The Bayesian Bite

Nikodem Lewandowski

From where do babies come from?

University of Gdańsk

Storks deliver babies!

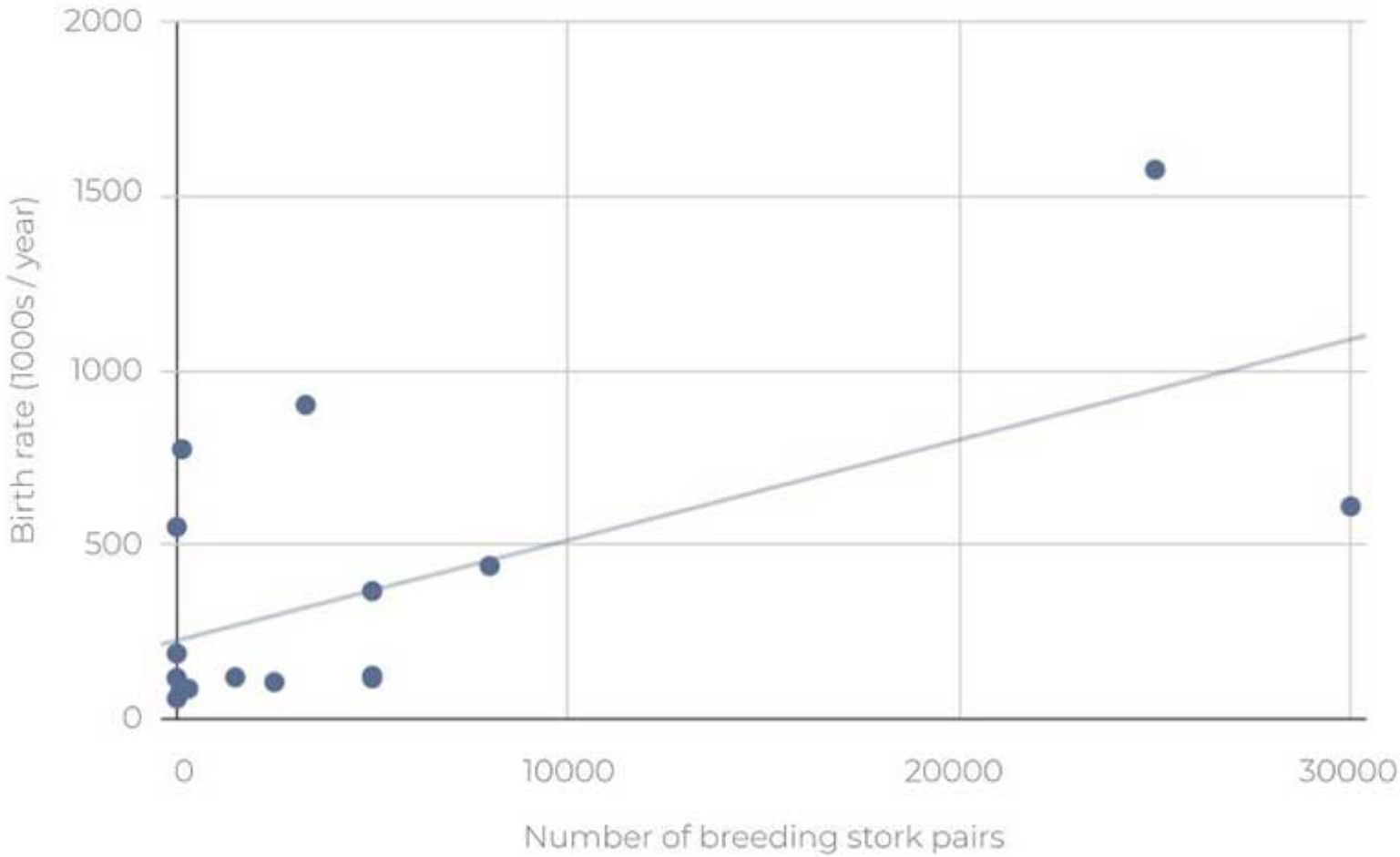
There is a puzzling article under the title *'Storks Deliver Babies (p= 0.008)'* by Robert Andrew Matthews

Wait, what!?

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

University of Gdańsk

Storks and Babies



University of Gdańsk

So, where is the catch?

Reality is complex

University of Gdańsk

French paradox

- French people have:
  - a relatively low incidence of coronary heart disease
  - a diet rich in saturated fats (which is a risk factor for coronary disease)
  - they drink a lot of wine

- Conclusion: Drinking wine promotes heart health (mainly because of resveratrol).

The lesson:

Humans have a tendency to explain complex phenomenons with overly simplified models

French paradox

- the newest evidence says that:
  - even the smallest amount of alcohol is harmful
  - the amount of resveratrol in wine is not significant for the heart health

- possible explanations of french paradox:
  - it's an illusion, created by differences in the way that French authorities collect health statistics
  - french people: walk a lot, don't eat a lot of processed food, eat fruits and vegetables in large quantity

University of Gdańsk

Purpose of This Course

- In today's science we use statistical **models** to represent reality
  - the model is as smart as a statistician that made it
  - it's very easy to get confused in interpreting the various classical parameters (p-values, means, etc.)

- Generative models that represent reality with modest assumptions may be an incredibly useful tool in any empirical discipline
  - with the obtained datasets, we can conduct analyses that shed light on important social phenomena

- the Bayesian setting (often referred to as the new statistics) provides us with more intuitive and modest methods that can be used to build such models

Conclusion:

Let's learn Bayesian methods

University of Gdańsk

Roadmap

Lectures

- Representing reality through data and probability
- Learning the differences between statistical correspondence (correlation) and causality
- Building and evaluating Bayesian statistical models, including:
  - Bayesian linear models
  - Bayesian logistic models

Tutorials

- Elements of Bayesian data analysis:
  - practical abilities of using R studio
  - exploratory data analysis
  - manipulating data
  - creating data visualizations
  - building the models introduced on lectures

The Goal

After this course you will be able to analyze a dataset and create a Bayesian model that will represent some kind of relation between variables. It will give you a set of abilities to conduct a scientific research e.g. writing a paper

Organizational info

Contact Details:
Nikodem Lewandowski
email: nikodemlewandowski@gmail.com
website: https://niklewa.github.io/teaching.html

How to pass?

- There is one grade for both the lectures and the tutorials
- You will be graded by the amount of points you collect, between 0 to 100
- You can collect points by:
  - writing an exam (max 60 points)
  - completing assignments on the tutorials (max 60 points)
  - making a project (max 60 points)*

Tips

- Make notes and read supplementary materials to understand lectures
- Attend tutorials (the best predictor of passing)

University of Gdańsk

More info

- Learning Bayesian statistics and programming is hard
  - don't even think that you will understand everything right away
  - you need to give yourself time and put some effort into it
- Feeling lost and unsure when learning that stuff is normal
  - it's just your brain being a jackass
- You can contact me at any time if you need guidance or have questions
  - especially if you become interested in that topic and want to learn more!

Disclaimer

Please DO NOT come to my classes if you only want to sit on your phone or disturb in any other way.

University of Gdańsk

Why Bayesian Methodology?

- It's superior to classical statistics.
- **Classical** approach, relies on trying to falsify a null hypothesis:
  - uses point estimates with problematic interpretations
  - follows a 'problem x -> test y' approach
  - assumes a normal distribution of variables
- **Bayesian** approach, relies on interpreting the outcome distributions:
  - enables sampling from generative models
  - customizable prior distributions
  - its interpretation is far more intuitive

Of course classical tools in many concepts ares still useful, but the problem is not really about the tool, but rather about the way it is used.

University of Gdańsk

Models?

- A statistical model is a mathematical representation of a phenomenon or process in the form of a set of statistical assumptions and equations.
- It is used to describe the relationship between variables and to make predictions or inferences about data

- It's a small world, that enable us to make low cost experiments
- The model is as smart as a statistician that made it

University of Gdańsk