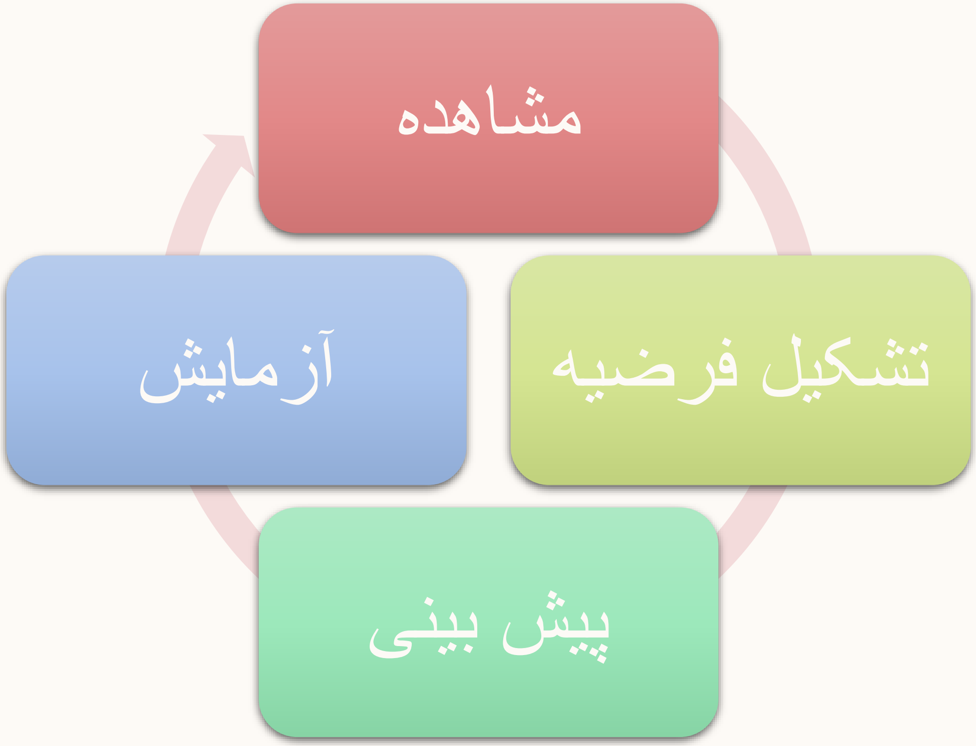
STATISTICS IN ELECTROPHYSIOLOGY

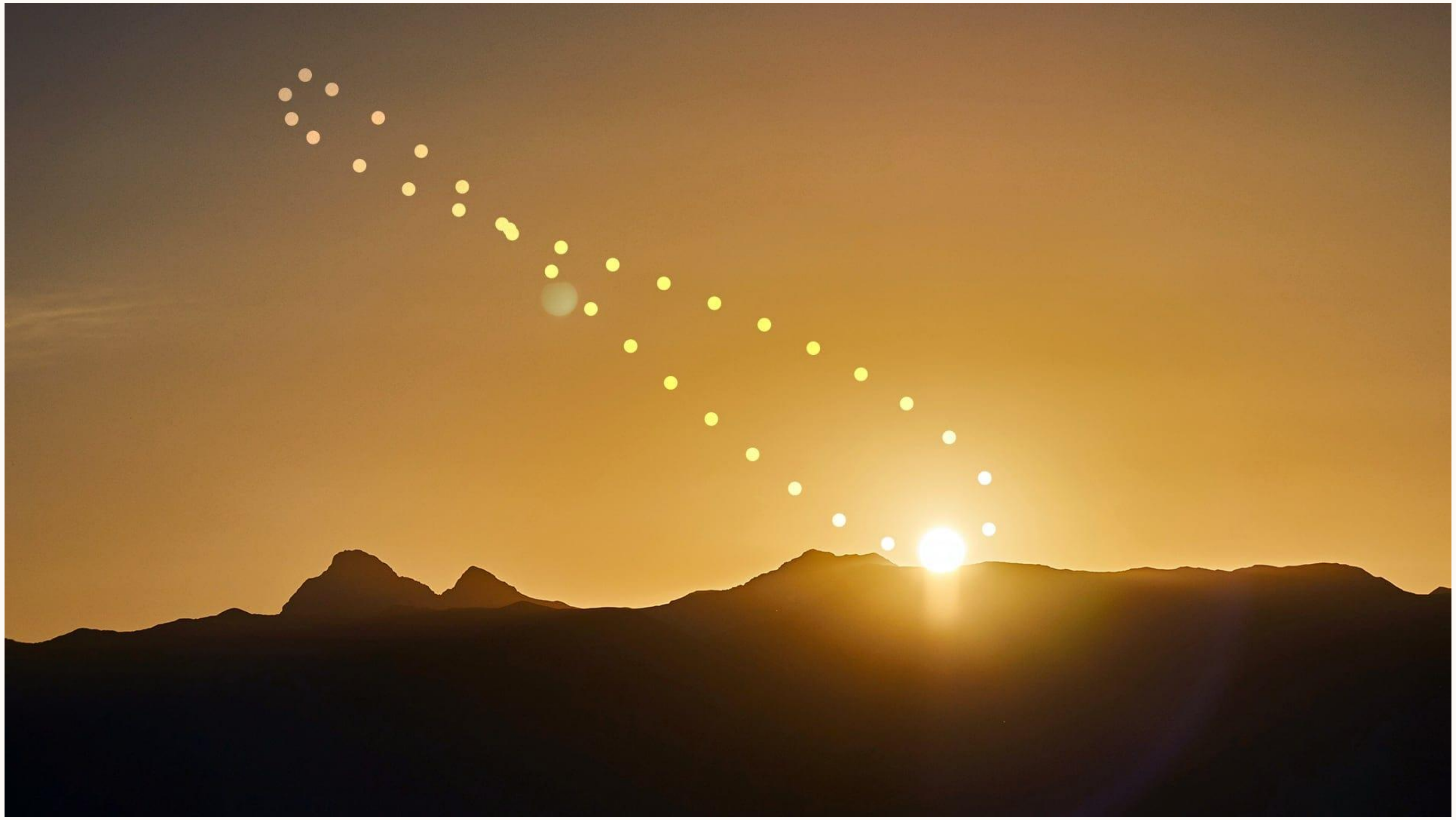
Babak.Aliyari.edu@gmail.com

https://interdisciplinaryschools.com/

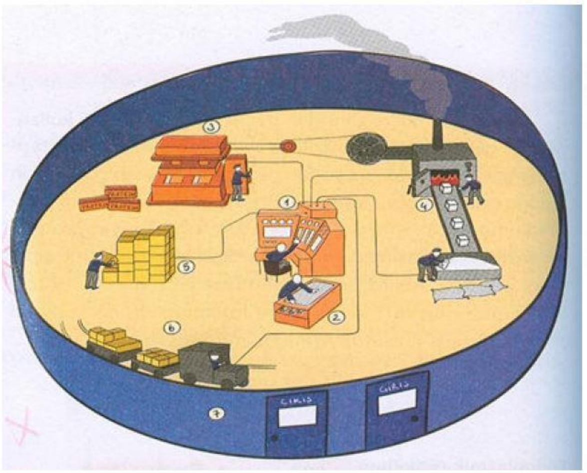
SCIENCE

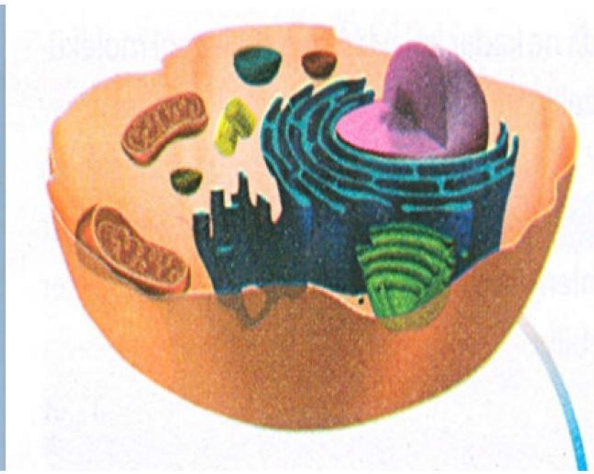


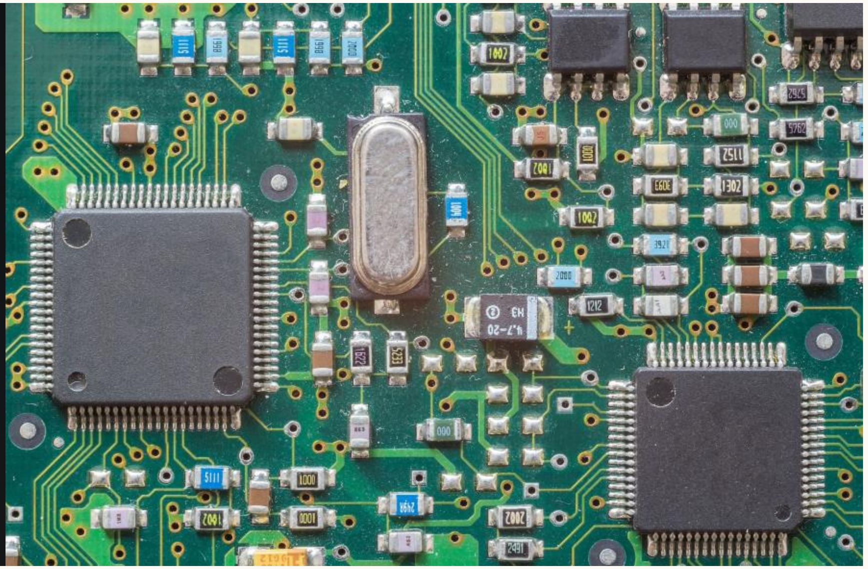
ANALEMMA



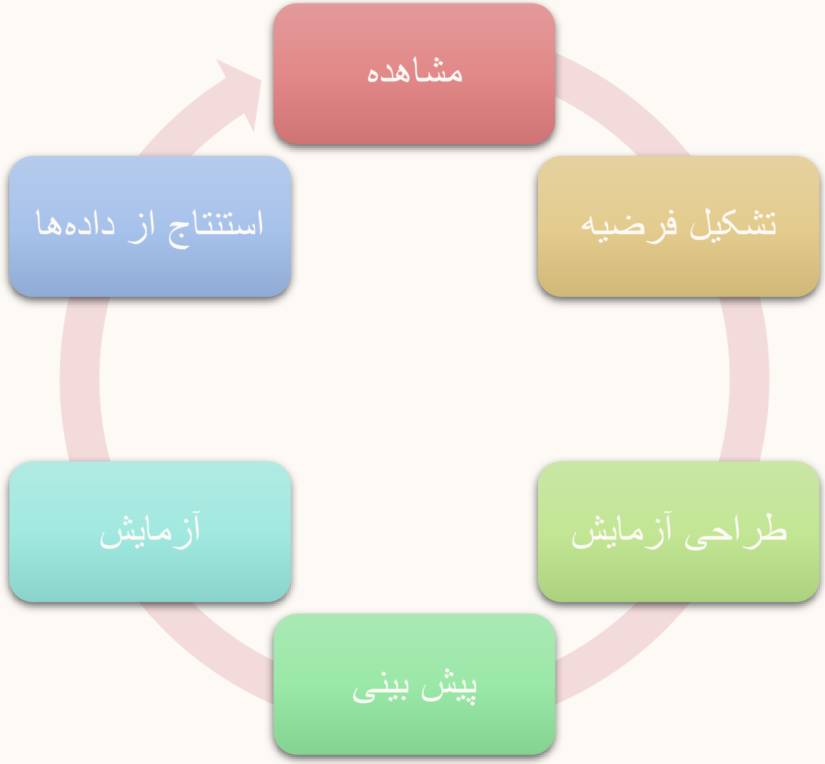
ANALOGY







SCIENCE



طراحی آزمایش: چگونه داده‌ها را به گونه‌ای جمع‌آوری کنیم که بتواند به طور معتبر به پرسش‌های تحقیق پاسخ دهد؟

استنتاج از داده‌ها: پس از جمع‌آوری داده‌ها، چگونه می‌توانیم از آنها نتیجه‌گیری معقولی داشته باشیم؟

STATISTICS AND ETHICS

- **ابطال‌پذیری و انتشار نتایج:** تمایل به انتشار فقط نتایج «مثبت» و معنادار (یعنی آنهایی که p-value پایینی دارند) می‌تواند منجر به تورش انتشار (Publication Bias) شود و تصویر نادرستی از واقعیت ارائه دهد.

- **اهمیت تکرارپذیری:** یک یافته علمی تنها زمانی قابل اعتماد است که در مطالعات مستقل مختلف تکرار شود. بحران «تکرارپذیری» در بسیاری از علوم، اهمیت طراحی دقیق آماری و گزارش شفاف را نشان می‌دهد.

- **تفاوت همبستگی و علّیت:** آمار به ما می‌گوید دو متغیر با هم ارتباط دارند، اما به خودی خود دلیل علّی این ارتباط را نشان نمی‌دهد. تشخیص این تفاوت نیازمند تفکر عمیق فلسفی و طراحی‌های تحقیقاتی قوی‌تر (مانند آزمایش‌های تصادفی‌شده کنترل‌شده) است.

- **مسئولیت اخلاقی:** سوء استفاده از ابزارهای آماری برای دستکاری ادراک عمومی (مثلاً با انتخاب گزینشی داده‌ها یا استفاده نادرست از نمودارها) یک نگرانی همیشگی است. آمار نه تنها یک ابزار فنی، بلکه یک ابزار اخلاقی است که برای کشف حقیقت باید به درستی به کار رود.

CORRELATION AND CAUSATION

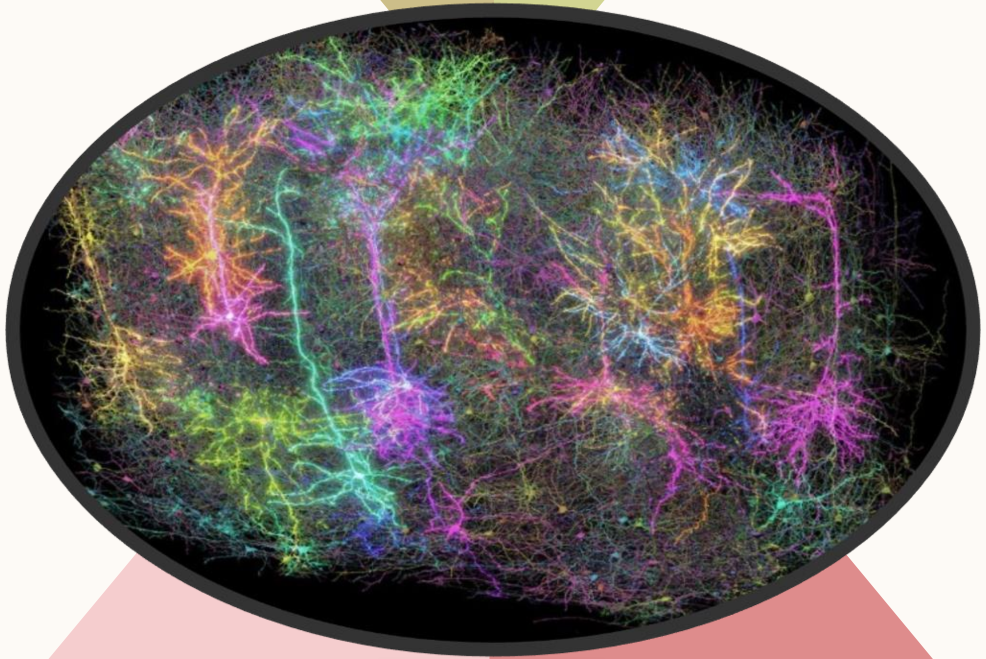
همبستگی و علّیت

Basic Concepts

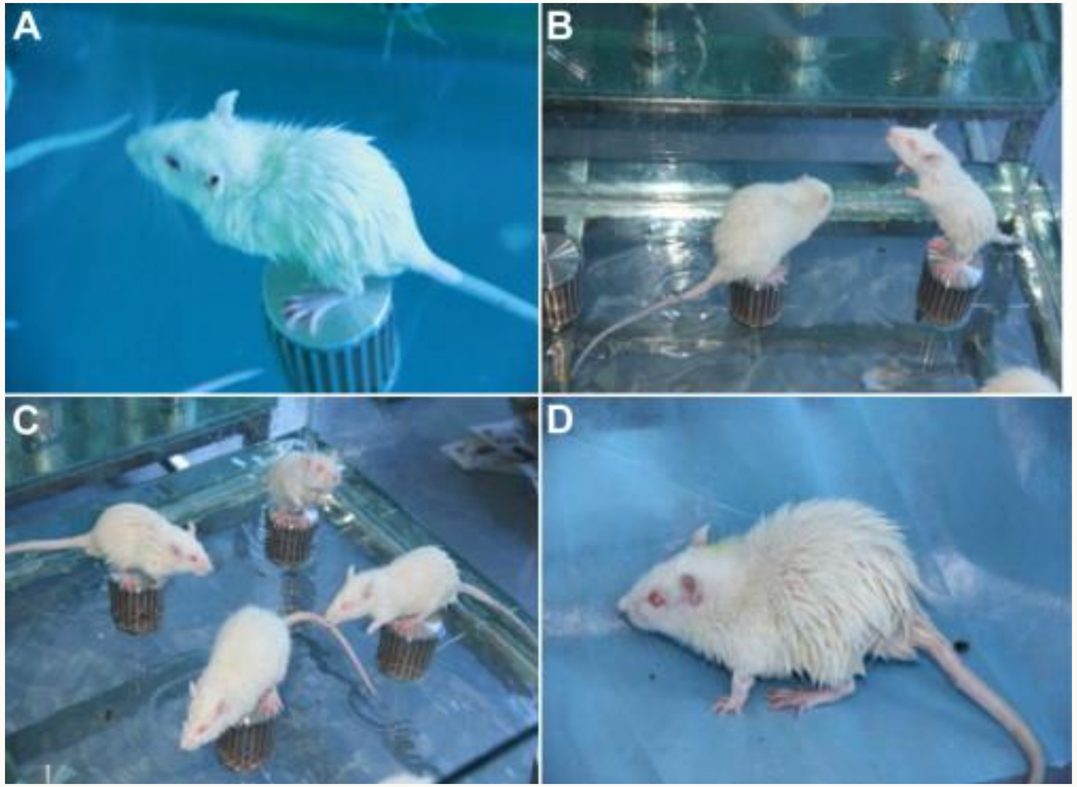
100000000000

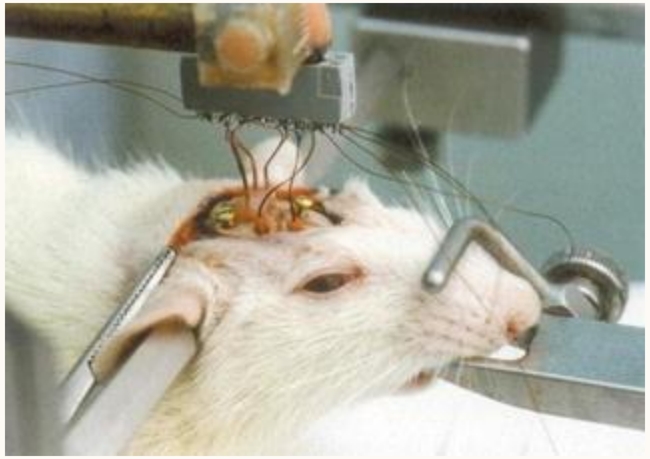
64

ϒ

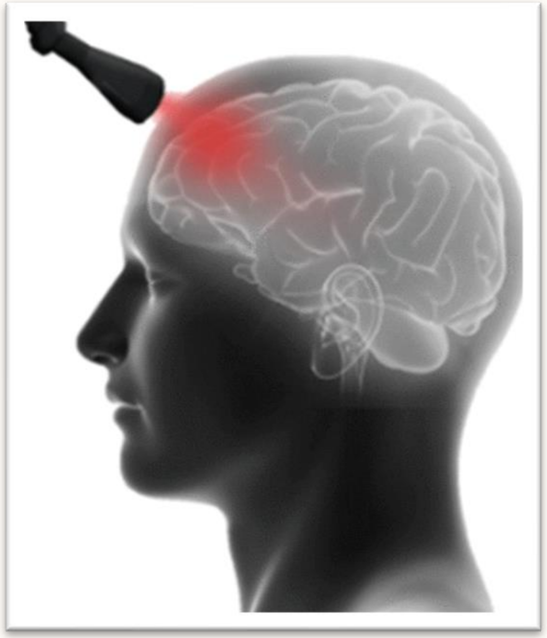


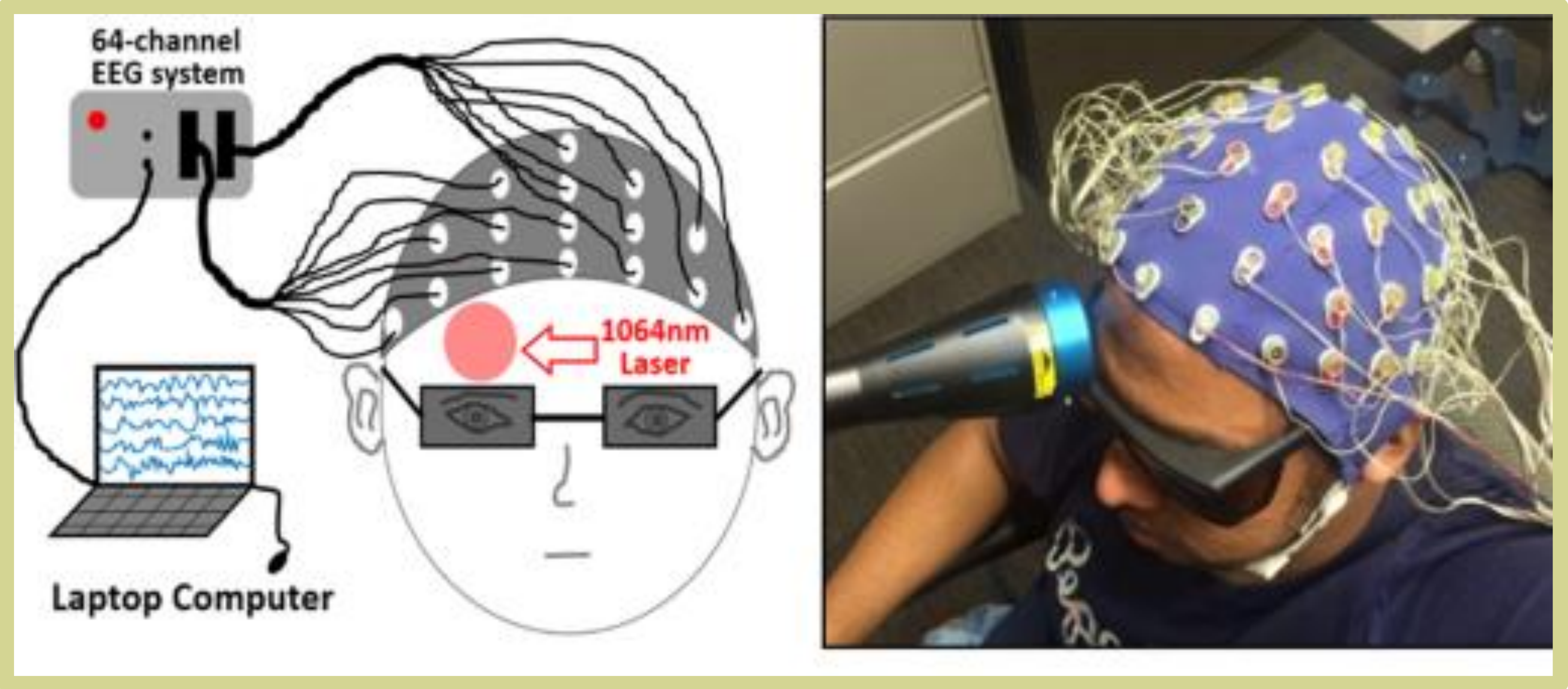
TYPES OF VARIABLES

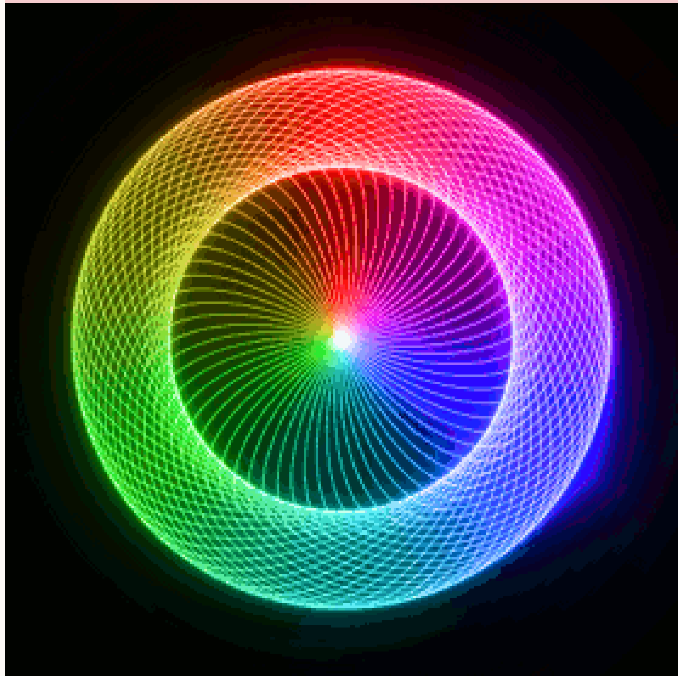




Types of variables







Observational measurement and its types

Nominal: allows assigning a name or category to the subject without any specific order between the categories.
Example: Brain Frequency
Type of electrode used (e.g., surface electrode, needle electrode, microelectrode)

Ordinal: not only classifies subjects but also establishes a meaningful order among them. However, the intervals between categories are not necessarily equal.
Example: Grading the severity of neuropathy on a scale like mild, moderate, severe

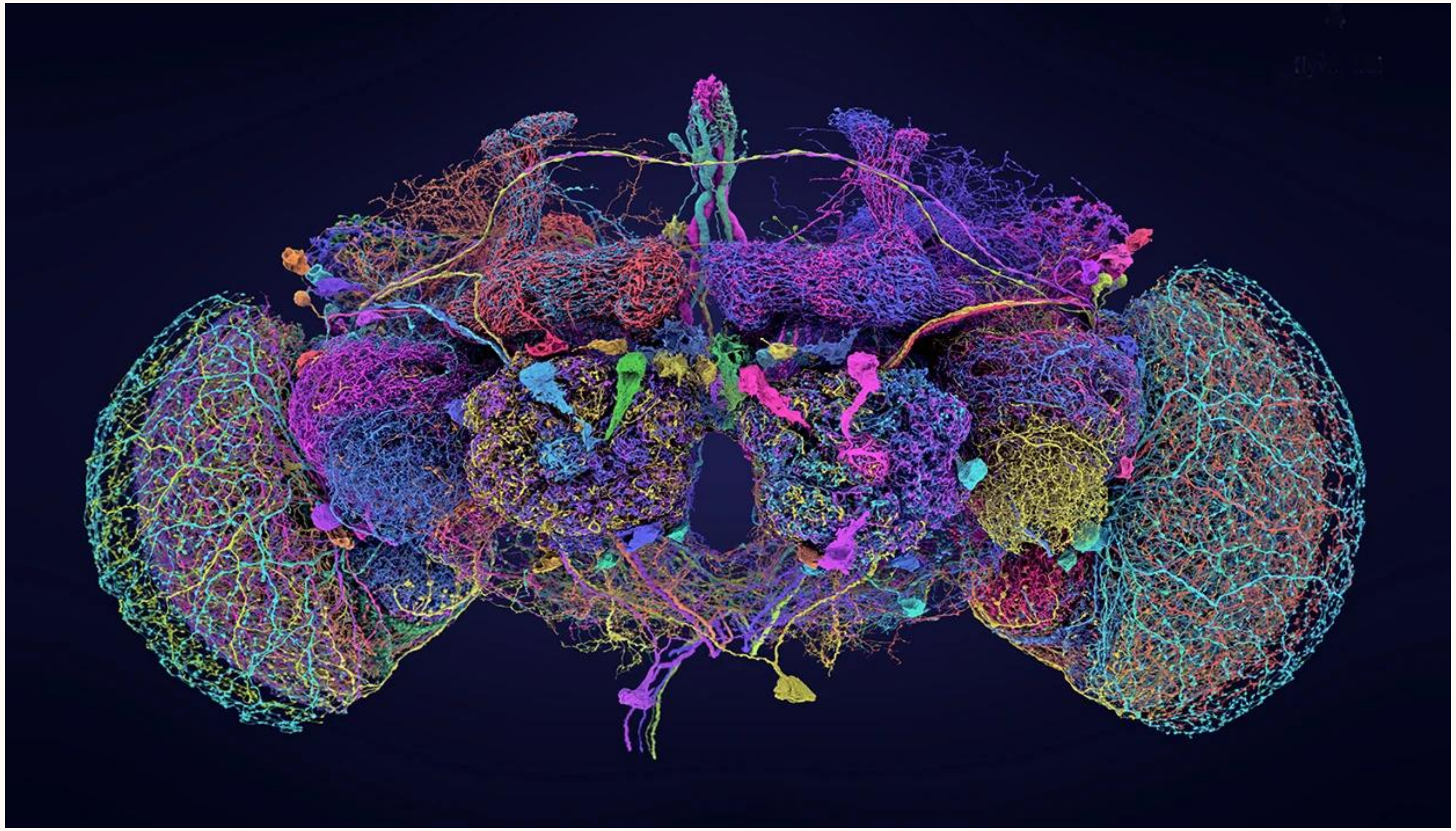
Interval: In this type of measurement, data not only have order but also equal intervals between values. However, the zero point is arbitrary and does not indicate the complete absence of the characteristic. Example: Measurement of nerve conduction latency in milliseconds (ms)

Ratio: In this type of measurement, in addition to having order and equal intervals, the zero point is absolute, indicating the complete absence of the characteristic. As a result, ratio calculations are meaningful. Example: Amplitude of evoked potentials in microvolts (µV)

Types of Variables

Main Type	Subtypes	Description	Examples
Categorical	Nominal	Categories without order	Alpha, Beta, Gamma
	Ordinal	Categories with order/rank	Brain State
	Binary	Two categories	Yes/No, Pass/Fail
Numeric	Discrete	Countable numbers, typically integers	Number of Electrodes
	Continuous	Measurable quantities allowing any value within a range	Frequency

DESCRIPTIVE VS. INFERENTIAL STATISTICS



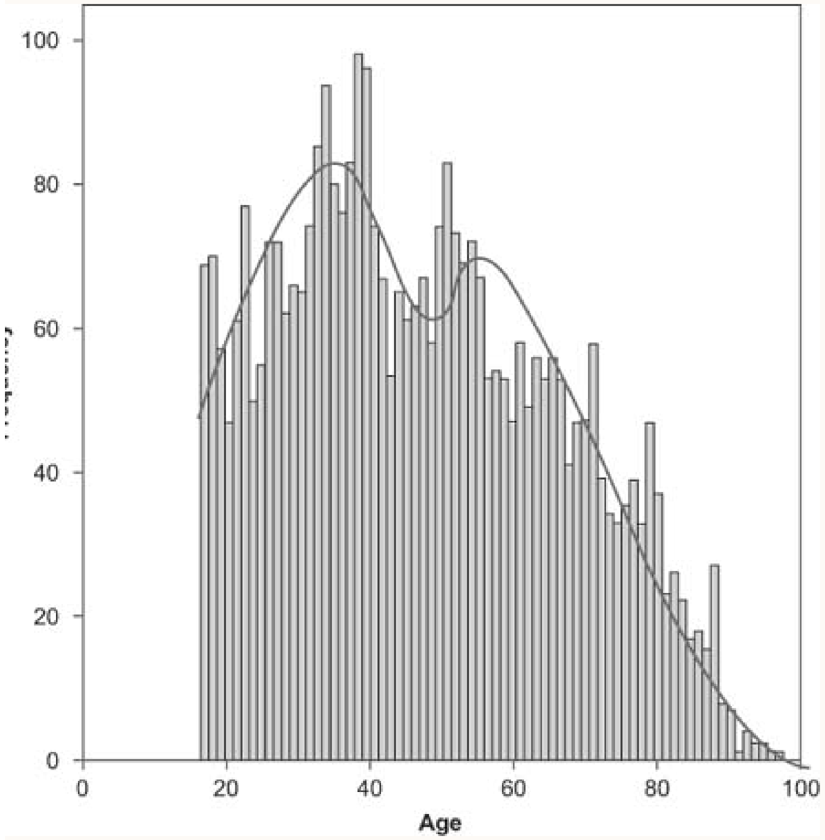
INTERDISCIPLINARY SCHOOLS

DESCRIPTIVE VS. INFERENTIAL STATISTICS

ویژگی	آمار استنباطی (Inferential)	آمار توصیفی (Descriptive)
هدف اصلی	نتیجه‌گیری، پیش‌بینی یا آزمون فرضیه درباره‌ی یک **جامعه‌ی بزرگتر** بر اساس یک **نمونه**.	خلاصه‌سازی، سازماندهی و توصیف ویژگی‌های اصلی یک **داده‌ی مشخص (نمونه یا جامعه)**.
کاری که انجام می‌دهد	استنتاج **«چه چیزی به احتمال زیاد درست است»** فراتر از داده‌های فعلی.	توصیف **«داده‌های فعلی چه شکلی هستند»**.
تمرکز	**تعمیم نتایج** از نمونه به جامعه.	توصیف **نمونه/جامعه‌ی فعلی** تحت مطالعه.
مفاهیم کلیدی	احتمال، توزیع‌های نمونه‌گیری، برآورد (فاصله اطمینان)، آزمون فرضیه (مقدار p)، آزمون t، ANOVA، کای-دو)، تحلیل همبستگی/رگرسیون برای استنتاج.	گرایش مرکزی (میانگین، میانه، نما)، پراکندگی (دامنه، واریانس، انحراف معیار)، توزیع (شکل، چولگی، کشیدگی)، فراوانی.
نمونه‌های خروجی	«با ۹۵٪ اطمینان، میانگین قد همه دانش‌آموزان مدرسه بین ۱۶۲ تا ۱۶۸ سانتیمتر است.» «تفاوت معنی‌داری در نمرات گروه الف و ب وجود دارد (p < ۰٫۰۵).» «سن پیش‌بین معنی‌داری برای درآمد است (p < ۰٫۰۱).»	«میانگین قد دانش‌آموزان این کلاس ۱۶۵ سانتی‌متر است.» «انحراف معیار نمرات آزمون ۱۰ است.» «توزیع حقوق‌ها به راست چوله است.»
با چه چیزی سروکار دارد	**عدم قطعیت**: برآورد پارامترهای جامعه و کمّی‌سازی خطا/معنی‌داری.	**قطعیت**: توصیف داده‌های واقعی جمع‌آوری‌شده.
داده‌های مورد استفاده	یک **نمونه** استخراج شده از جامعه‌ی بزرگتر.	**تمام داده‌ها** (یا زیرگروه مشخص) که تحلیل می‌شوند.
هدف نهایی	**استنتاج**: فراتر رفتن از داده‌های نمونه برای قضاوت درباره‌ی جامعه‌ی هدف.	**ارائه**: ساده‌سازی داده‌های بزرگ به صورت خلاصه‌های قابل‌درک (جداول، نمودارها، اعداد کلیدی).
قیاس	استفاده از یک تکه پارچه برای استنتاج الگو/کیفیت کل پارچه.	گرفتن عکس از یک گروه خاص – آن گروه را در آن لحظه ثبت می‌کند.

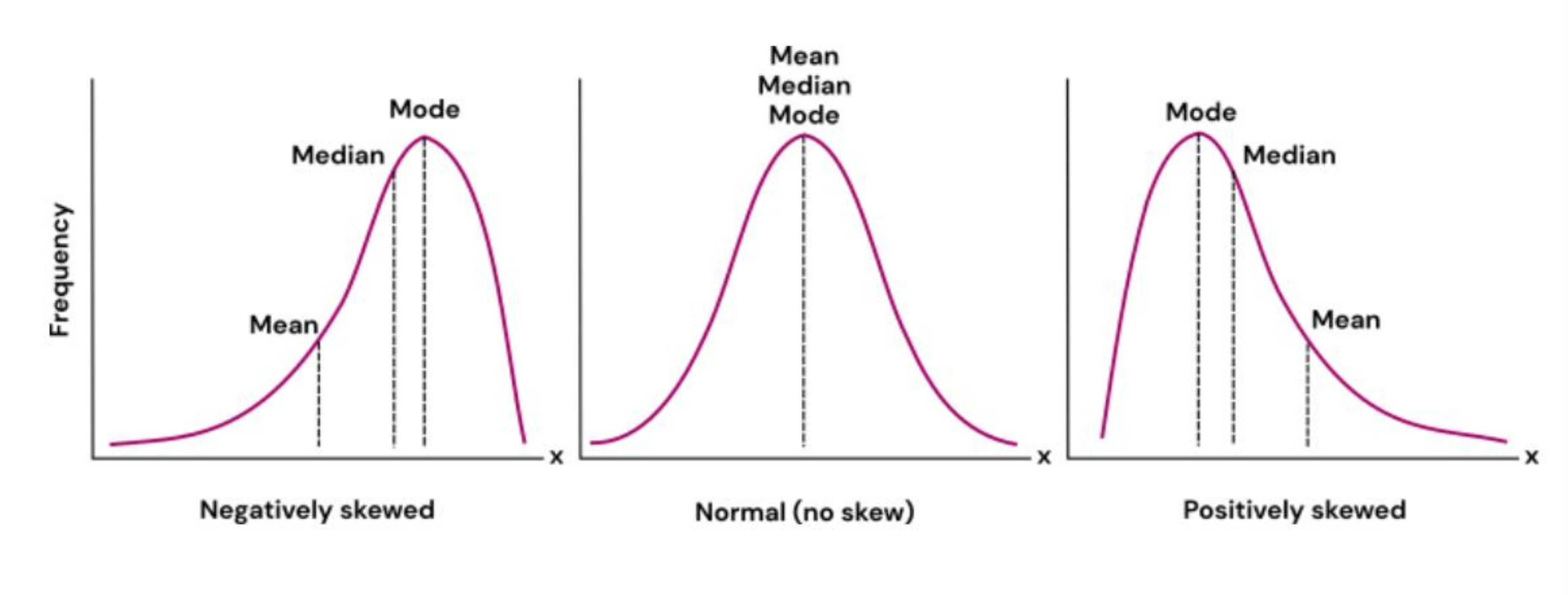
MEASURES OF VALUES TO IMAGINE DATASET:

Types of Descriptive Statistics		
Measure	**Types**	**Application**
Central tendency	Mean, mode, median.	Used to demonstrate distribution of values.
Frequency	Frequency ratio, rate, and proportion.	Utilized to show how often a value occurs.
Measures of dispersion	Range, variance, standard deviation.	Applied to identify the spread of values.
Position	Percentiles, quartiles, and z-scores.	To highlight where the values fall in relation to one another.



MEASURES OF CENTRAL VALUES:

1. Mean: For the dataset {2, 4, 6, 8, 10}, the mean is (2 + 4 + 6 + 8 + 10) / 5 = 6.
2. Median: For the dataset {3, 5, 7, 9, 11}, the median is 7 (the middle value).
3. Mode: For the dataset {1, 2, 2, 3, 4}, the mode is 2 (since it occurs most often).



MEASURES OF DISPERSION

Quantifying how spread out data points are in a dataset.

1. Range

Definition: The difference between the maximum and minimum values.

Formula: Range = Max - Min

Example: For {3, 7, 2, 10}:

Max = 10, Min = 2

Range = 10 - 2 = 8

2. Variance

Variance measures the average squared deviation of each data point from the mean.

$$\text{Variance}(\sigma^2) = \frac{\sum(x_i - \bar{x})^2}{n}$$

$$\text{Variance for population} = \frac{\sum(x_i - \bar{x})^2}{n}$$

$$\text{Variance for sample} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}$$

3. Standard Deviation (SD)

Definition: The standard deviation is the square root of variance and shows the dispersion of the dataset.

$$\text{Standard Deviation}(\sigma) = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

4. Coefficient of Variation

Definition: To standardize Measures of Dispersion

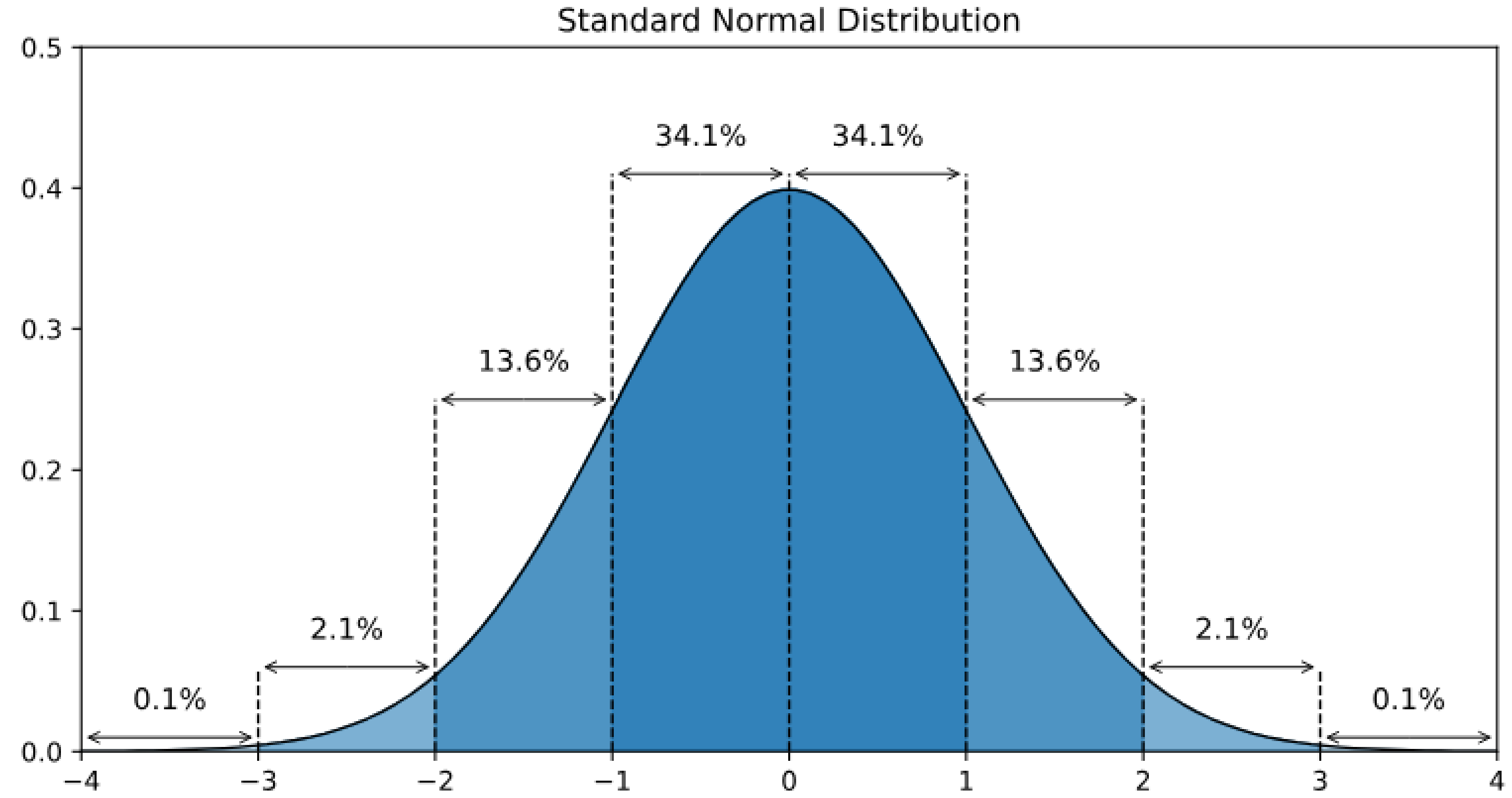
$$CV = \frac{SD}{\mu}$$

5. Standard Error (SE)

The standard error indicates the accuracy of the sample mean.

$$\text{Standard Error(SE)} = \frac{SD}{\sqrt{n}}$$

In probability theory and statistics, a normal distribution or Gaussian distribution is a type of continuous probability distribution for a real-valued random variable.



6. Quartile (Q)

Divides data into four equal parts:

- Q1 (First quartile - P25): 25% of data falls below this value.
- Q2 (Second quartile - P50): 50% of data falls below (also known as the Median).
- Q3 (Third quartile - P75): 75% of data falls below this value.

Interquartile Range (IQR):

- The difference between the third and first quartile: IQR = Q3 - Q1
- Represents the spread of the middle 50% of data and is resistant to outliers.

Numerical Example:

Given the dataset [10, 15, 20, 25, 30, 35, 40, 45, 50]:

- Q1 = 17.5, Q2 =30, Q3 = 42.5
- IQR = Q3 - Q1 = 42.5 – 17.5 = 25

7. Confidence Interval (CI):

a range of values that, with a specified confidence level, is likely to contain the true population parameter (such as the mean or proportion).

$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

CI = confidence interval

$\bar{x}$ = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

Confidence level	Critical (z) value to be used in confidence interval calculation
50%	0.67449
75%	1.15035
90%	1.64485
95%	1.95996
97%	2.17009
99%	2.57583
99.9%	3.29053

Q

Suppose the average test score in a sample of 30 students is 75, with a standard deviation of 10. Calculate the 95% confidence interval for the true population mean.

Suppose the average test score in a sample of 30 students is 75, with a standard deviation of 10. Calculate the 95% confidence interval for the true population mean:

Suppose the average test score in a sample of **30 students** is **75**, with a standard deviation of **10**. Calculate the **95% confidence interval** for the true population mean:

$$CI = 75 \pm (1.96 \times \frac{10}{\sqrt{30}})$$

$$CI = 75 \pm (1.96 \times 1.83)$$

$$CI = 75 \pm 3.58$$

Thus, the **95% confidence interval** is:

$$(71.42, 78.58)$$

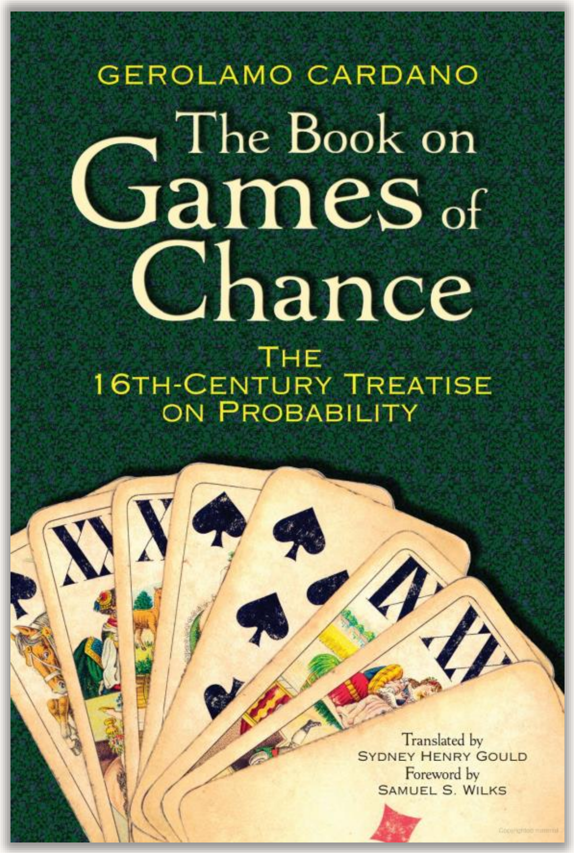
This means we are **95% confident** that the true population mean lies within this range.

PROBABILITY

It is one of the oldest statistical concepts: the concept of probability was a tool used to solve problems related to games of chance, employed as far back as the 14th century.

GIROLAMO CARDANO

(1501–1576)



THE FIVE APPROACHES TO PROBABILITY

1. Classical (Games of Chance)

Definition: Probability is defined as the ratio of the number of favorable outcomes to the total number of possible equally likely outcomes.

Formula: P(A) = (Number of favorable outcomes for A) / (Total number of possible outcomes)

Example: The probability of rolling a 4 on a fair six-sided die is 1/6.

Requirement: This approach only works when all outcomes are symmetrical and equally likely (like dice, coins, and fair decks of cards).

2. Empirical (Relative Frequency)

Definition: Probability is defined as the long-term relative frequency of an event occurring after many repetitions or trials.

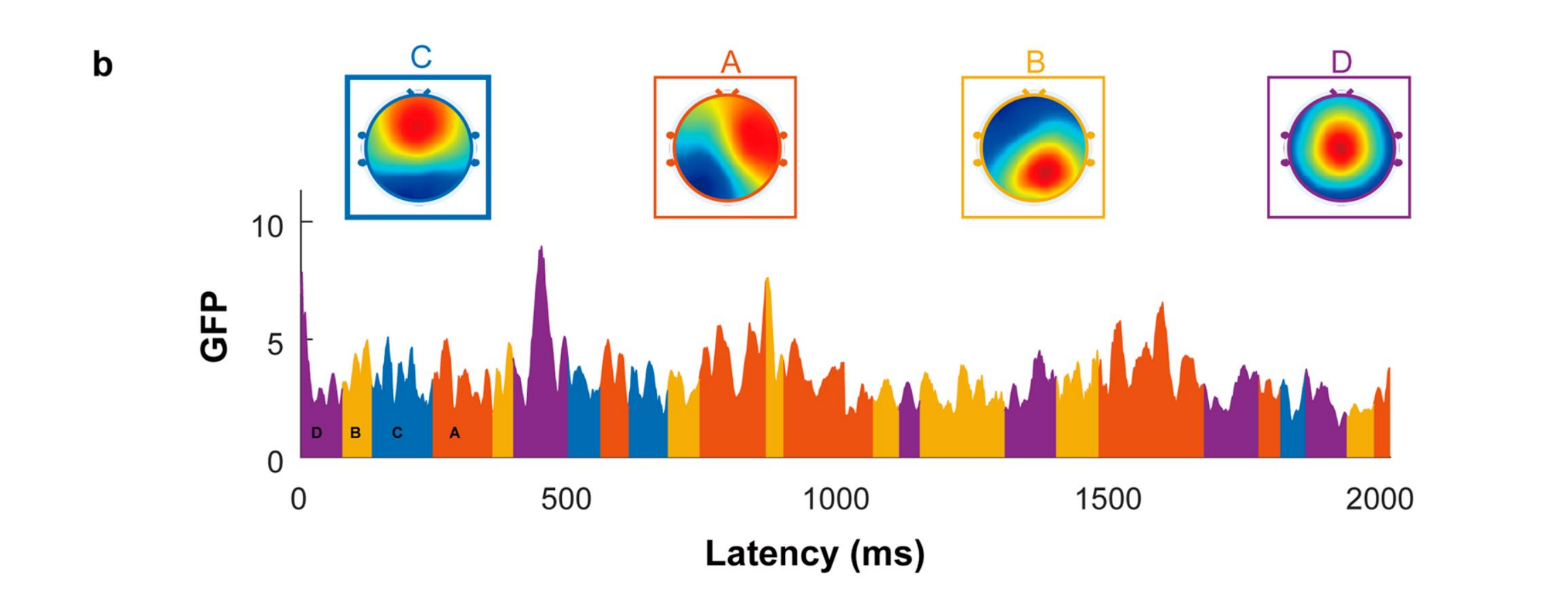
Formula: P(A) = (Number of times event A occurs) / (Total number of trials)

Example: If a factory produces 1,000 light bulbs and 15 are defective, the empirical probability of a bulb being defective is 15/1000 = 0.015.

Key Idea: It is based on observed data rather than theoretical symmetry.

THE FIVE APPROACHES TO PROBABILITY

2. Empirical (Relative Frequency)



THE FIVE APPROACHES TO PROBABILITY

3. Subjective

Definition: Probability represents a personal degree of belief or confidence that a particular event will happen. It is not based on precise calculations or long-term frequencies but on individual judgment, experience, and intuition.

Example: A doctor might say, "Based on my expertise, I think there is a 90% chance this patient's illness is viral." A sports fan might say, "My team has a 70% chance of winning the championship this year."

4. Axiomatic (Mathematical)

Definition: This is a rigorous mathematical approach that defines probability based on a set of axioms (rules) from which all other probability rules can be derived. It was formalized by the mathematician Andrey Kolmogorov.

Key Idea: It doesn't concern itself with how probabilities are assigned or interpreted; it only provides the mathematical framework that all probability must follow (e.g., probabilities must be between 0 and 1, and the probability of a certain event is 1).

THE FIVE APPROACHES TO PROBABILITY

5. Bayesian

Definition: This approach interprets probability as a degree of belief. Its core idea is that beliefs (probabilities) can be updated as new evidence is acquired, using a mathematical formula called Bayes' Theorem.

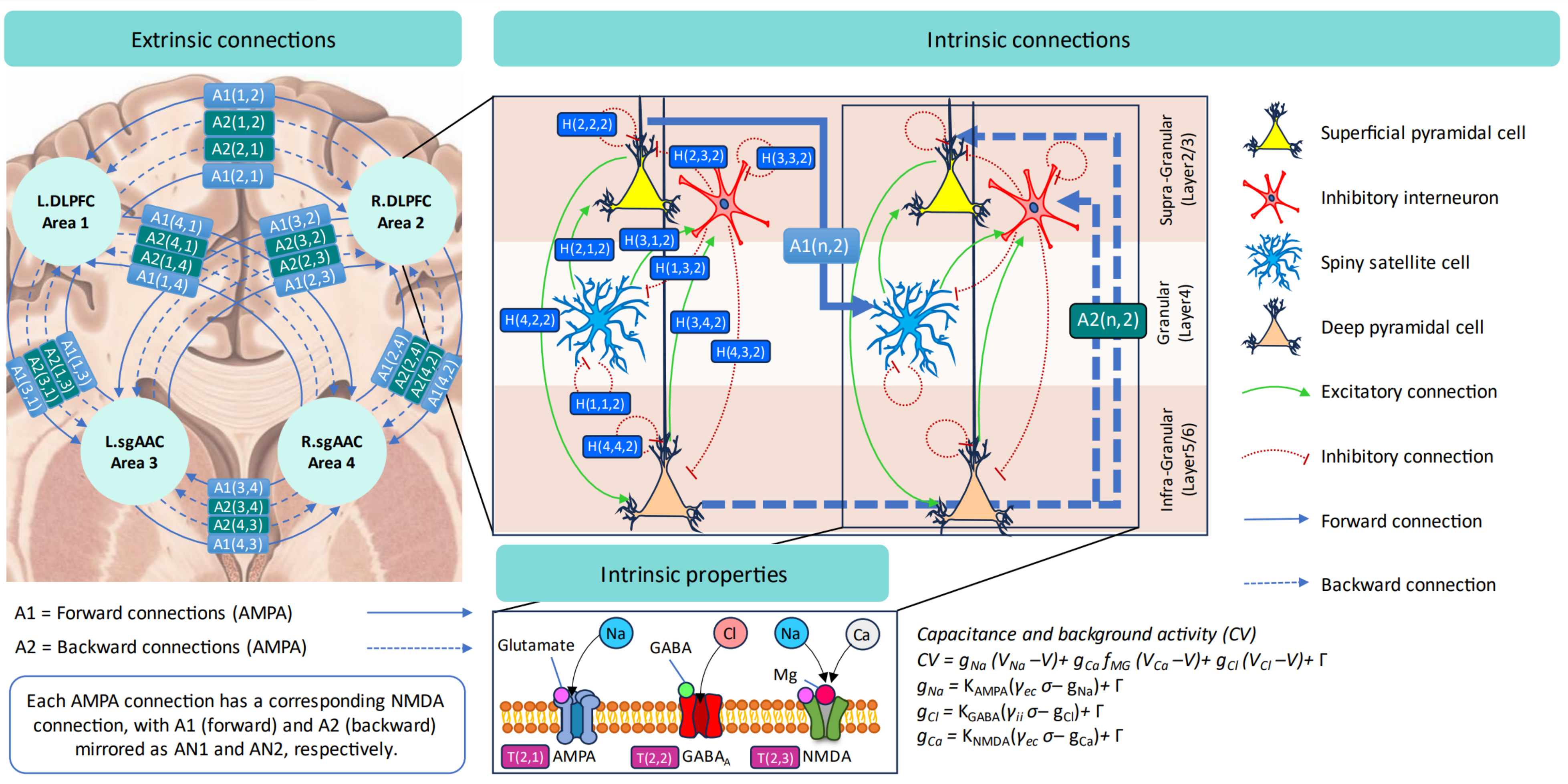
Formula: $P(A|B) = [P(B|A) * P(A)] / P(B)$

Key Idea: You start with a prior probability (an initial belief). Then you collect data and use Bayes' Theorem to update that belief into a posterior probability.

Example: You might have a prior belief about how effective a drug is (e.g., a 20% chance it works). After conducting a clinical trial (collecting data), you use Bayes' Theorem to update that probability to, say, an 80% chance it works (the posterior probability).

THE FIVE APPROACHES TO PROBABILITY

5. Bayesian



PROBABILITY DISTRIBUTIONS

Binomial: Models the number of successes in a fixed number of independent trials (e.g., number of heads in 10 coin flips).

Poisson: Models the number of events occurring in a fixed interval of time or space (e.g., number of customers arriving at a store in an hour).

Normal (Gaussian): The famous "bell curve." Models many natural phenomena and is fundamental in statistics due to the Central Limit Theorem.

Chi-square: Used primarily for hypothesis tests, such as testing for independence in categorical data or goodness-of-fit.

t-distribution: Similar to the normal distribution but with thicker tails. Used for estimating population means when the sample size is small.

F-distribution: Used primarily for comparing two variances (e.g., in ANOVA or regression analysis).