Open and reproducible science: dependable computations and statistics - Introduction

Eva Furrer

Code of conduct for scientific integrity

"Reliability, honesty, respect, and accountability are the basic principles of scientific integrity. They underpin the independence and credibility of science and its disciplines as well as the accountability and ***reproducibility*** *of research findings and their acceptance by society. As a system operating according to specific rules, science has a responsibility to create the structures and an environment that foster scientific integrity."*

https://akademien-schweiz.ch/en/uber-uns/kommissionen-und-arbeitsgruppen/wissenschaftliche-integritat/

UNESCO Recommendation on Open Science

"Building on the essential principles of academic freedom, research integrity and scientific excellence, ***open science sets a new paradigm that integrates into the scientific enterprise practices for reproducibility, transparency, sharing and collaboration*** *resulting from the increased opening of scientific contents, tools and processes."*

https://en.unesco.org/science-sustainable-future/open-science/recommendation

Five selfish reasons to work reproducibly

F Markowetz

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0850-7

"Working transparently and reproducibly has a lot to do with empathy: put yourself into the shoes of one of your collaboration partners and ask yourself, would that person be able to access my data and make sense of my analyses. Learning the tools of the trade will require commitment and a massive investment of your time and energy. A priori it is not clear why the benefits of working reproducibly outweigh its costs."

Dependable computations and statistics

In order to adhere to the principles of reliability, honesty, respect, and accountability we propose to work

- openly/transparently
- reproducibly

along the entire research process, specifically in the

- computations underlying any results
- statistical approaches, assumptions and techniques underlying the computations.

Example: research question

Are soccer players with dark skin tone more likely than those with light skin tone to receive red cards from referees?

Consider for a moment how you would test this research hypothesis using a complex archival data set including referees' decisions across numerous leagues, games, years, referees, and players and a variety of potentially relevant control variables that you might or might not include in your analysis.

Example: data

Data and profile photos from all soccer players (N = 2053) playing in the first male divisions of England, Germany, France and Spain in the 2012-2013 season and all referees (N = 3147) that these players played under in their professional career.

In all, the dataset has a total of 146028 dyads of players and referees including among other information the number of red cards given to a player by a particular referee throughout all matches the two encountered each other.

Player photos were available for 1586. Players skin tone was coded by two independent raters blind to the research question who categorized players on a 5-point scale (from very light skin to very dark skin with neither dark nor light skin as the center value).

Example: variables

- player: ID, name, club, country of club, birthday, height (in cm), weight (in kg), position
- player-referee dyad: number of games, victories, ties , losses, goals scored by a player
- number of yellow cards player received from referee, number of yellow-red cards player received from referee, number of red cards player received from referee
- ID of player photo (if available), skin rating of photo by rater 1, skin rating of photo by rater 2

Example: variables

- referee: ID number (referee name removed for anonymizing purposes), country ID number (country name removed for anonymizing purposes), mean implicit bias score for referee country (higher values correspond to white | good, black | bad associations)

- sample size for race IAT in that particular country, standard error for mean estimate of race IAT, mean explicit bias score for referee country (higher values correspond to greater feelings of warmth toward whites versus blacks), sample size for explicit bias in that particular country, standard error for mean estimate of explicit bias measure

Example: analysis decisions

- Would you treat each red-card decision as an independent observation?
- How would you address the possibility that some referees give more red cards than others?
- Would you try to control for the seniority of the referee?
- Would you take into account whether a referee's familiarity with a player affects the referee's likelihood of assigning a red card?
- Would you look at whether players in some leagues are more likely to receive red cards compared with players in other leagues, and whether the proportion of players with dark skin varies across leagues and player positions?

Garden of the forking path

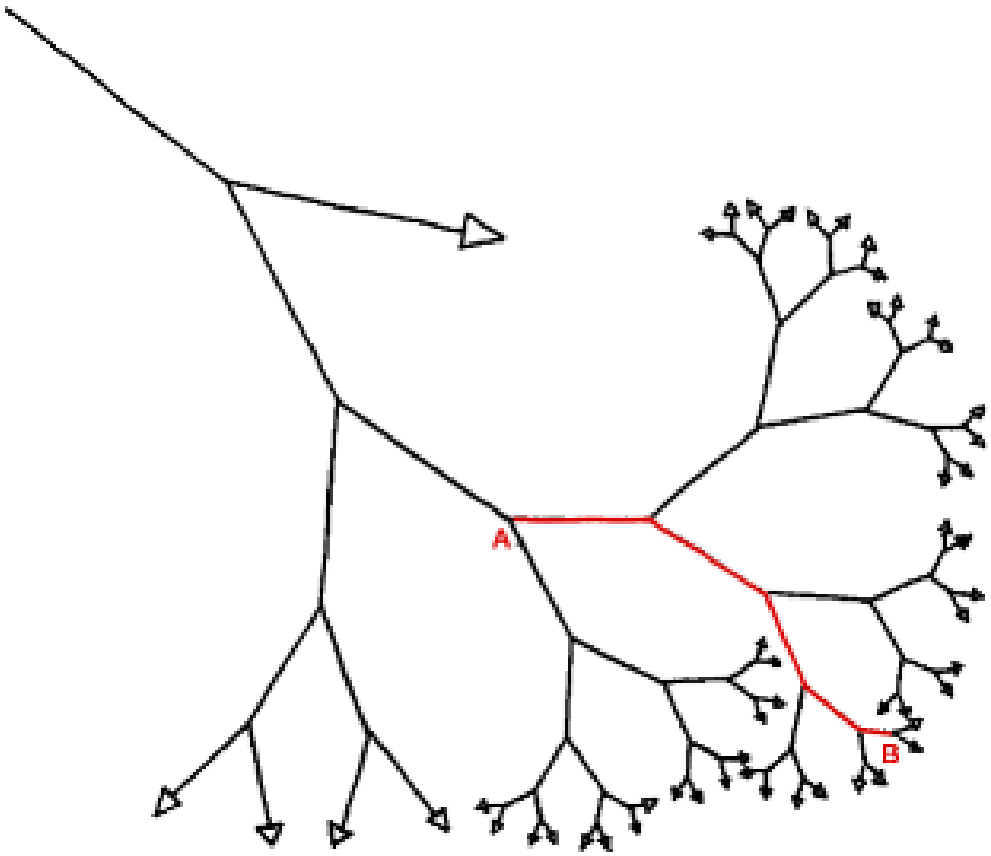

Image source unknown, reproduced from D. Bishop

https://promotion.charite.de/fileadmin/user_upload/microsites/ohne_AZ/sonstige/gss/K

Task: article and data source

The example is from the paper by Silberzahn et al.
https://journals.sagepub.com/doi/10.1177/2515245917747646
which provides all material in the OSF repository
https://osf.io/gvm2z/.

Task: what to do

The paper is based on the analysis of 32 teams. Assigning the numbers 1 to 26 to the letters of the alphabet, determine your team number from the starting letter of your last name.

Use the OSF repository to find out for **your team**

1. What exclusions of data points were used?
2. What main techniques/models/approaches were used?
3. Which software was used?

Enter your answers in the OpenEdx course.