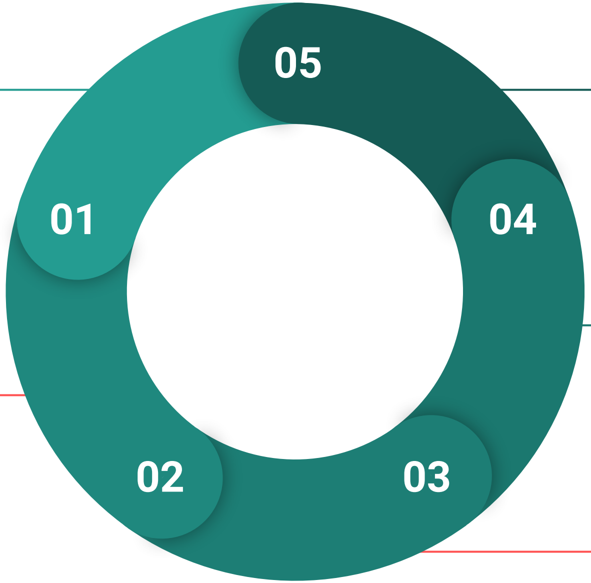
Statistical Underpinnings

BIPN 162

The life cycle of neural data

Experimental Design
(Often out of the hands of neural data scientists!)

Collecting measurements & preprocessing
Data wrangling, source separation, image filtering, spike sorting, dimensionality reduction



Hypothesis testing & deriving scientific conclusions
Bootstrap, permutation, multiple comparisons, interpretation

Model building, optimization, and parameter estimation
Dimensionality reduction, neural coding, decoding

Exploratory data analysis
Dimensionality reduction, data management, visualization

What we've covered so far / are currently covering

Neuroscience relies on probability & statistics

Probabilistic models can be built to explain

- Single neurons
- Neural networks
- Animal behavior
- Group behavior

$$\mathcal{D} = \{X_1, X_2, X_3, \ldots\}$$

(a sample of random variables)

Experimental $\boxed{\text{data}}$ is analyzed using statistics

Models are evaluated using statistics

Nervous systems are $\boxed{\text{stochastic}}$

Randomly determined; having a random probability distribution or pattern that may be analyzed statistically but *may not be predicted precisely.*
(Oxford Dictionary)

Because of this, we need to take into account **uncertainty**.

In a stochastic environment, we use a **probabilistic framework**.

Put into math, we can say there *isn't* a perfect mapping between our observed data (x) and a world property (Θ): $\Theta_1 \rightarrow x_1$

... we need **inference**!

Inference: using mathematical models to make general claims from particular data

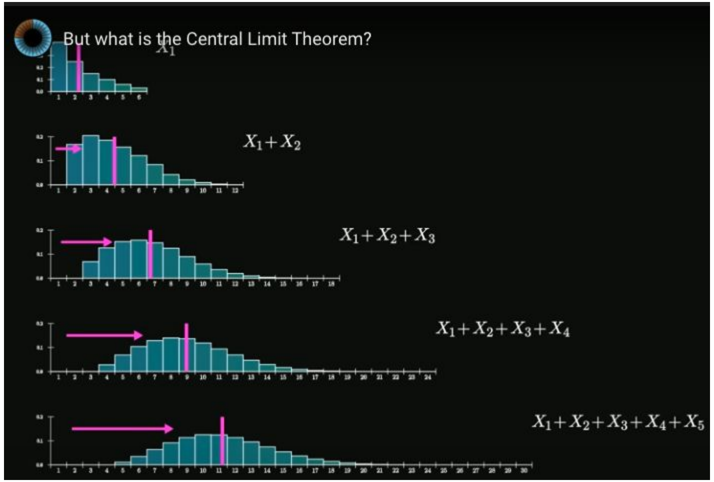
Three challenges of statistical inference:

1. Generalizing **from sample to population**, a problem that is associated with survey sampling but actually arises in nearly every application of statistical inference
2. Generalizing **from treatment to control group**, a problem that is associated with causal inference, which is implicitly or explicitly part of the interpretation of most regressions we have seen
3. Generalizing **from observed measurements to the underlying constructs of interest**, as most of the time our data do not record exactly what we would ideally like to study

These are problems of prediction

Adapted from Regression and Other Stories

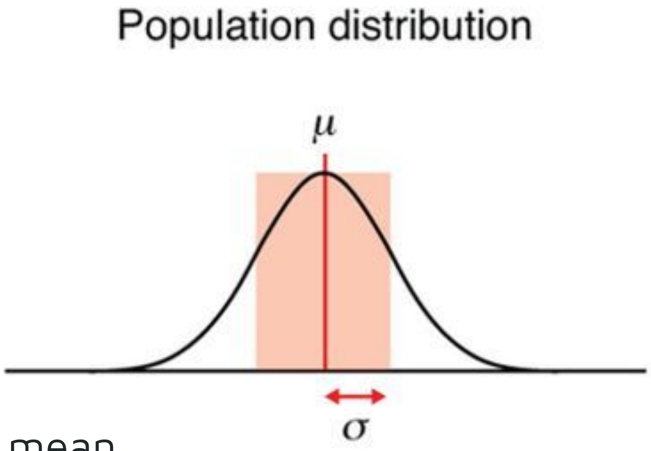
Take Home Viewing: Central Limit Theorem

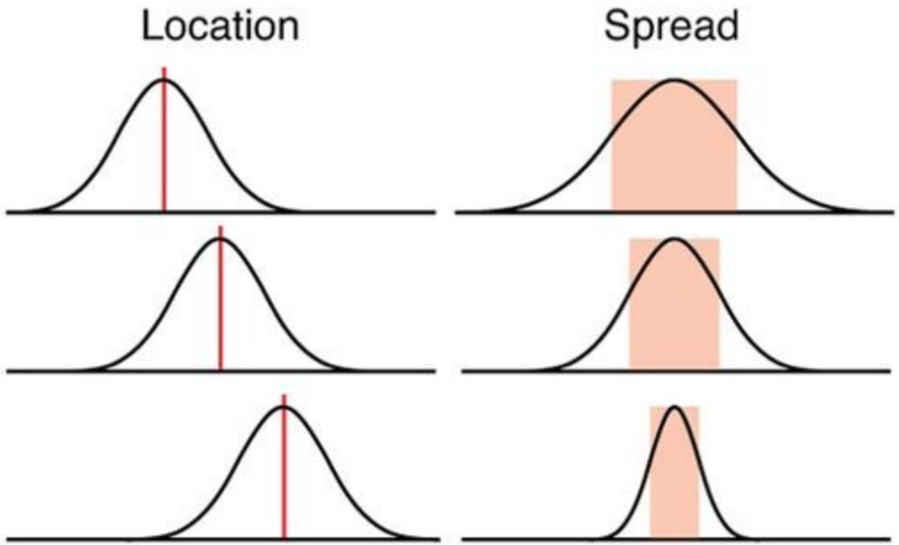


- Define the **mean**, **variance**, and **standard deviation**
- Use the central limit theorem to predict the distribution of means from a given sample
- Identify definitive features of the **standard normal distribution**

Populations differ by location & spread

By plotting frequency histograms for a dataset,
we can see how they differ in their **location** (i.e. central tendency) & **spread** (i.e. distribution, dispersion)





μ = mean
σ = standard deviation

Put in mathematical notation…

Descriptive statistics summarize some property of your collection of points

sample mean $$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{k=1}^{n} X_k$$

sample variance $$s_X^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

See also The Map of Mathematics & Math Notation Cheat Sheet

The **central limit theorem**: the distribution of sample means will become increasingly close to a normal distribution as the sample size increases, regardless of the shape of the population distribution

Risk with small samples:
You could still get a rogue mean for your sample

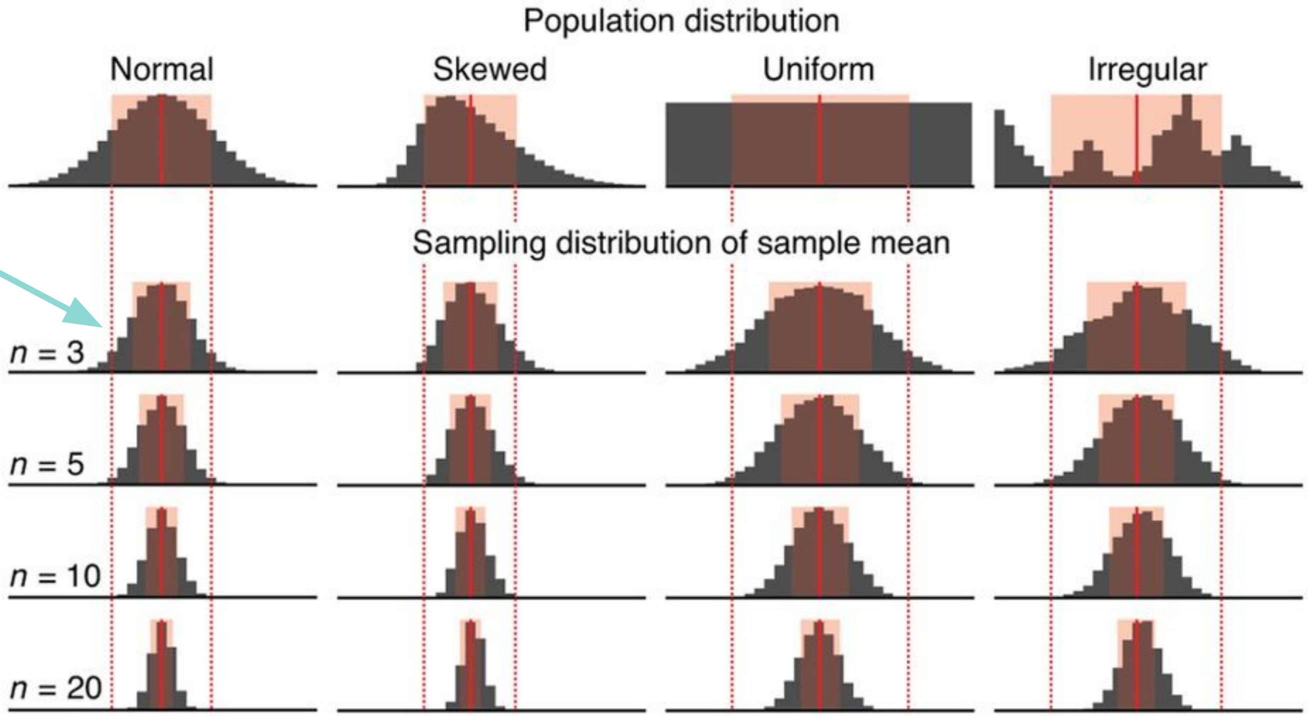
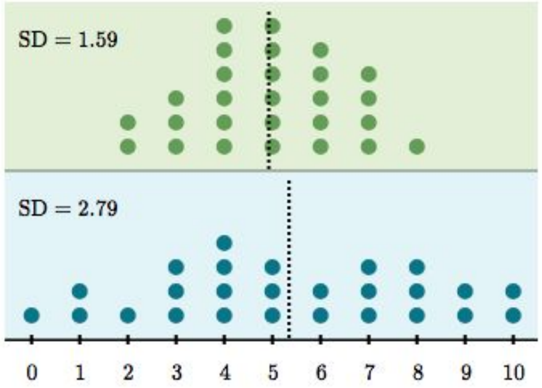


Image: https://www.nature.com/articles/nmeth.2613; watch this live! https://www.esci.thenewstatistics.com/esci-dances.html

Describing the distribution of a dataset

- **Range**: lowest & highest values
- **Variance (σ^2)**: mean of the squared difference from the mean (**M**); measures variability from mean but doesn't match the units of the data
- **Standard deviation (S.D., σ) = $\sqrt{\sigma^2}$**
  - A measure of how far a typical data point lies from the mean ; reflects the spread of the population
- **Standard Error of the Mean (S.E.M.) = SD/$\sqrt{N}$**
  - N = # data points
  - Reflects the uncertainty in the mean & its dependency on the sample size (measure of sampling variability)
- **Confidence intervals**: Interval inside which the true value is likely to lie (https://rpsychologist.com/d3/CI/)

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum (X_i - M)^2}{N-1}}$$



By the end of this lecture you’ll be able to:

- Explain the difference between:
  - **univariate** and **multivariate** distributions
  - **discrete** and **continuous** distributions
- Identify multiple ways to generate and describe distributions
- Implement multiple measures to describe:
  - the **correlation** between two continuous variables
  - tthe difference between two distributions

Types of data

Univariate data: observations about **one** attribute

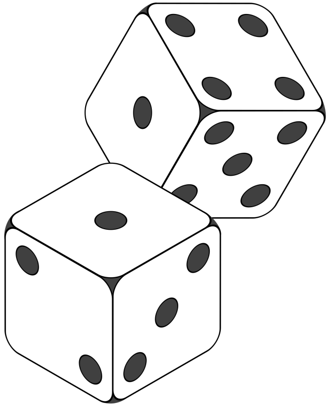
Bivariate data: observations about **two** attributes

Multivariate data: observations about **multiple** attributes

Most common in data science!

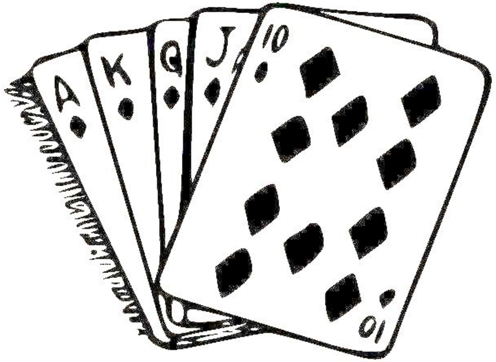
Kinds of (univariate) probability distributions

Dependence and independence



Each time you roll a dice, that is an **independent event**.

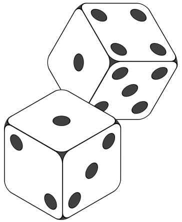
In other words, what happened previously *does not* impact future events.

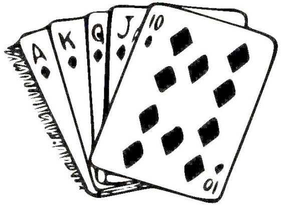


Each time you draw from a deck of cards, you change the likelihood of drawing a particular card on the next draw (assuming non-replacement).

What happened previously *does* impact future events.

Dependence and independence





Mathematically, two events (here, E & F) are independent if the probability that they *both* happen is the product of the probabilities that each of them happen:

$$P(E,F) = P(E) * P(F)$$

(this is a **generative** model)

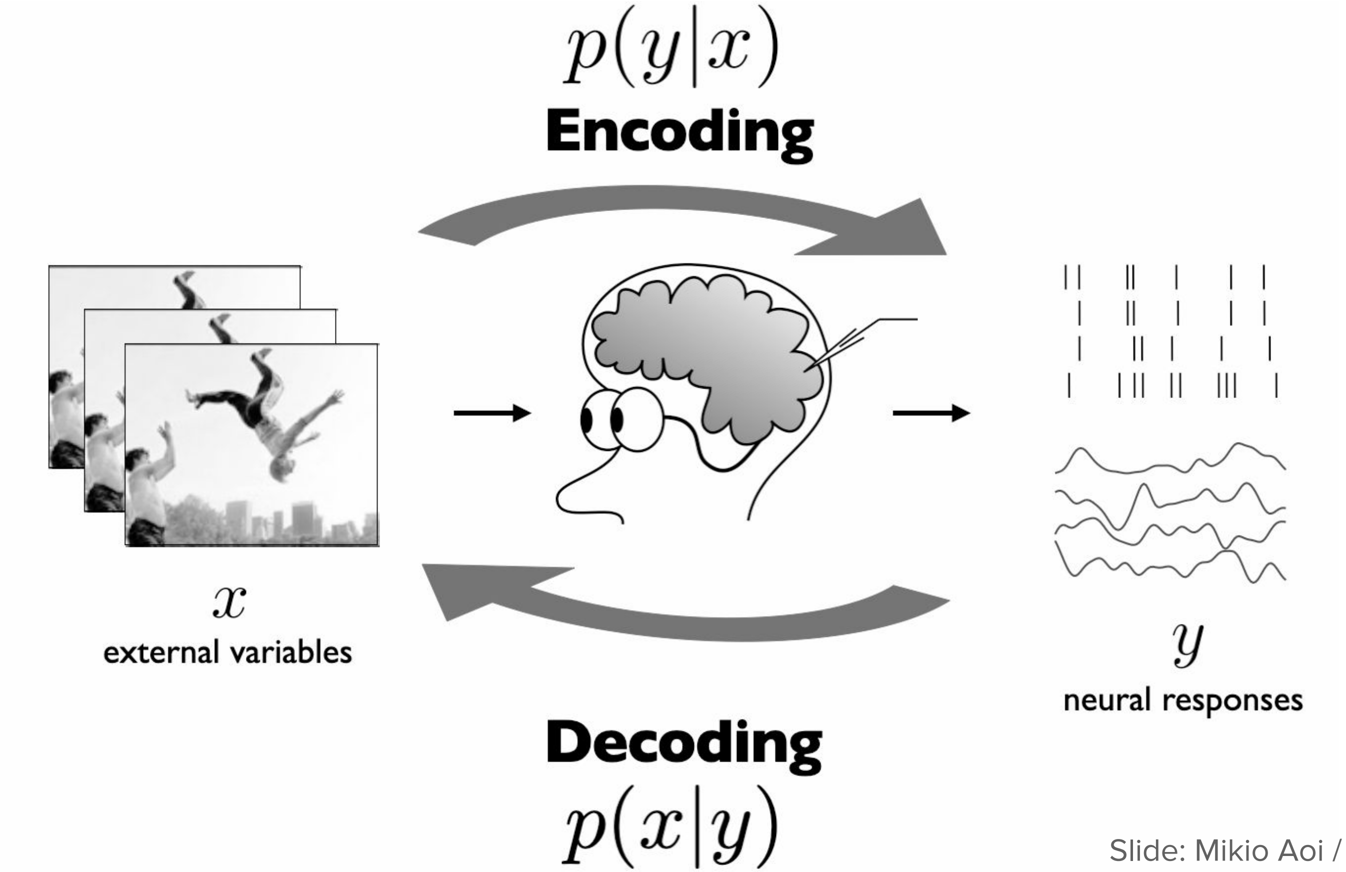
If they are *not* necessarily independent, we define the probability of E "conditional on F" as:

$$P(E|F) = P(E,F) / P(F)$$

or

$$P(E,F) = P(E|F) * P(F)$$

In terms of neural coding...



Slide: Mikio Aoi / Marcus Benna

Discrete distributions:

Binomial distribution

- Modeled by a Bernoulli distribution with two outcomes, left & right.
- ***P(Left) = p ; P(Right) = 1 - p***
- ***p*** is the probability of turning left, **n** is the number of binary events, or trials, and ***k*** is the number of times the rat turned left.

$$P(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

binomial coefficient
(represents all of the possible combinations of k out of n; "n choose k")

See this video: [Binomial distributions | Probabilities of probabilities, part 1]()

Discrete distributions:

Multinomial distributions

- Modeled by a **categorical distribution**
- Generalization of binomial distribution
- $\mathbf{n}$ possible outcomes, each with their own probability

See https://en.wikipedia.org/wiki/Multinomial_distribution

Discrete distributions:

Poisson distribution

- Gives the probability of an event happening a certain number of times (k) within a given interval of time or space
- Used for count data (e.g., # of cases of cancer; number of spikes)
- Every event is *independent* of the previous one (and should happen at a constant average rate)
- $\boldsymbol{\lambda}$ is a parameter corresponding to the average outcome of $\mathbf{x}$ (sometimes $\mathbf{k}$)

  (In neural modeling, $\boldsymbol{\lambda}$ is often the overall spike rate)

Example: If you know 750 people, 1% of all people in the population are named Chris, and you are as likely to know Chris' as anyone else, the number of Chris' you know could be modeled as a Poisson distribution with expectation ($\boldsymbol{\lambda}$) = 7.5.

To find the probability of knowing exactly 5 people named Chris out of 750, you would plug in $\boldsymbol{\lambda}$ = 7.5 and $\mathbf{x}$ = 5 into the formula:

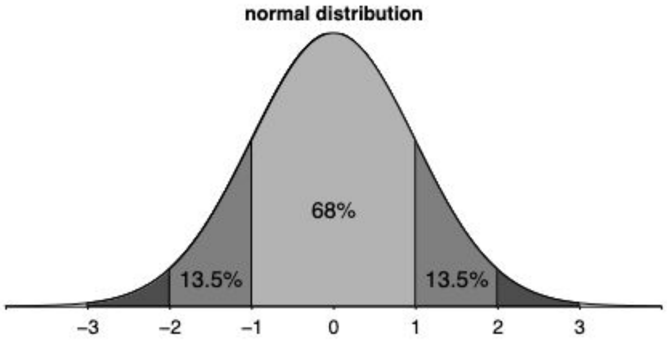
$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Continuous distributions

- Uniform distribution
- Gaussian (Normal) distribution
  - Because of the central limit theorem, many quantities are Gaussian distributed!

$$X_i \sim \mathcal{N}(\mu, \sigma^2)$$

While for discrete outcomes, we can ask about the probability of a specific event ("What is the probability this neuron will fire 4 times in the next second"), this is not defined for a continuous distribution ("What is the probability of the BOLD signal being exactly 4.000120141…"). Hence we need to focus on **intervals** when calculating probabilities from a continuous distribution.



Approximately 50% of the mass of the normal distribution falls within 0.67 STDEV from the mean, 68% of the mass falls within 1 STDEV from the mean, 95% within 2 STDEV of the mean, and 99.7% within 3 STDEV

Content adapted from NMA

How should we model the probability of making a 3-point basket?

I. Binomial distribution
II. Poisson distribution
III. Normal distribution
IV. Uniform distribution

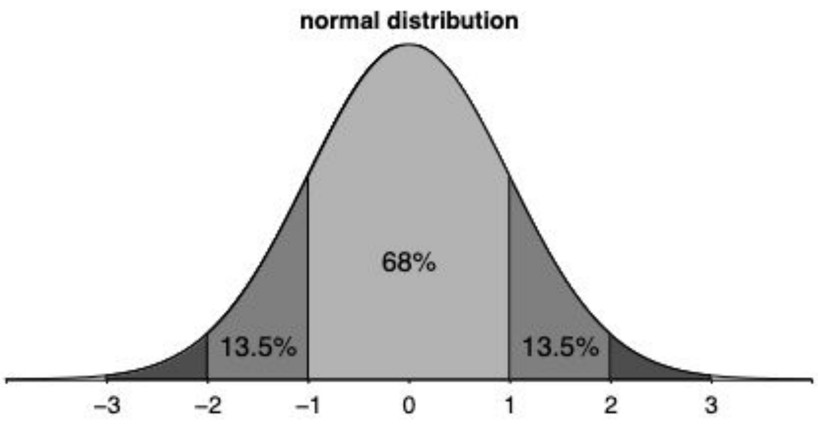
Image: NYT

Basketball players and fans commonly believe that players tend to shoot in streaks and that the chances of hitting a shot are greater following a hit than following a miss. Analyses of both professional and college basketball reveal that, contrary to common belief, the outcomes of successive shots are largely independent. The "detection" of streaks in random sequences is traced to an erroneous intuition that the law of large numbers applies to small samples as well.

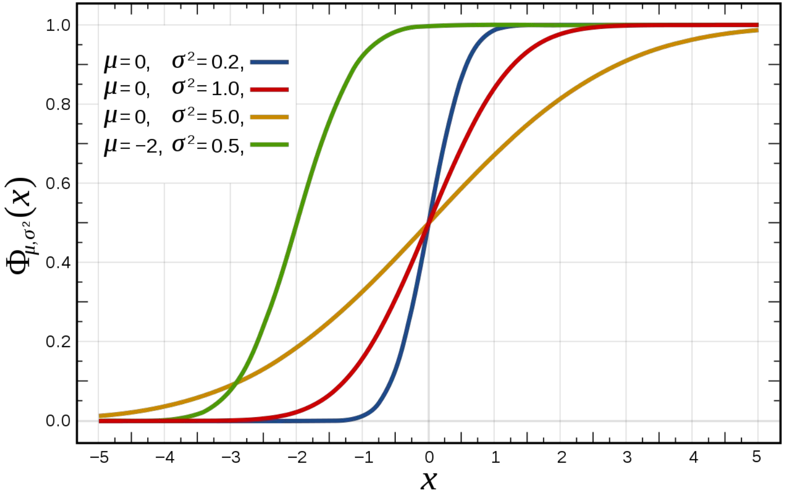
Binomial shots explanation: [Binomial probability example (video) | Khan Academy]
Quote from Tversky et al., [Misperceptions of chance processes in basketball]
See also [Probability of Independent Events with Steph Curry | by Hunter Carver]

Probability distributions tell us about the likelihood that a value falls between two values

Probability density functions (PDFs)



Cumulative density functions (CDFs)



Transformations of data

- **Linear transformations** do not change shape of distribution
  - Linearly transformed normal distributions are still normal
  - Converting inches to centimeters, for example.
  - **Standardization** using a z-score

$$Z = \frac{x - \mu}{\sigma}$$

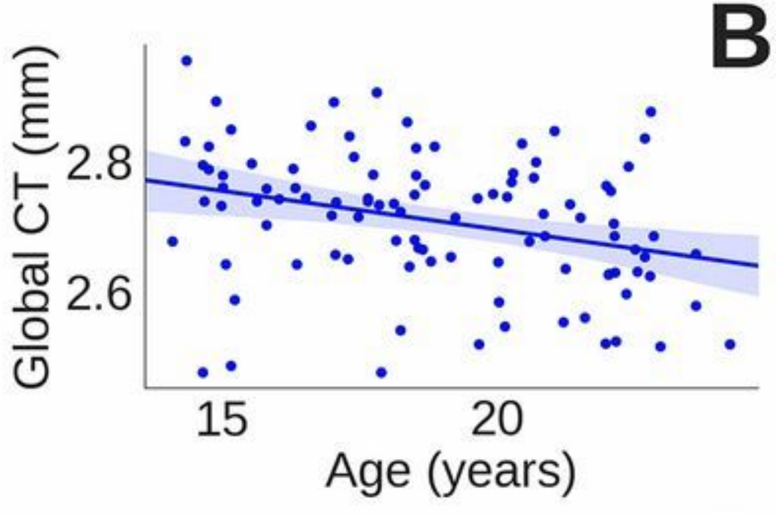
- **Nonlinear** transformations do change the shape (often intentionally)
  - E.g., log transforms

By the end of this lecture you'll be able to:

- Explain the difference between:
  - **univariate** and **multivariate** distributions
  - **discrete** and **continuous** distributions
- Identify multiple ways to generate and describe distributions
- **Implement multiple measures to describe**:
  - the **correlation** between two continuous variables
  - the difference between two distributions

Inferential statistics
generalize from observed data
to the world at large

With **bivariate** data, we are often interested in whether or not there is a relationship between two variables



Describing the relationship between two variables

In the simplest relationship between two variables, they're **proportional**.

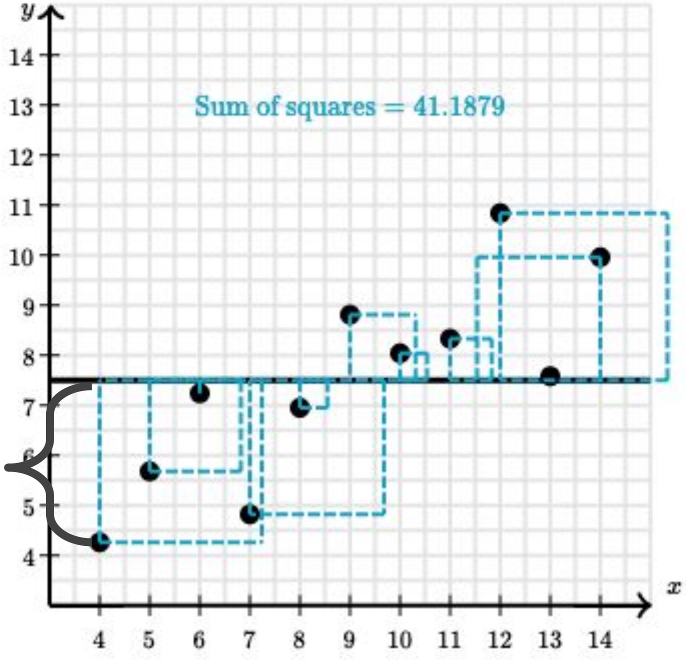
Linear equation $y = \beta x$

Linear combination $y = x_1\beta_1 + x_2\beta_2$

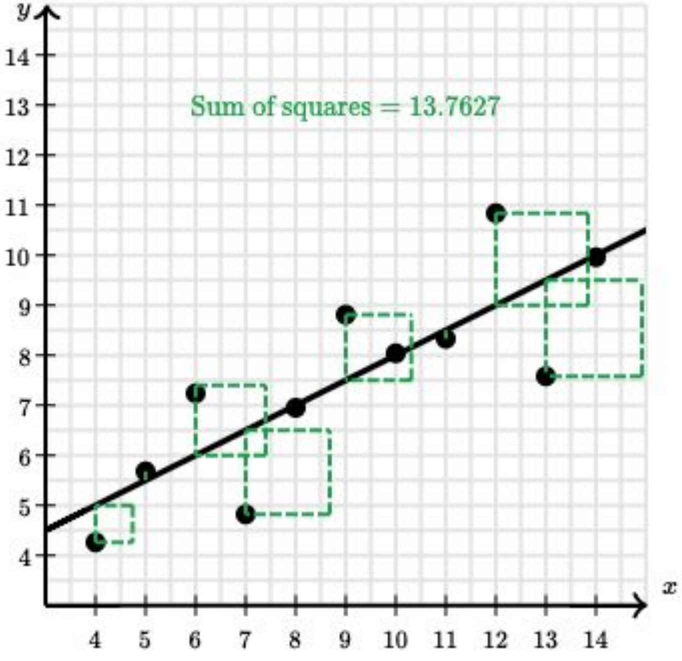
An extension of this is that they're **linearly related**.

R^2 measures how much prediction error we eliminated

residual : distance from our measured point to the trendline



trendline is simply the mean

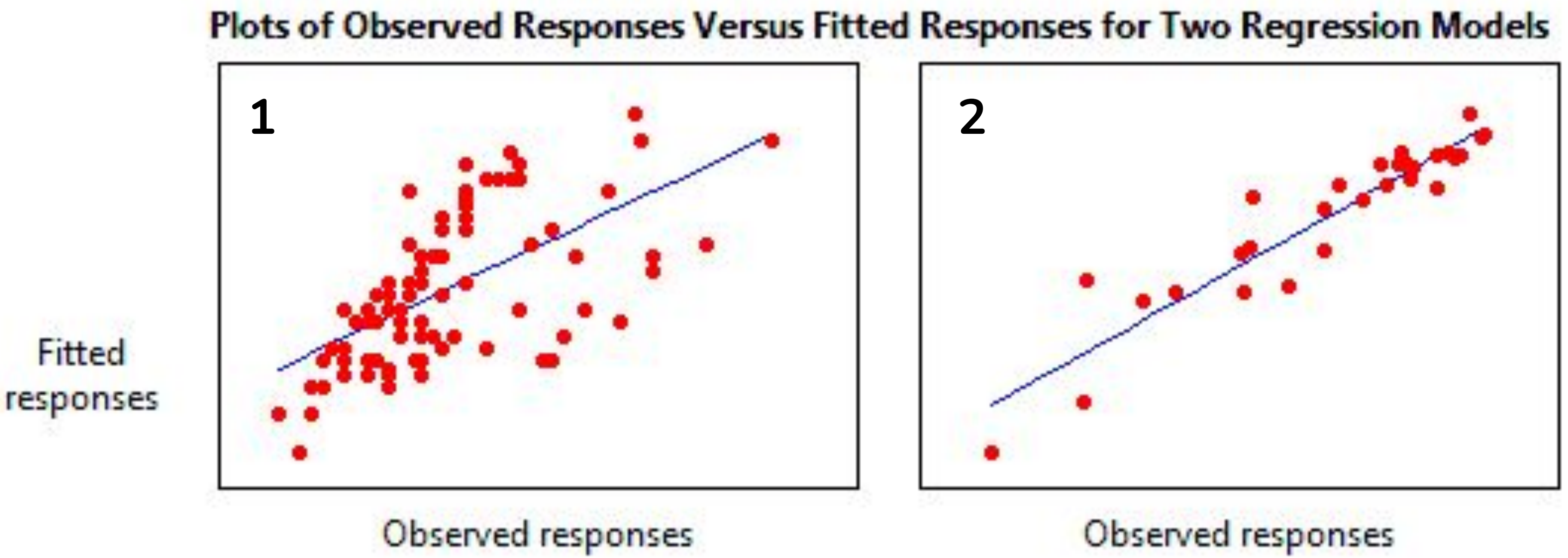


trendline with **least squares regression**

$$\frac{\boxed{41.1879} - \boxed{13.7627}}{\boxed{41.1879}} = \frac{27.4252}{41.1879} \approx 66.59\%$$

= the R^2 value!

Another explanation: https://www.youtube.com/watch?v=PaFPbb66DxQ

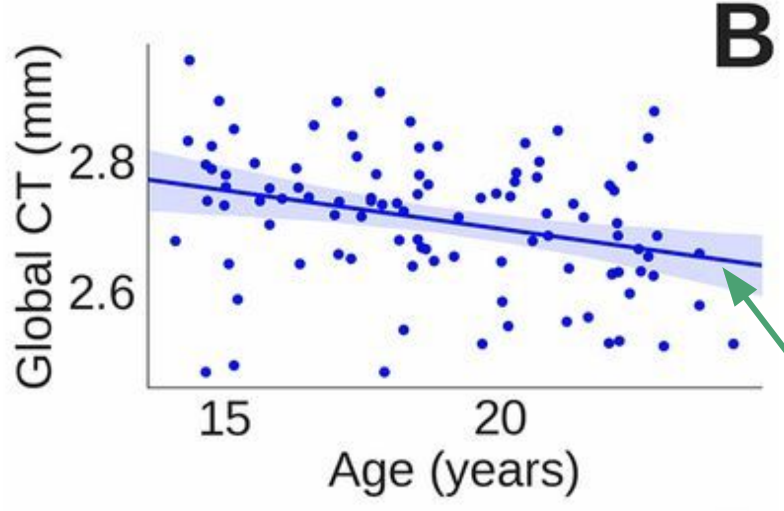


Which of these have a higher R^2 value?

What's the difference between r and R^2?

r	**R^2**
correlation coefficient (ρ, rho)	coefficient of multiple determination
measure of the strength of the **linear** relationship between two variables	measure of the strength of the **regression** equation, or how well we can predict *y*
ranges from -1 to +1	ranges from 0 to +1
Differences are less intuitive	Differences are more intuitive, in other words, $R^2 = 0.6$ is twice of good as a fit as $R^2 = 0.3$

Additional explanation: https://www.youtube.com/watch?v=2AQKmw14mHM
Play with r values: https://www.esci.thenewstatistics.com/esci-correlation.html



"In the adolescent period from 14 to 24 y old, there was evidence of significant cortical shrinkage (r^2 = 0.10; *P* = 0.006; estimated global rate of shrinkage; ΔCT = −0.011 mm/y) (Fig. 1B)."

Linear regression line

(r^2 tells you how well line captures data)

Covariance vs. correlation

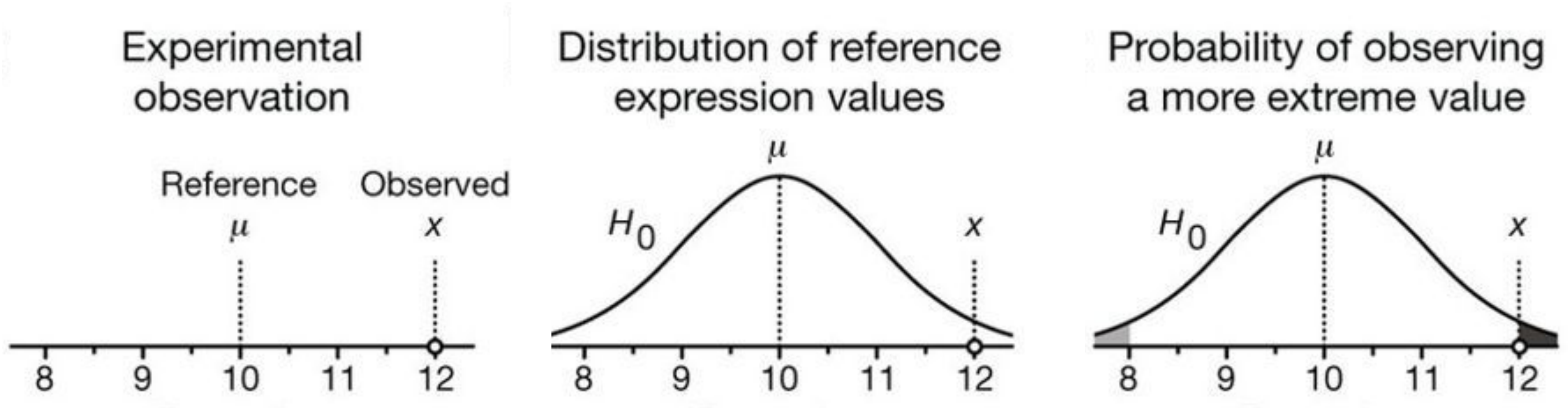
- Covariance measures how two variables vary in tandem from their means
  - The variance formula, but for 2 variables!
  - Difficult to interpret... and also sometimes papers *say* covariance but they're actually calculating a correlation...
- Correlation is a much more common metric, and more interpretable
  - Compute covariance and divide out the standard deviations of both variables

More information: Pearson's Correlation, Clearly Explained!!!
What is COVARIANCE? What is CORRELATION? Detailed video!

Hypothesis testing allows us to ask whether two distributions of data are similar or different

The type of test we run depends on whether our data is *normally distributed* or not

Testing significance for one sample from a normal distribution

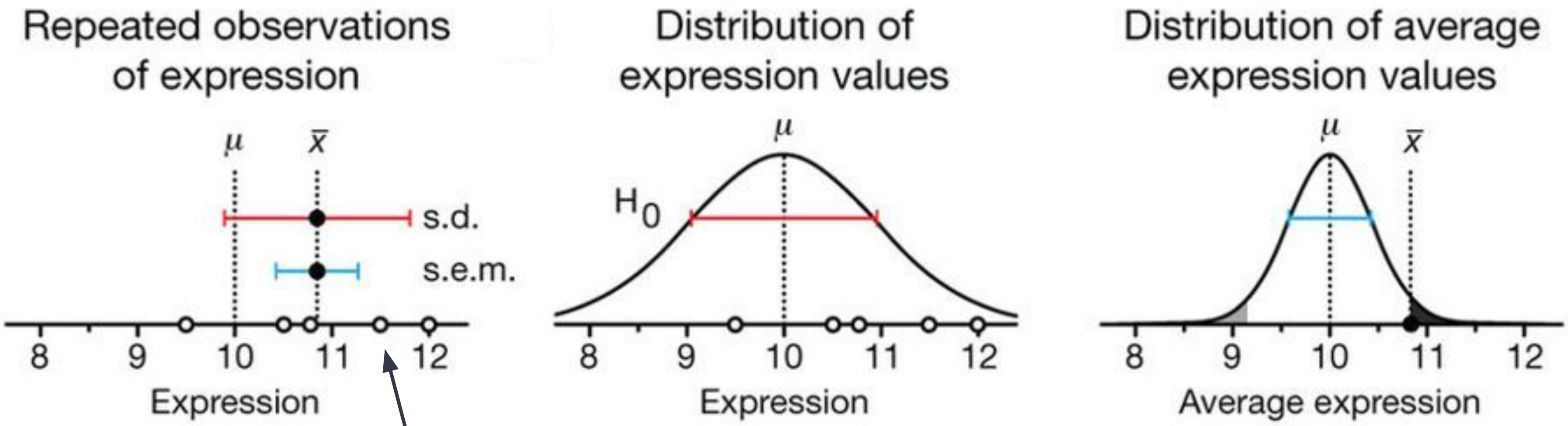


H_0: the null distribution
(variability is due to noise in our measurement)

The statistical significance of the observation *x* is the probability of sampling that value from the distribution, given by the shaded areas (one or two-tailed) under the distribution curve.

https://www.nature.com/articles/nmeth.2698

How can we estimate the spread of the null distribution?

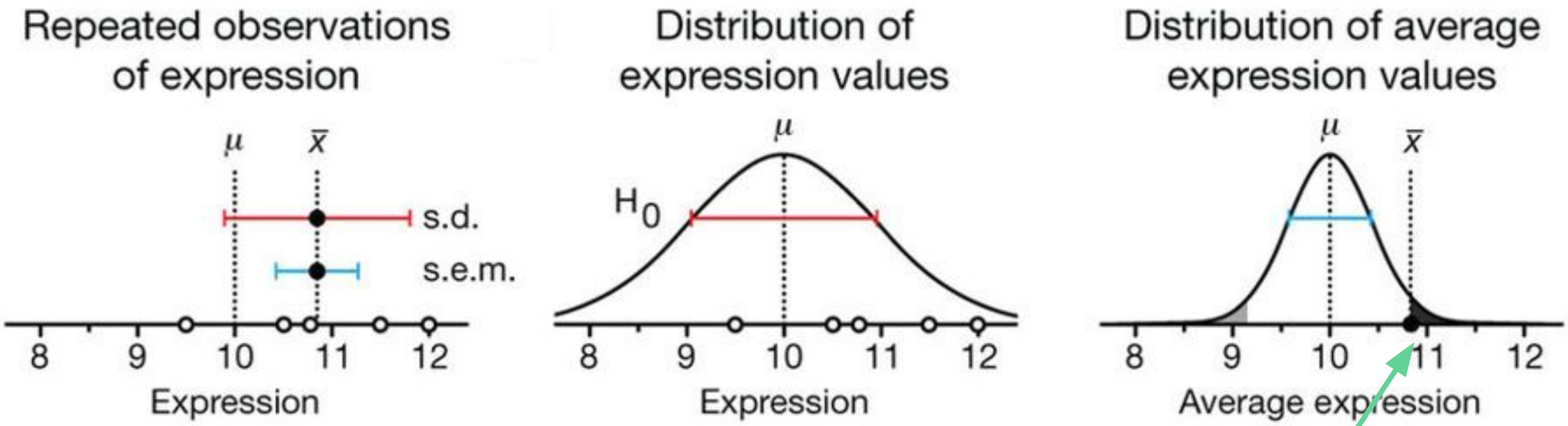


5 observations
= sample

First assumption: s.d. of null distribution is = to s.d. of sample
(equal variances)

Sampling distribution of means is also normal, estimated by s.d./√n = s.e.m.

How can we estimate the spread of the null distribution?

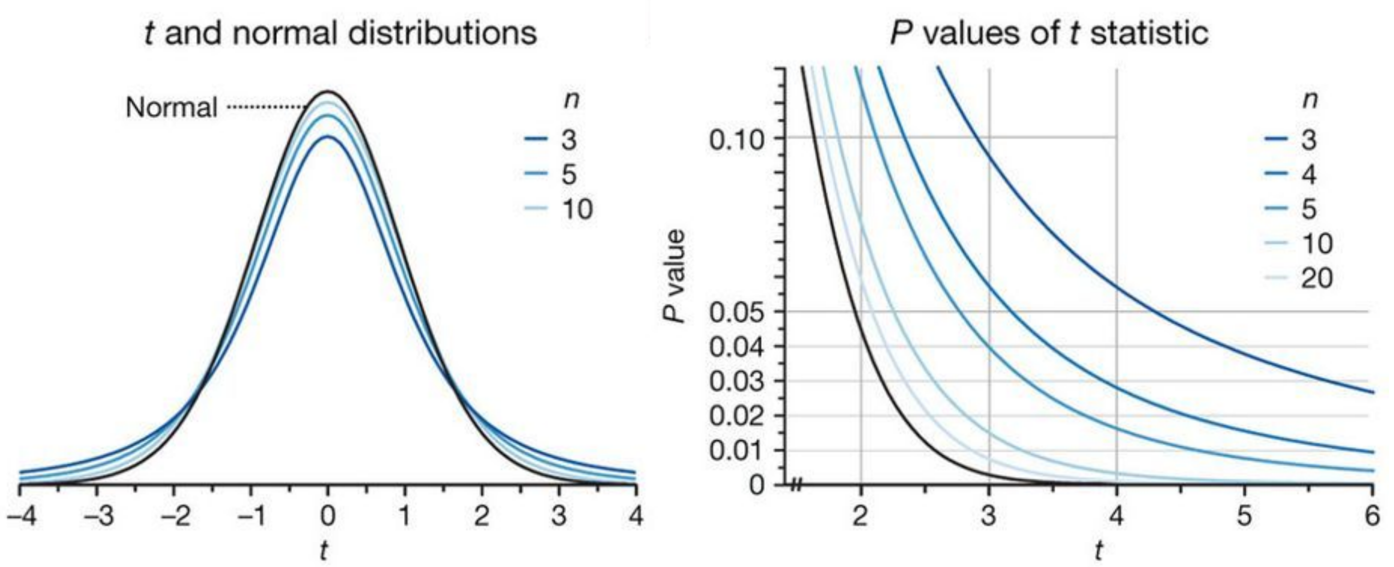


Repeated independent observations are used to estimate the s.d. of the null distribution and derive a more robust *P* value.

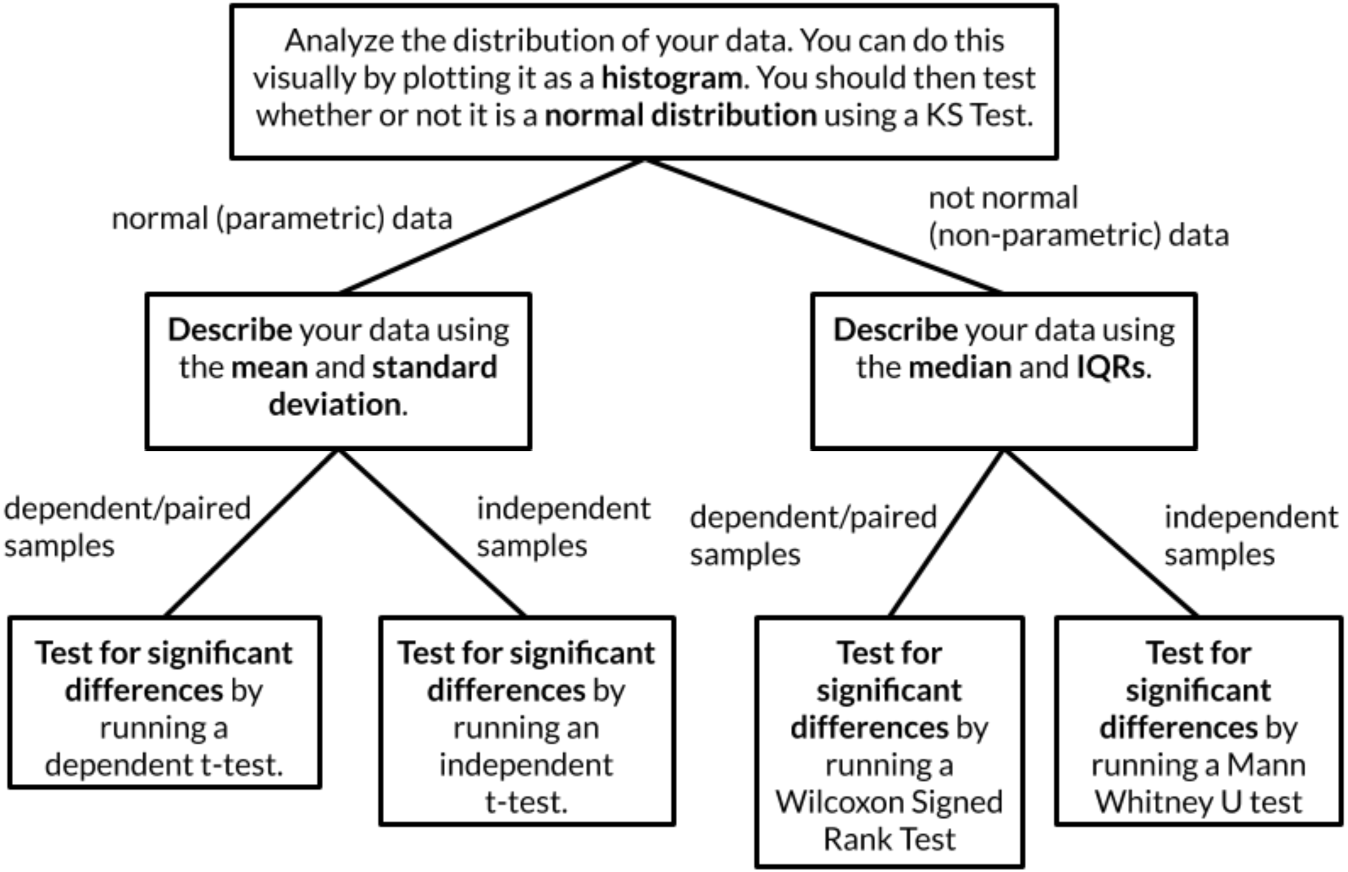
The *P* value of $\bar{x}$ is the shaded area under this curve.

We'll use all of these assumptions to generate a test statistic:

$$t = \frac{\text{sample mean - population mean}}{\text{standard error}}$$ **= one sample t-test**

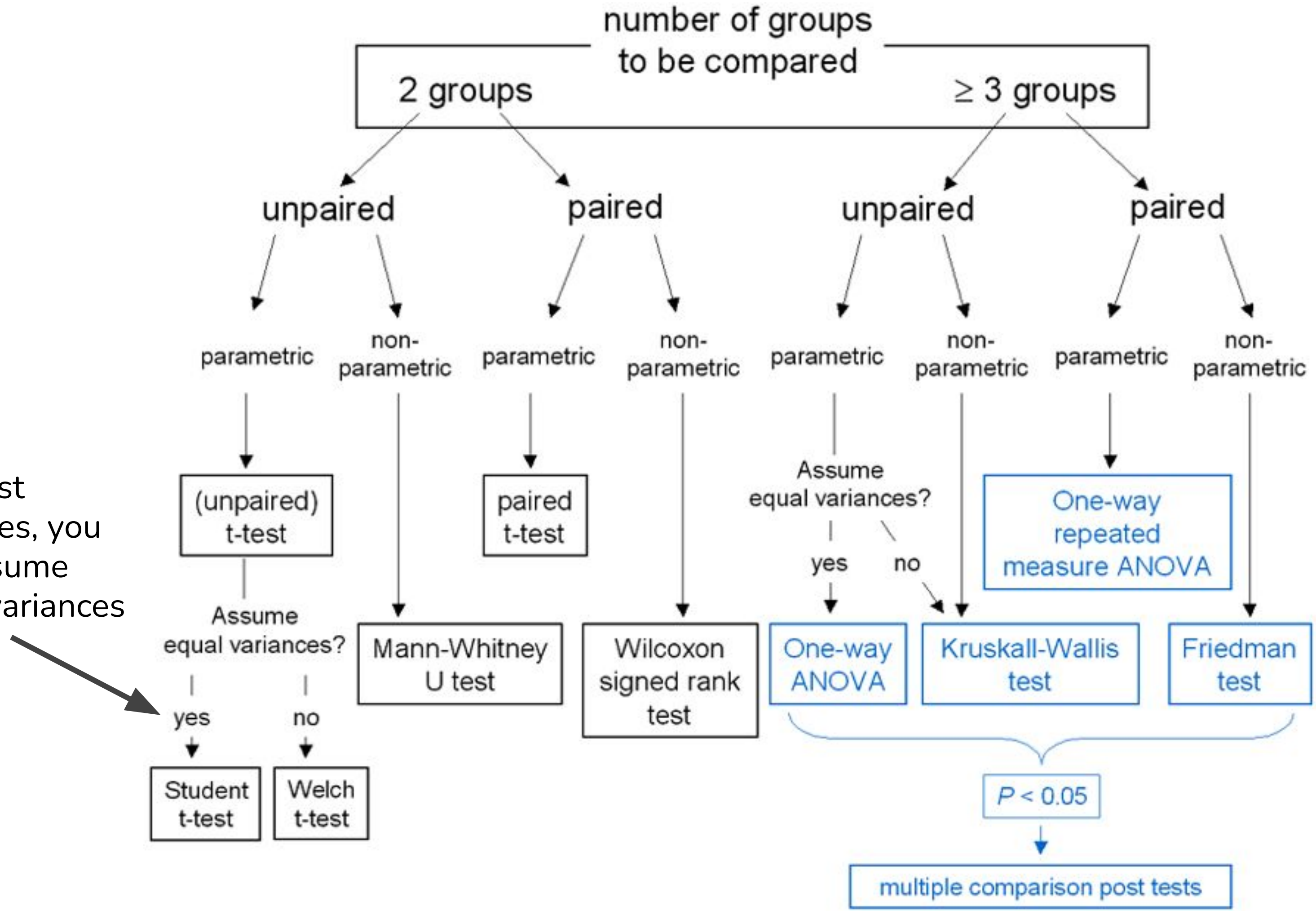


https://www.nature.com/articles/nmeth.2698



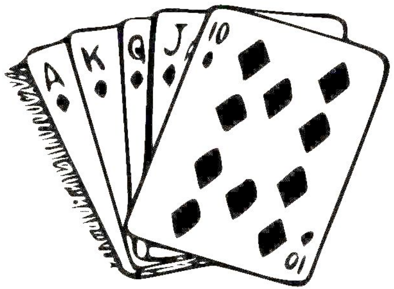
Statistics workflow for two groups of data

For most purposes, you can assume equal variances



More common in data science: **permutation** tests

- "Proof by contradiction"
- Null hypothesis = all samples come from the same distribution
- Shuffle the labels! (aka a **shuffle test**)



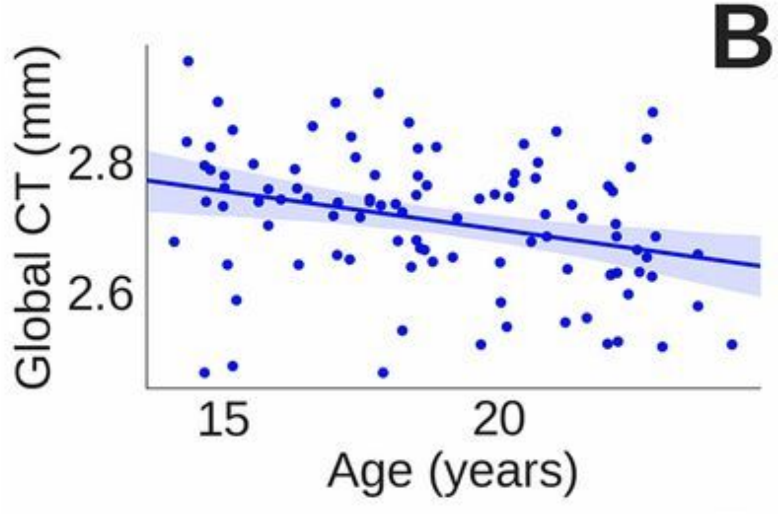
$$ {}_{n}P_{r} = \frac{n!}{(n-r)!} $$

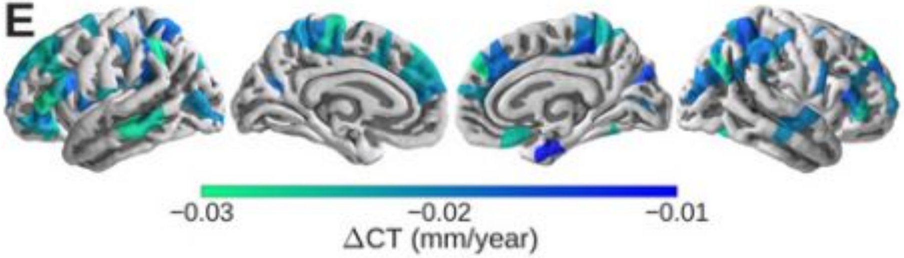
${}_{n}P_{r}$ = permutation

n = total number of objects

r = number of objects selected

For an example: https://inferentialthinking.com/chapters/12/1/AB_Testing.html#permutation-test





(B) CT decreased linearly with increasing age in the range 14–24 y old, although (E) there were regional differences in the rate of cortical shrinkage (ΔCT), with significantly nonzero rates of shrinkage (**permutation test**; FDR = 0.05)

… The statistical significance of the goodness of fit of the first two PLS components was tested with two-tailed α = 0.05 by 1,000 **permutations** of the response variables.”

Terms from today

Continuous

Discrete

Binomial

Multinomial

Poisson

Bivariate

Multivariate

Inferential statistics

Hypothesis testing

Probability density function

Cumulative density function

Permutation