



# Correlation vs. Causation

# Meeting Name: 'CvC'



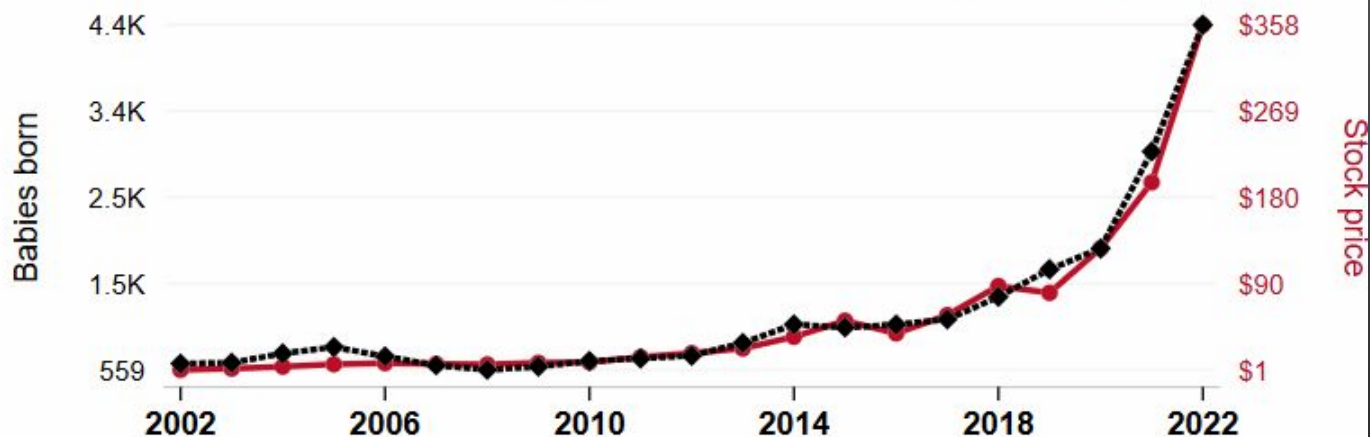
Today can be summed  
up in 5 words:  
Correlation does NOT  
imply causation!!!

## How Correlation and causation are Different:

### Popularity of the first name Walker

correlates with

### Old Dominion Freight Line's stock price (ODFL)



◆ Babies of all sexes born in the US named Walker · Source: US Social Security Administration

● Opening price of Old Dominion Freight Line (ODFL) on the first trading day of the year · Source: LSEG Analytics (Refinitiv)

2002-2022,  $r=0.992$ ,  $r^2=0.985$ ,  $p<0.01$  · [tylervigen.com/spurious/correlation/2159](https://tylervigen.com/spurious/correlation/2159)

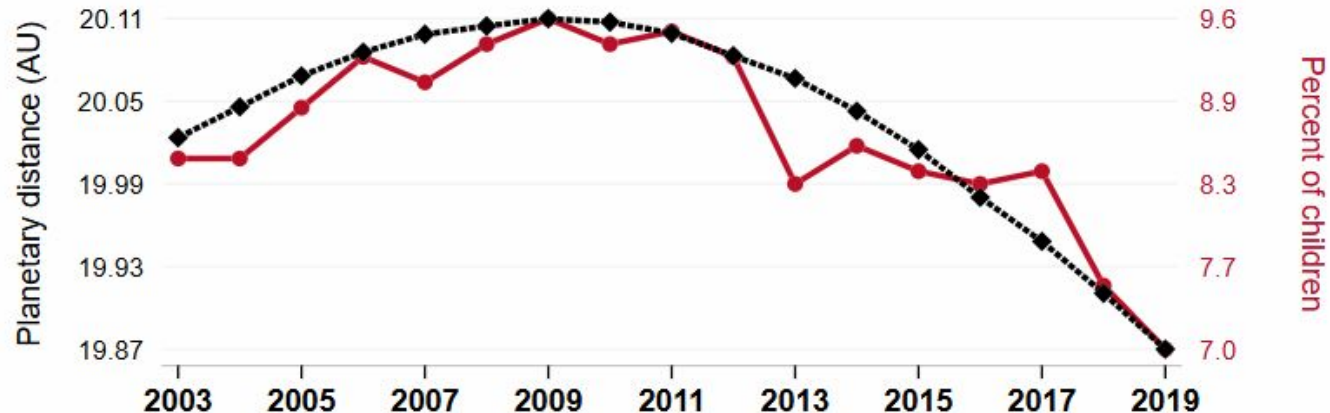
[tylervigen.com/spurious-us-correlations](https://tylervigen.com/spurious-us-correlations)

## How Correlation and causation are Different:

### The distance between Uranus and Earth

correlates with

### Asthma prevalence in American children



◆ The average distance between Uranus and Earth as measured on the first day of each month · Source: Caculated using Astropy

● Percent of children in the US with asthma · Source: National Center for Health Statistics

2003-2019,  $r=0.932$ ,  $r^2=0.869$ ,  $p<0.01$  · [tylervigen.com/spurious/correlation/2592](http://tylervigen.com/spurious/correlation/2592)

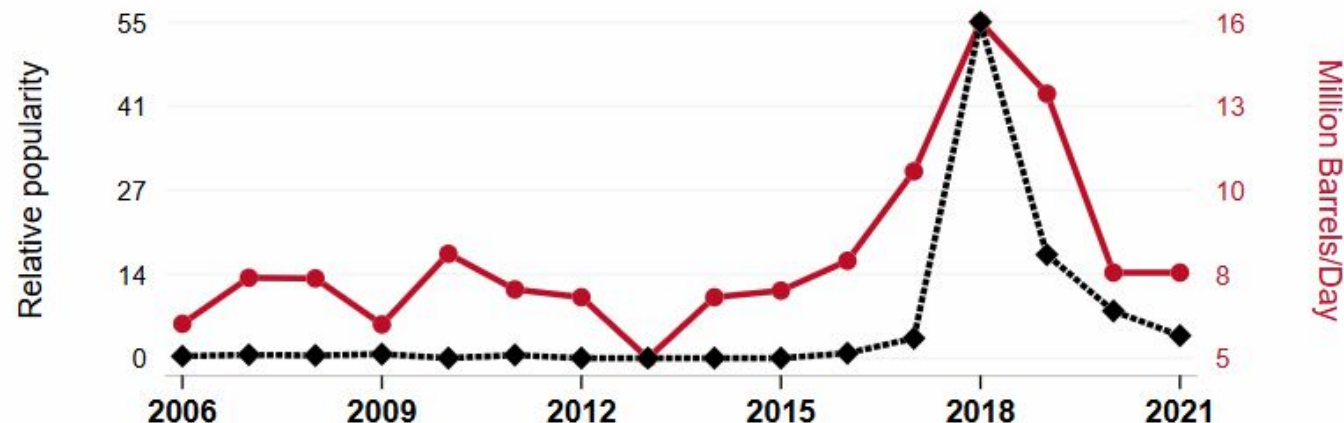
[tylervigen.com/spurious-us-correlations](http://tylervigen.com/spurious-us-correlations)

## How Correlation and causation are Different:

### Popularity of the 'floss dance' meme

correlates with

### Jet fuel used in Kazakhstan



◆ Relative volume of Google searches for 'floss dance' (without quotes, in the United States) · Source: Google Trends

● Volume of jet fuel used consumed in Kazakhstan in millions of barrels per day · Source: Energy Information Administration

2006-2021,  $r=0.850$ ,  $r^2=0.723$ ,  $p<0.01$  · [tylervigen.com/spurious/correlation/5133](https://tylervigen.com/spurious/correlation/5133)

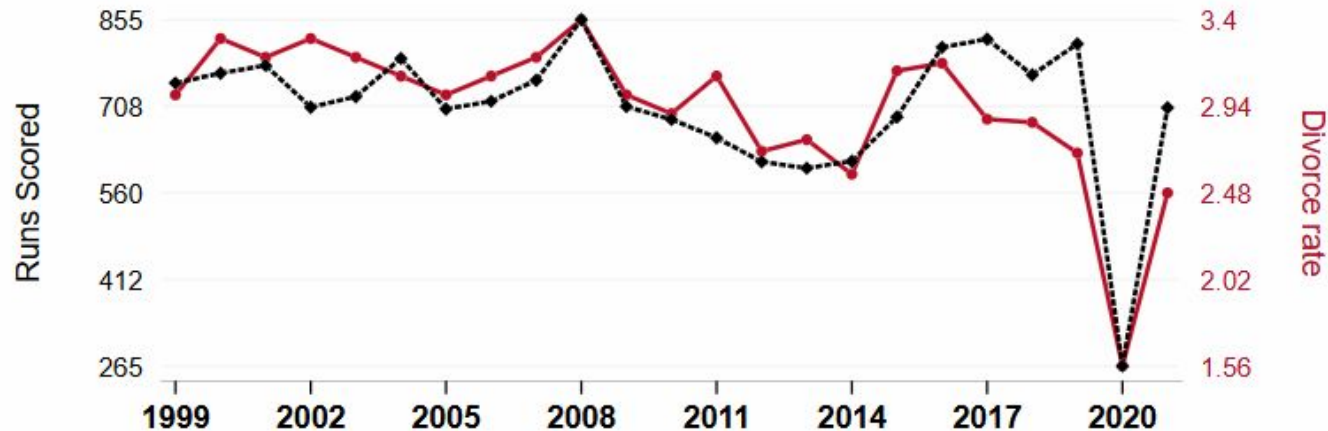
[tylervigen.com/spurious-correlations](https://tylervigen.com/spurious-correlations)

## How Correlation and causation are Different:

### Total Runs Scored by Chicago Cubs Team in National League (Central and East Division)

correlates with

### The divorce rate in Connecticut



◆--- Total Runs Scored by Chicago Cubs Team in National League (Central and East Division) · Source: Baseball Reference

●— The divorce rate in Connecticut · Source: CDC National Vital Statistics

1999-2021,  $r=0.799$ ,  $r^2=0.638$ ,  $p<0.01$  · [tylervigen.com/spurious/correlation/1173](https://tylervigen.com/spurious/correlation/1173)

[tylervigen.com/spurious-us-correlations](https://tylervigen.com/spurious-us-correlations)



# Algorithmic bias threatens legal fairness in AI era

SMALL BUSINESS

## How AI's Hidden Data Bias Can Impact Your Data Privacy Program (And What To Do About It)

By [Jodi Daniels](#), Forbes Councils Member.

for [Forbes Business Council](#), [COUNCIL POST](#) | Membership (fee-based)

Published Jun 12, 2025, 09:45am EDT

## Job Applicant's Algorithmic Bias Discrimination Lawsuit Survives Motion to Dismiss

Law and the Workplace on **July 25, 2024**



# QR Code for Lab:



# Hypothesis Testing

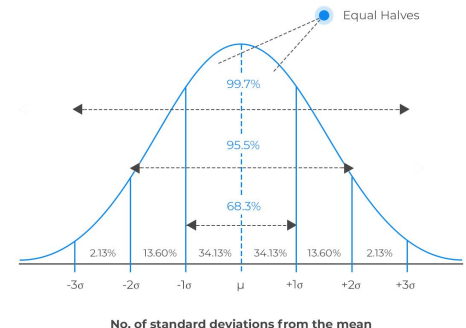
- $H_0$  - Null Hypothesis, there is no difference or change
  - Ex: The average height of a male in the US is equal to 70 inches.
- $H_a$  - Alternative Hypothesis, there is a significant difference or change
  - Ex: The average height of a male in the US is not equal to 70 inches.

# Hypothesis Testing (test statistic, z-value, t-value)

- Test statistic: quantity that is measured and compared
  - Mean, proportion, etc.
  - t-value, z-value, ...
- t-tests for means
- z-tests for proportion
- p-value: probability that the null hypothesis would produce the observed test statistic
- $\alpha$ : significance level
- Critical value: cutoff for test statistic



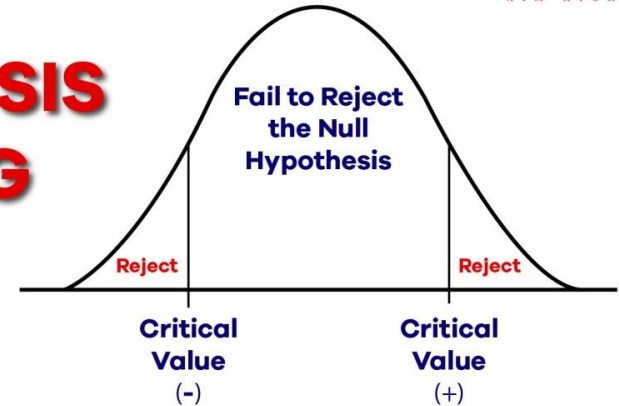
Shape of the normal distribution



# Drawing conclusions from a hypothesis test

- Use normal table or calculator to find p-value from the test statistic
- If  $p\text{-value} < \alpha$ , reject  $H_0$
- Otherwise, you do NOT reject  $H_0$

## HYPOTHESIS TESTING



## Type I and Type II errors

Type I error: You incorrectly reject the null hypothesis

Type II error: You incorrectly accept the null hypothesis

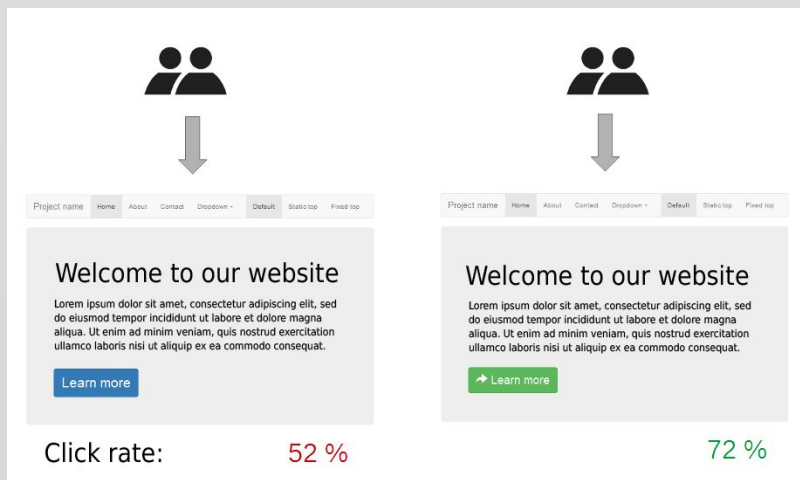
|                                 | Fail to Reject $H_0$                  | Reject $H_0$                         |
|---------------------------------|---------------------------------------|--------------------------------------|
| $H_0$ is True                   | Correct Decision                      | <b>Type I Error</b> (False Positive) |
| $H_0$ is False ( $H_1$ is True) | <b>Type II Error</b> (False Negative) | Correct Decision                     |

**IMPORTANT:**

Hypothesis testing can  
find correlation but not  
causation

# A/B testing

- Application of two-sample hypothesis testing to user-experience research
  - Testing subject's response to two variants of something





## Resources

<https://www.tylervigen.com/spurious-correlation>

<https://statisticsbyjim.com/basics/correlation-vs-causation/>

<https://www.statology.org/hypothesis-testing/>

<https://www.scribbr.com/statistics/type-i-and-type-ii-errors/>

**Thank You**