

Introduction into Biostatistics

Anna Poetsch, Biotechnology Center, TU Dresden

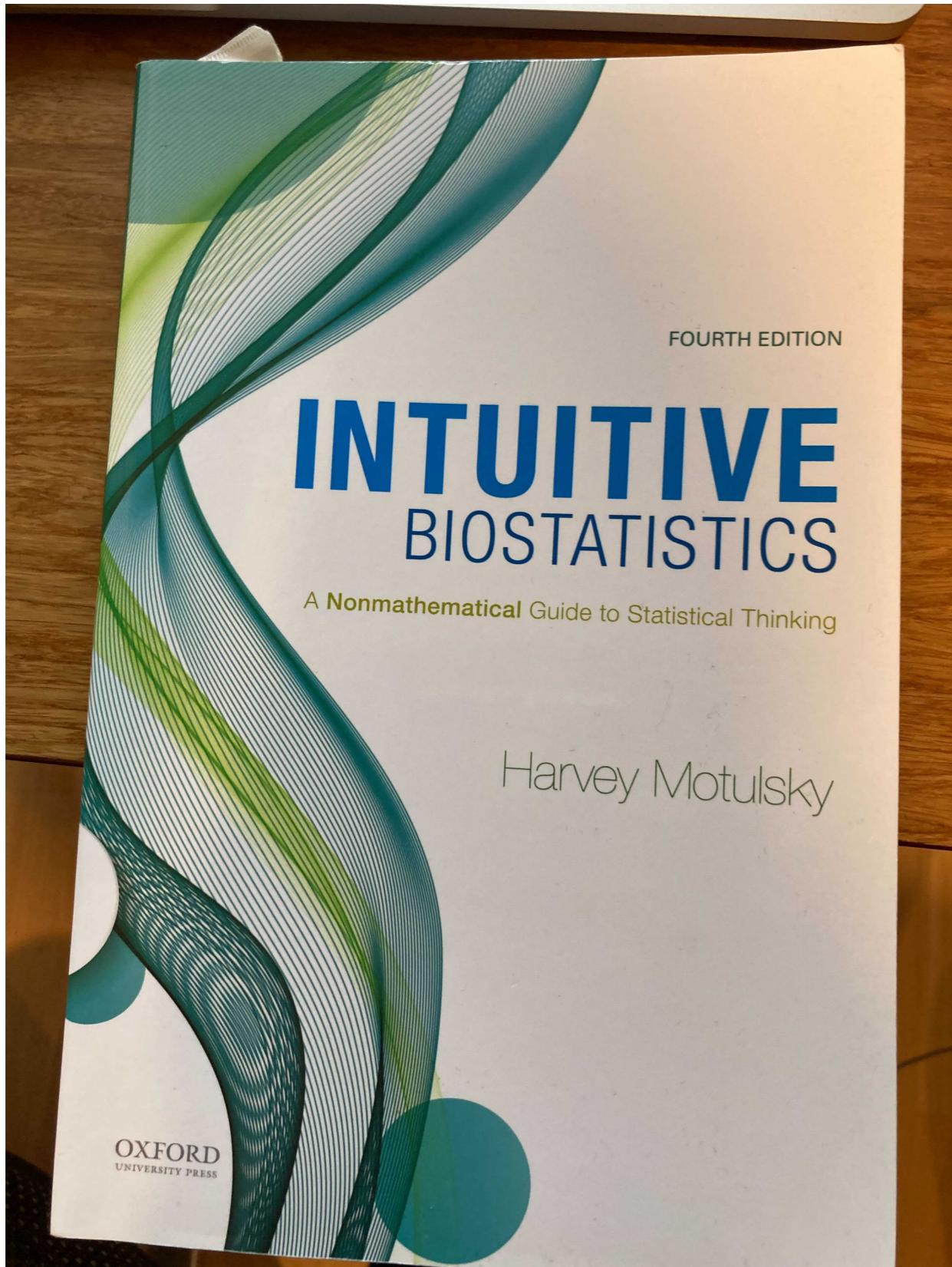
Organisation

- 16.5. Introduction to biostatistics
- 14.6. Hypothesis testing
- 21.6. Multiple comparisons and correlations
- 28.6. Big data, clustering, dimensionality reduction



by Melissa
Sanabria

Sources



& “the internet”

Introduction

- Intuition and bias
- Accuracy and precision
- Probability
- Confidence

Quiz

Please answer with 90% confidence!

- Martin Luther King Jr.'s age at death
- Length of the Danube [km]
- # of biomedical publications published by 2020 (listed in PubMed)
- # countries in Europe
- Diameter of the moon [km]
- # cars produced in 2020
- Year Mozart was born
- Distance Dresden-Tokyo [km]
- Gestation period of a Asian Elefant [days]
- Height of the Frauenkirche [m]

Quiz

Please answer with 90% confidence!

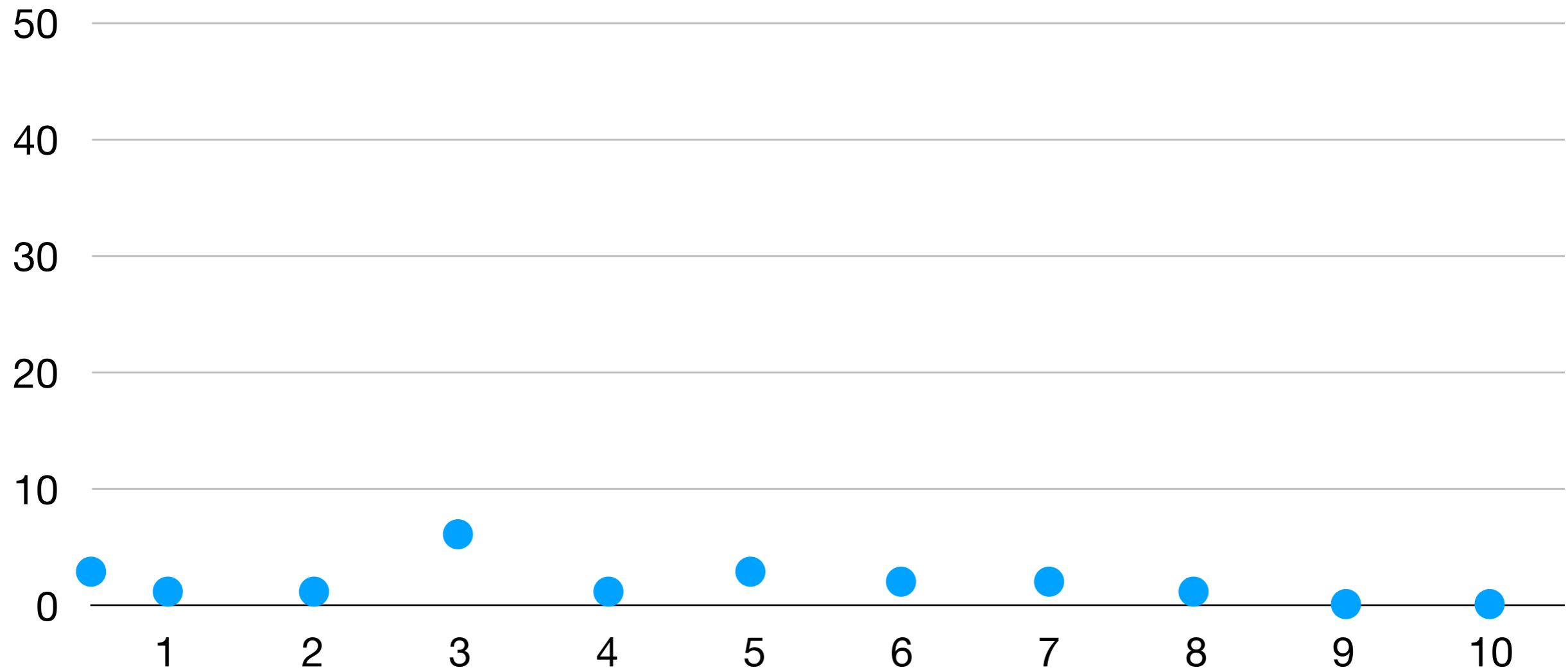
- Martin Luther King Jr.'s age at death **39**
- Length of the Danube [km] **2.850**
- # of biomedical publications published by 2020 (listed in PubMed) **31.563.992**
- # countries in Europe **51**
- Diameter of the moon [km] **3.476**
- # cars produced in 2020 **78 Mio**
- Year Mozart was born **1756**
- Distance Dresden-Tokyo [km] **9.020**
- Gestation period of an Asian Elephant [days] **617**
- Height of the Frauenkirche [m] **91**

Please mark your count of right answers in

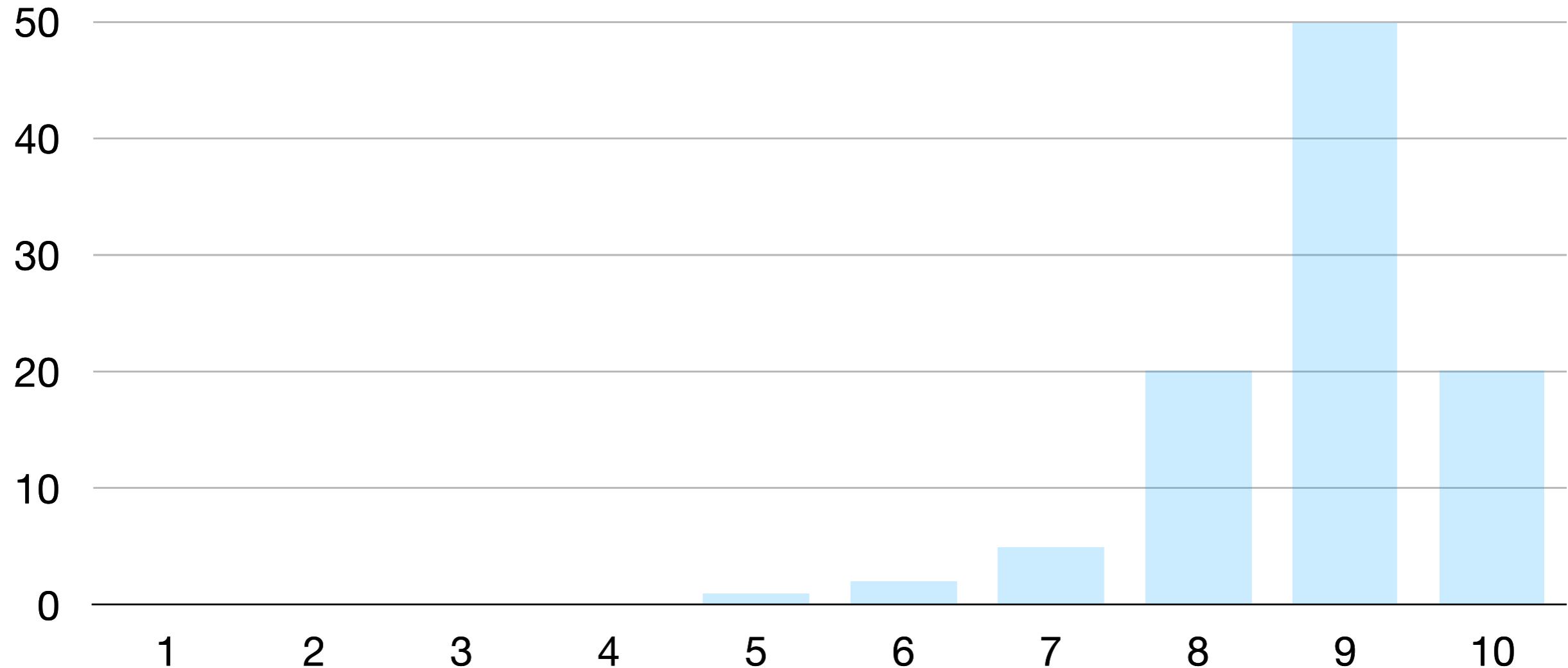
<https://dud-poll.inf.tu-dresden.de/stats/>

(No need to add your name)

Stats



Expected Stats

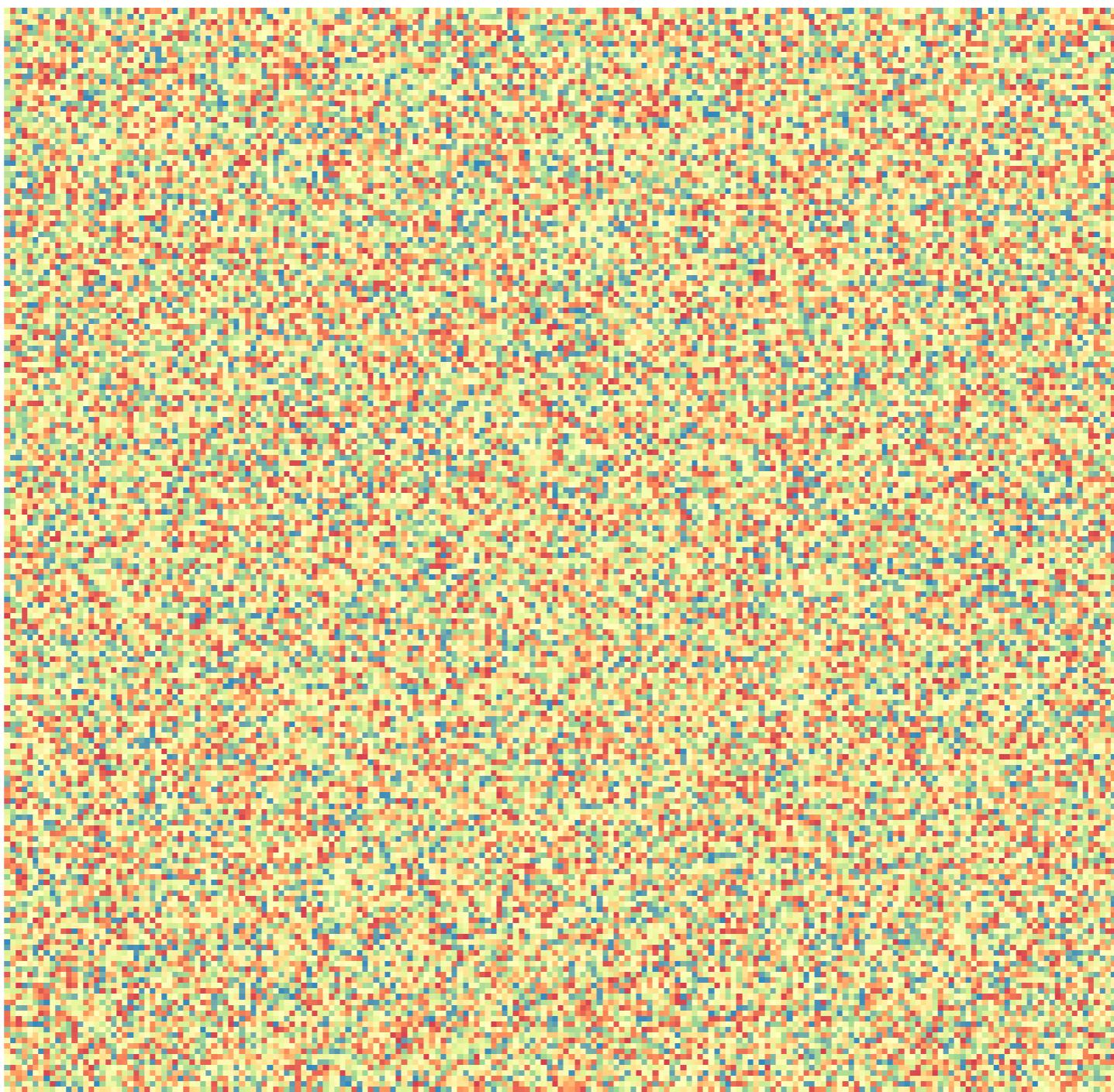


We tend to fool ourselves!

- We tend to be overconfident

- We tend to be overconfident
- We tend to jump to conclusions

- We tend to be overconfident
- We tend to jump to conclusions
- We see patterns in random data



- We tend to be overconfident
- We tend to jump to conclusions
- We see patterns in random data
- We don't realise that coincidences are common

Enzo Ferrari



Mesut Özil



Died in 1988

Born in 1988

- We tend to be overconfident
- We tend to jump to conclusions
- We see patterns in random data
- We don't realise that coincidences are common
- We don't expect variability to depend on sample size

Small countries and COVID19 incidence

<input type="checkbox"/> Tanzania	8.52
<input type="checkbox"/> Micronesia (country)	8.69
<input type="checkbox"/> Vanuatu	13.02
<input type="checkbox"/> Samoa	15.12
<input type="checkbox"/> Kiribati	16.74
<input type="checkbox"/> Solomon Islands	29.12

Cumulative incidence/ 1M inhabitants
<https://ourworldindata.org>; 31.05.2021

Small countries and COVID19 incidence

<input type="checkbox"/> Tanzania	8.52
<input type="checkbox"/> Micronesia (country)	8.69
<input type="checkbox"/> Vanuatu	13.02
<input type="checkbox"/> Samoa	15.12
<input type="checkbox"/> Kiribati	16.74
<input type="checkbox"/> Solomon Islands	29.12

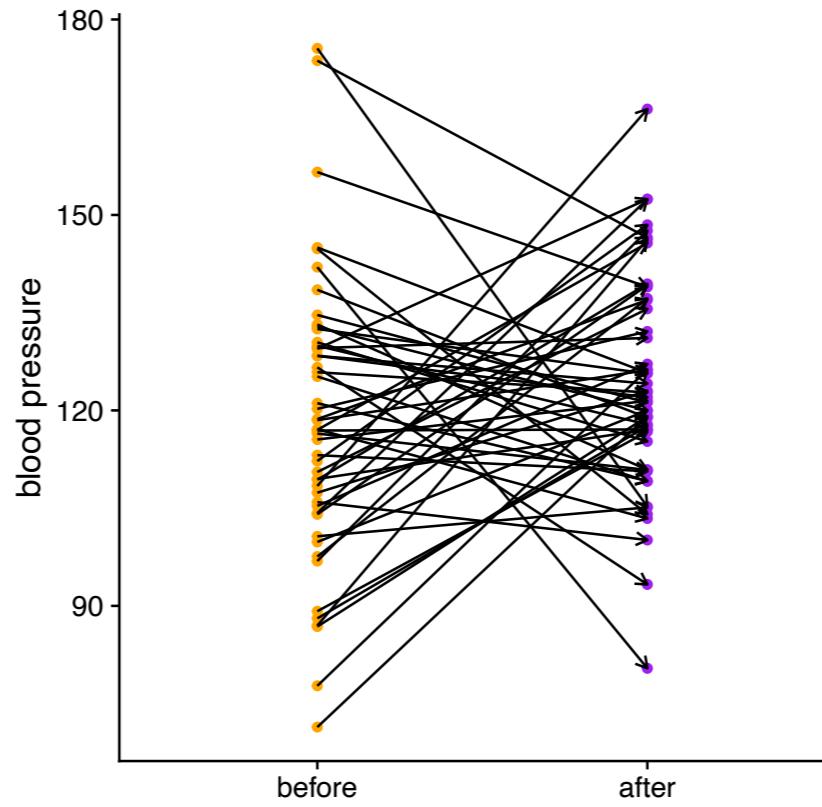
<input type="checkbox"/> Andorra	177,221.25
<input type="checkbox"/> Montenegro	158,591.03
<input type="checkbox"/> Czechia	155,118.29
<input type="checkbox"/> San Marino	149,949.91
<input type="checkbox"/> Bahrain	139,961.44
<input type="checkbox"/> Slovenia	122,000.62

The extremes are likely linked to small sample sizes!

- We tend to be overconfident
- We tend to jump to conclusions
- We see patterns in random data
- We don't realise that coincidences are common
- We don't expect variability to depend on sample size
- We are fooled by multiple comparisons

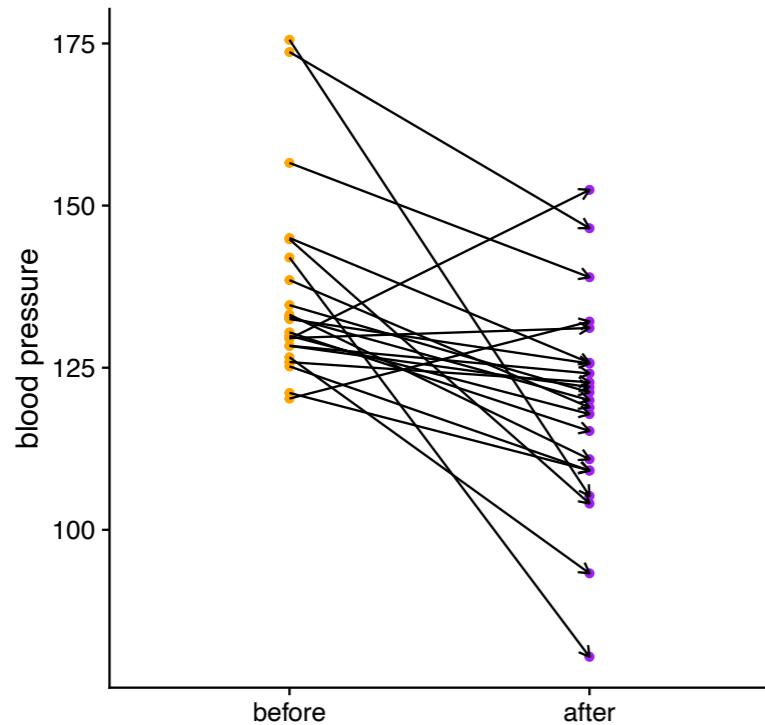
- We tend to be overconfident
- We tend to jump to conclusions
- We see patterns in random data
- We don't realise that coincidences are common
- We don't expect variability to depend on sample size
- We are fooled by multiple comparisons
- We intuitively follow logic that is in fact dictated by regression to the mean

Regression to the mean

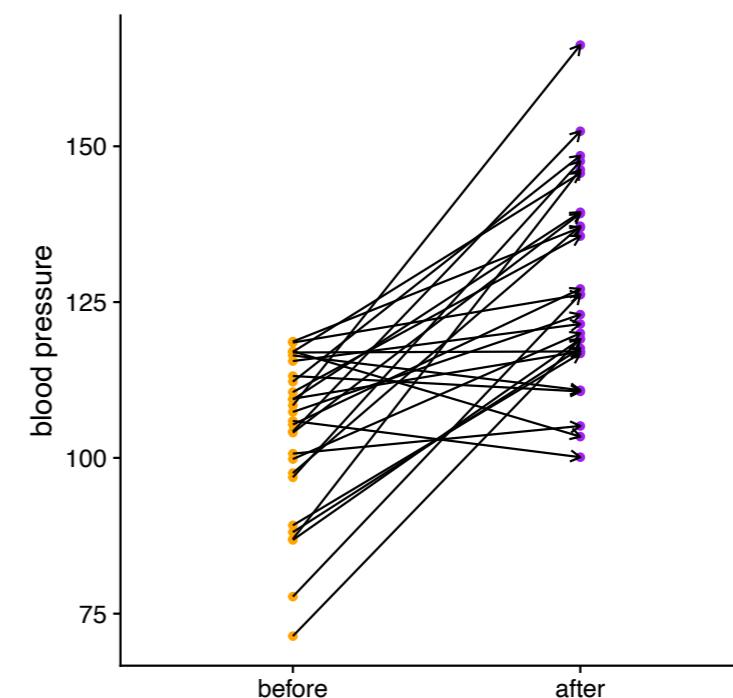


Using a drug to lower blood pressure

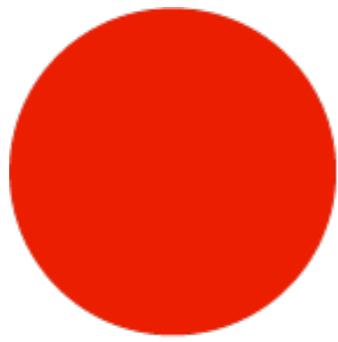
Patient selection for high blood pressure



The rest



Regression to the mean or: The curse of being on the cover



**USA
TODAY
SPORTS
WEEKLY™**
A GANNETT COMPANY



By Gannett Corporation - www.usatoday.com, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=48586574>

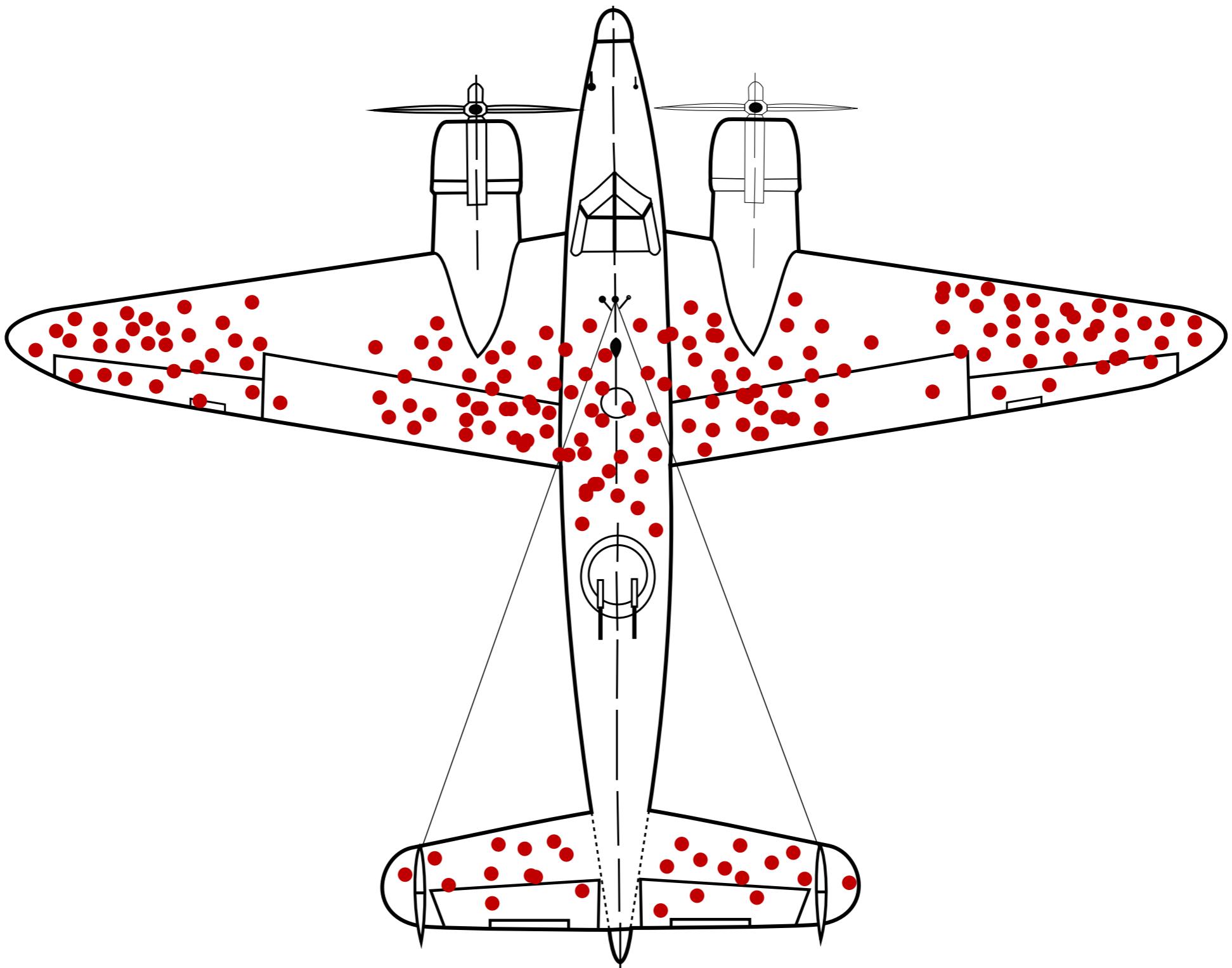
By All-Pro Reels - <https://www.flickr.com/photos/joeglo/51684898697/>, CC BY-SA 2.0, <https://commons.wikimedia.org/w/index.php?curid=112650621>

- We tend to be overconfident
- We tend to jump to conclusions
- We see patterns in random data
- We don't realise that coincidences are common
- We don't expect variability to depend on sample size
- We are fooled by multiple comparisons
- We intuitively follow logic that is in fact dictated by regression to the mean
- We are biased

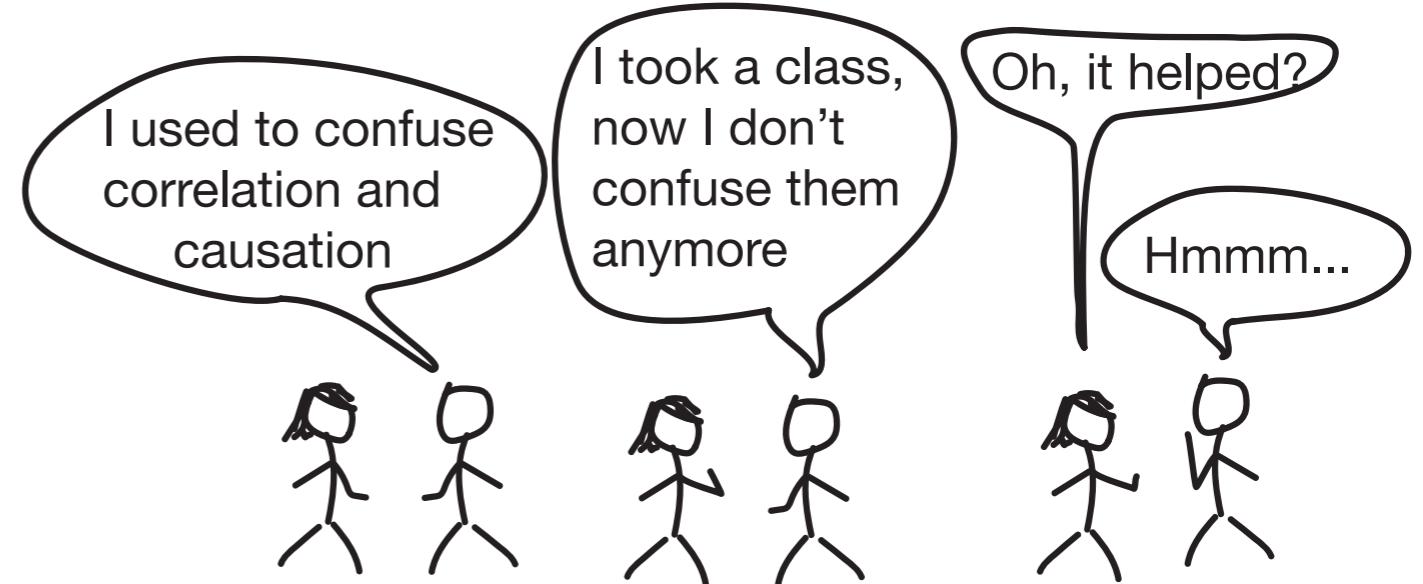
Representation bias



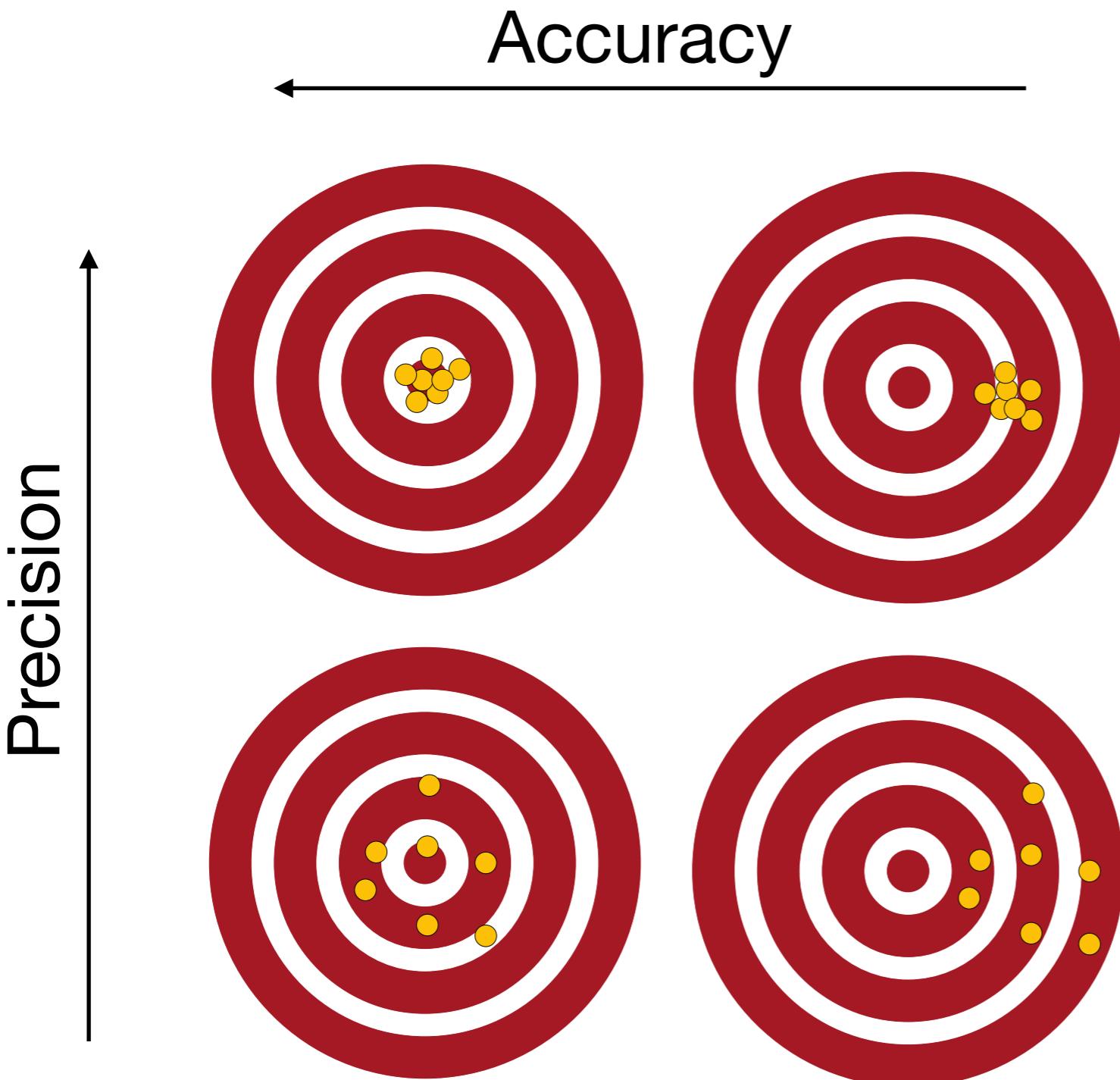
Survivorship bias



- We tend to be overconfident
- We tend to jump to conclusions
- We see patterns in random data
- We don't realise that coincidences are common
- We don't expect variability to depend on sample size
- We are fooled by multiple comparisons
- We intuitively follow logic that is in fact dictated by regression to the mean
- We are biased
- We confuse correlation with causation



Accuracy and precision



"Accuracy" is the ability to hit a target,
"Precision" is the ability to achieve the same results over and over.

You can be accurate without being precise and precise without being accurate.

We crave easy explanations,
follow intuitions,
and aim for certainty.

Statistics offers probabilities!

Probability

Probabilities range from 0 to 1

They are displayed as proportions or in %

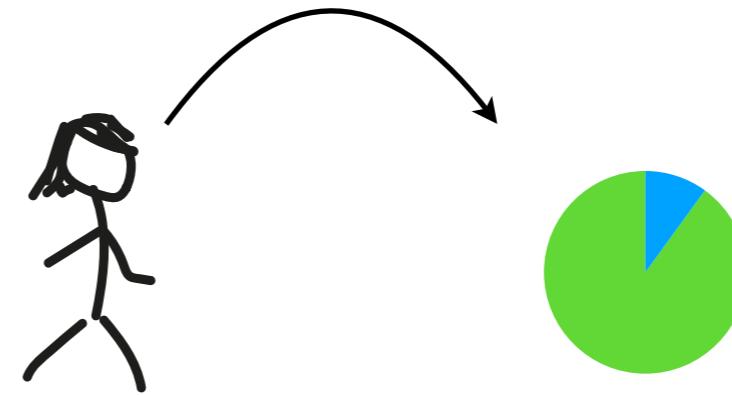
“Used to quantify a prediction on future events,

describe past events,

or the certainty of a belief. “

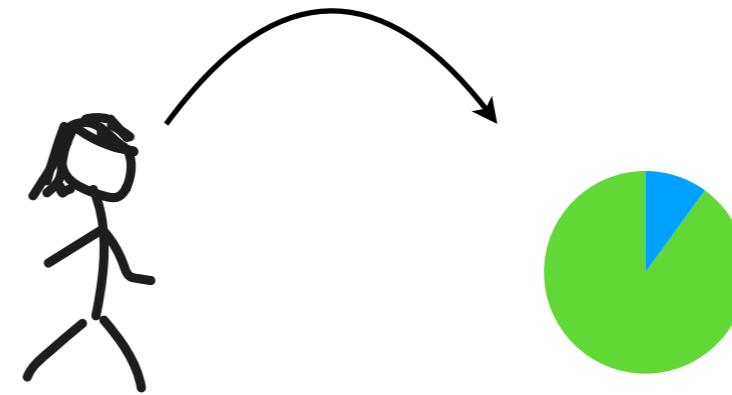
Views on probability

A model

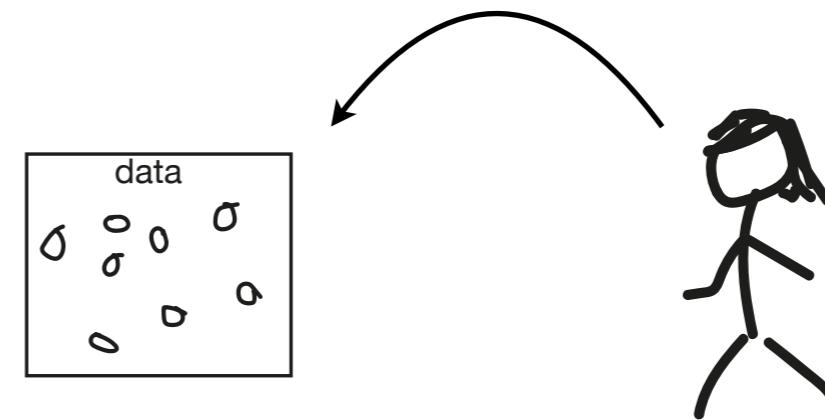


Views on probability

A model

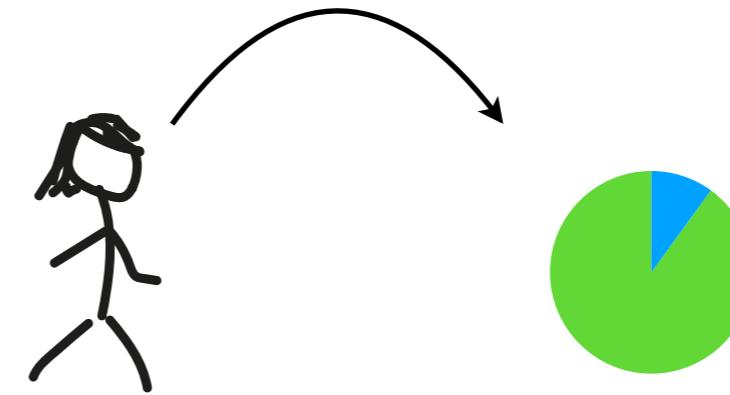


Data

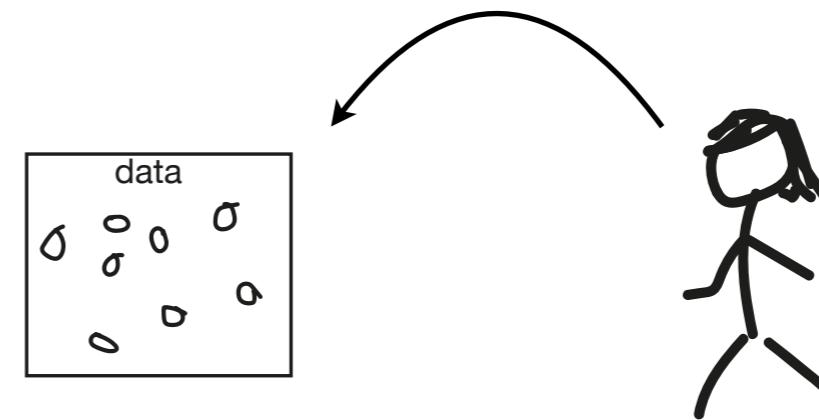


Views on probability

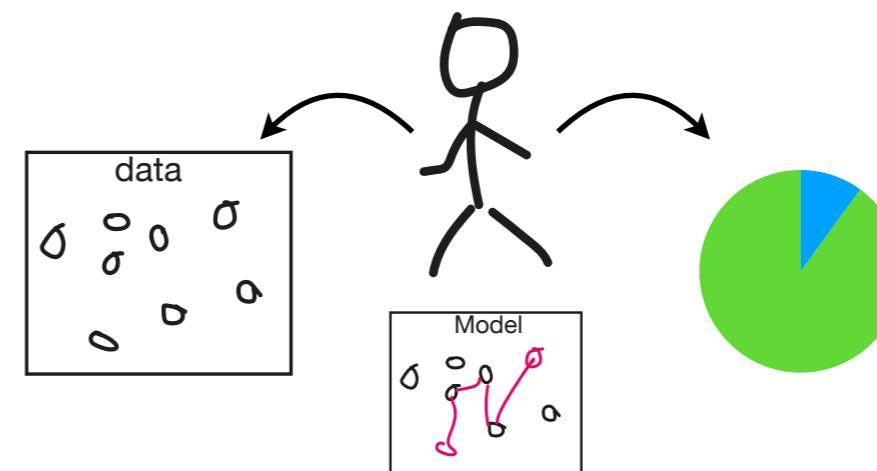
A model

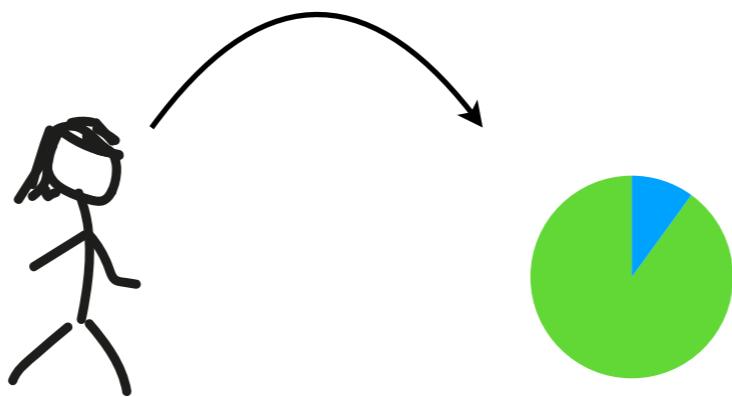


Data

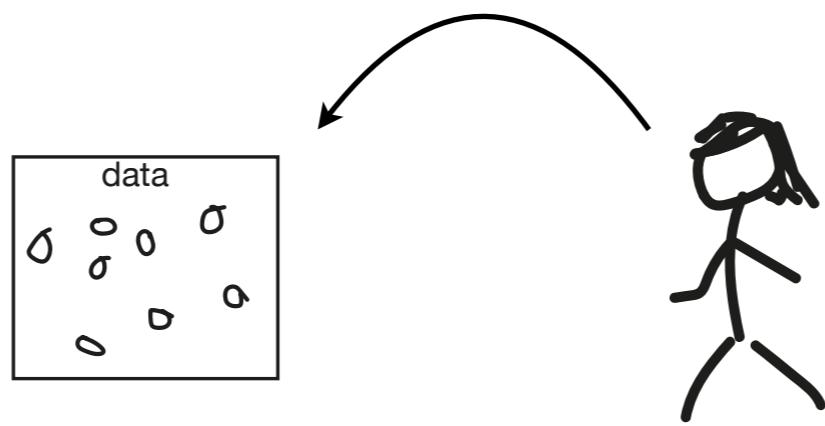


A model based on data

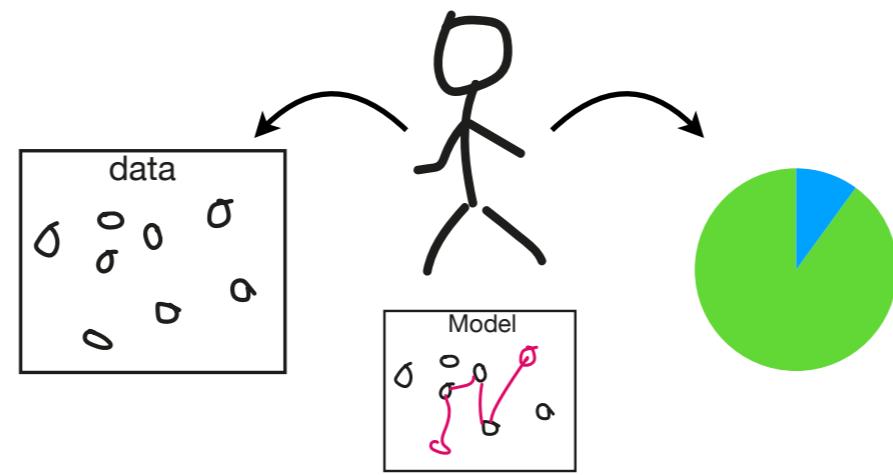




probability of male baby = 50%



probability of male baby (in 2011) = 51.7%



probability of male baby = it depends

Be aware of assumptions!

We are assuming that:

- there is only one baby
- gender is binary
- there are no differences between countries and ethnic groups
- sex ratios are equal over time and seasons
- there are no sex-selective abortions or miscarriages

Be aware of reversing probabilities!

Most COVID19 patients have a cough, it is still unlikely that someone who coughs in a tram will have COVID19

If you find that statistics books are boring, it does not mean that every boring book is about statistics

When you find a good cancer drug, you can kill your cell culture with it.

If your cells die, you might want to think for a moment before you ring Stockholm.

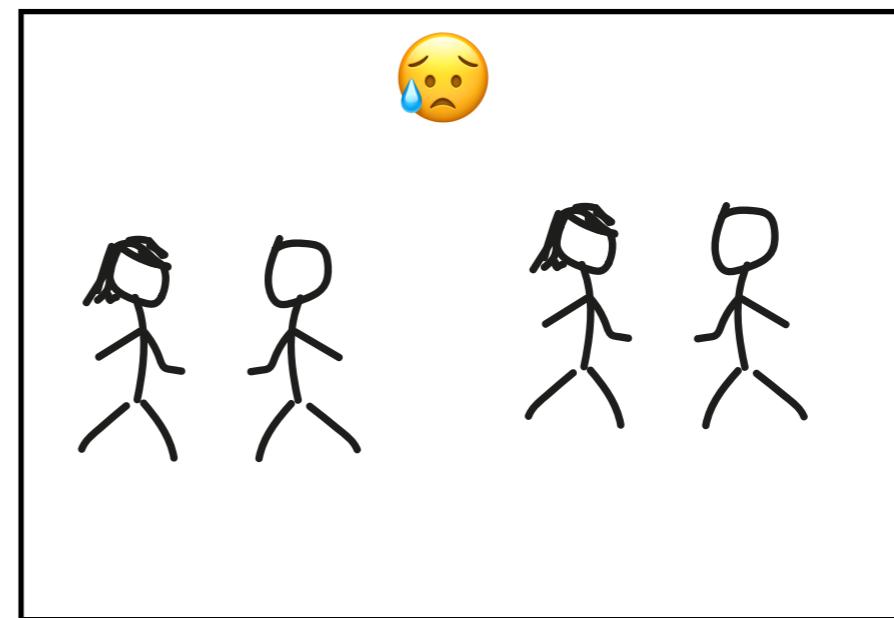
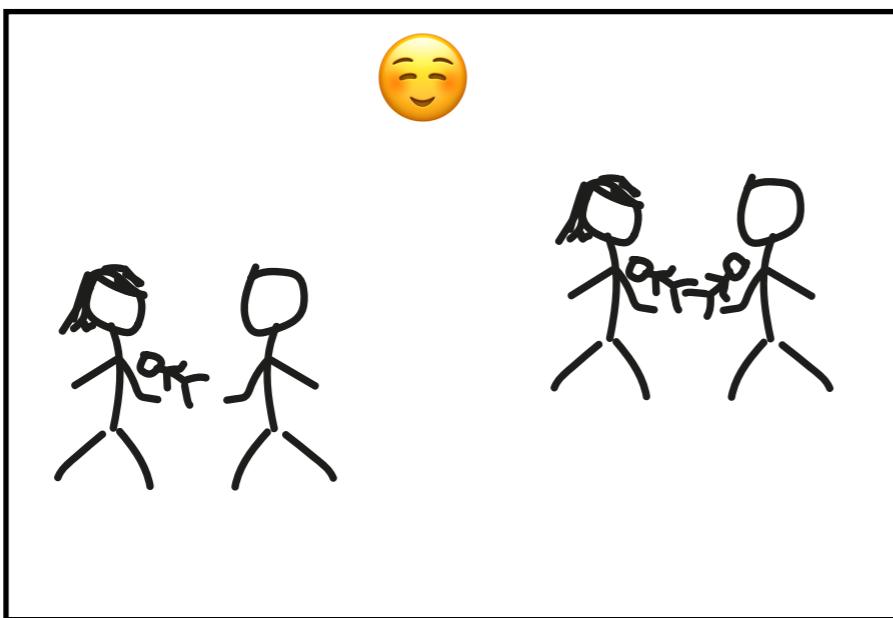
Don't assume probability has a memory



Clearly define the numerator and denominator

Example: Success rate in in-vitro-fertilisation clinic

Do we count the babies or the mothers?



4 trials



3 babies:



2 successful attempts:



What is the success rate?

Summary

- Statistics is not always intuitive
- Multiple testing
- Regression to the mean
- Bias
- Accuracy
- Precision
- Probability
- Confidence

Confidence

-> Jupyter notebook