Causation vs Correlation

Nikodem Lewandowski

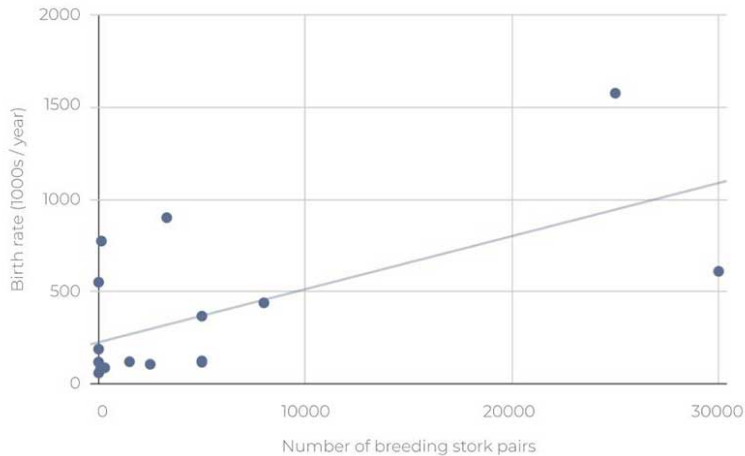**University of Gdańsk**

Introduction

There is a puzzling article under the title *'Storks Deliver Babies (p=
0.008)'* by Robert Andrew Matthews

Storks deliver babies?

It turns out that there is a high correlation between the number of storks
and the number of babies born in a given region.

Storks and Babies

Explanation

- The author aimed to show that correlation-measuring methods do not represent causation.

- Correlation measures the statistical association between two variables

- Causation implies a cause-and-effect relationship between two variables. Changes in one variable directly lead to changes in the other.

Correlation does not imply causation

is a phrase used in science and statistics to emphasize that a correlation between two variables does not imply that one causes the other.

University of Gdańsk

Causation and Correlation

- In contemporary statistics, there is a widely spread belief that statistics is a science of correlation and there are no tools for measuring causation.

- Although the development of data science is proving this wrong

- There are many methods of creating causal models that represent cause-and-effect relationships between variables. Under some assumptions, these models can be used to estimate causal effects.

University of Gdańsk

Final example

- Research in the 1930s and 1940s started to highlight a statistical association between cigarette smoking and lung cancer

- There was a serious debate about whether smoking or stress causes cancer

Which one is true?

smoking -> cancer

stress -> cancer

University of Gdańsk

Suplementary materials

Cool videos about causation and correlation:

- How Ice Cream Kills! Correlation vs. Causation
- Causal Inference - EXPLAINED!