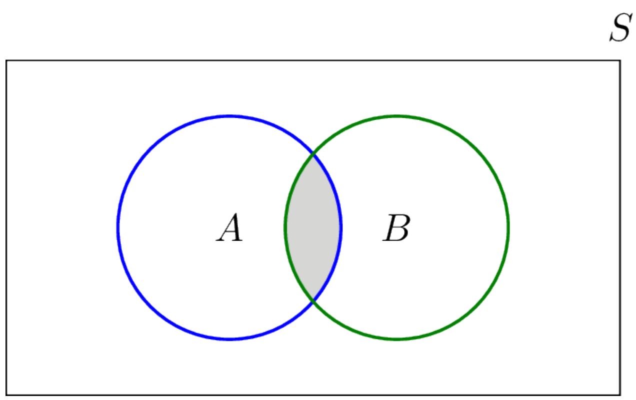
If A and B are events, then $A \cup B$ and $A \cap B$ are events too.

- $\cup$ - Union occurs iif A **or** B occurs
- $\cap$ - Intersection occurs iif A **and** B occurs

Some properties can be easily visualized with *Venn diagrams*:

$P(A^\complement)=1-P(A)$

$P(A \cup B)=P(A)+P(B)-P(A \cap B)$



Conditional Probability

As you obtain **additional information**, how would you update probabilities of events?

Defitinion

Let A and B be two events.

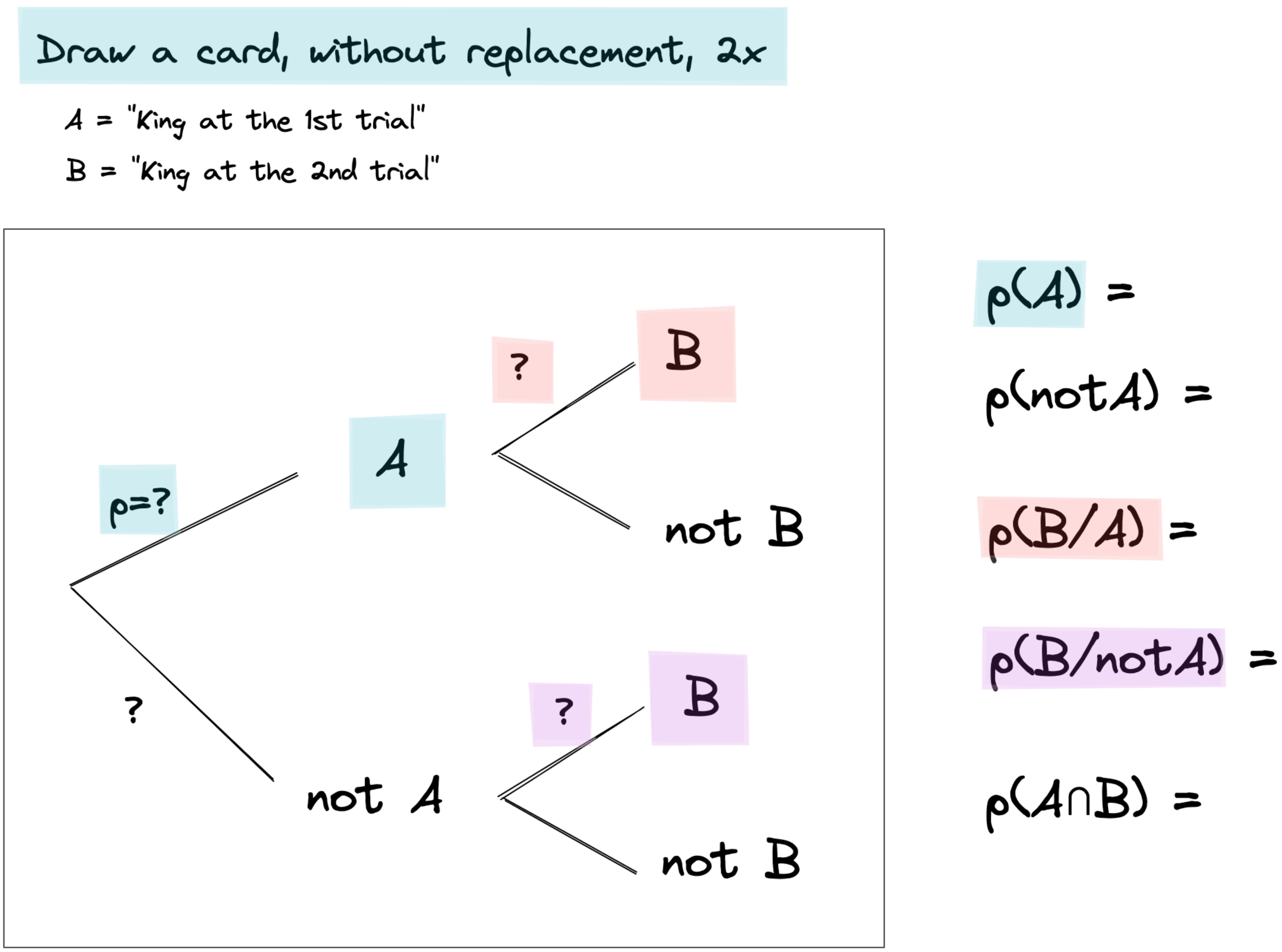
By **definition** (nothing to demonstrate):
$P(A \cap B) = P(A)\cdot P(B\mid A)$

$P(B\mid A)$ is called the **conditional probability of B given A**, sometimes also noted $P_A(B)$

👩‍🏫 **Example random experiment: Draw two cards, one at a time, without replacement, in a deck of 52 cards.**

You draw the first card: it's a King (event A).

❓ What's the probability that the second card is also a King (event B)?



Solution

The first card is not placed back in the deck which means A and B are **dependent**.

$$P(A) = \frac{4}{52} \approx 0.077$$

$$P(B \mid A) = \frac{3}{51} \approx 0.059$$

$$P(A \cap B) = P(A) \cdot P(B\mid A) = \frac{1}{221} \approx 0.005$$

Statistical Independence

Events A and B are independent iif:
$P(A\mid B) = P(A)$

[Bayes' Theorem (https://en.wikipedia.org/wiki/Bayes%27_theorem)](https://en.wikipedia.org/wiki/Bayes%27_theorem)

🤔 Suppose we know $P(B \mid A)$ and we want to compute $P(A \mid B)$

We know that, *by definition*

$$P(A \cap B) = P(A) \cdot P(B\mid A) \\ P(B \cap A) = P(B) \cdot P(A\mid B)$$

With a simple symmetry argument, we have $P(A \cap B) = P(B \cap A)$ so...

🔥 **Bayes Theorem** $P(A\mid B)={\frac {P(B\mid A) \cdot P(A)}{P(B)}}$ for any two events A and B, where $P(B) \neq 0$:

👩‍🔬 Example - Testing a Disease

A disease affects about 1 out of 1000 people.
$P(Sick) = 0.001$

There is a test to check whether the person has the disease and we know that:

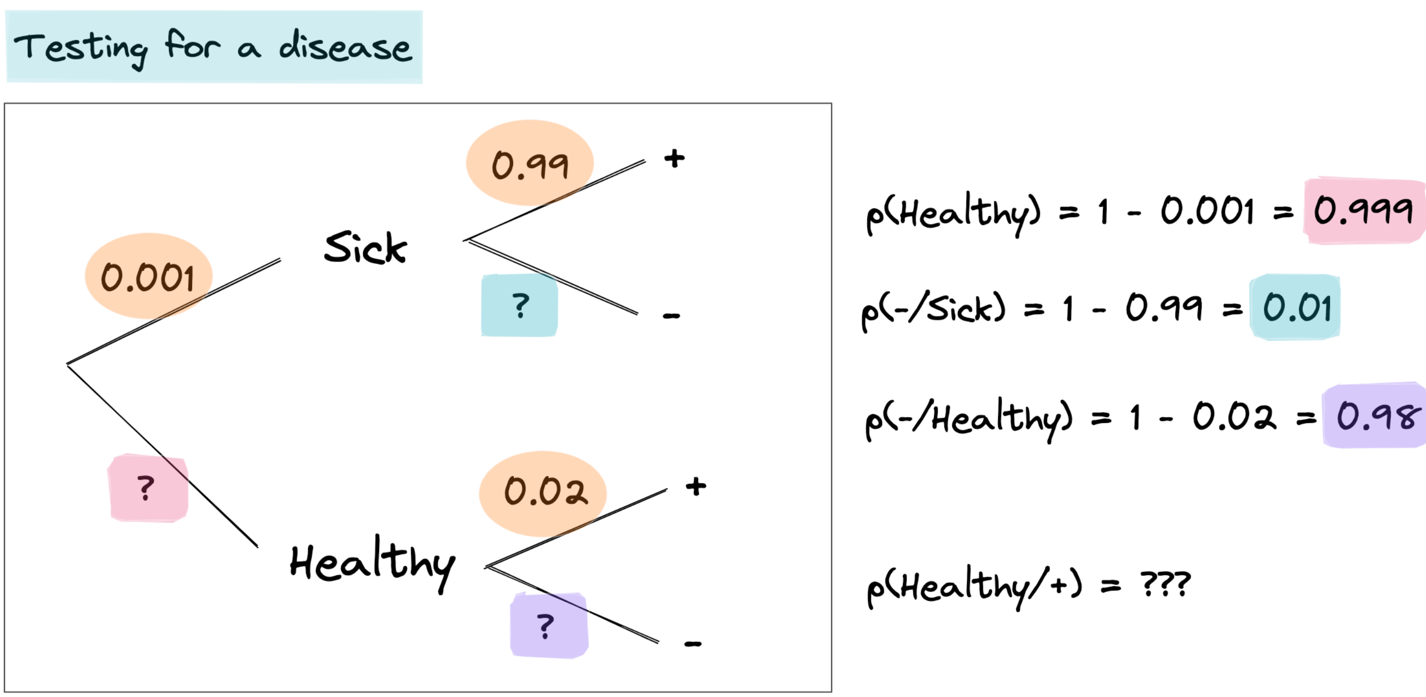
The probability that the test result is **positive** ($\pmb{+}$) given that the person **is sick** is 99% (**true positive**):
$$P(\pmb{+}\mid S_{ick}) = 0.99$$

The probability that the test result is **positive** ($\pmb{+}$) given that the person **is healthy** is 2% (**false positive**): ($H_{ealthy} = S_{ick}^\complement$)
$$P(\pmb{+}\mid H_{ealhy}) = P(\pmb{+}\mid S_{ick}^\complement) = 0.02$$

A random person gets tested for the disease and the result comes back positive 😱.

🤔 What is the probability that the person **actually** has the disease?

$$P(S_{ick} \mid \pmb{+}) = \text{?}$$



Bayes' theorem can be written as:
$$P(S_{ick}\mid \pmb{+})={\frac {P(\pmb{+}\mid S_{ick}) \cdot P(S_{ick})}{P(\pmb{+})}}$$

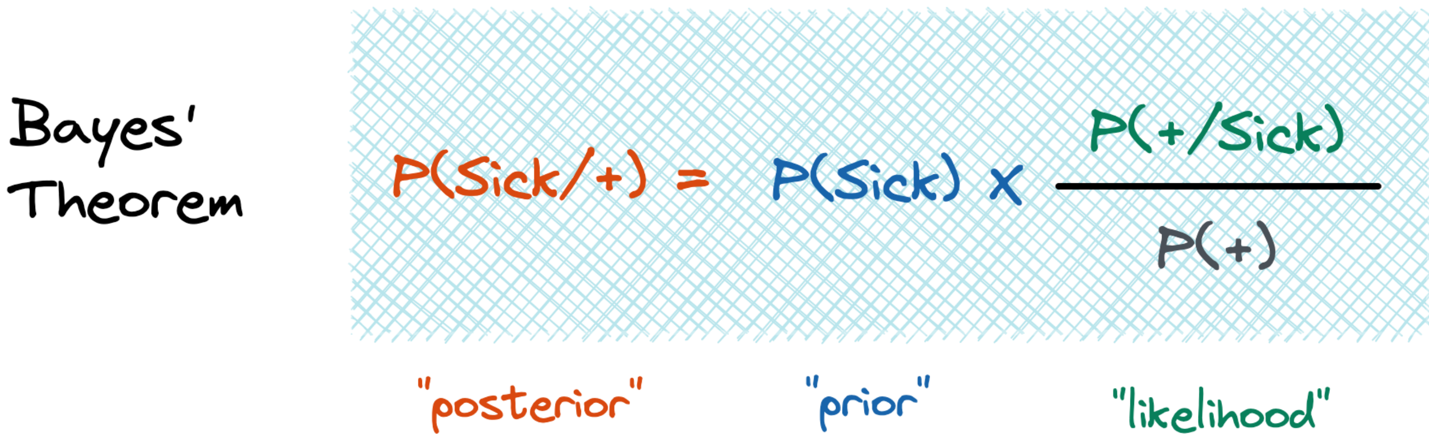
$$\begin{align} P(\pmb{+}) &= P(\pmb{+} \cap S_{ick}) + P(\pmb{+} \cap S_{ick}^\complement) \end{align}$$

$$\begin{align} P(\pmb{+}) &= P(\pmb{+} \mid S_{ick}) \cdot P(S_{ick}) + P(\pmb{+} \mid H_{ealthy}) \cdot P(H_{ealthy})\\ &= P(\pmb{+} \mid S_{ick}) \cdot P(S_{ick}) + P(\pmb{+} \mid H_{ealthy}) \cdot (1 - P(S_{ick}))\\ &= 0.99 \times 0.001 + 0.02 \times (1 - 0.001) \\ &= 0.02097 \end{align}$$

We can now compute $P(S_{ick} \mid \pmb{+})$:

$$\begin{align} P(S_{ick}\mid \pmb{+}) &={\frac {0.99 \times 0.001}{0.02097}} \\ &\approx 0.047 \end{align}$$

💡 Less than 5% of people positively tested with test $\pmb{+}$ actually have the disease. This is called the **false positive paradox** (https://en.wikipedia.org/wiki/Base_rate_fallacy#False_positive_paradox)



👉 Bayes Theorem (https://www.youtube.com/watch?v=HZGCoVF3YvM) on Youtube, by 3blue1brown 👌

4️⃣ Random variable

Statistics ❤️ Probabilities

To analyze random experiments, we focus on some **numerical** aspects of the experiment.

For example, in a soccer game we may be interested in the number of goals, shots, etc...

Example of a random experiment

Let's toss a fair coin **twice**.

Sample space is $S = \{ (H, H), (H, T), (T, H), (T, T) \}$

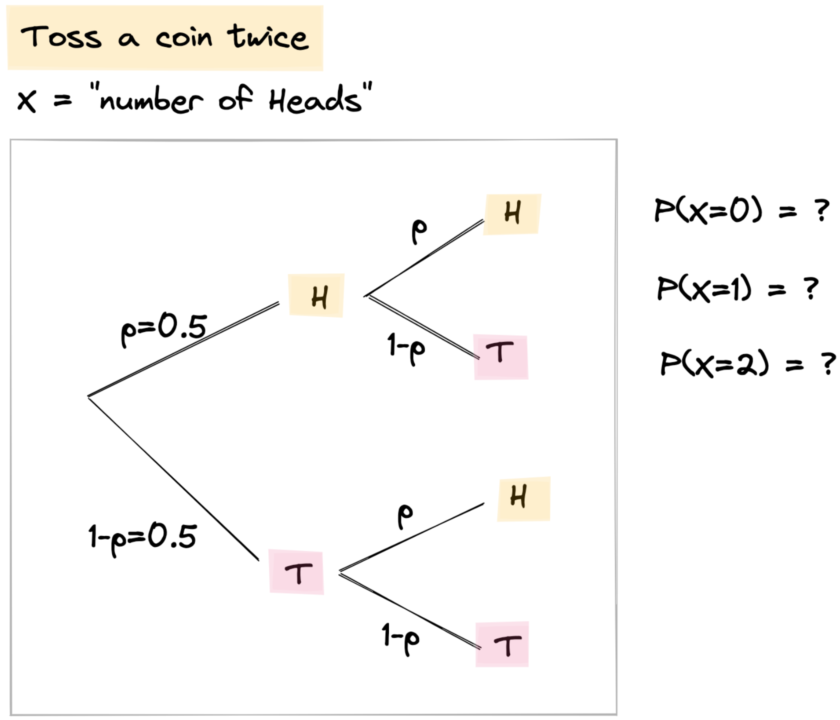
Let's define the random variable X as the **number of heads**.

Definition

A random variable X is a function from the sample space to the real numbers:
$X:S\rightarrow \mathbb{R}$

The range of a random variable X, shown by $Range(X)$, is the set of possible values of X.

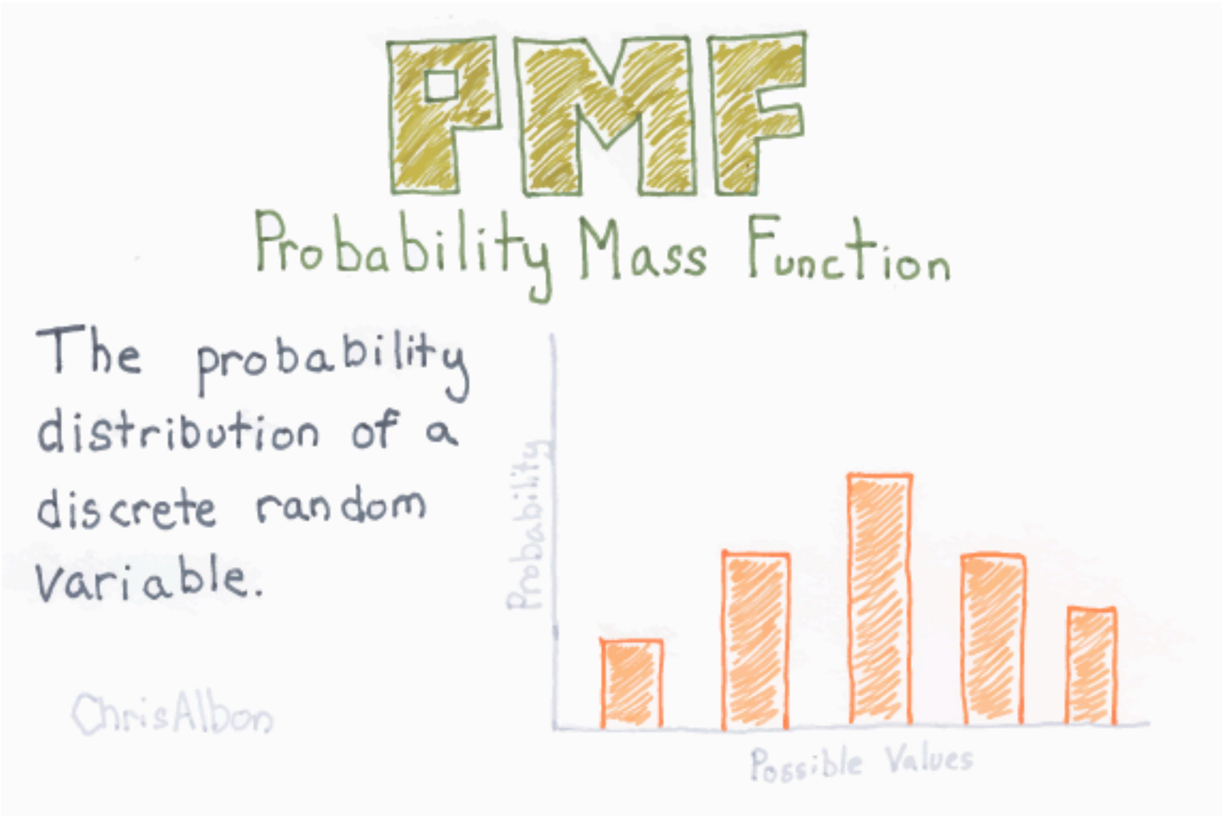
In our example (X as the **number of heads**):
$Range(X) = \{ 0, 1, 2 \}$



Probability Mass Function (PMF (https://en.wikipedia.org/wiki/Probability_mass_function))

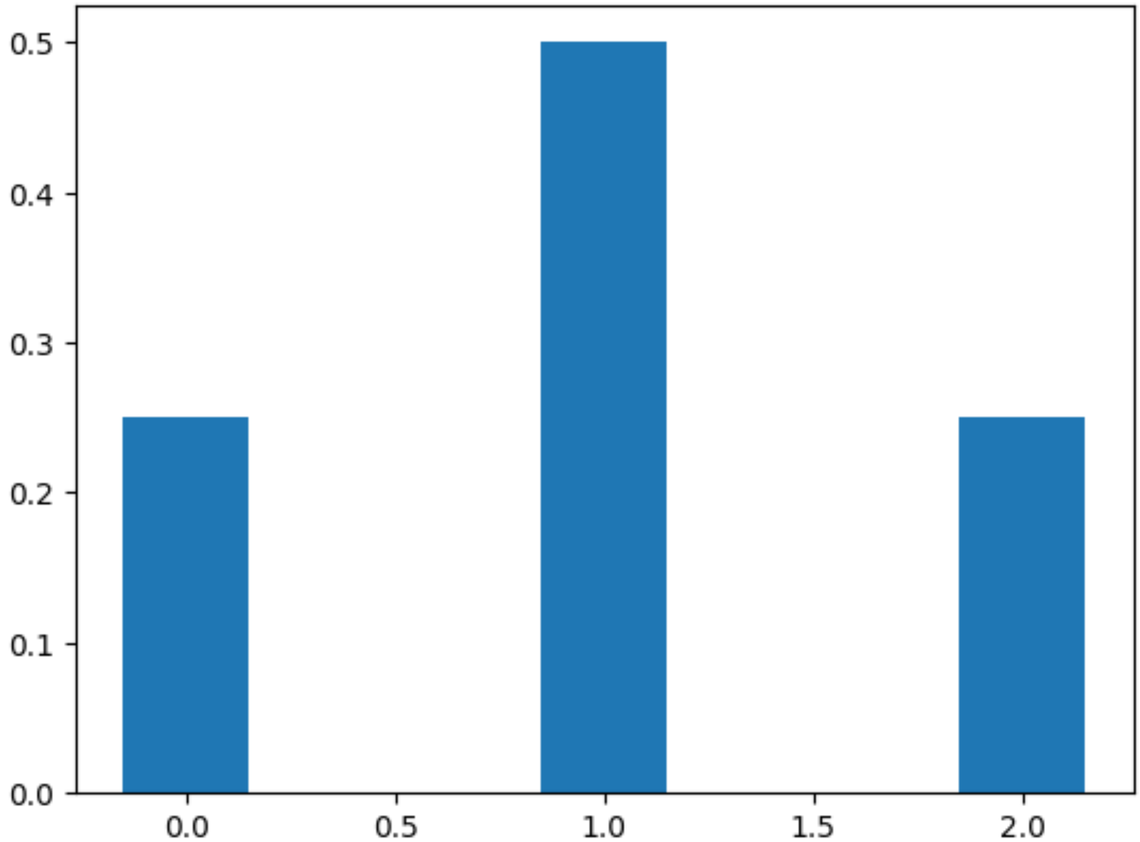
The PMF for X is defined as:

$\forall x_i \text{ in } Range(X),\ \text{pmf}_{X}(x_i) = P(X = x_i)$



✏️ Let's draw the PMF of $X =$ **"Number of Heads"** for the sample space:
$S = \{ (H, H), (H, T), (T, H), (T, T) \}$

In [ ]:
```
plt.bar(x=[0,1,2], height=[0.25, 0.5, 0.25], width=0.3);
```



Expected value (https://en.wikipedia.org/wiki/Expected_value) E[X]

Intuitively, a random variable's expected value $E[X]$ represents the **average** of a large number of independent realizations of the random variable X.

X being a **discrete** random value:
$${\displaystyle \operatorname {E} [X]=\sum _{i=1}^{n }x_{i}\,p(x_i).}$$

$$E[\textit{"Number of heads in 2 tosses"}] = 0*0.25 + 1*0.5 + 2*0.25 = 1$$

Bernoulli process (https://en.wikipedia.org/wiki/Bernoulli_process)

Take a random experiment with exactly **two possible outcomes** and repeat this experiment multiple times

✏️ e.g let's define our Bernoulli experiment as:

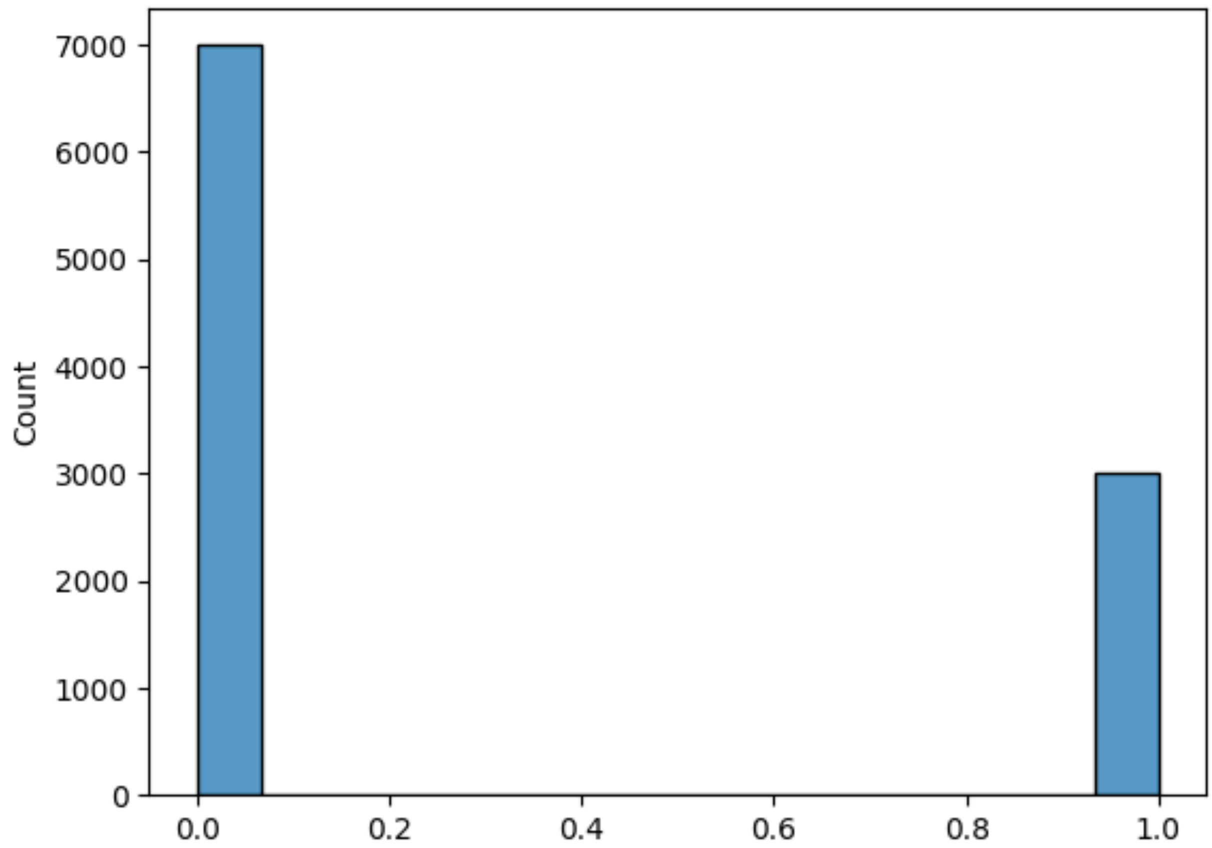
> toss a coin once
>
> with a proba `p = 0.3` of getting 1 (e.g. *Heads*) at each toss

and repeat this experiment a `size = 10000` times (large to smooth-out noise)

In [ ]:
```
size = 10000
p = 0.3
np.random.binomial(n=1, p=p, size=size)
```

Out[ ]:
```
array([0, 0, 1, ..., 1, 0, 0])
```

In [ ]:
```
# Plot results after 10000 repetitions
sns.histplot(np.random.binomial(n=1, p=p, size=size), kde=False);
```



☝️ The associated PMF is called a **Bernoulli Distribution** $B(p)$.

[Binomial distribution (https://en.wikipedia.org/wiki/Binomial_distribution)](https://en.wikipedia.org/wiki/Binomial_distribution)

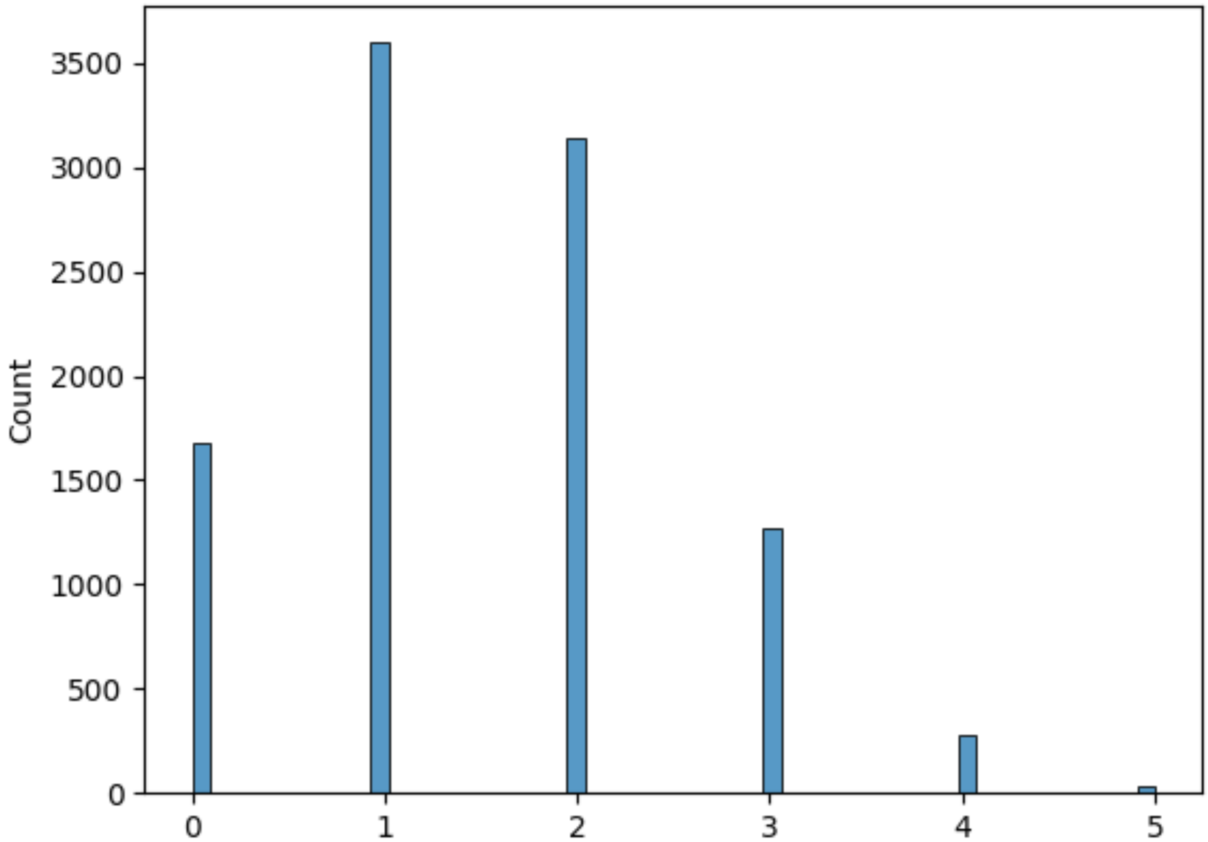
🧐 What if we repeat the tossing of the coin `n = 5` times and **count the number of heads in the 5 trials** ?

This is called a **Binomial experiment**

In [ ]:
```
n = 5
np.random.binomial(n=n, p=p, size=size)
```

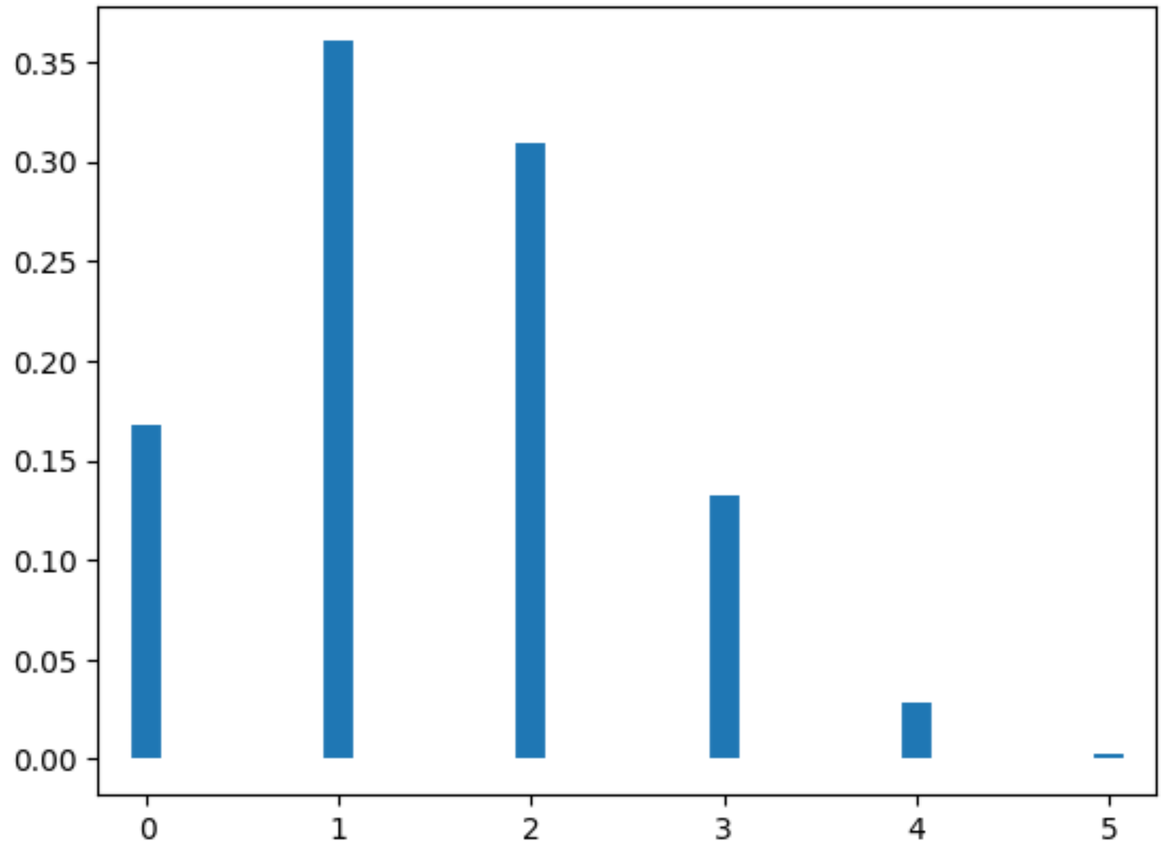
Out[ ]:
```
array([1, 0, 4, ..., 2, 2, 0])
```

In [ ]:
```
# Plot results after counting the number of heads ("5 trials")
n = 5
sns.histplot(np.random.binomial(n=n, p=p, size=size), kde=False);
```



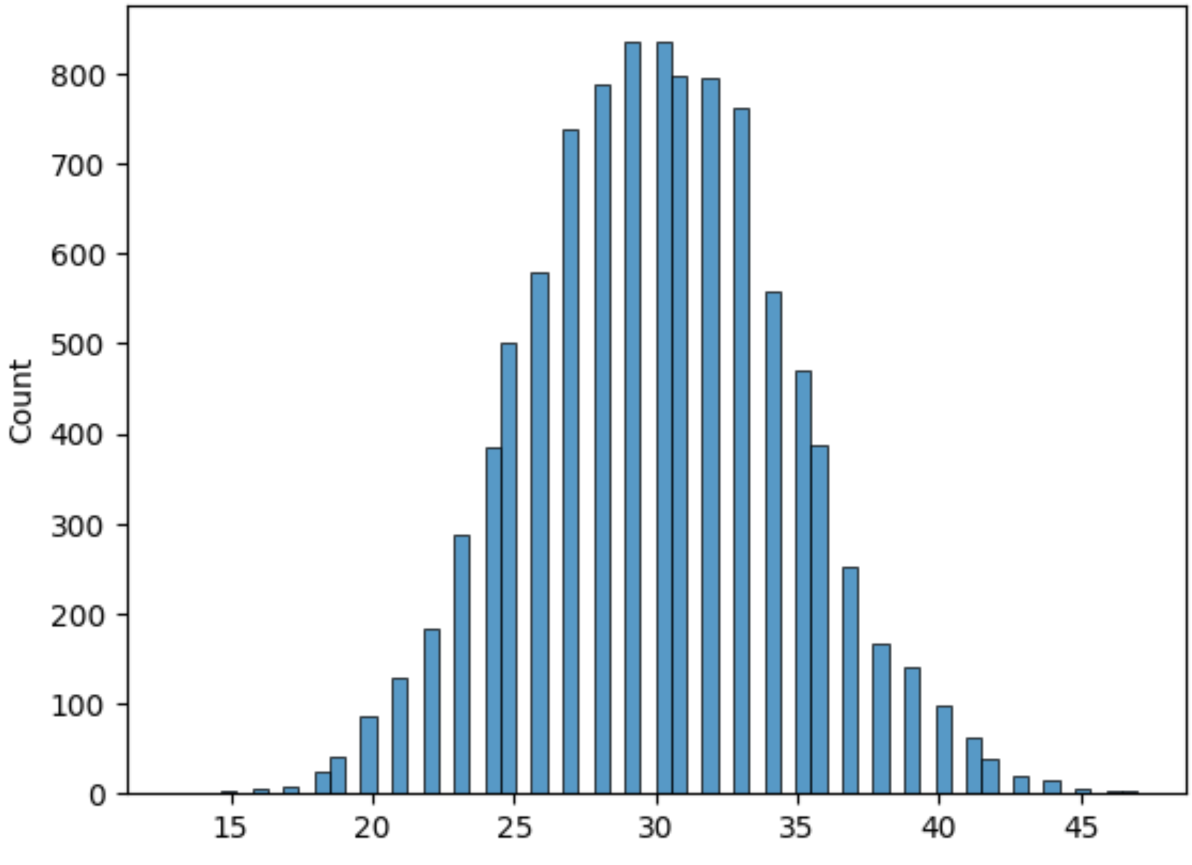
☝️ The **sum** of n **Bernoulli Distributions** B(p) is called a **Binomial Distribution** \mathcal B(n, p).

In [ ]:
```
# Below is the theoretical "perfect" binomial B(n=5, p=0.3) distributi
on (obtained if "size" --> infinity)
n = 5
x = np.arange(n + 1)
pmf = stats.binom.pmf(x, n, p)
plt.vlines(x, 0, pmf, linewidth=10);
```



As n increases, **Binomial Distribution** B(n, p) approximates a [normal distribution (https://en.wikipedia.org/wiki/Normal_distribution)](https://en.wikipedia.org/wiki/Normal_distribution) \mathcal N centered on n*p

In [ ]:
```
n = 100
sns.histplot(np.random.binomial(n=n, p=p, size=size), kde=False);
```

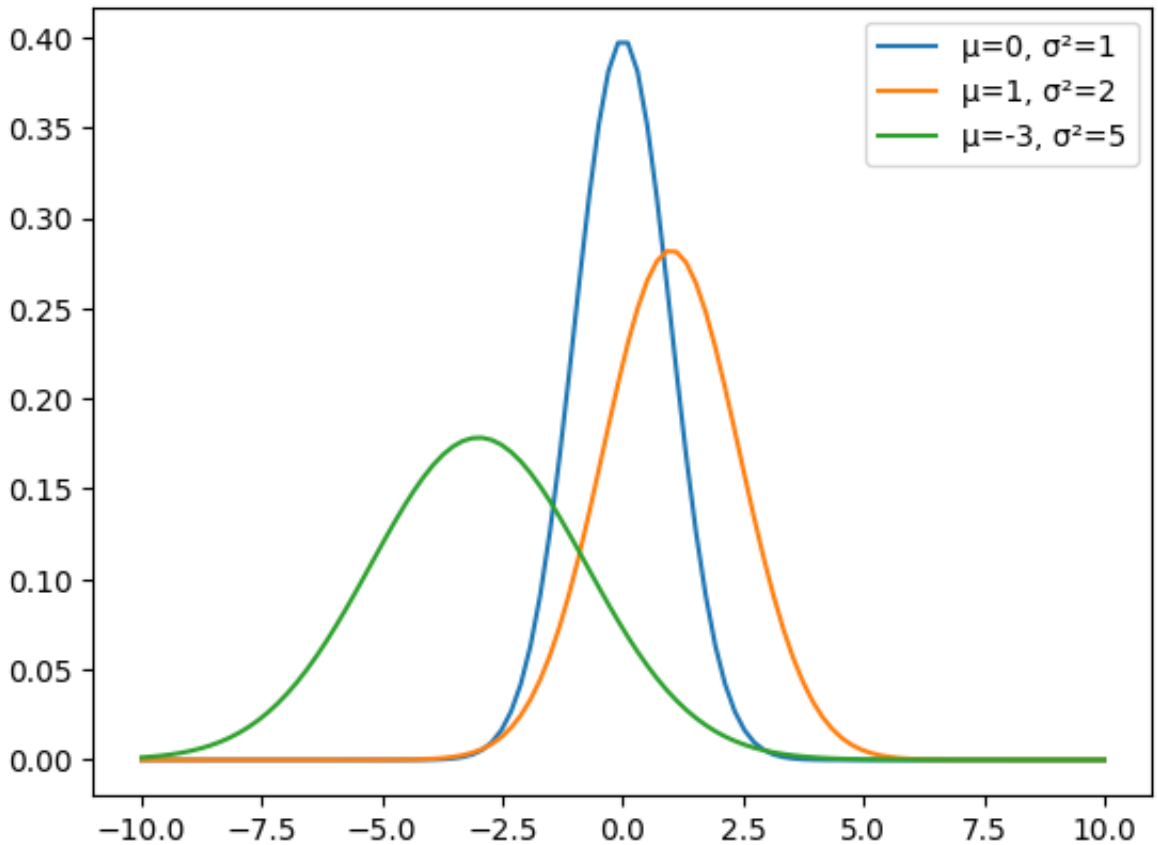


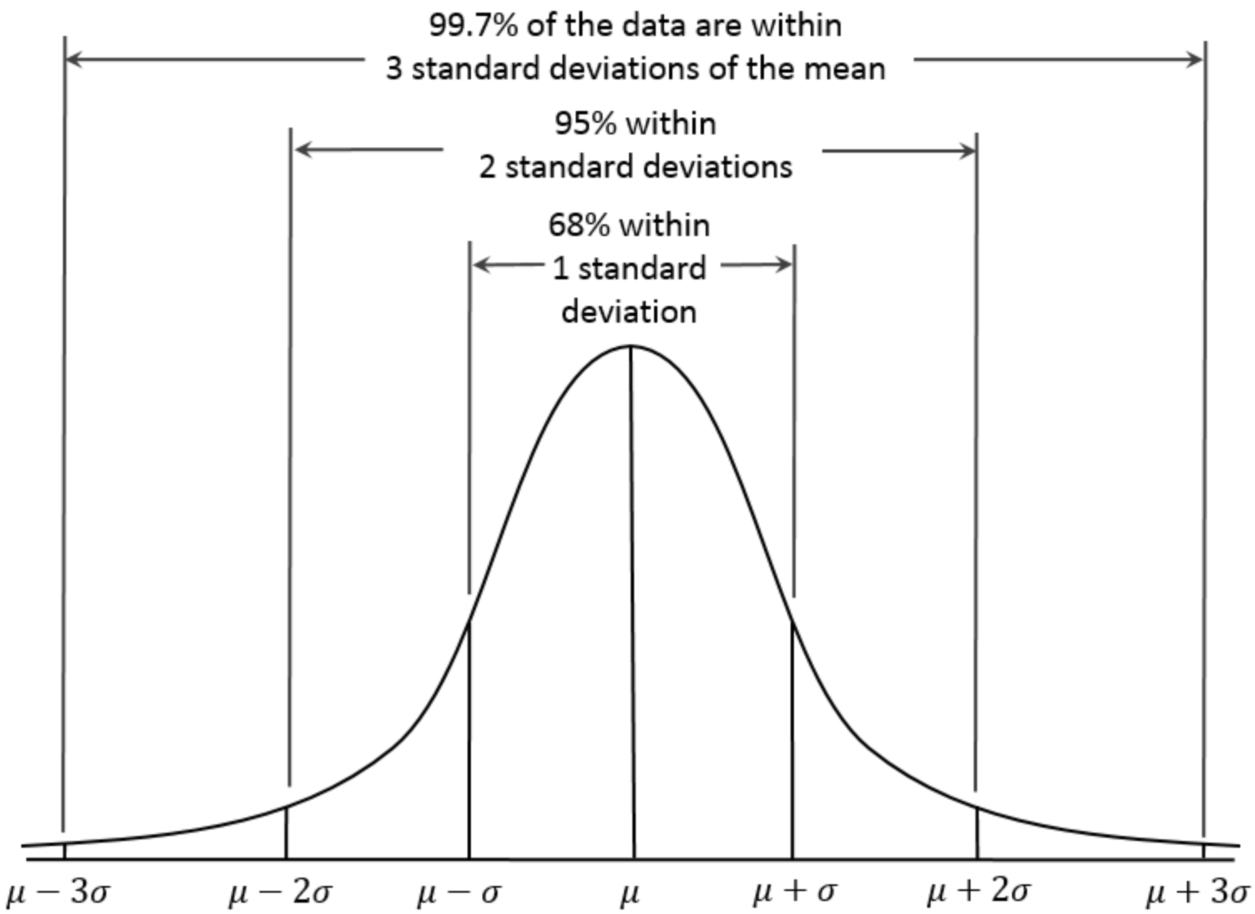
🔥 Normal Distribution 🔥

PDF (Probability Density Function) is:

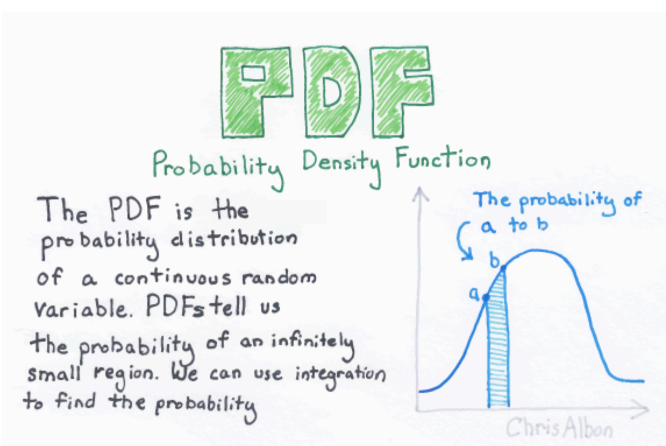
{\mathcal N(\mu, \sigma) ={\frac {1}{\sqrt {2\pi \sigma ^{2}}}}e^{-{\frac {(x-\mu)^{2}}{2\sigma ^{2}}}}}

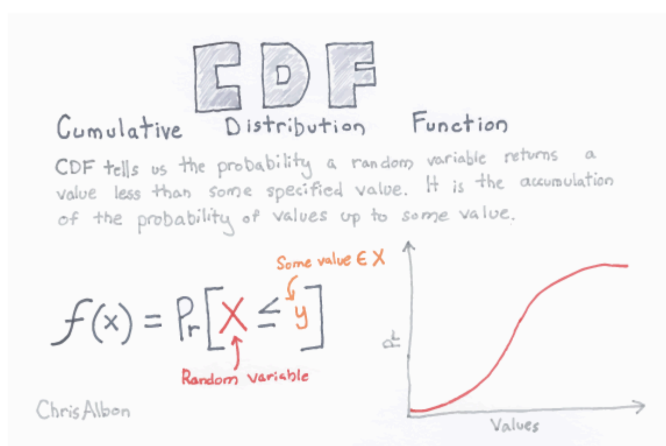
In [ ]:
```
def plot_normal_distribution(mu, variance):
    sigma = math.sqrt(variance)
    x = np.linspace(-10, 10, 100)
    plt.plot(x, stats.norm.pdf(x, mu, sigma), label=f"μ={mu}, σ²={vari
ance}")

plot_normal_distribution(0, 1)
plot_normal_distribution(1, 2)
plot_normal_distribution(-3, 5)
plt.legend()
plt.show()
```



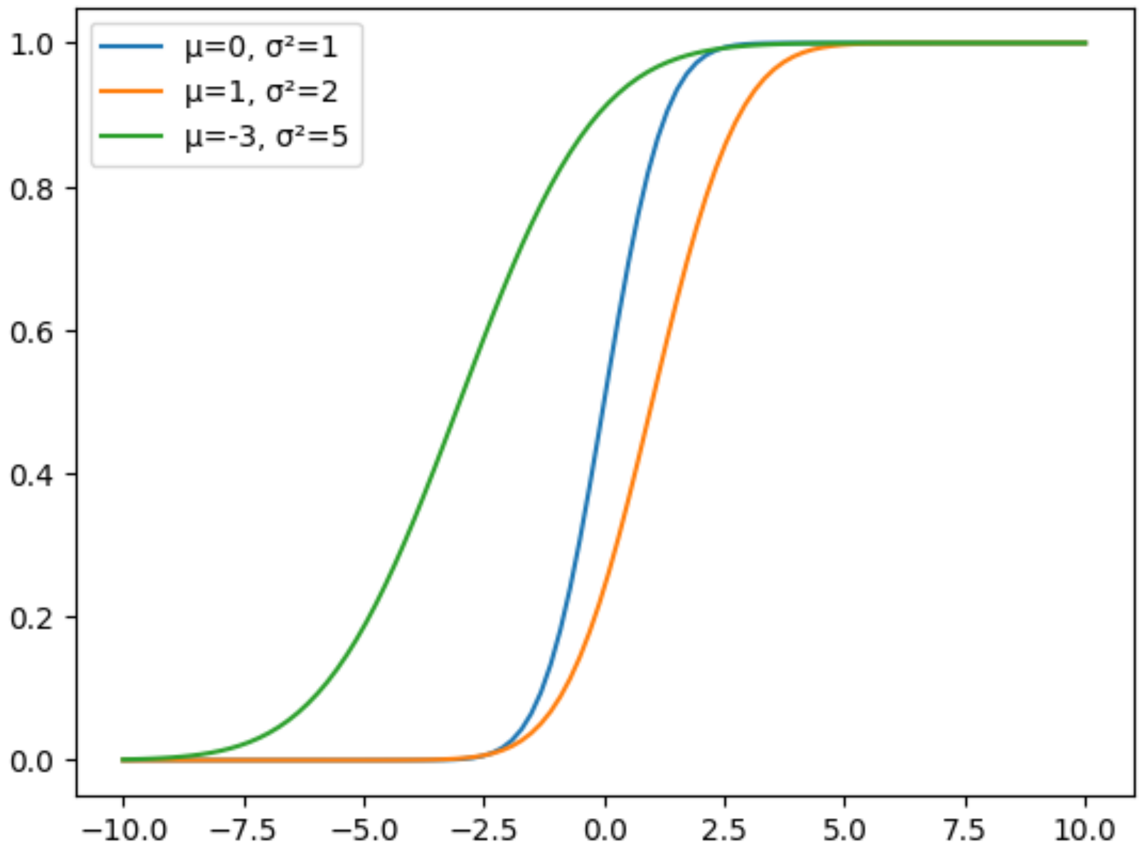

99.7% of the data are within
3 standard deviations of the mean
95% within
2 standard deviations
68% within
1 standard
deviation
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$

PDF vs CDF?


PDF
Probability Density Function
The PDF is the probability distribution of a continuous random variable. PDFs tell us the probability of an infinitely small region. We can use integration to find the probability
The probability of a to b
a
b
Chris Albon


CDF
Cumulative Distribution Function
CDF tells us the probability a random variable returns a value less than some specified value. It is the accumulation of the probability of values up to some value.
Some value ∈ X
$f(x) = P_r[X \leq y]$
Random variable
Values
Chris Albon

In [ ]:
```
def plot_cumulative_normal_distribution(mu, variance):
    sigma = math.sqrt(variance)
    x = np.linspace(-10, 10, 100)
    plt.plot(x, stats.norm.cdf(x, mu, sigma), label=f"μ={mu}, σ²={vari
ance}")

plot_cumulative_normal_distribution(0, 1)
plot_cumulative_normal_distribution(1, 2)
plot_cumulative_normal_distribution(-3, 5)
plt.legend()
plt.show()
```



5️⃣ Central Limit Theorem 💪

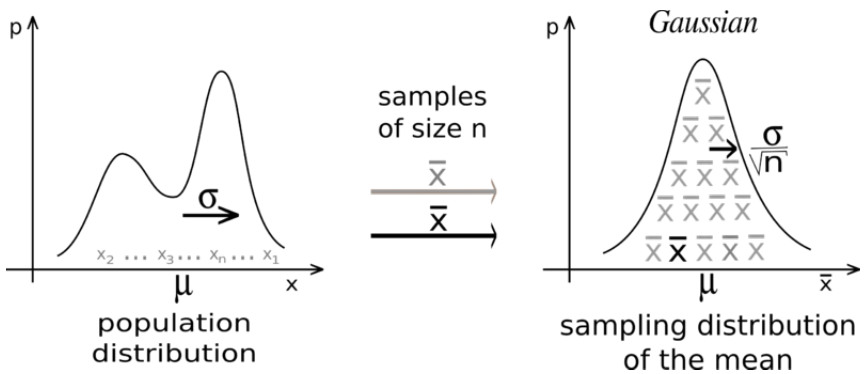
We will cover the [Central Limit Theorem (https://en.wikipedia.org/wiki/Central_limit_theorem)](https://en.wikipedia.org/wiki/Central_limit_theorem) extensively tonight during the RECAP. Here are the main ideas for future reference:

- We saw that the sum/mean of a **Bernoulli process** converges towards a \mathcal N distribution
- Actually, this holds true for **any** random process!

When **independent** random variables $X_1\dots X_n$ with **common** probability distribution (with mean μ and standard devation σ) are **added**:

- Their mean $\overline X$ converges towards a normal distribution as the number of samples n increases
- centered on the common mean μ
- with standard deviation ${\frac{\sigma}{\sqrt n}}$

This holds true **whatever** the form of the common distribution is



[Z-score (https://en.wikipedia.org/wiki/Standard_score)](https://en.wikipedia.org/wiki/Standard_score)

If x is an observation derived from a random variable $X(\mu,\sigma)$
$z={x-\mu \over \sigma }$

z = value of x expressed in **number of std above/below the mean**

Then CLT can be re-written as:
${\textstyle Z=\left({\frac {{\overline{X}}-\mu }{\sigma /\surd n}}\right)}\to \mathcal N(0,1) \ \text{as } n\to \infty$

Cheat Sheet

Summary statistics
· Mean μ, Median, Mode
· Standard deviation σ, Variance σ^2, IQR
· Correlation $r = Corr(X,Y)$

Probability
· Conditional probability $P(B|A)$
· Independence $P(B|A) = P(B)$
· Bayes Theorem $P(B|A) = P(A|B) {\frac {P(B)}{P(A)}}$

Random variables X (numerical outcome of a random experiment)
Random process (repeated sequence of random variable trials)

Distribution of probability
· Binomial $\mathcal B(n, p)$ from Bernoulli (0/1) processes
· Normal $\mathcal N(\mu, \sigma^2)$ from sum of idd random variables (https://en.wikipedia.org/wiki/Independent_and_identically_distributed_random_variables)

Central Limit Theorem
· $\overline{X} = {\frac{X_1+...+X_n}{n}} \xrightarrow[n \to \infty]{} \mathcal N(\mu,{(\frac{\sigma}{\surd n})^2})$
· ${\textstyle Z=\left({\frac {{\overline{X}}-\mu }{\sigma /\surd n}}\right)} \xrightarrow[n \to \infty]{} \mathcal N(0,1)$

Going Further

Statistics:

- Sensitivity & Specificity (https://en.wikipedia.org/wiki/Sensitivity_and_specificity)
- Central Limit Theorem (https://en.wikipedia.org/wiki/Central_limit_theorem)
- Skew and Kurtosis (https://codeburst.io/2-important-statistics-terms-you-need-to-know-in-data-science-skewness-and-kurtosis-388fef94eeaa)
- Sampling & Statistical Inference (https://www.probabilitycourse.com/chapter8/8_1_0_intro.php)
- Confidence Interval (https://en.wikipedia.org/wiki/Confidence_interval)
- Simpsons' Paradox (https://en.wikipedia.org/wiki/Simpson%27s_paradox)
- Bootstrapping (https://en.wikipedia.org/wiki/Bootstrapping_%28statistics%29)

Probabilities

- Probability Mass & Density Functions (https://hadrienj.github.io/posts/Probability-Mass-and-Density-Functions/) (Hadrien Jean 🙌)
- Marginal & Conditional Probabilities (https://hadrienj.github.io/posts/Marginal-and-Conditional-Probability/) (🙌)
- 📺 3Blue1Brown - Central Limit Theorem - Inutuitive understanding (https://www.youtube.com/watch?v=zeJD6dqJ5lo)
- 📺 3Blue1Brown - Bayes Theorem - Inutuitive understanding (https://www.youtube.com/watch?v=HZGCoVF3YvM)

Your turn!