Statistics Fundamentals

Providing universal access to AI education and practice

Disclosure

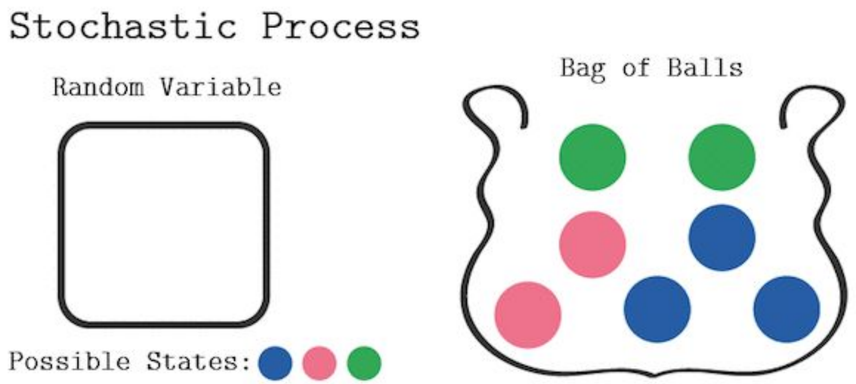
- This is a very condensed and simplified version of **statistics**. It is not comprehensive, and is absolutely not a substitute for a one-year university course -we would be very rich otherwise-.

- You are strongly encouraged to do the included Exercises and review or even research some of the topics in order to reinforce the main ideas.

Ready?

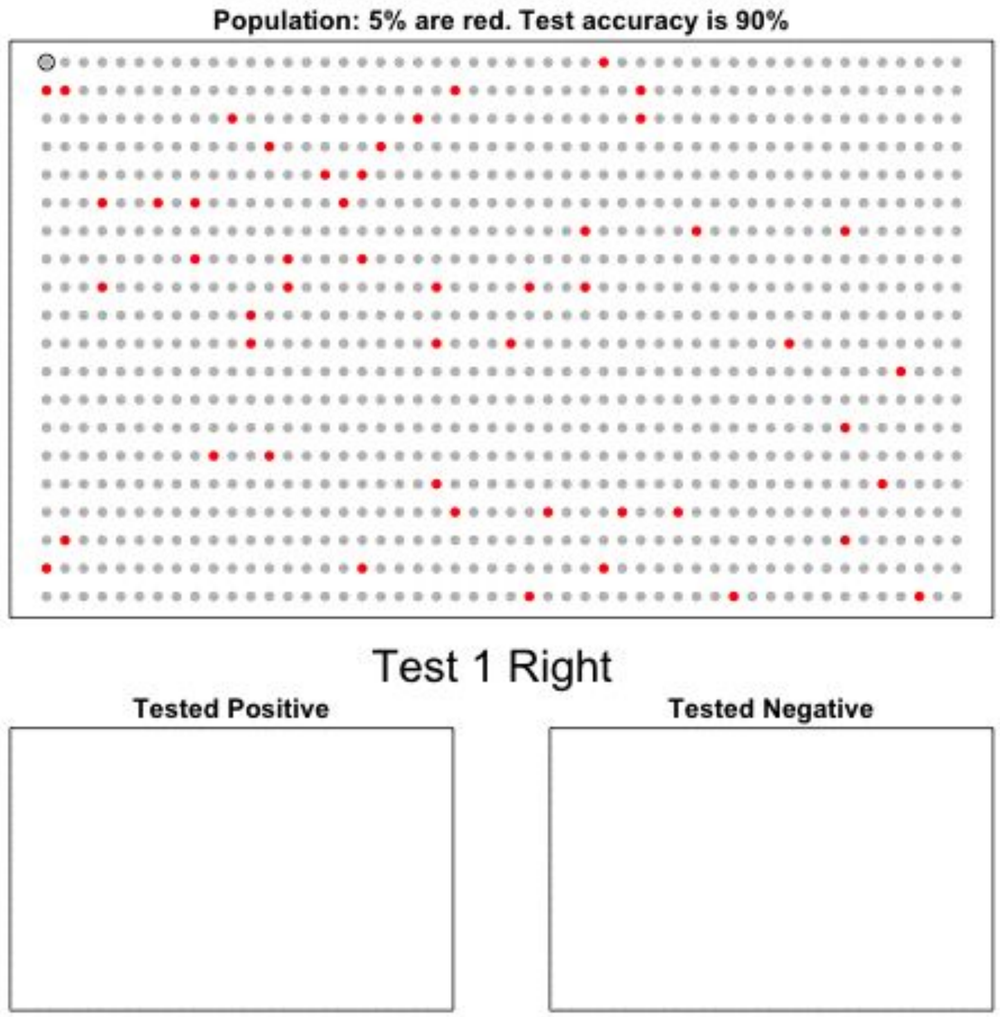
What you should already know: **Probability**

- Axioms of Probability
- Marginal Probability
- **Joint Probability**
- **Conditional Probability**
- **Bayes Rule**
- Permutations and Combinations



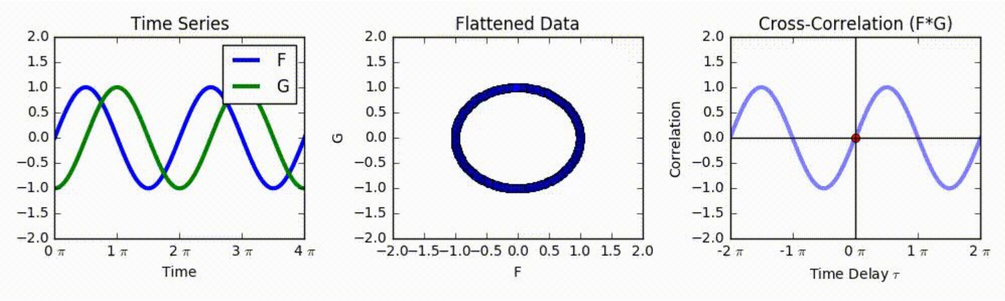
What you should already know: **Statistics**

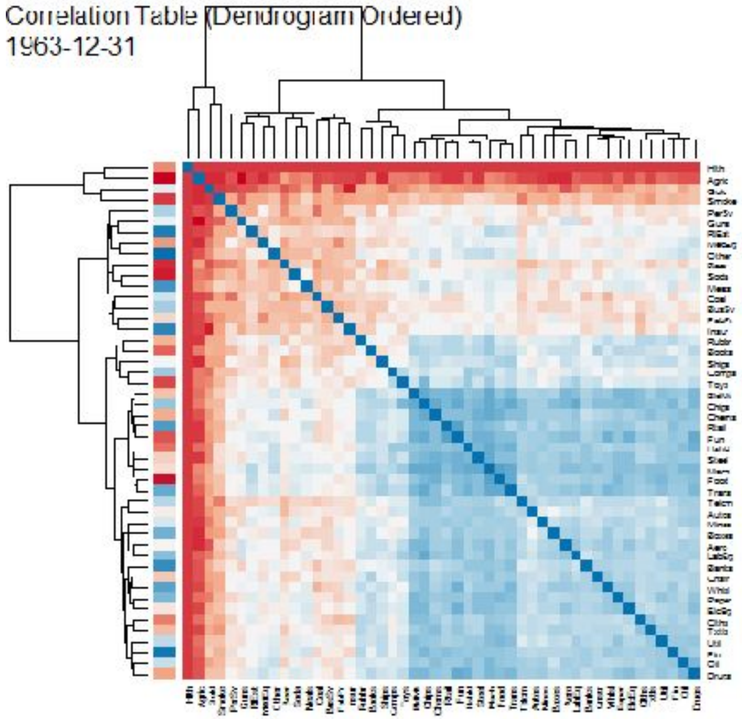
- Sampling
- Central Value
- **Mean and Std deviation**
- **Types of Bias**
- **Correlation** (actually explained last week!)
- **Normal Distribution**



What we will learn today: **Statistics**

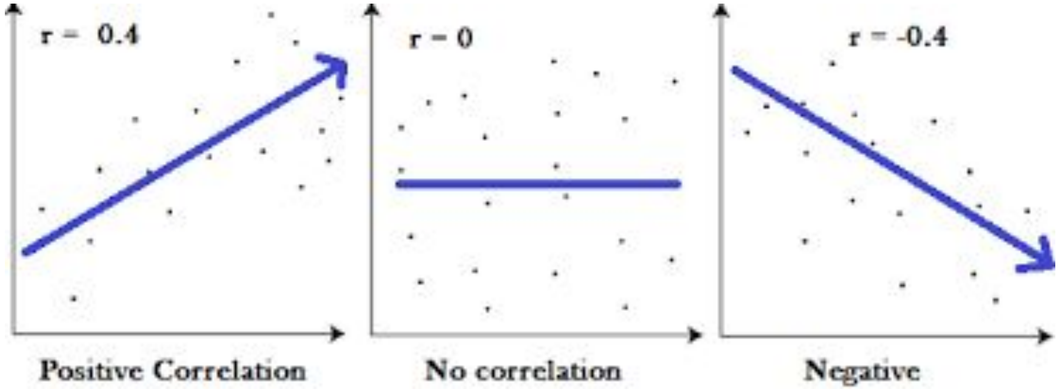
- **Correlation** (actually explained last week!)
- **Types of Distributions**
- **Central Limit Theorem**
- **Hypothesis Testing**
- **R // R^2**

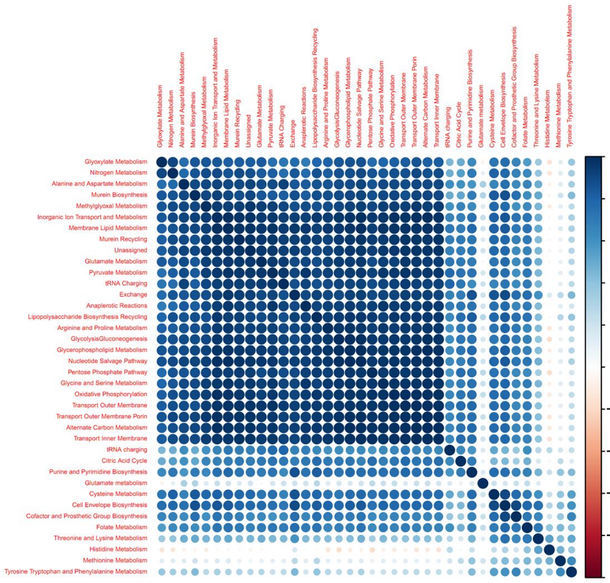


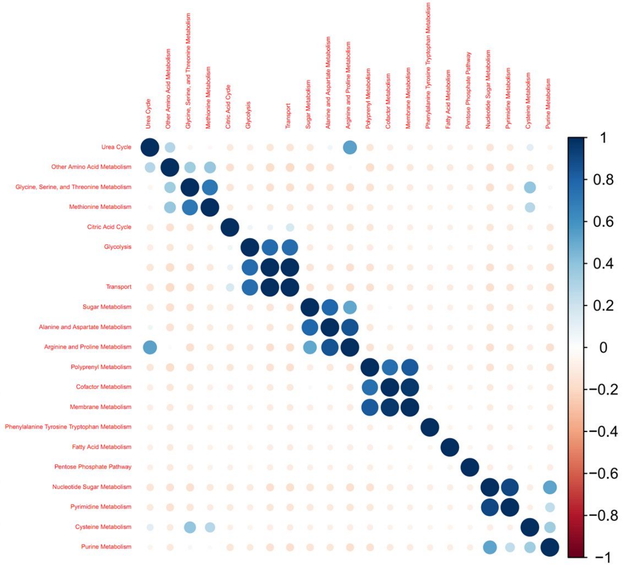


Correlation

- The correlation coefficient quantifies relationship between values.
- **R = 0** means there is no relationship between the variables at all.
- **0 < R <= 1** means that there is a positive correlation
- **-1 <= R < 0** means that there is a negative correlation
- **1** indicates that the **two variables are moving in unison**. They **rise and fall together** and have perfect correlation.

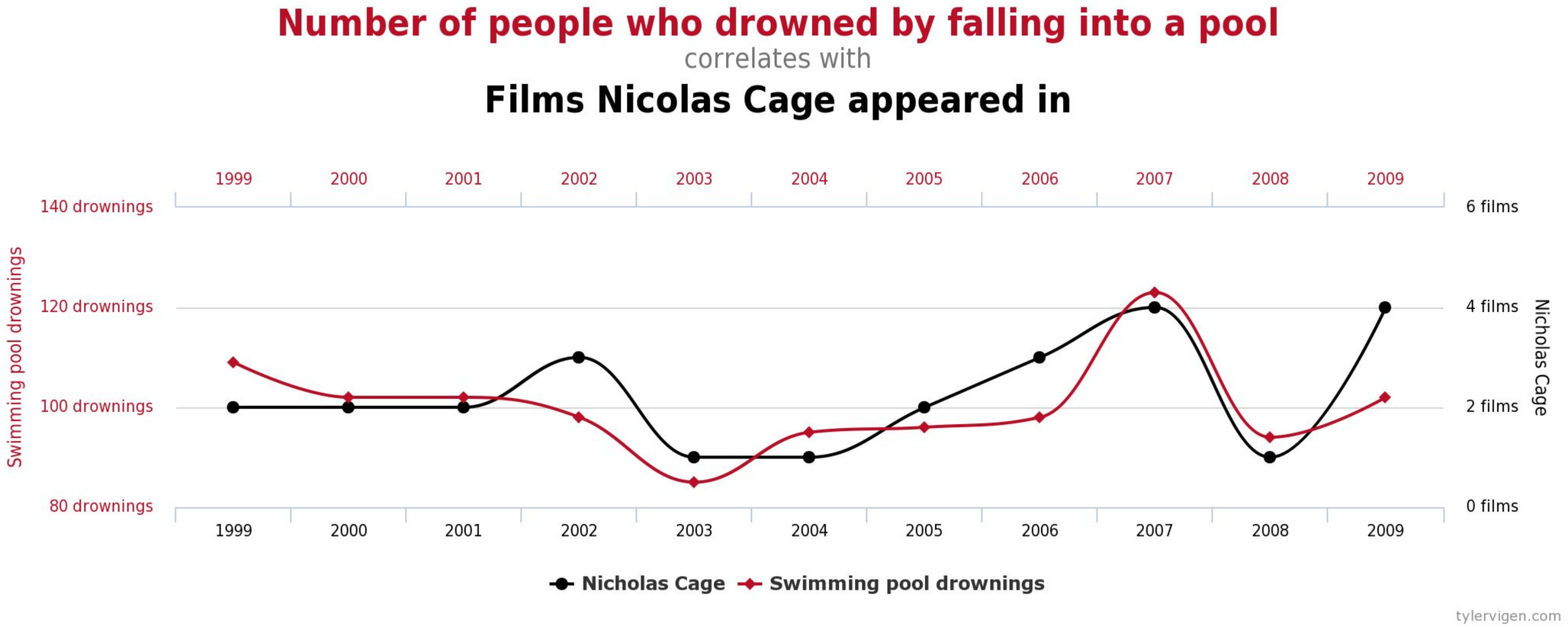




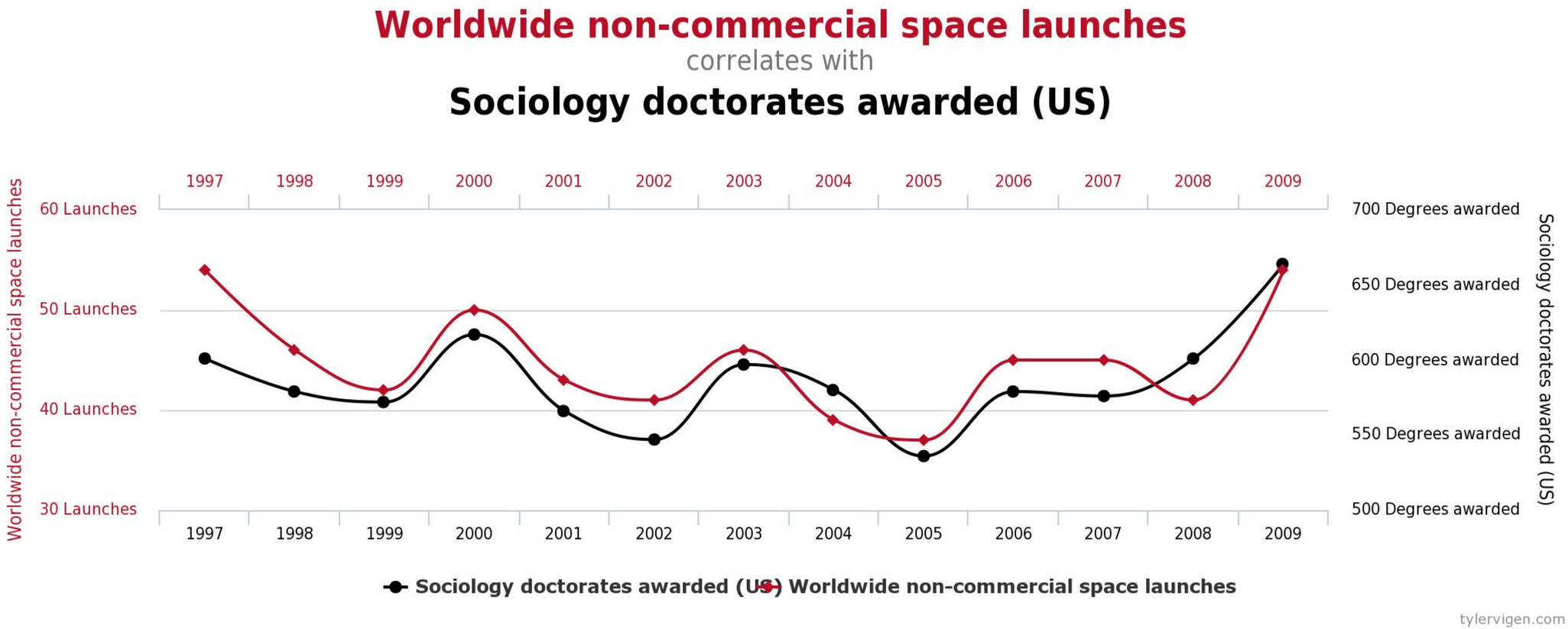


-1 Correlation explained with **cats** 1

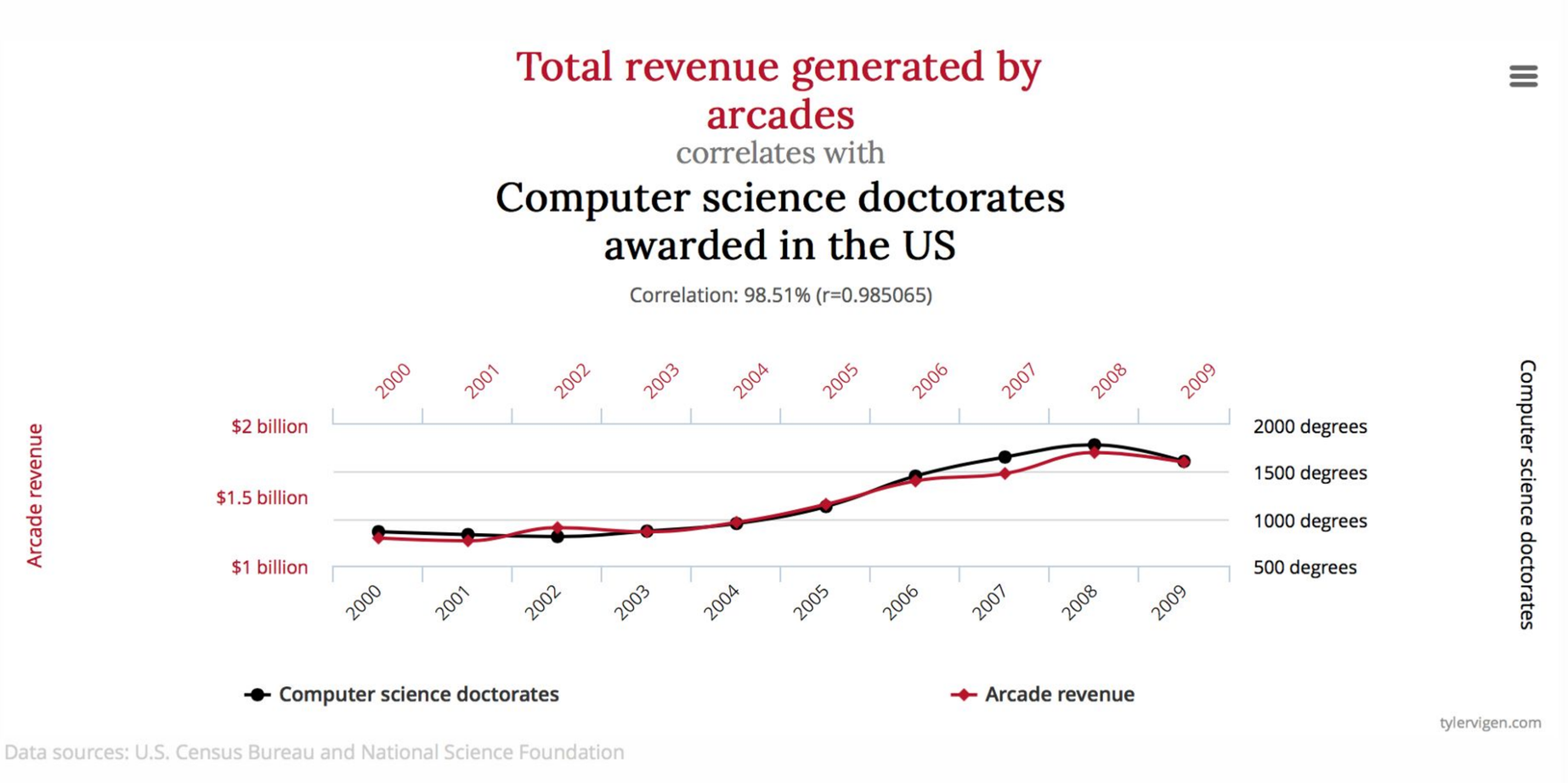
Correlation does not imply **Causation**



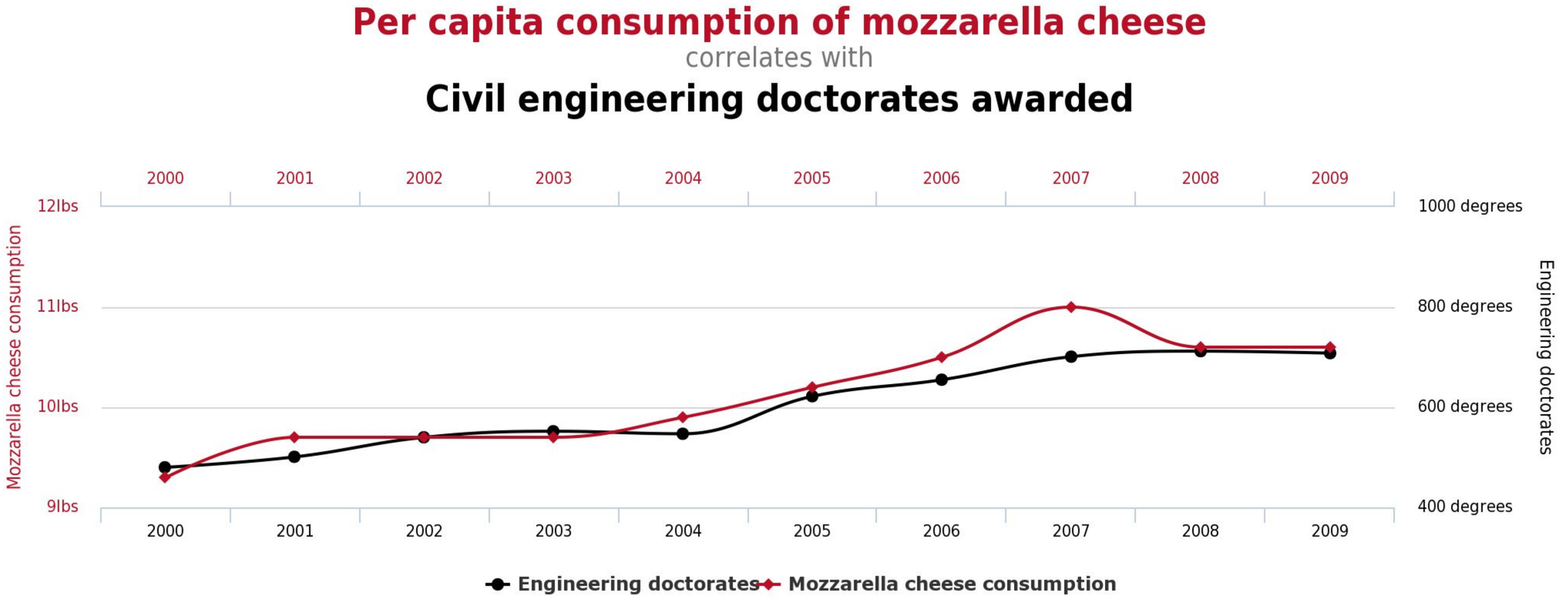
Correlation does not imply **Causation**



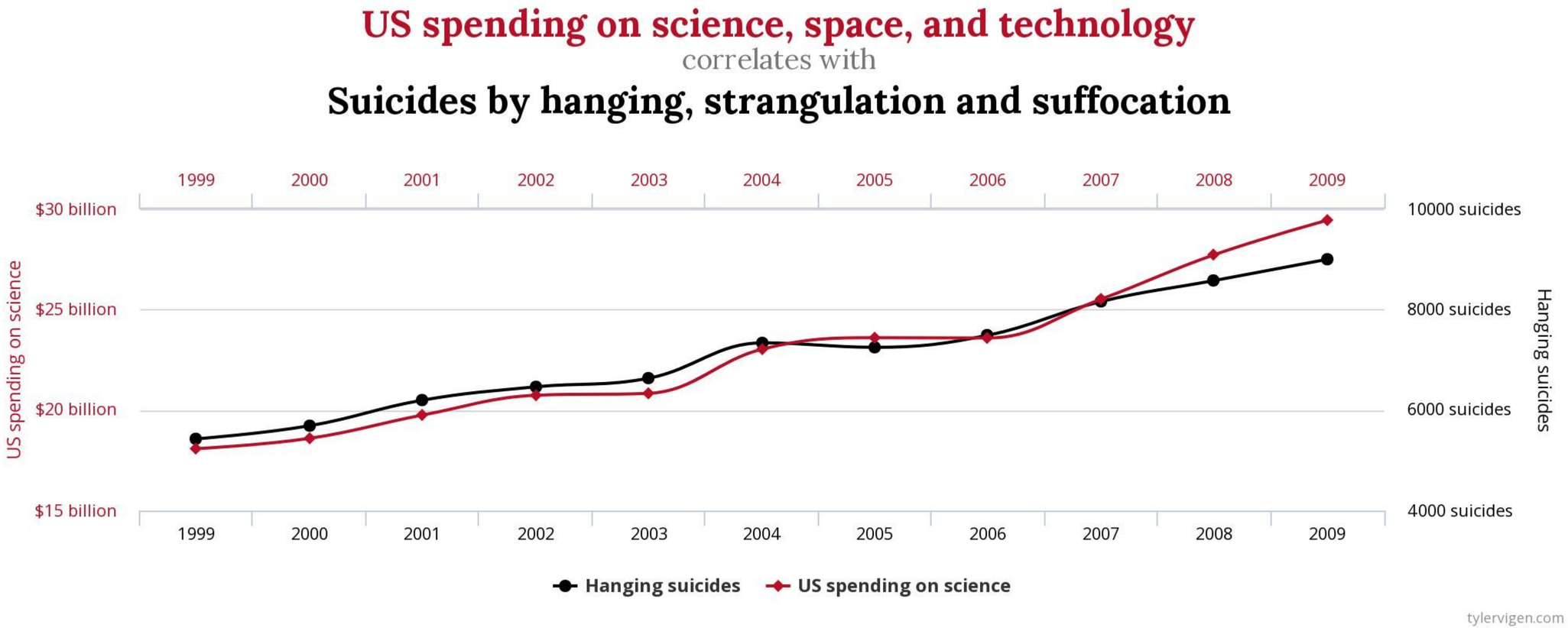
Correlation does not imply **Causation**



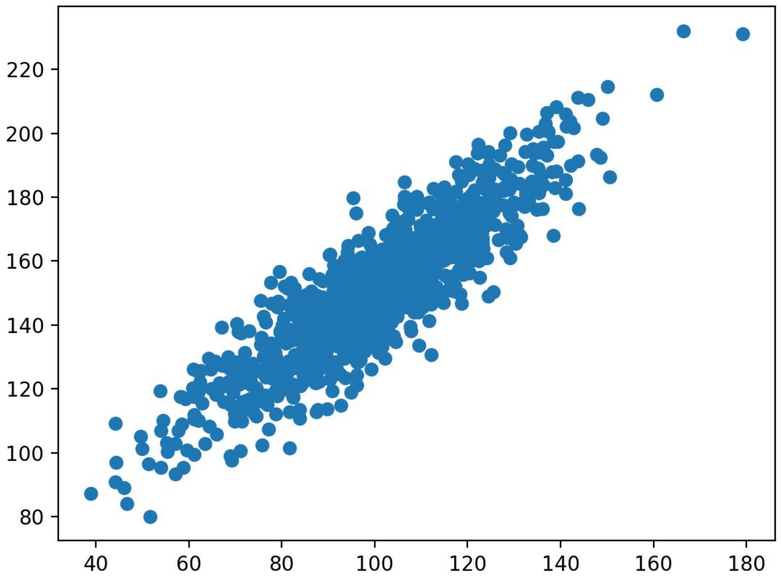
Correlation does not imply **Causation**

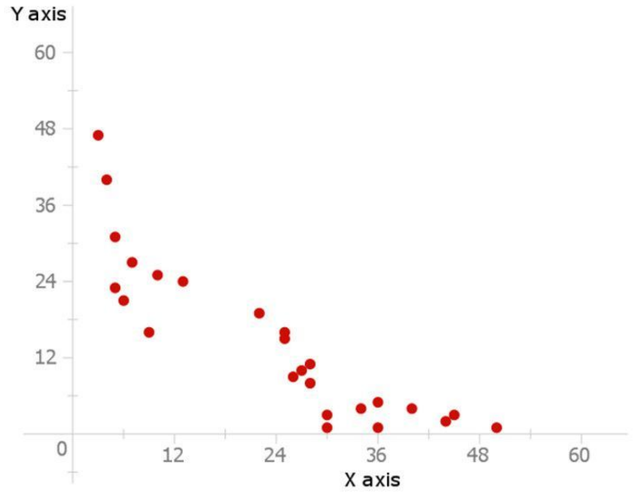


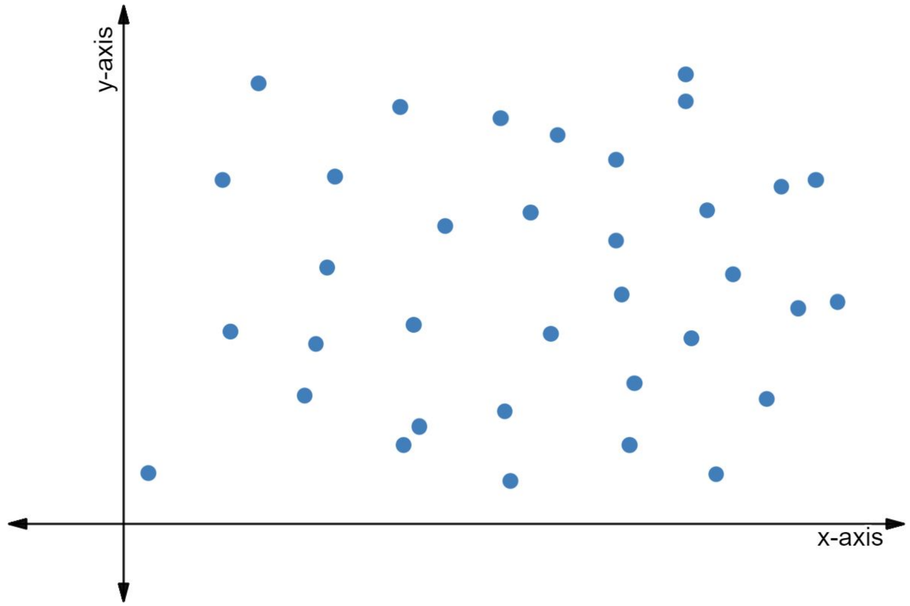
Correlation does not imply **Causation**



What is what?

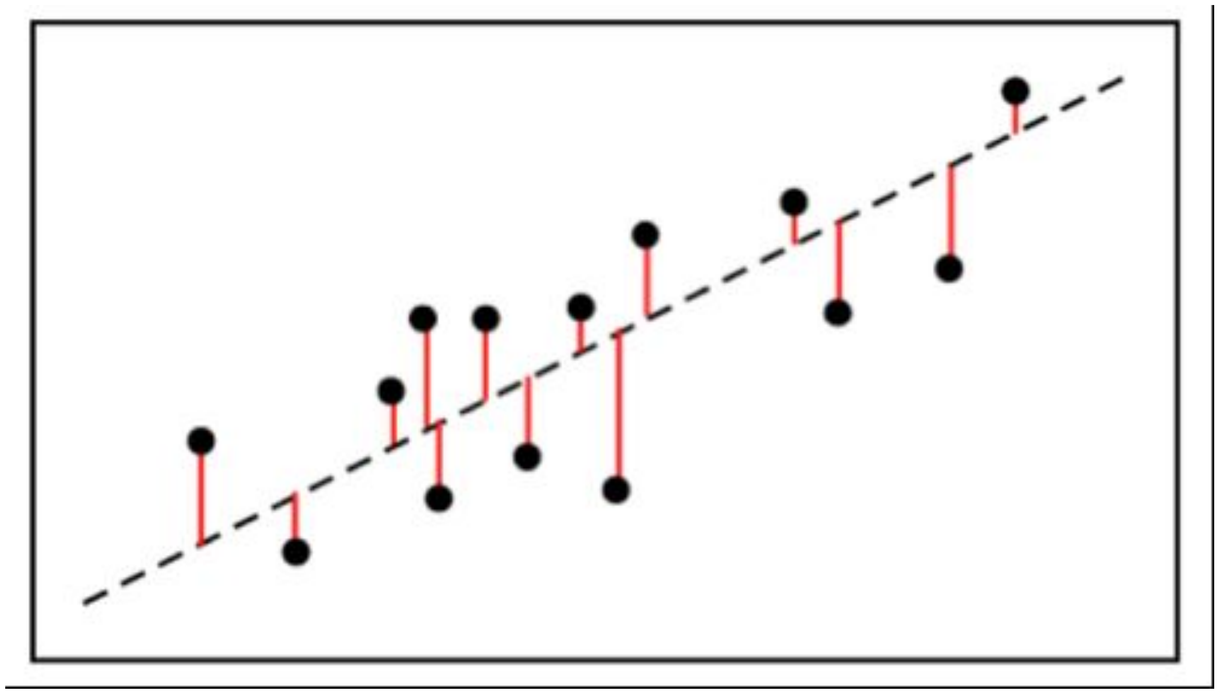

220
200
180
160
140
120
100
80
40
60
80
100
120
140
160
180


Y axis
60
48
36
24
12
0
12
24
36
48
60
X axis


y-axis
x-axis

R and R^2

- **Coefficient of Determination** is the square of Coefficient of Correlation.
- **R-squared** is a statistical measure that represents the **proportion of the variance** for a **dependent variable** that's explained by an **independent variable**



R^2 vs Adjusted R^2

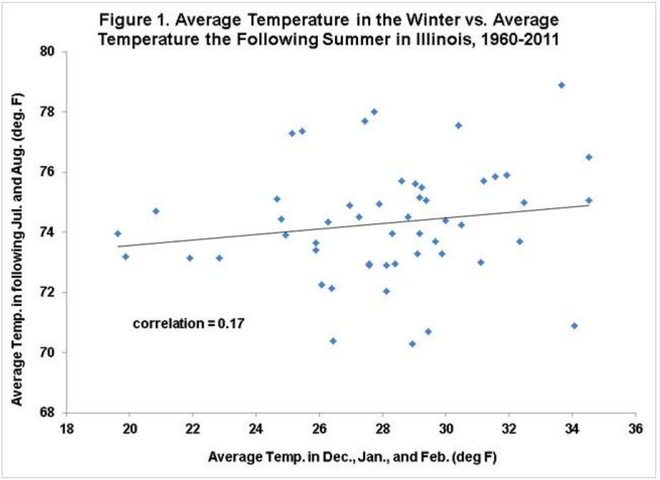
- Both **R^2** and the **adjusted R^2** give you an idea of how many **data points fall** within the **line** of the **regression equation**.

$$R^2 = 1 - \frac{SS_{residuals}}{SS_{total}}$$

- However, the main difference is that **R^2 assumes** that **every** single **variable** explains the **variation** in the **dependent variable**.

$$\text{Adjusted } R^2 = 1 - \frac{SS_{residuals} / (n-K)}{SS_{total} / (n-1)}$$

- The **adjusted R^2** tells you the **percentage of variation explained** by **only the independent variables** that actually affect the dependent variable.



R^2 vs Adjusted R^2

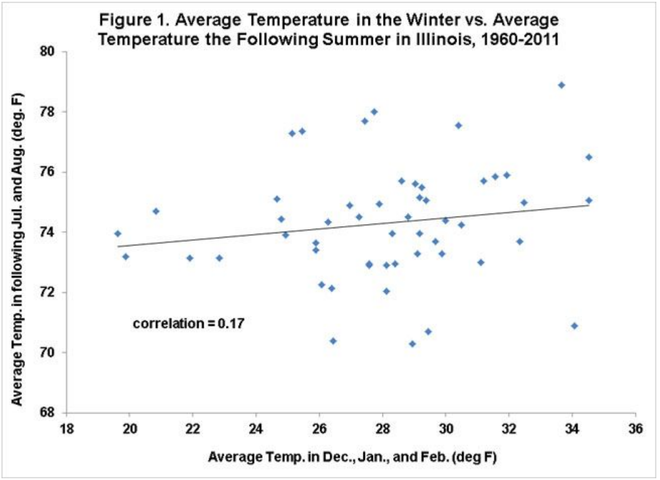
- In other words, the **adjusted R^2** penalized our choice of additional independent variables (or parameters) if that addition is not good enough.

$$R^2 = 1 - \frac{SS_{residuals}}{SS_{total}}$$

- We could get **too many variables** to explain the **weather**, which would lead to a higher R^2. The adjusted fixes this!

$$\text{Adjusted } R^2 = 1 - \frac{SS_{residuals} / (n-K)}{SS_{total} / (n-1)}$$

- In order to choose which model is better, we can see how the **Adjusted R^2** is better :)
- **SS** = Sum of Squared residuals.



Types of Distributions

- Bates Distribution.
- **Bernoulli Distribution**
- Beta Binomial Distribution
- **Beta Distribution**.
- **Binomial Distribution**.
- **Bimodal Distribution**.
- **Bivariate Normal Distribution**.
- Bradford Distribution
- Burr Distribution.
- **Categorical Distribution**
- Cauchy Distribution.
- Compound Probability Distribution
- Continuous Probability Distribution
- Cumulative Frequency Distribution
- Cumulative Distribution Function
- Degenerate Distribution.
- Dirichlet Distribution.
- **Discrete Probability Distribution**
- Empirical Distribution Function
- Erlang Distribution.

Types of Distributions

- **Exponential Distribution.**
- Extreme Value Distribution.
- **F Distribution.**
- Factorial Distribution
- **Fat Tail Distribution.**
- Fisk Distribution.
- Folded Normal / Half Normal Distribution.
- G-and-H Distribution.
- Generalized Error Distribution.
- Geometric Distribution.
- Gompertz Distribution.
- Heavy Tailed Distribution
- Hypergeometric Distribution.
- Inverse Gaussian Distribution.
- Inverse Normal
- J Shaped Distribution.
- Kent Distribution
- Kumaraswamy Distribution
- Laplace Distribution.

Types of Distributions

- Lévy Distribution.
- Lindley Distribution.
- Lognormal Distribution.
- Lomax Distribution.
- Long Tail Distribution.
- **Marginal Distribution**
- Mixture Distribution
- Multimodal Distribution.
- **Multinomial Distribution.**
- **Multivariate Normal Distribution.**
- Nakagami Distribution.
- Negative Binomial Distribution
- Normal Distribution.
- Open Ended Distribution
- **Pareto Distribution.**
- **Pearson Distribution.**
- PERT Distribution.
- **Poisson Distribution.**
- Power Law Distribution
- Rayleigh Distribution.

Types of Distributions

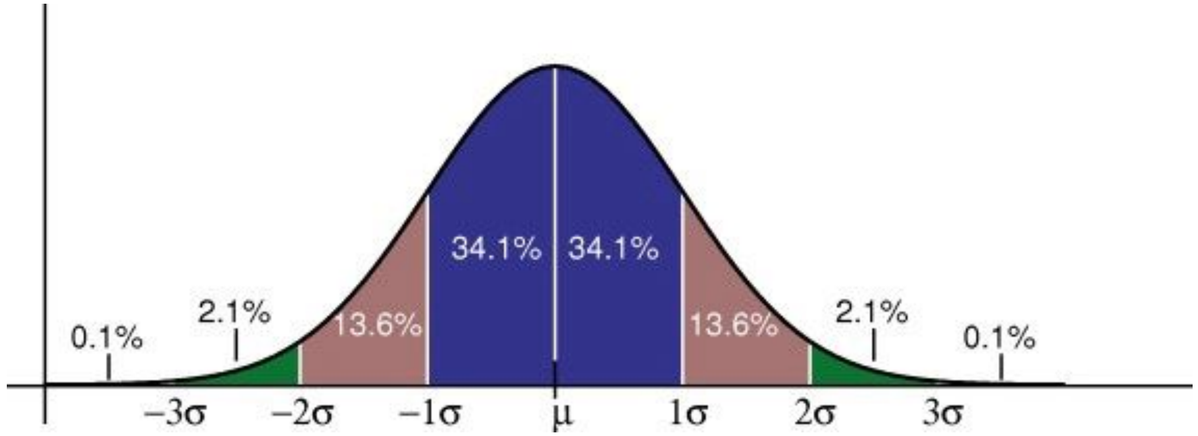
- Reciprocal Distribution.
- Relative Frequency Distribution
- Rician Distribution.
- Skewed Distribution
- Stable Distribution
- Symmetric Distribution
- **T Distribution.**
- Trapezoidal Distribution.
- Triangular Distribution.
- Truncated Normal Distribution.
- Tukey Lambda Distribution.
- Tweedie Distribution.
- Uniform Distribution.
- Unimodal Distribution.
- U-Shaped Distribution.
- Von Mises Distribution.
- Wallenius Distribution.
- Waring Distribution.
- Weibull Distribution.
- Wishart Distribution.
- Yule-Simon Distribution
- **Zeta Distribution**.

Know by heart **ONE DISTRIBUTION** (The Normal one)

Usage: Everything

Parameters

- Mean (mu)
- Variance (sigma)



Formula

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\mu =$ Mean

$\sigma =$ Standard Deviation

$\pi \approx 3.14159\cdots$

$e \approx 2.71828\cdots$

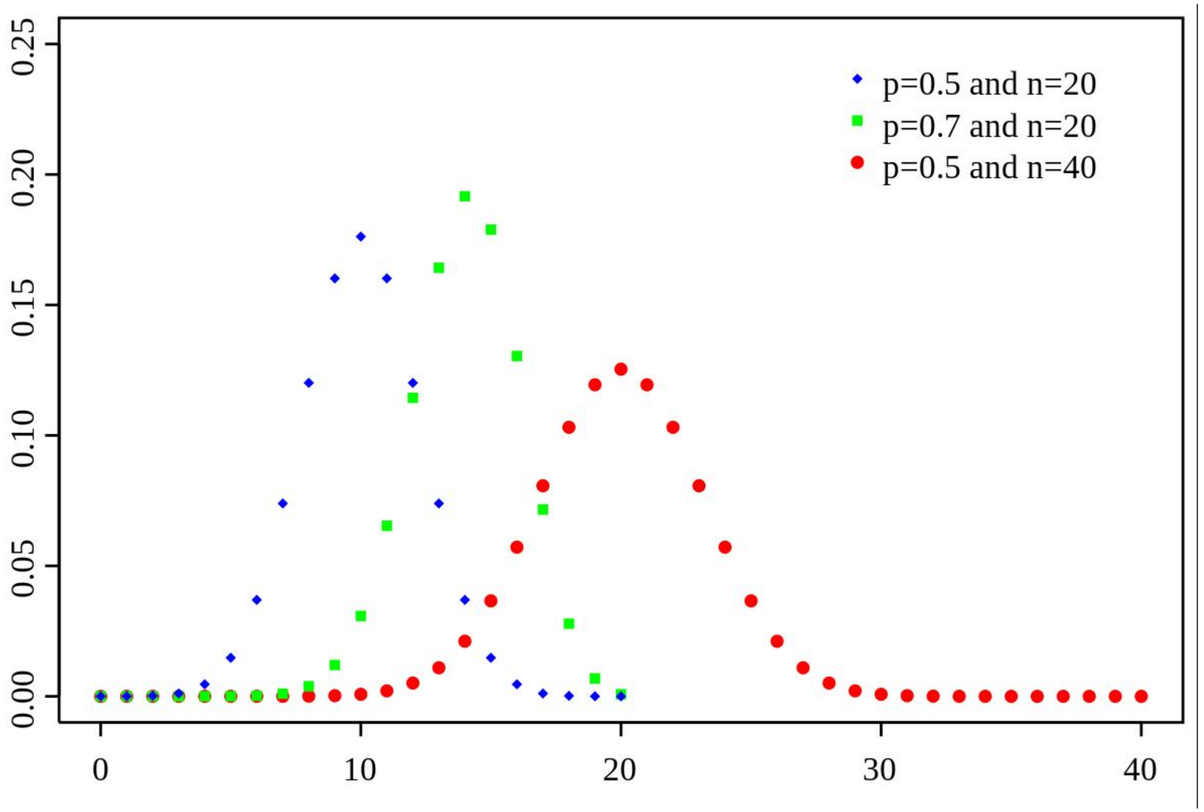
How to learn a new distribution: **Bionmial**

Usage: Discrete Outcomes (A or B)
Parameters

- p
- q

Formula

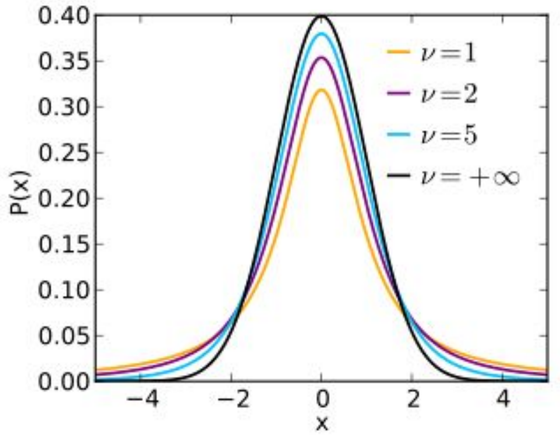
$$P(X) = \frac{n!}{(n-X)!\,X!} \cdot (p)^X \cdot (q)^{n-X}$$



How to learn a new distribution: **t-Student**

Usage: Small Sample Size
Parameters

- v = variance
- mu (normally 0)



Formula

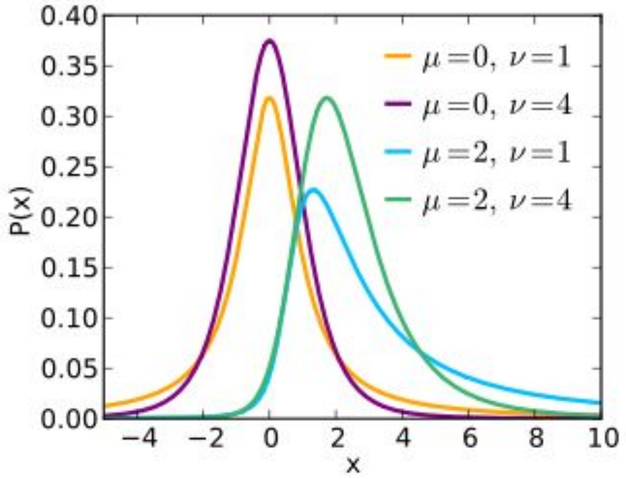
Probability density function [edit]

Student's ***t*-distribution** has the probability density function given by

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

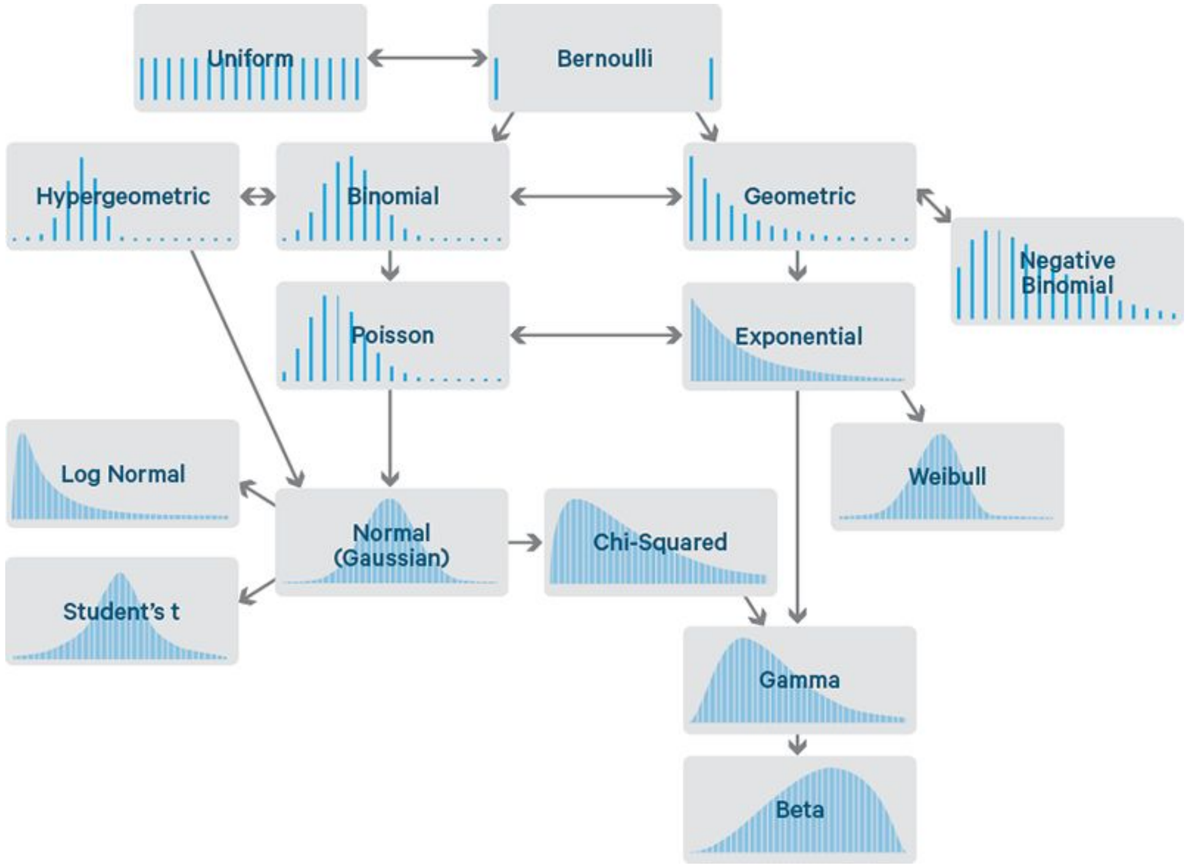
where ν is the number of *degrees of freedom* and Γ is the gamma function. This may also be written as

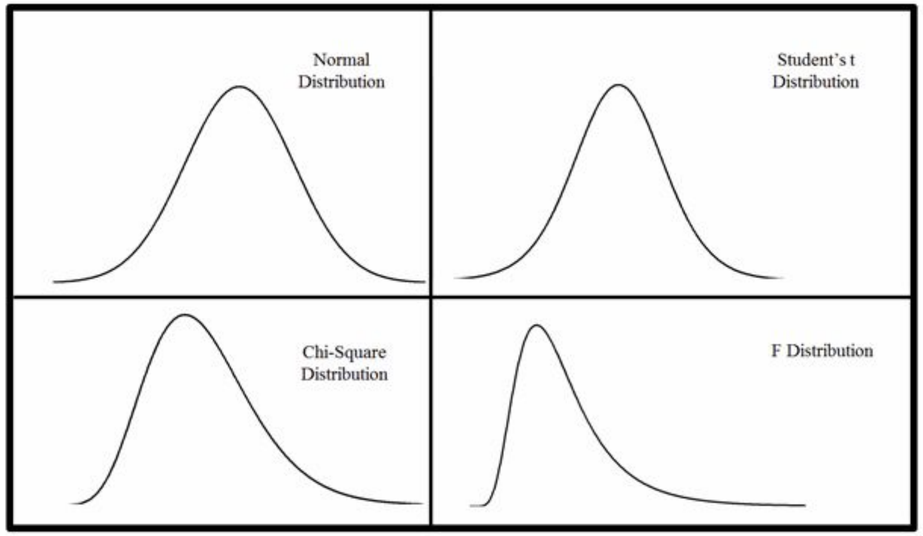
$$f(t) = \frac{1}{\sqrt{\nu}\,\mathrm{B}(\frac{1}{2},\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$



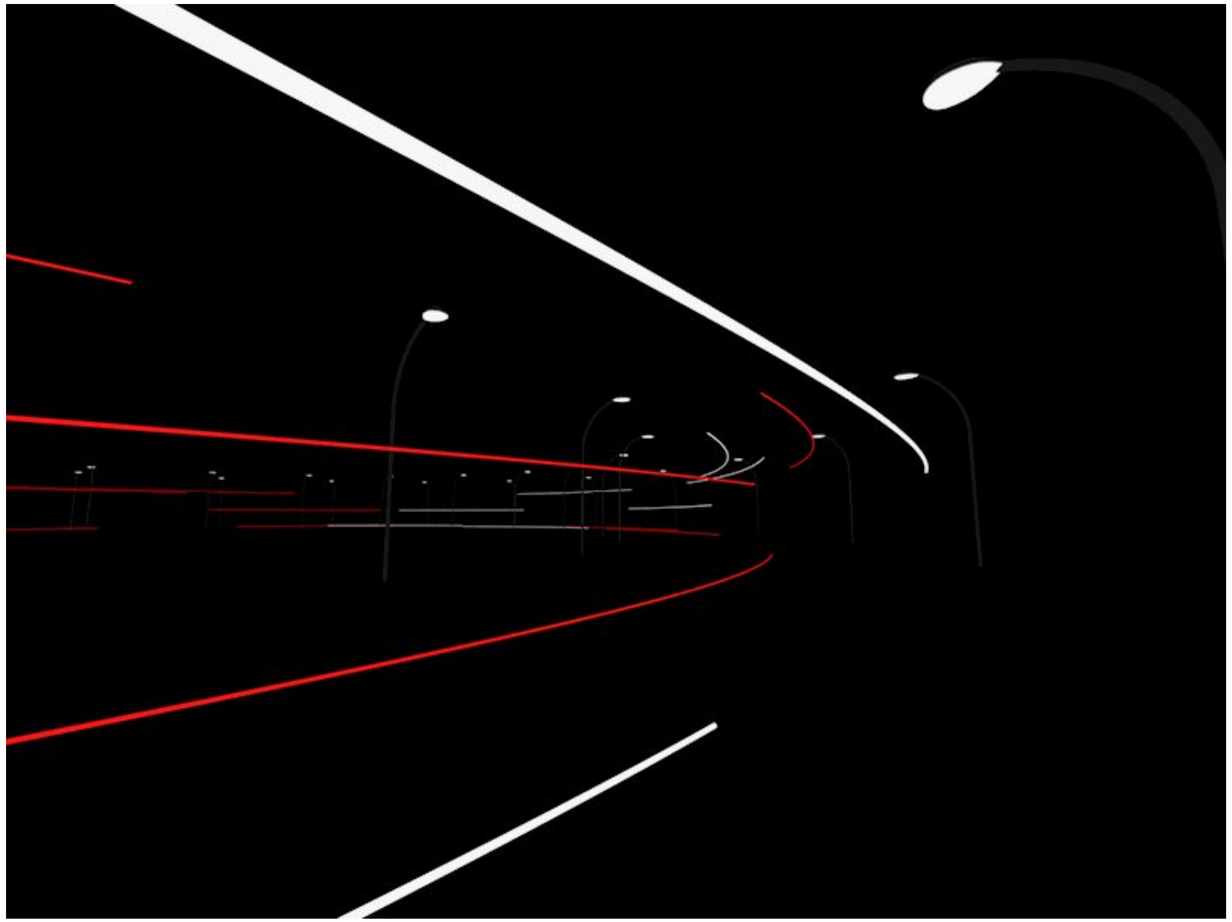
Family of Distributions

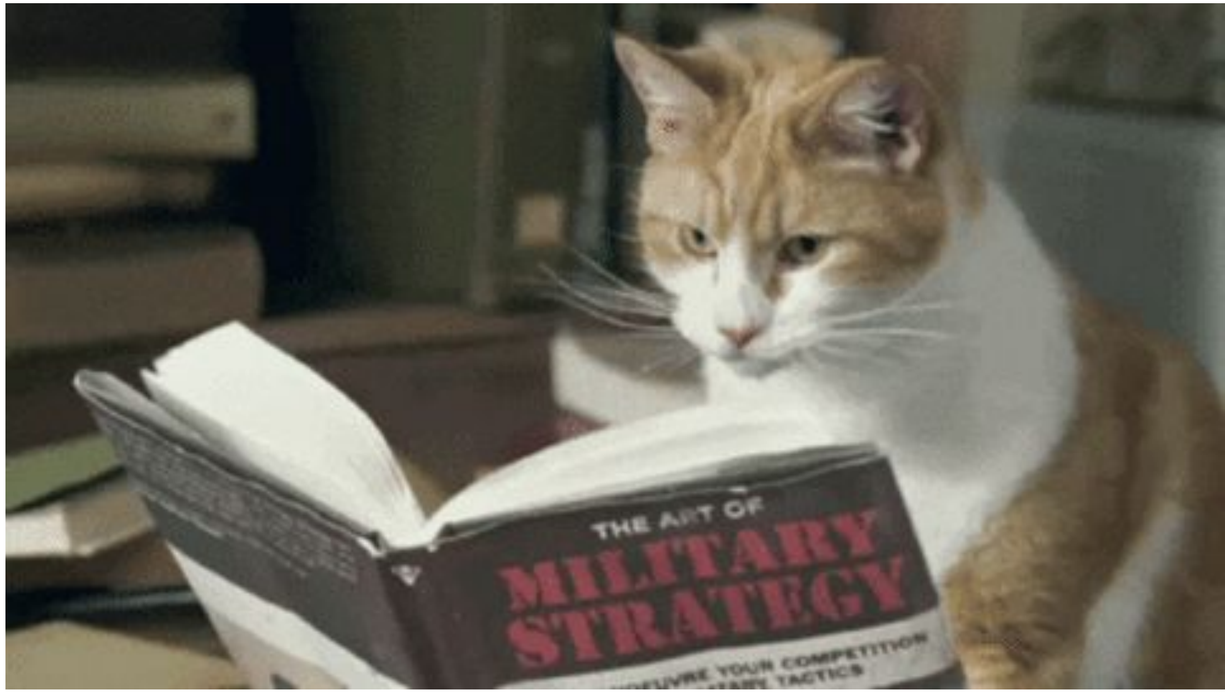
- Discrete Distributions
- Continuous Distributions
- Mixed (D+C) Distributions
- Joint Distributions (product of Distr)
- Non-numeric Distributions





You will have to learn on the job...



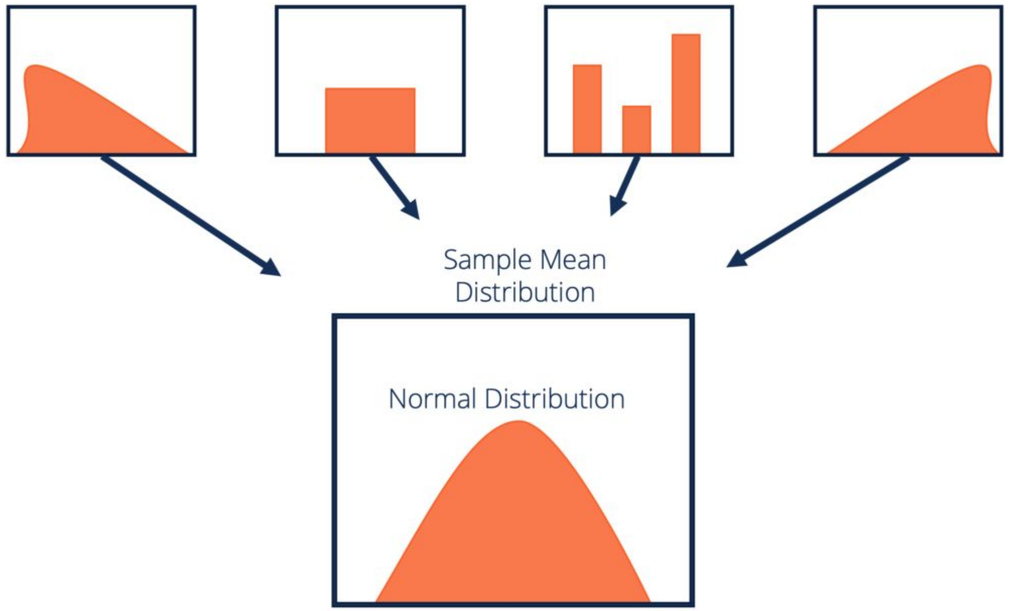


Central Limit Theorem

- It establishes that, in some situations, when **independent random variables** are added, their properly normalized sum **tends toward a normal distribution**.

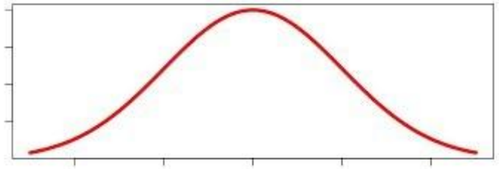
- This happens even if the original variables themselves are not normally distributed.

- The theorem is a key concept in probability theory because it implies that **probabilistic and statistical methods** that work for **normal distributions** can be **applicable to many problems** involving **other types of distributions**



Central Limit Theorem (Video)

The Normal Distribution…

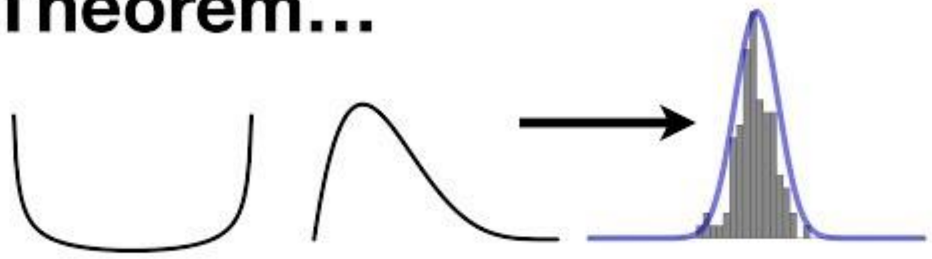


Clearly Explained!!!!!!

Sampling from a Statistical Distribution…

…Clearly Explained!!!

The Central Limit Theorem…



…Clearly Explained!!!

Hypothesis Testing

- A hypothesis test is a **technique** for using data to **validate or invalidate a claim about a population**. For example, a politician may claim that 80% of the people agree that volcanic bread is the best bread— is that really true?

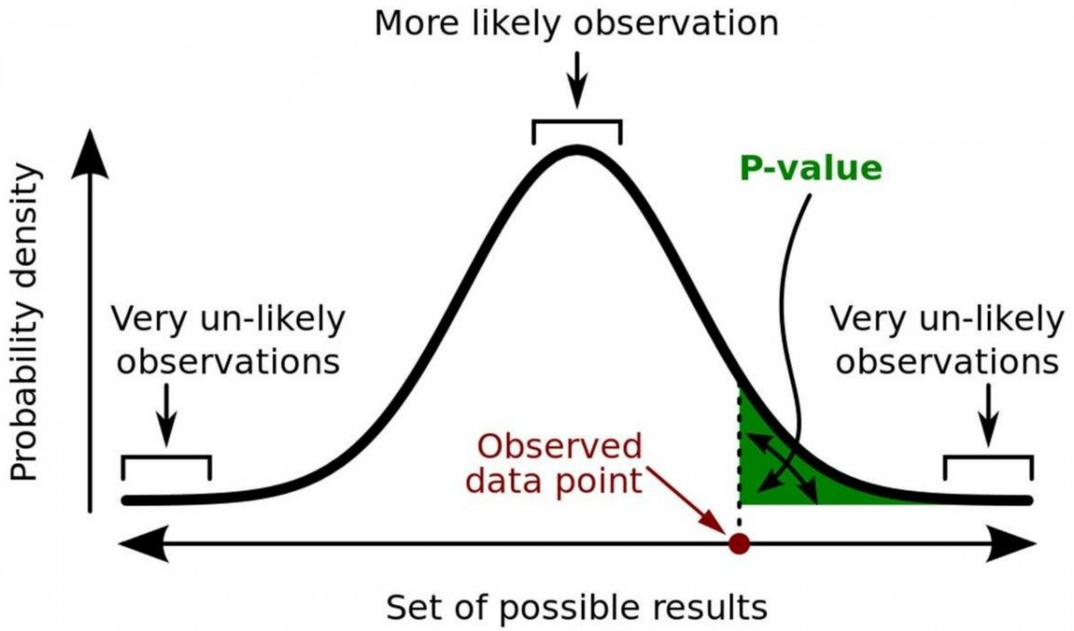
- The most common tested elements are:
  - The population mean
  - The population proportion
  - The difference in two population means or proportions (Is it true that the russians drink more vodka than their European counterparts? → Be careful w/ **sample size**

p - value

- When you perform a **hypothesis test** in statistics, a p-value helps you determine the **significance** of your results.

- The *alternative hypothesis* is the one you would believe if the **null hypothesis is concluded to be untrue**.

- A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis.
- A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis.
- Everything else is marginal. People report it to try to trick you into thinking they passed

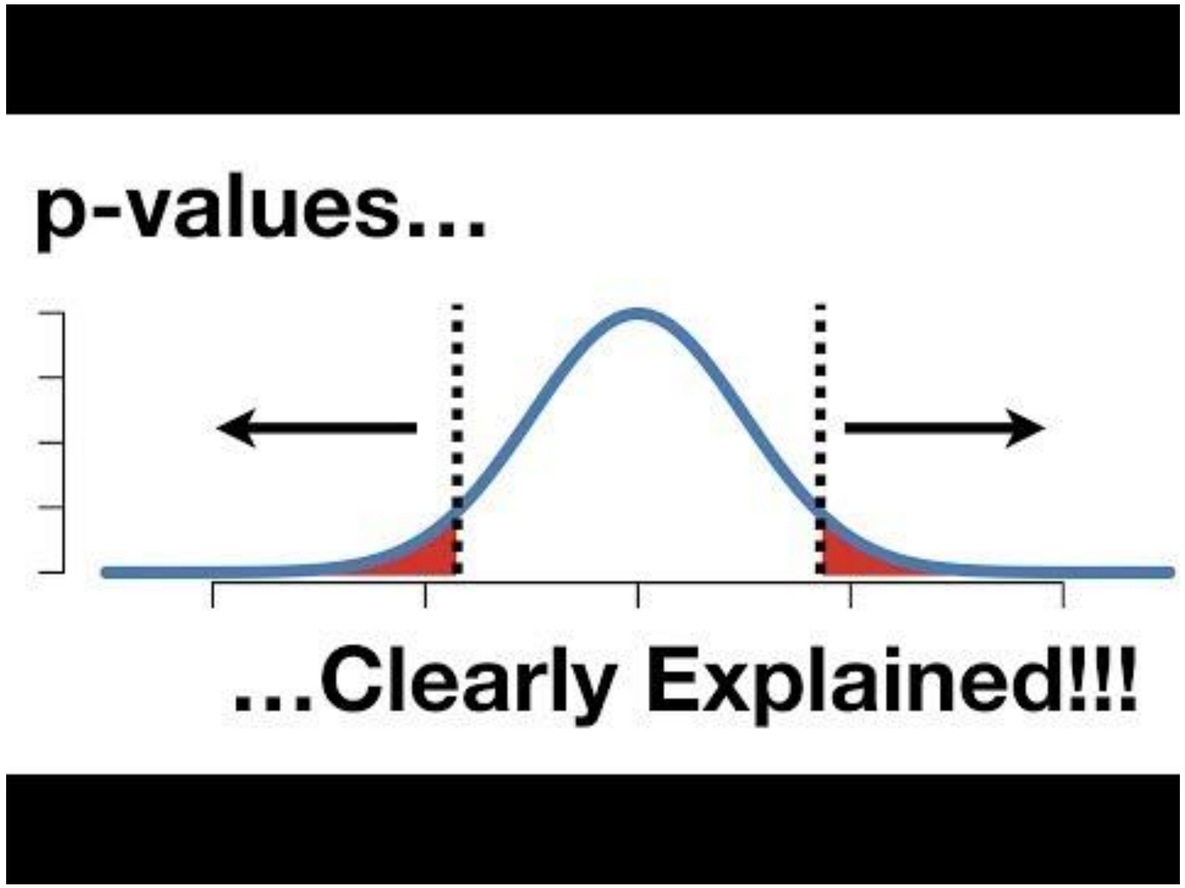


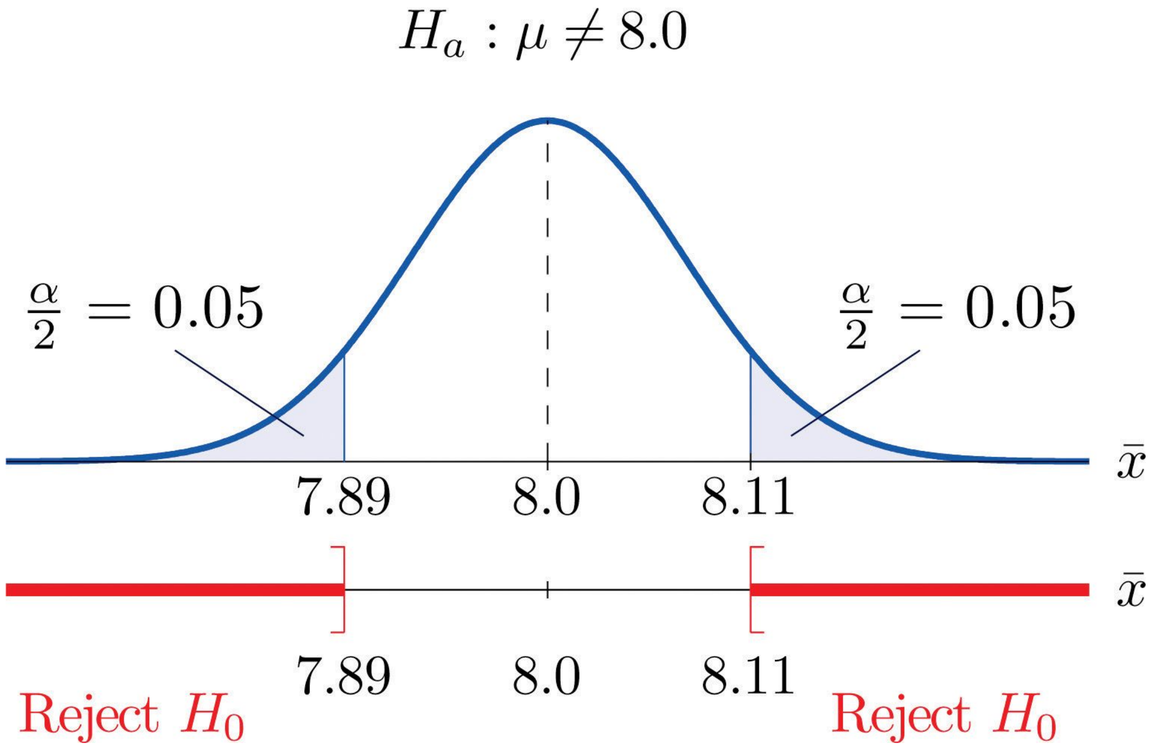
Summary

$p < 0.05$ → statistically significant difference

$p > 0.05$ → no statistically significant difference

Hypothesis Testing





Breath in

- Everything that you have done in life has prepared you to get to this exact point you are right now.

- Want to go from point **a** to point **a+1**?

Exercise **time**!

- Confidence interval notebook tutorial
- Hypothesis testing notebook tutorial
- Hypothesis testing implementation from towards data science