

AdTrans 2023-1-PL01-KA220-HED-000158917



Methods of sociology and demography

Data analysis and interpretation

Kristyna Rybova

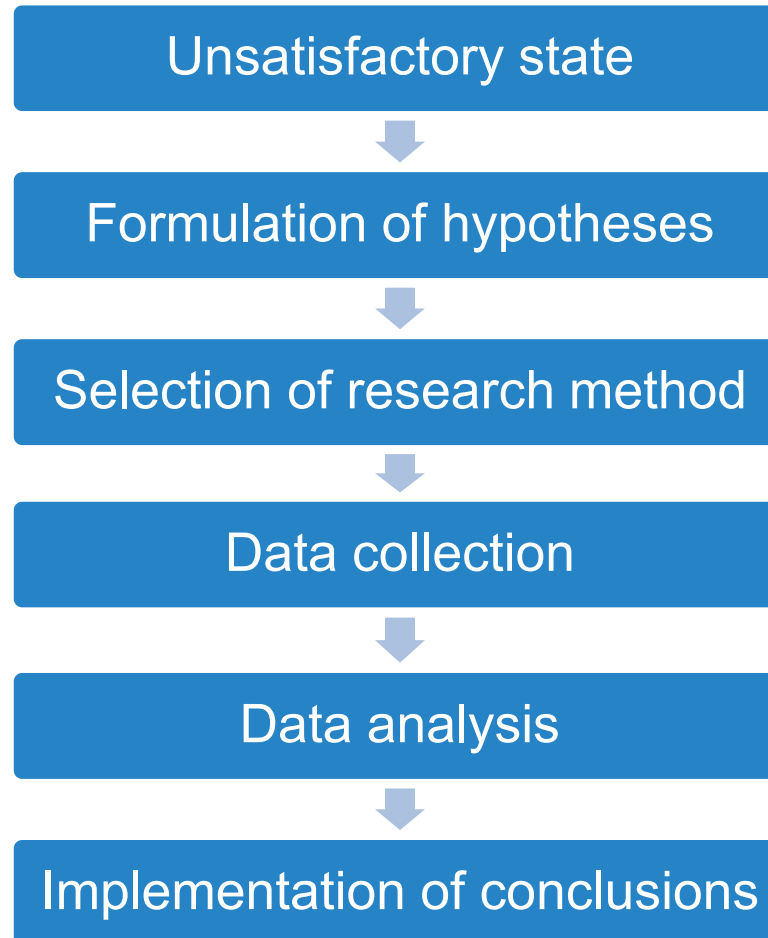


Co-funded by
the European Union



Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

RESEARCH SCHEME



DATA ANALYSIS

The form of the data can limit the types of analysis → think about data analysis before data collection to avoid collecting “bad” data (e.g. insufficient scope, unsuitable data collection method, ...)

Why analyze data?

- Summarizing large amounts of data using descriptive statistics
- Explore data with visualization
- Draw conclusions about a population from observations of a sample of the population
- Create models of how the real world works

SUMMARIZING LARGE AMOUNTS OF DATA USING DESCRIPTIVE STATISTICS

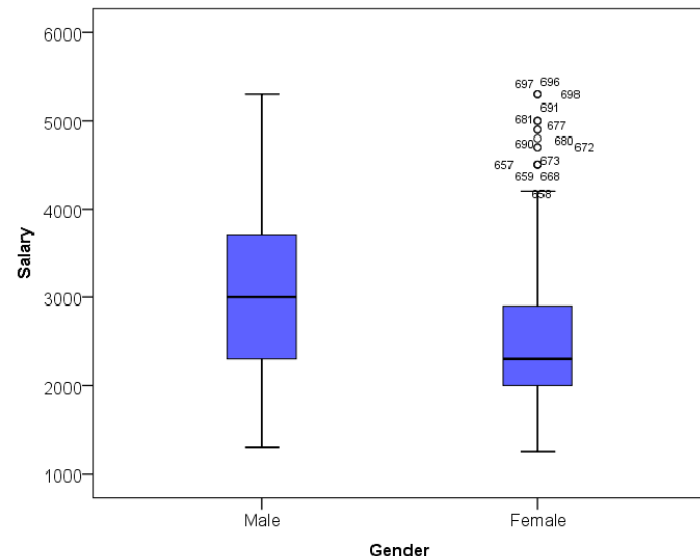
Average values

Description of variability

- How representative are the mean values?
- There are differences in data between groups

Outliers

- Some analysis methods are extremely sensitive

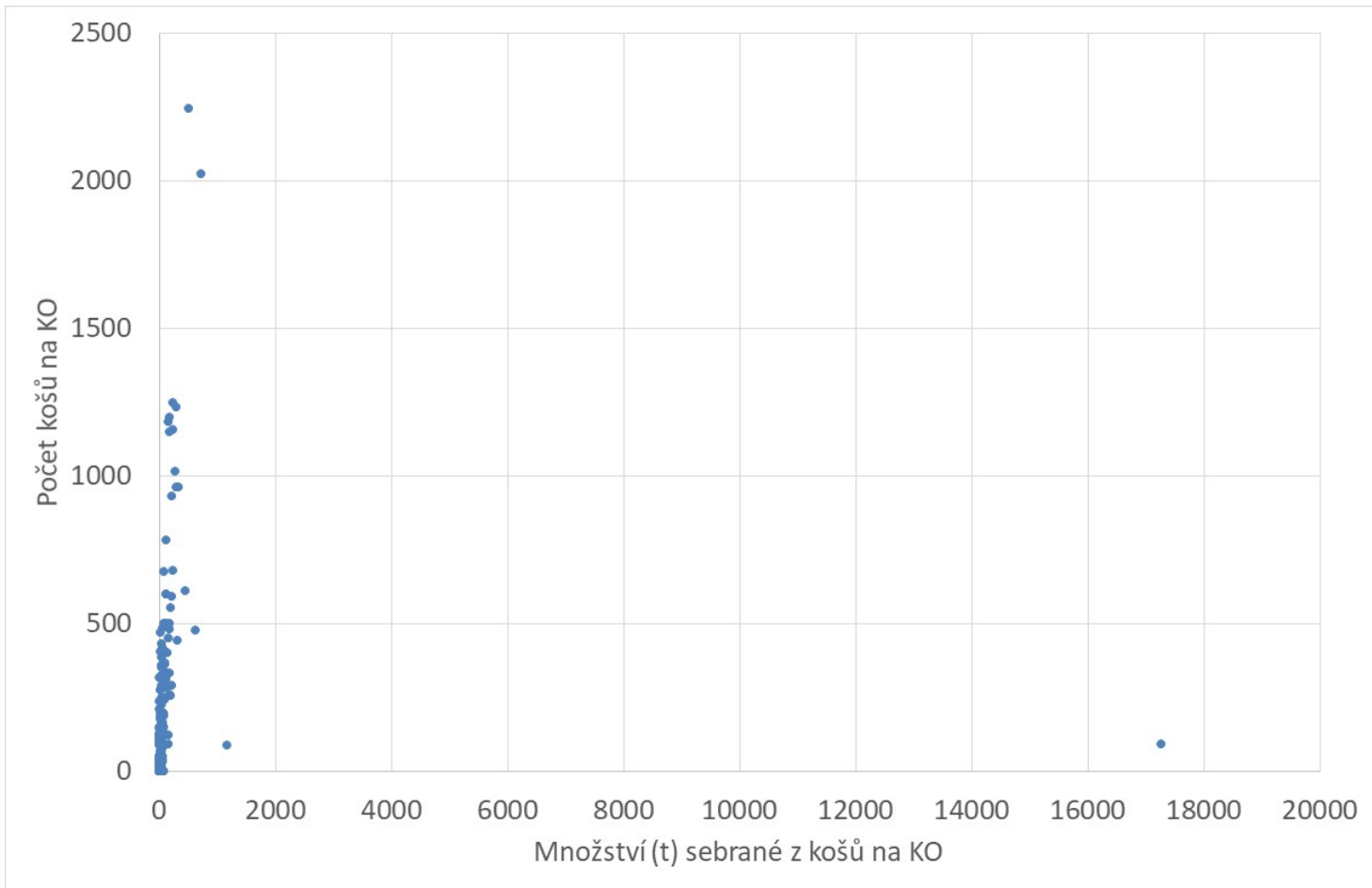


https://www.researchgate.net/publication/275329571_Quantitative_Methods_for_Communication_Students

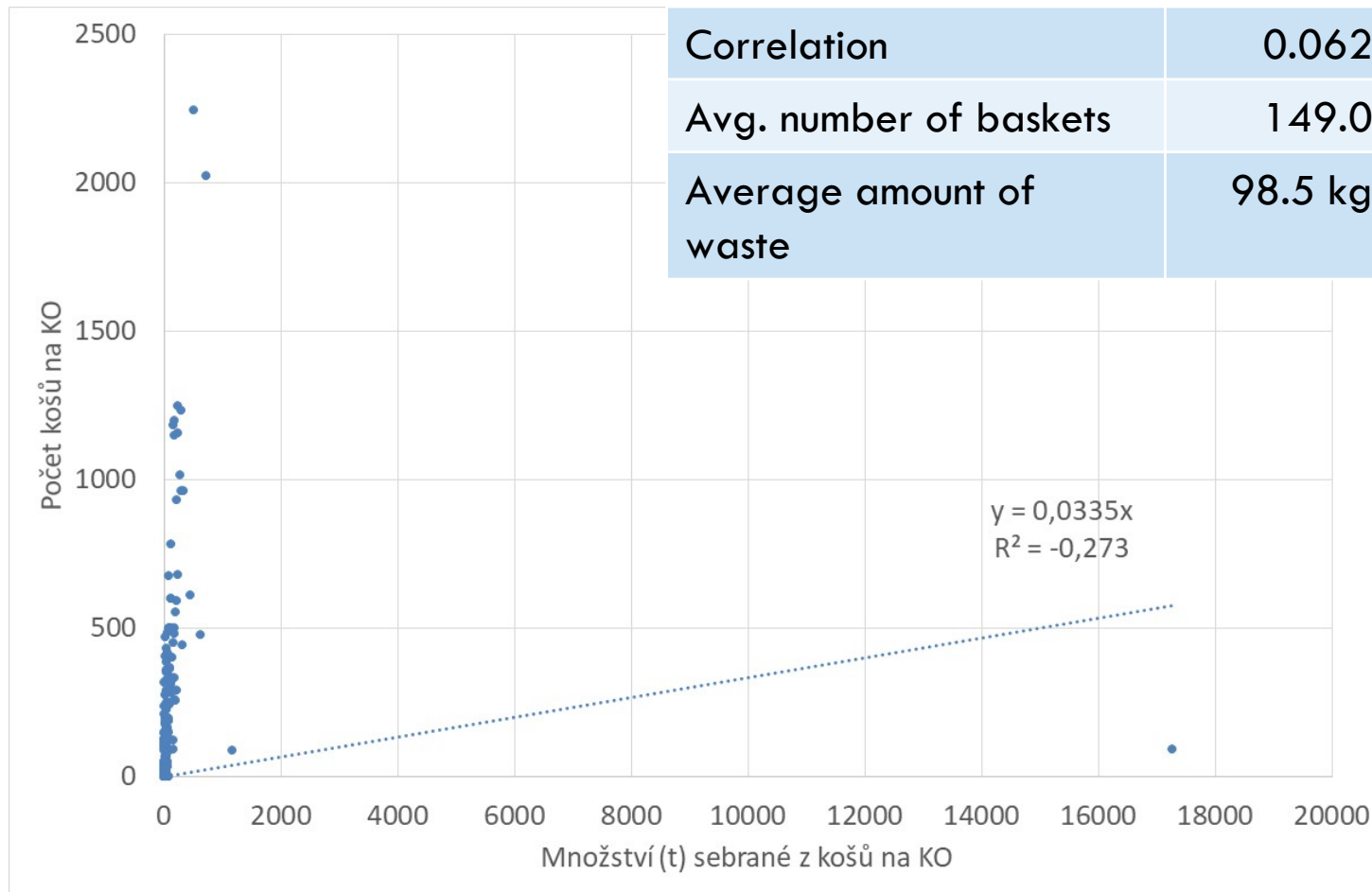
POPULATION HOMOGENEITY

Individual	Population A Number of beers	Population B Number of beers
Individual 1	8	0
Individual 2	8	0
Individual 3	8	0
Individual 4	8	0
Individual 5	8	40
Sum	40	40
Arithmetic mean	8	8

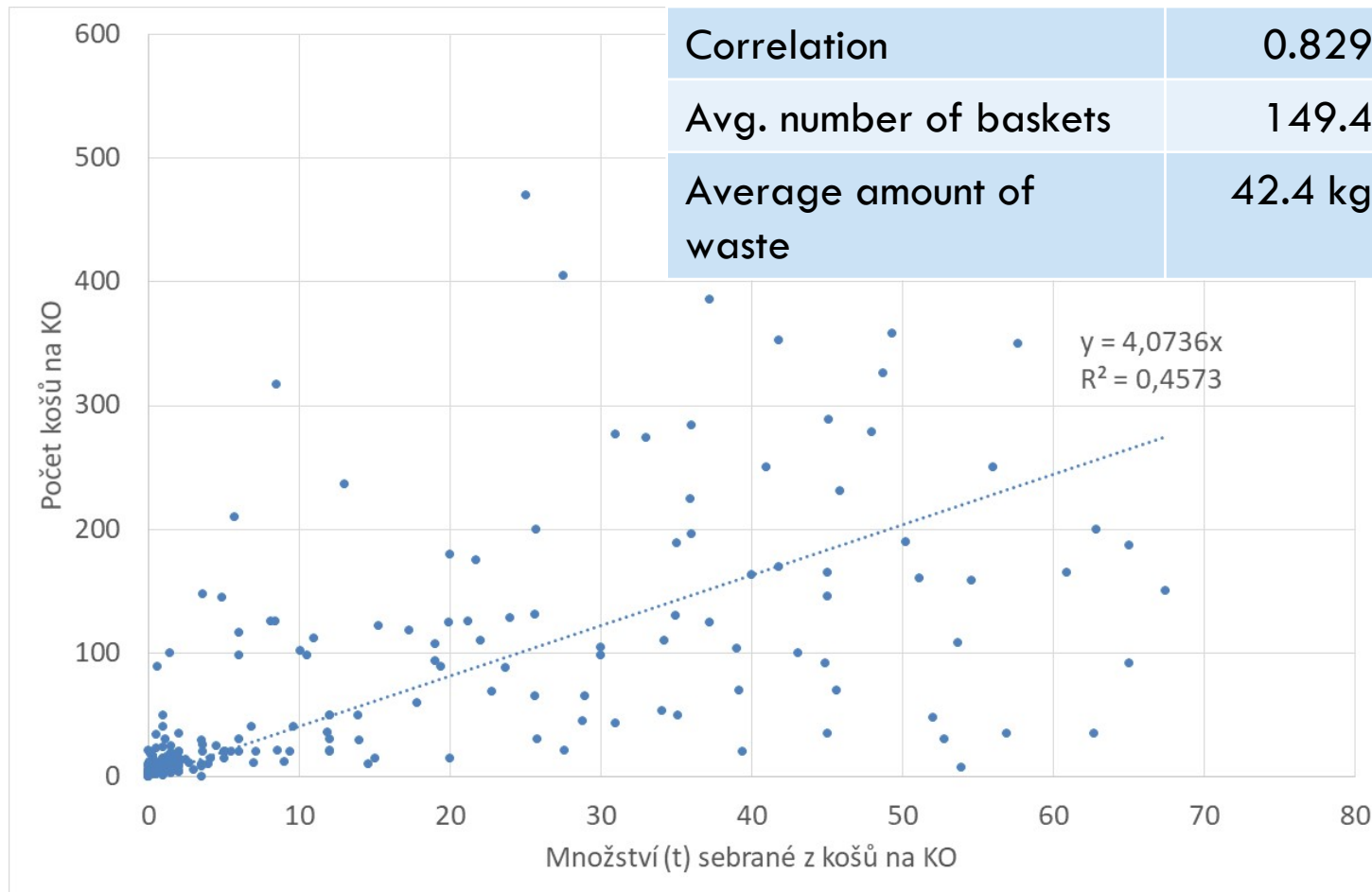
OUTSTANDING VALUES



OUTSTANDING VALUES



OUTSTANDING VALUES



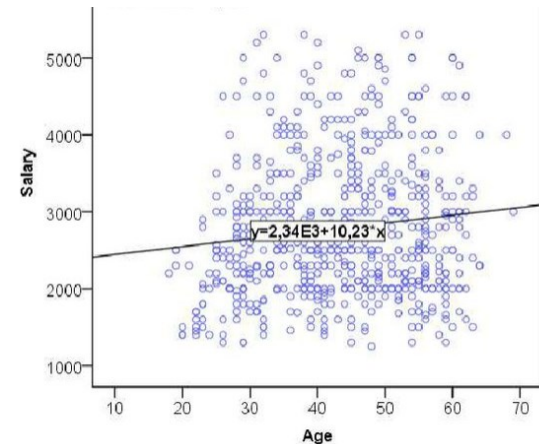
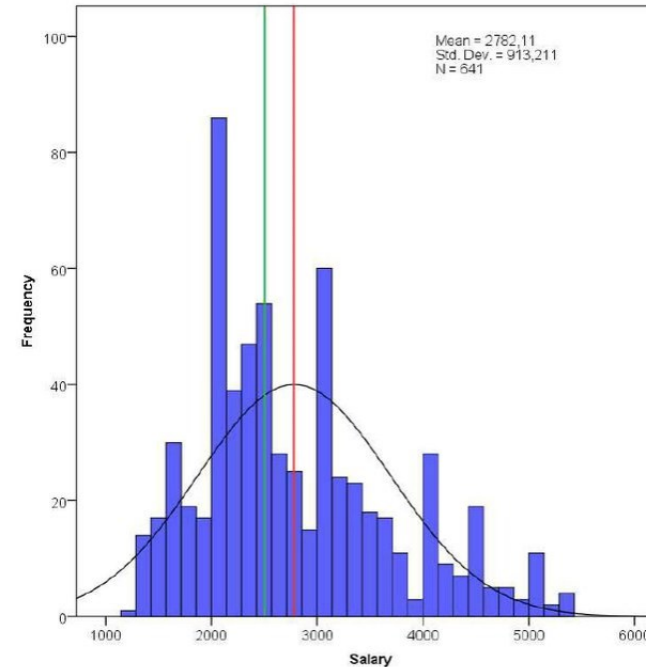
EXPLORE DATA WITH VISUALIZATION

Exploratory analysis – aims to propose hypotheses about data and relationships

Confirmatory analysis – hypothesis testing

A whole range of methods and techniques

- Box plot (box plot)
- Histogram
- Scatter plot (correlation diagram)



HYPOTHESIS TESTING

The basic goal – to draw conclusions about a wider set of unobserved data (population) based on information about the observed sample

- How likely is it that the conclusions obtained by analyzing a sample are the same as those that would be reached by observing the entire population?
- How likely is it that the observed distribution of a phenomenon (in space) arose by chance?

MODEL CREATION

The goal is to create a mathematical model representing our understanding of the situation and to determine how accurately this model describes the given situation.

Show information about the nature of relationships between variables

For example, parameter estimation using regression

- The dependent variable is a function of one or more independent variables
- The goal is to estimate parameters that reflect the relationship between the dependent and a given independent variable in the model

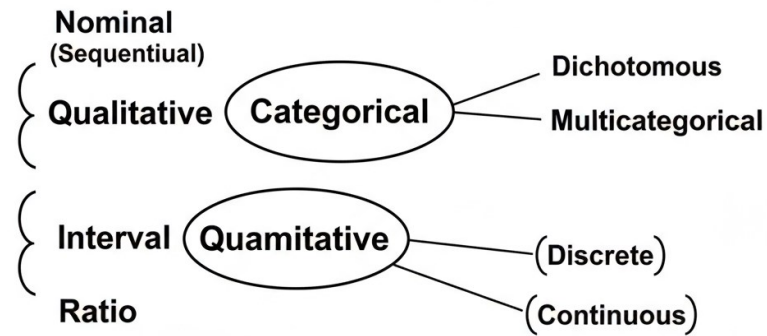
Forecasting a future situation or estimating values for a unit where data is not available

METHODS OF ANALYSIS

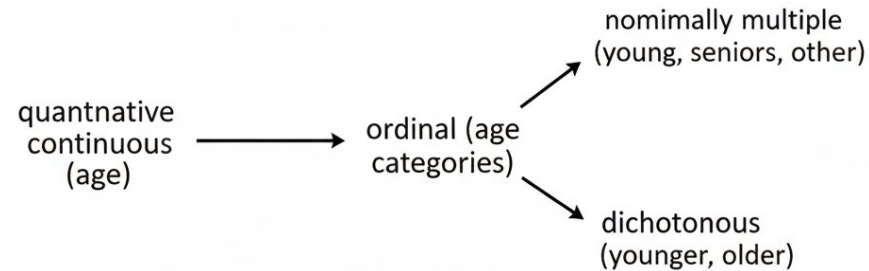
Many statistical methods, many applicable to the same data

2 properties of the analyzed data – type of variables and number of selections

VARIABLE TYPES



Variable type conversion options



METHODS OF ANALYSIS

Many statistical methods, many applicable to the same data

2 properties of the analyzed data – type of variables and number of selections

Nominal + ordinal = categorical variables

- Limited use of some analysis methods (e.g., no average)

Number of selections (data sets)

- 1 selection – e.g. analysis of how values in the region differ from the situation at the state level
- Multiple selections – e.g. analyzing how values differ across multiple regions

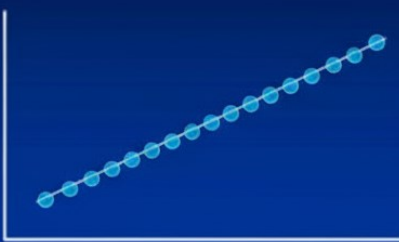
FINDING RELATIONSHIPS BETWEEN VARIABLES

	Mathematics				
		AND	B	C	D
descriptive	AND	50%	35%	10%	0%
	geometry				
	B	45%	55%	25%	10%
	C	5%	8%	55%	10%
	D	0%	2%	10%	80%
	N	100%	100%	100%	100%
		150	360	400	50

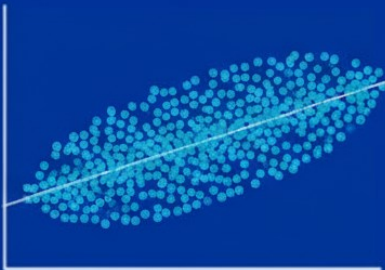
FINDING RELATIONSHIPS BETWEEN VARIABLES

	Variable X				
Variable Y		AND	B	C	D
	J				
	TO				
	L				
	M				

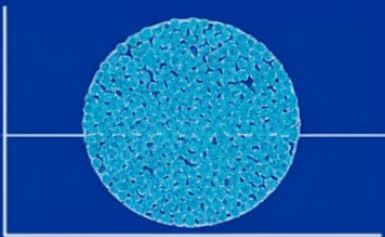
FINDING RELATIONSHIPS BETWEEN QUANTITATIVE (CARDINAL) VARIABLES



PERFECT POSITIVE CORRELATION.
Y increase directly with X.
 $r = +1$



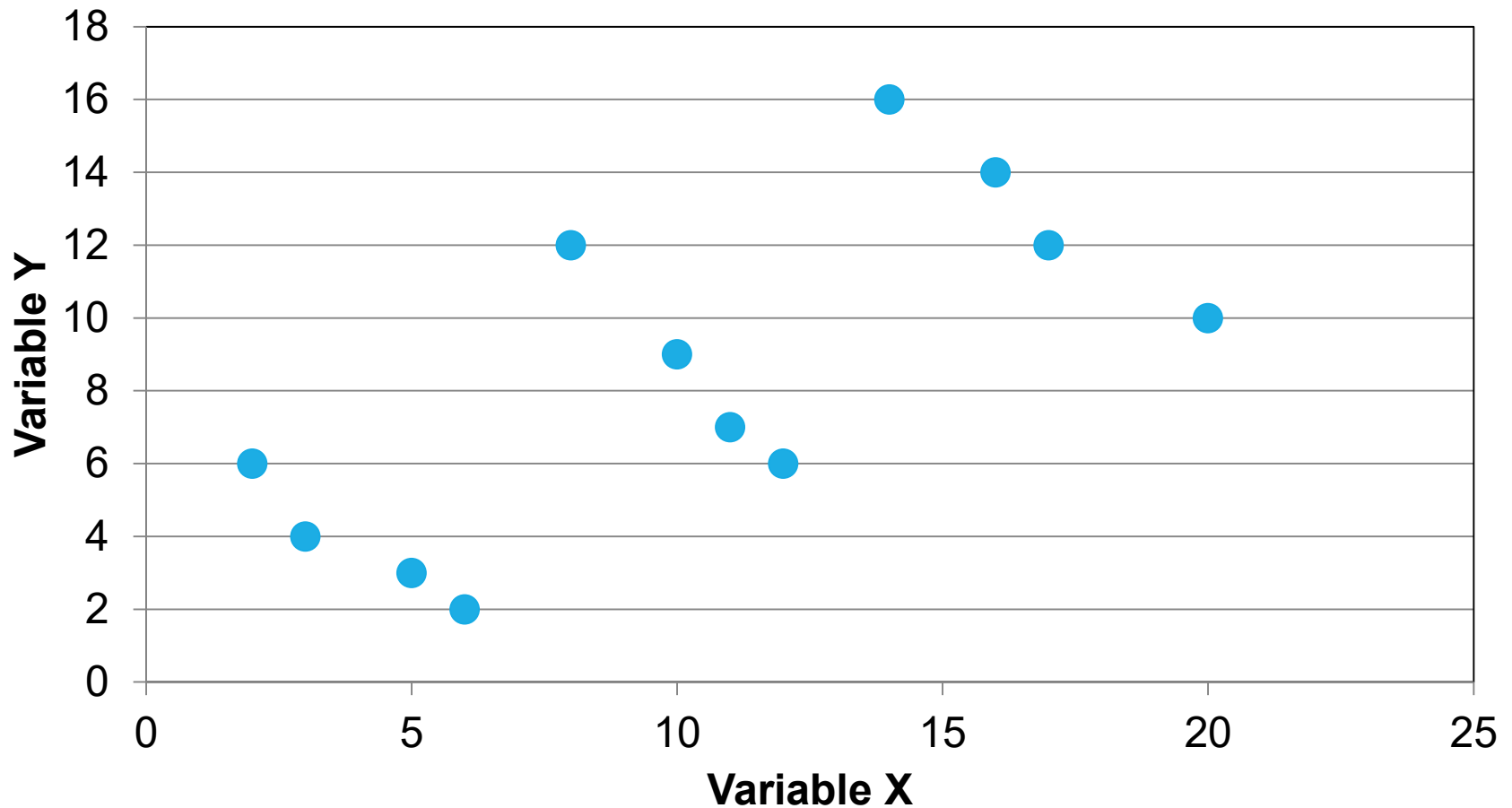
STRONG CORRELATION. X and Y are closely related, but



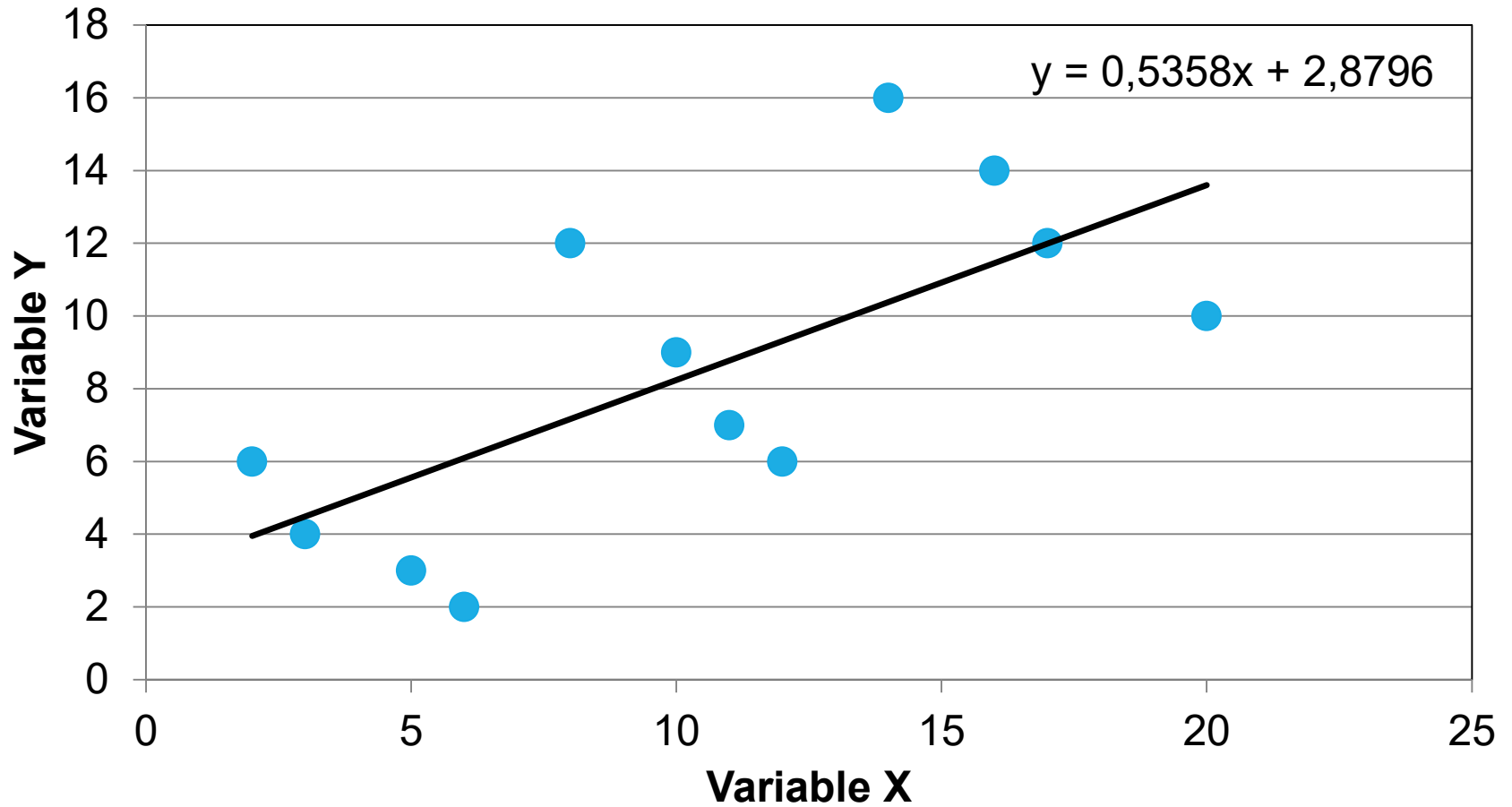
PERFECT INDEPENDENCE. No relationship relationship between X and Y, $r = 0$.

EXAMPLE: CORRELATION OF VARIABLES

WHAT WILL THE REGRESSION LINE LOOK LIKE? WHAT WILL THE VALUE OF THE CORRELATION COEFFICIENT BE?



EXAMPLE: CORRELATION OF VARIABLES



CORRELATION DOES NOT MEAN CAUSALITY!

If two phenomena follow one after the other, or appear to be related, this does not mean that one results from the other, or that one causes the other.

Correlation means a mutual relationship, but not necessarily that phenomena depend on each other, and it cannot be claimed that one of them is the cause and the other the effect.

When two events occur simultaneously, it does not mean that one of them causes the other.

DISTORTION

How it appears to us: $X \rightarrow Y$

False correlation: $X \leftarrow Z \rightarrow Y$

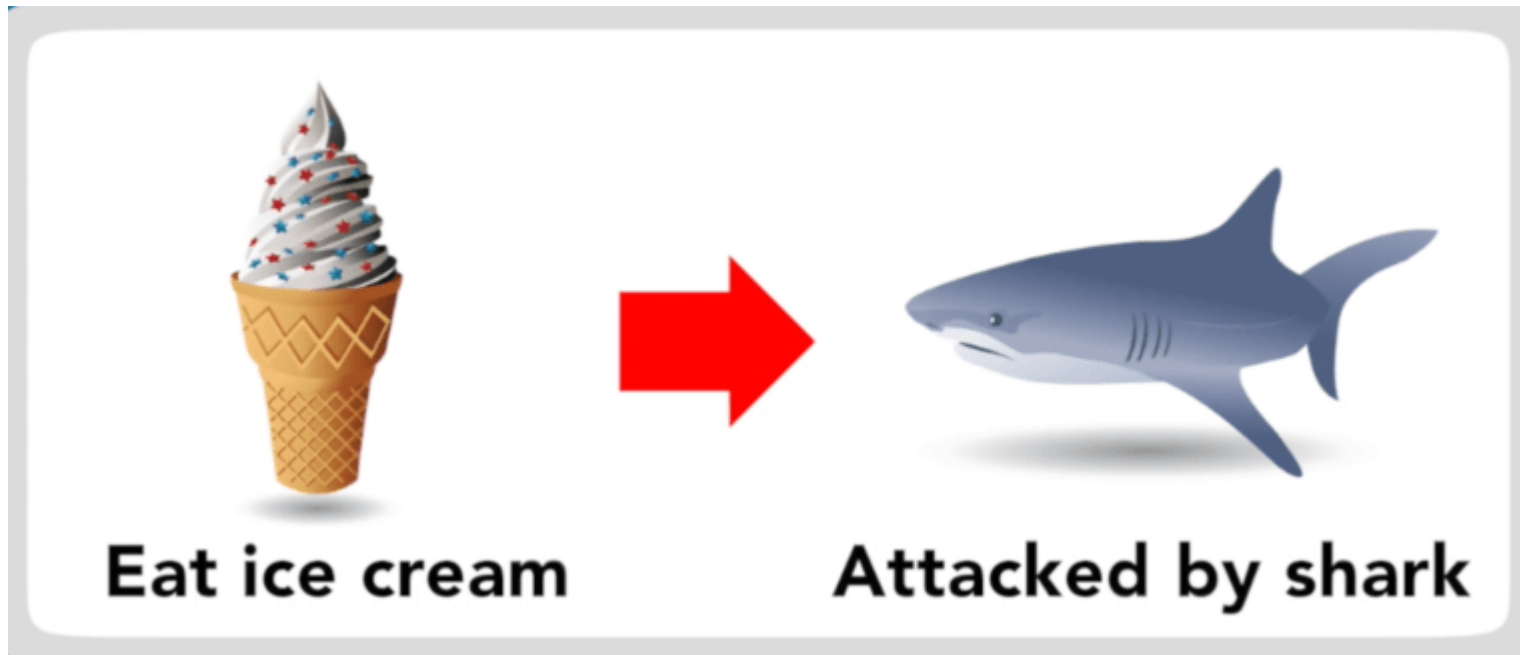
Development sequence: $Z \rightarrow X \rightarrow Y$

Missing middle term: $X \rightarrow Z \rightarrow Y$

Double cause: $X \rightarrow Y \leftarrow Z$

DISTORTION

When ice cream sales increase, the number of shark attacks on humans will also increase.



DISTORTION

Since the 1950s, both atmospheric CO₂ levels and the incidence of obesity have risen sharply. Does rising CO₂ concentrations cause obesity?

DISTORTION

Since the 1950s, both atmospheric CO₂ levels and the incidence of obesity have risen sharply. Does rising CO₂ concentrations cause obesity?

→ both as a result of richer populations eating more and consuming more energy

DISTORTION OF "THE THEORY OF THE STORK"

M. Disman: In a number of European regions, it was found that the more storks live in a certain landscape, the higher the birth rate there. The correlation coefficients were so significant that it is very unlikely that the connection found is random. So are we willing to accept the hypothesis that storks do carry babies after all?

DISTORTION OF "THE THEORY OF THE STORK"

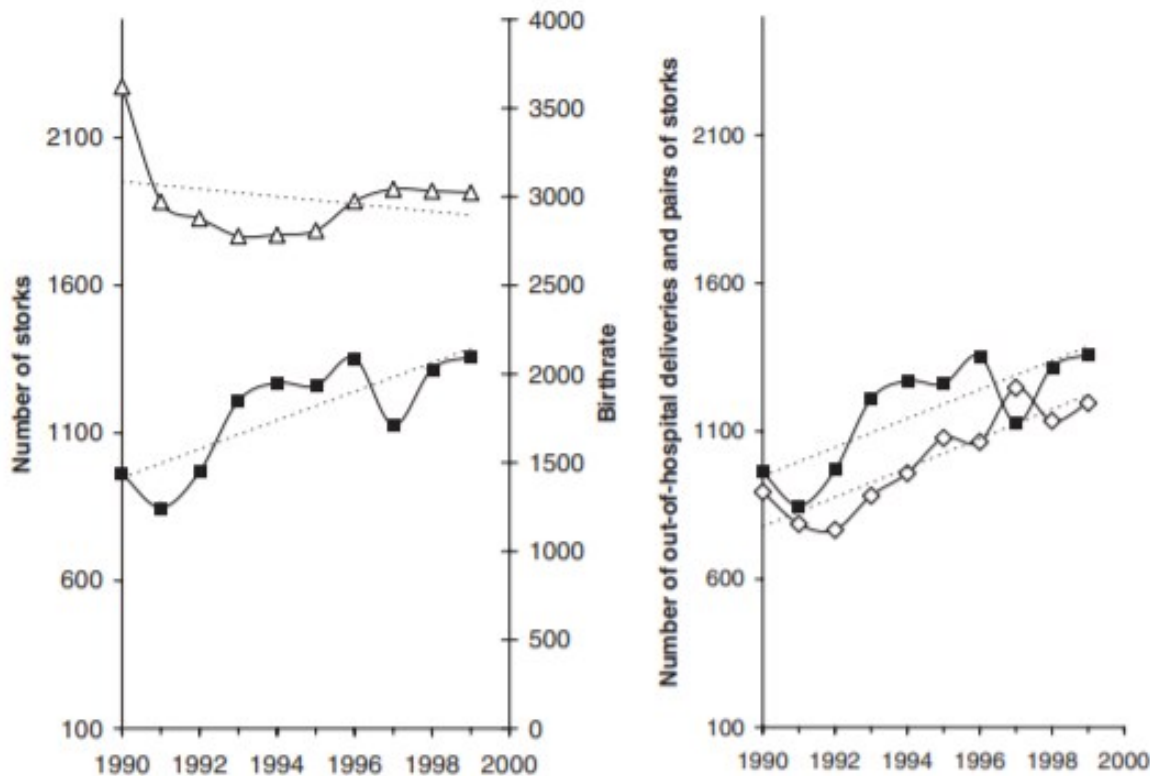


Figure 2. Storks in Brandenburg and the birthrates in Berlin, Germany (1990–99). Open triangles show number of clinical deliveries per year in Berlin. Open diamonds show number of out-of-hospital deliveries per year in Berlin. Number of pairs of storks are shown as full squares. Dotted lines represent linear regression trend ($y = mx + b$). For the convenience of the readers, two figures are presented. Left graph shows clinical deliveries against pairs of storks using two scalings, right graph shows numbers of out-of-hospital deliveries and pairs of storks both on the same scale. In both figures, data are from the years 1990–2000.

DISTORTION OF "THE THEORY OF THE STORK"

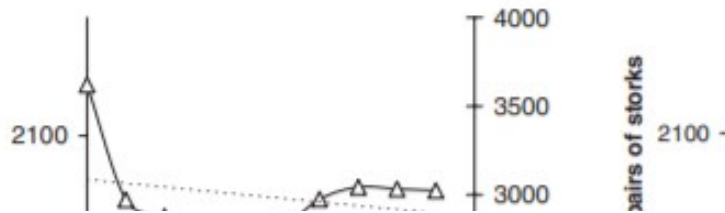
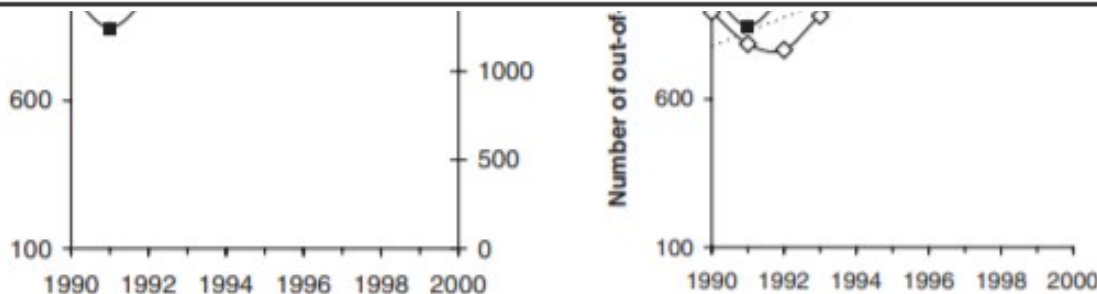


Figure 2. Storks in Brandenburg and the birthrates in Berlin, Germany (1990–99). Open triangles show number of clinical deliveries per year in Berlin. Open diamonds show number of out-of-hospital deliveries per year in Berlin.

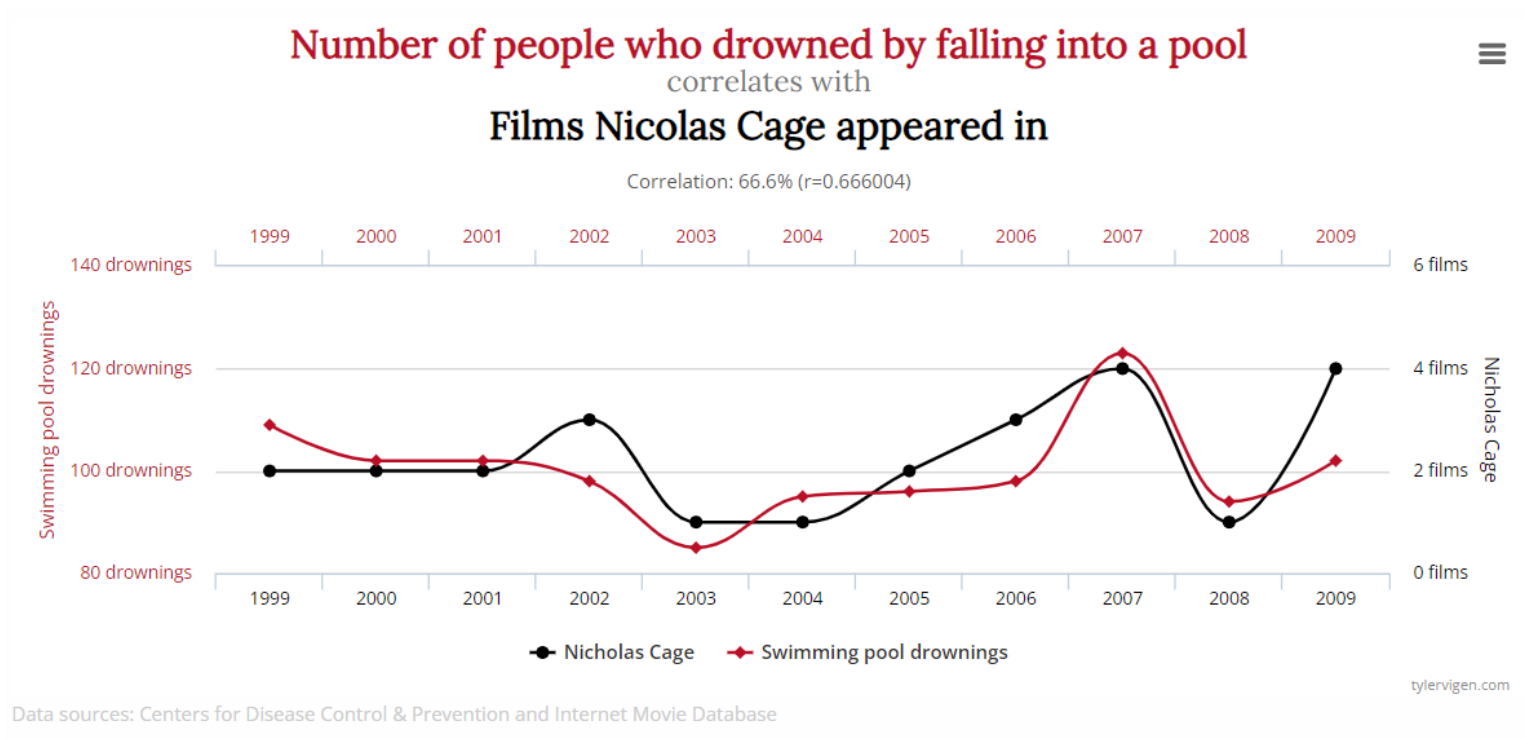
The intended value (disclaimer): This article is not intended to disprove the value of serious epidemiological investigations. It is an example of how studies based on popular belief and unsubstantiated theory, seconded by low quality references and supported by coincidental statistical association could lead to apparent scientific endorsement. Insofar it is a humorous case study for education in perinatal epidemiology.



shows numbers of out-of-hospital deliveries and pairs of storks both on the same scale. In both figures, data are from the years 1990–2000.

CORRELATION DOES NOT MEAN CAUSALITY

"Nicolas Cage movies are often so bad that they often make people go drown in a swimming pool."



WHEN READING STATISTICS, PAY ATTENTION TO:

"exact" numbers in statistical reports

percentages versus absolute numbers

comparable data

do not confuse indicators

Is the comparison appropriate?

for withdrawals (if not made by professionals)

form of questions

own preparation of the investigation (experiment)

"people die more often in the hospital than at home"

RESEARCH ETHICS

Science is part of the reality being researched → the researcher is responsible for how he enters the world and what consequences his research and the report about it have in this world

Research is conceived, prepared and implemented in such a way that it is ethical in all its steps and outputs

So-called procedural ethics

- ❑ basic ethical principles that are the same for all researchers
- ❑ general rules formulated by law and rules agreed upon by the wider academic community – codes of ethics

RESEARCH ETHICS

- respect the privacy and autonomy of research participants

- ☐ Including data anonymization

- prevent the risk of exploitation or abuse of the subjects studied

- ☐ During the research and after its completion - risk of misuse of results, e.g. by the sponsor or another entity

- not to harm or injure research participants

- ☐ Mainly research on sensitive, intimate topics; often e.g. psychological research

- ☐ Do not promote your own opinions as better, more legitimate, more rational, or of higher quality.

- take into account the impacts of one's own research on other researchers in the same field

- ☐ Through the terrain and, for example, the archive

- research must not cause harm to the researcher

- ethical behavior in relation to the academic community

- ☐ Including publication ethics

COPYRIGHT, PUBLICATION ETHICS

If parts or entire works of other authors are used for research and publication purposes, the author's name and a reference to the work must always be provided according to the relevant standards of bibliographic citations.

Plagiarism

- ❓ Taking and using ideas, data, measurement tools, etc. from other authors and not citing the original source of information
- ❓ Gross violation of publication ethics

Counterfeiting

- Data distortion

Fabrication

- ❓ Inventing data

TO EXPAND THE TOPIC OF RESEARCH ETHICS

Examples of ethically controversial experiments include the Milgram experiment and the Stanford Prison Experiment.

For example, articles by psychiatrist Radkin Honzák

<https://psychologie.cz/milgramuv-brutalni-experiment/>

<https://psychologie.cz/stanfordsky-experiment/>