

Winter semester 2019-2020, Mondays 9:30-14:00

Research Methods and Ethics

HOW TO LIE WITH STATISTICS?

Oslo Metropolitan University

2019.09.23, Pedro G. Lind

What is statistics?

“It is the practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample.”

Oxford English Dictionary

And it deals with a fundamental concept...

The concept of PROBABILITY

What is probability?

“The quality or state of being probable; the extent to which something is likely to happen or be the case.”

Oxford English Dictionary

*“The quality or state of being probable; the extent to which something **is likely** to happen or be the case.”*

Oxford English Dictionary

Why is probability so important?

Because the universe is uncertain...

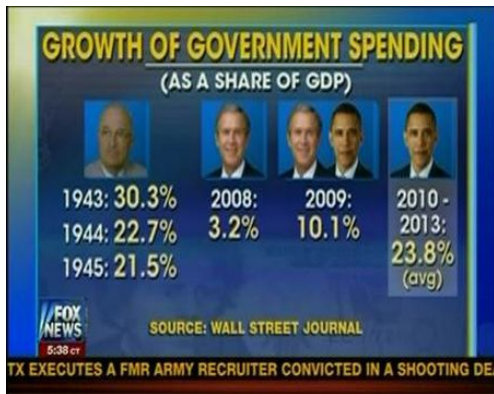
... and probability is a weapon for surviving in it!

- ▶ Important to know the weather tomorrow.
- ▶ Important to invest in the stock market.
- ▶ Important to evaluate risk in security systems.
- ▶ Important to handle big data in clouds.
- ▶

What do we do with probability and statistics?

- ▶ Probability is what we want to calculate for making decisions: How probable it is that our decision will not lead to a catastrophe?
- ▶ **PROBLEM:** we cannot measure probabilities directly.
- ▶ **SOLUTION:** We do statistics with data!
- ▶ **TAKE-AWAY:** From statistical analysis of real data we can estimate probabilities and make decisions!

The dark side of the force...



...7 tips to better lie with statistics...

Sources: “Lessons from how to lie with statistics”

available at <https://towardsdatascience.com/>

Tip 1 to better lie with statistics

Use big names to support your claims
... don't let a fact prove an argument wrong!

According to repeated nationwide surveys,

More Doctors Smoke **CAMELS** than any other cigarette!

Doctors in every branch of medicine were asked, "What cigarette do you smoke?" The brand named most was Camel!

You'll enjoy Camels for the same reasons so many doctors enjoy them. Camels have cool, mild, soothing, pack after pack, and a flavor unmatched by any other cigarette. Make this sensible test: Smoke only Camels for 30 days and see how well Camels please your taste, how well they suit your throat as your steady smoke. You'll see how enjoyable a cigarette can be!

THE DOCTORS' CHOICE IS AMERICA'S CHOICE!



ROBERT TRUMP says, "I smoke Camels. They taste right, they smoke easy and clean!"

WILL BOWEN says, "I am a great pleasure from Camels that I smoke every day!"

WILLIAM HILLMAN says, "Camels will get you and done. I've smoked 'em for years!"

For 30 days, test Camels in your "T-Zone" ("T" for Throat, "T" for Taste).

E=mc²



„Wir nutzen nur 10% unseres geistigen Potentials“
A. Einstein

Wie kann ich die in mir schlummernden Fähigkeiten und Energien ausschöpfen und meine Intelligenz voll nutzen?

Durch Einsteins Relativitätstheorie wissen wir mehr über die Energien dieses Universums. Aber was wissen wir schon über die in uns schlummernden Kräfte und Energien?

Warum nutzen wir nur 10% unserer geistigen Möglichkeiten? In dem Buch „Dianetik“ (jetzt auch als Video) von U. Ron Hubbard finden Sie die Antworten und ein tiefes Verstehen Ihres eigenen Verhaltens. Die Reaktionen anderer werden plötzlich klar. Aber noch mehr: Es eröffnet Ihnen den gesamten Horizont Ihrer Möglichkeiten – eine ausgeglichene Persönlichkeit mit einem positiven Blick in die Zukunft. Dieses Buch zeigt Ihnen, wie Sie die in Ihnen schlummernden Fähigkeiten und Energien besser aktivieren und voll ausschöpfen. Entdecken Sie, was in Ihnen steckt! Dieses faszinierende Buch bestellen Sie mit unschätzbaren Antworten!

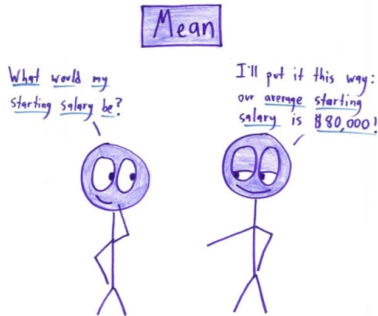
Auch als Video!

MIT 118 S. 82 farbigen Illustrationen

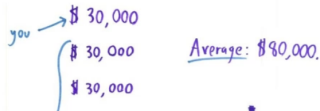
Tip 2 to better lie with statistics

Use the mean as an authority!

HOW TO LIE WITH STATISTICS



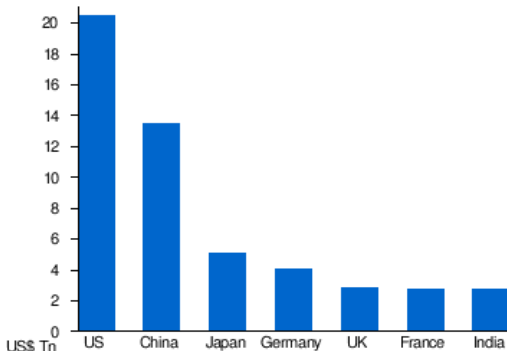
HOW TO LIE WITH STATISTICS



Tip 3 to better lie with statistics

Use numbers without comparing them with a reference!

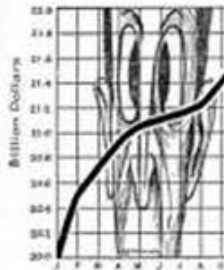
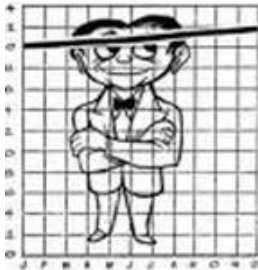
Estimated GDP for Norway in 2019: ~ 0.427 US\$Trillion



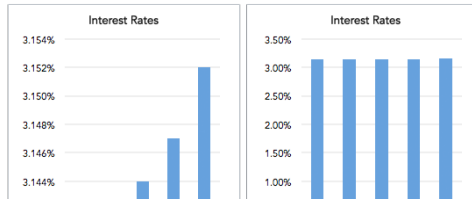
(Percentual variation would be the correct approach...)

Tip 4 to better lie with statistics

Show your data according the “needs”...



Same Data, Different Y-Axis



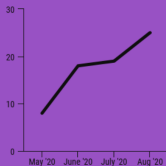
Tip 5 to better lie with statistics

Select data that best serve your purposes

3

CHERRY PICKING DATA

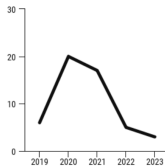
Writers may only include certain data points on their graphs to reinforce their narratives. This can create a false impression of the data.



MISLEADING

- Only a few months out of the year are graphed, depicting an upward trends

VS



ACCURATE

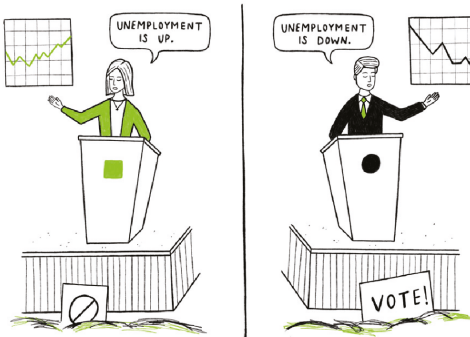
- A much wider date range is graphed, revealing an overall downward trend
- This graphs shows the bigger picture

Tip 5 to better lie with statistics

Select the data that best serve your purposes

Use *special* samples

... and confirm them with your own (always biased) experience...

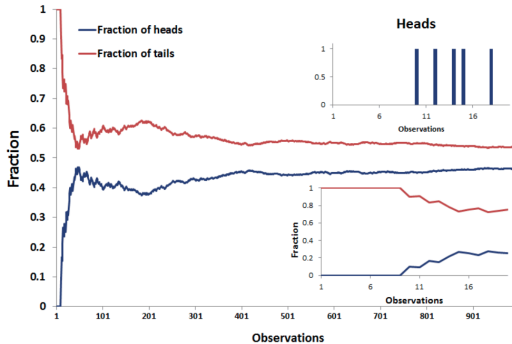


Tip 5 to better lie with statistics

Select the data that best serve your purposes

Use *small* samples

Errors will be so large that any outcome is possible!



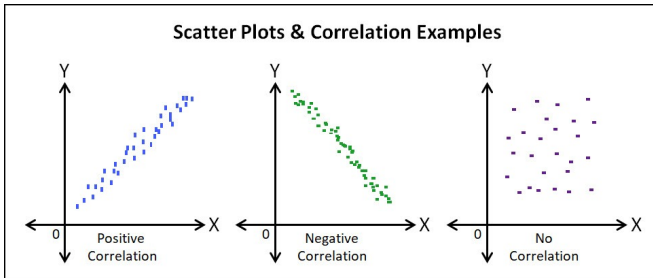
Tip 6 to better lie with statistics

Use CORRELATION as a synonym of CAUSATION

- ▶ Assume X and Y are correlated.

- ▶ **What does that mean?**

It *only* means that they increase together or decrease together (positive), or one goes up as the other does down (negative).

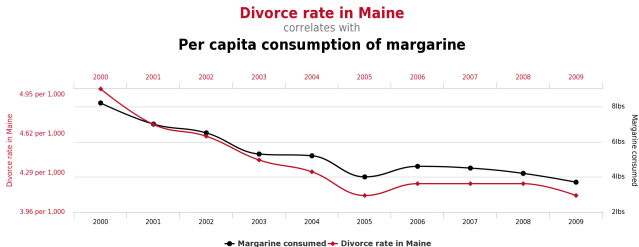


Tip 6 to better lie with statistics

Use CORRELATION as a synonymous for CAUSATION

Three different explanations are possible:

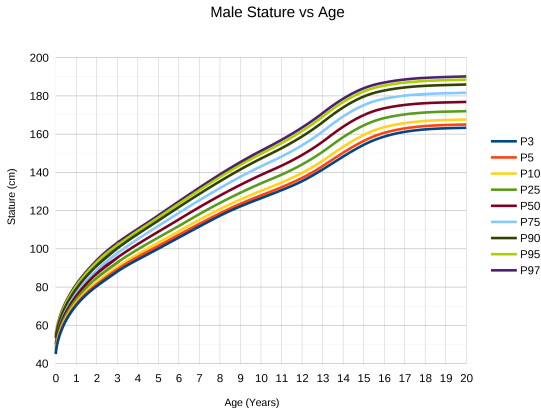
- ▶ One **causes** the other ($X = \text{clouds}$ and $Y = \text{rain}$).
- ▶ Both are caused by one **third thing** ($X = \text{sun glasses}$ and $Y = \text{ice-creams}$).
- ▶ Both things are completely **unrelated**.



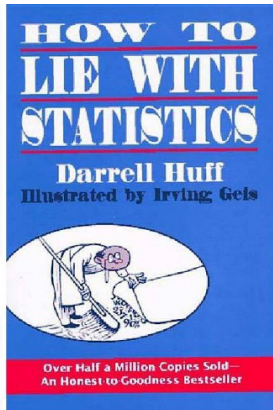
Tip 7 to better lie with statistics

Extrapolate as much as you can!

- ▶ In statistics no relationship lasts forever!
- ▶ Extrapolations are **very dangerous** in a serious statistical analysis.



How to lie with statistics?



Take-home messages part I

If you want to lie to other:

- ▶ As future producers of tables and graphs, follow these tips for misleading your statistical results to your peers or costumers.
- ▶ But to succeed on that, **develop your “data literacy”**: learn how to properly read, understand and explain data.

Take-home messages part II

If you do not want to be tricked by other:

- ▶ Do not forget: statistical models are possible interpretations of uncertain data, sometimes designed to inform, sometime to persuade. Try to **distinguish between them**.
- ▶ Be as **objective** as possible while analyzing your data.
- ▶ Be **critical** about the data-gathering process.
- ▶ Check the values and their ranges.
- ▶ Report the **confidence intervals** (details next week...)