Destatis
Statistisches Bundesamt

Quality Dimensions of Machine Learning in Official Statistics

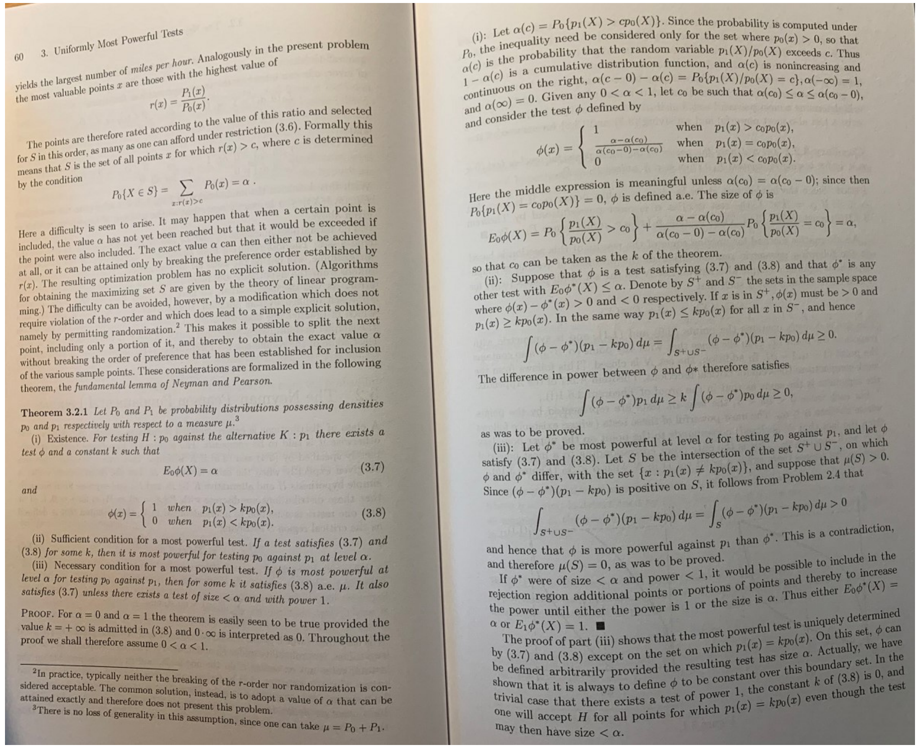
Florian Dumpert, Younes Saidani

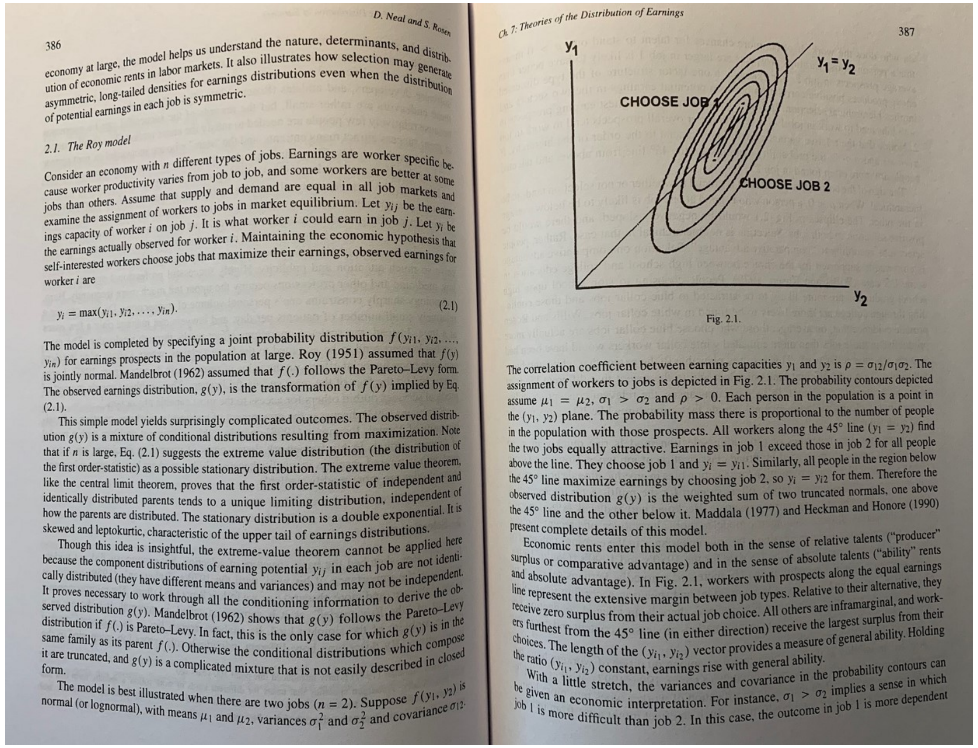
Federal Statistical Office of Germany

Joint work with: Christian Borgs, Alexander Brand, Andreas Nickl, Alexandra Rittmann, Johannes Rohde, Christian Salwiczek, Nina Storfinger, Selina Straub

Official Statistics

It is not about …





Lehmann EL, Romano JP (2005) Testing Statistical Hypotheses, 3rd edition, Springer
Atkinson AB, Bourguignon F (2000) Handbook of Income Distribution Vol. 1, North Holland

Official Statistics

It is about …

Die Statistik für Bundeszwecke (Bundesstatistik) hat im föderativ gegliederten Gesamtsystem der amtlichen Statistik die Aufgabe, **laufend Daten über Massenerscheinungen zu erheben, zu sammeln, aufzubereiten, darzustellen und zu analysieren. Für sie gelten die Grundsätze der Neutralität, Objektivität und fachlichen Unabhängigkeit. Sie gewinnt die Daten unter Verwendung wissenschaftlicher Erkenntnisse und unter Einsatz der jeweils sachgerechten Methoden und Informationstechniken.** Durch die Ergebnisse der Bundesstatistik werden gesellschaftliche, wirtschaftliche und ökologische Zusammenhänge für Bund, Länder einschließlich Gemeinden und Gemeindeverbände, Gesellschaft, Wirtschaft, Wissenschaft und Forschung aufgeschlüsselt. **Die Bundesstatistik ist Voraussetzung für eine am Sozialstaatsprinzip ausgerichtete Politik**. […]

In the federally structured overall system of official statistics, statistics for federal purposes (federal statistics) have the task of **continuously collecting, collating, processing, presenting and analysing data on mass phenomena. It is governed by the principles of neutrality, objectivity and professional independence. It collects data using scientific knowledge and appropriate methods and information technology**. The results of federal statistics provide a breakdown of social, economic and ecological relationships for the Federation, the Länder including municipalities and municipal associations, society, the economy, science and research. **Federal statistics are a prerequisite for a policy oriented towards the welfare state principle**. […]

§ 1 BStatG (https://www.gesetze-im-internet.de/bstatg_1987/__1.html) and an unauthorised translation by the author of the slides

Official Statistics

How we work: The Generic Statistical Business Process Model (GSBPM)

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

https://statswiki.unece.org/display/GSBPM

„Bad quality reduces trust very, very fast.“

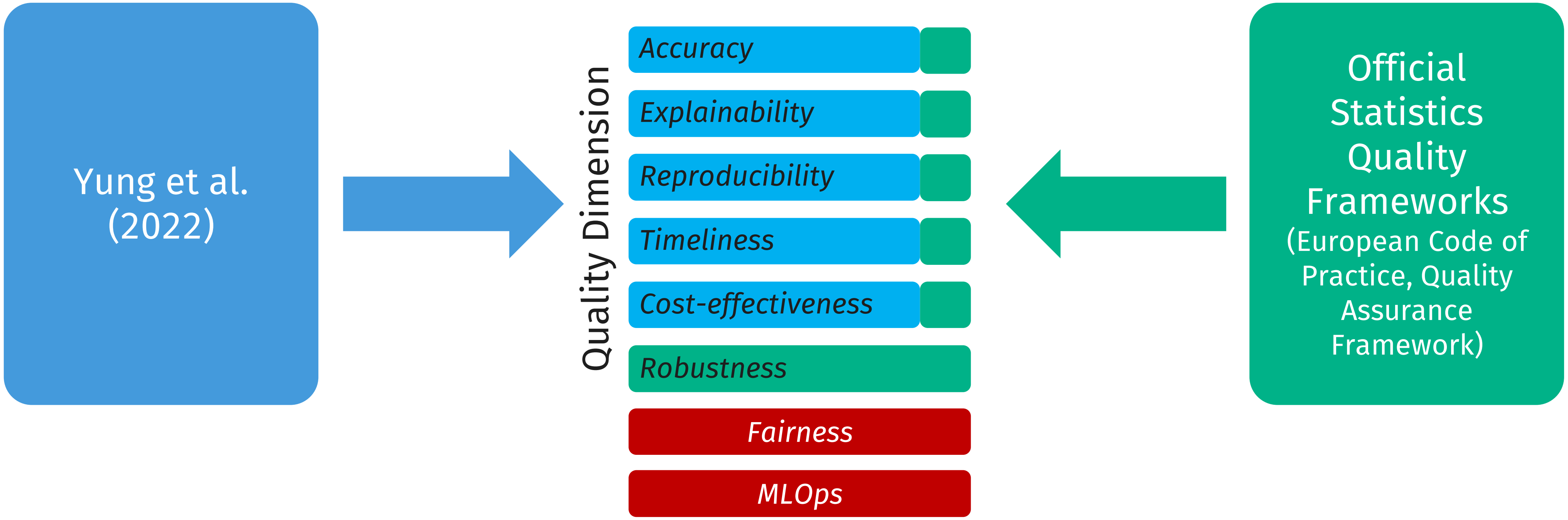
Walter Radermacher, 2022

Starting Points

- **Quality Framework for Statistical Algorithms**[1)] with its dimensions
  - Accuracy
  - Explainability
  - Reproducibility
  - Timeliness
  - Cost-effectiveness
- **Quality frameworks in official statistics**[2)]

1) Yung W et al (2022) A quality framework for statistical algorithms. Statistical Journal of the IAOS, 38(1), 291–308
2) https://www.destatis.de/DE/Methoden/Qualitaet/qualitaetshandbuch.html, https://ec.europa.eu/eurostat/web/quality/european-quality-standards/quality-assurance-framework, https://unstats.un.org/unsd/unsystem/documents/UNSQAF-2018.pdf

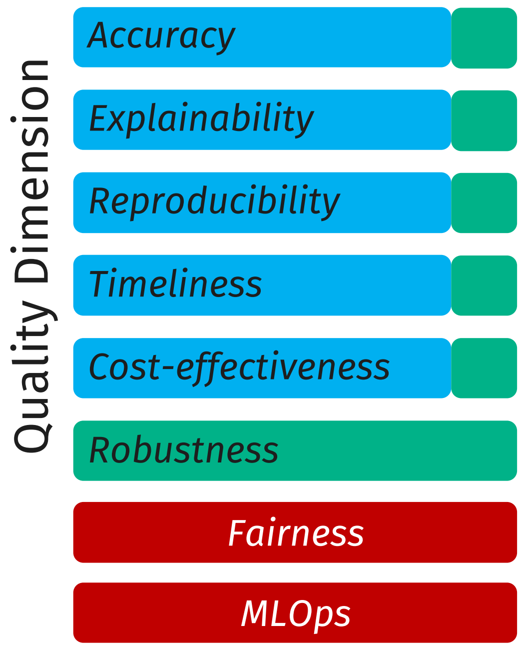
Quality Dimensions



Yung W et al (2022) A quality framework for statistical algorithms. Statistical Journal of the IAOS, 38(1), 291–308
https://ec.europa.eu/eurostat/web/quality/european-quality-standards/quality-assurance-framework
Saidani Y et al (2023) Qualitätsdimensionen maschinellen Lernens in der amtlichen Statistik. AStA Wirtschafts- und Sozialstatistisches Archiv, 17(3-4), 253–303

Our Three Main Contributions

- **Robustness** is proposed as a stand-alone quality dimension
- Machine learning operations (**MLOps**) and **fairness** are discussed as two cross-cutting issues
- Suggestions are made how quality assurance can be conducted **in practice** for each quality dimension



Saidani Y et al (2023) Qualitätsdimensionen maschinellen Lernens in der amtlichen Statistik. AStA Wirtschafts- und Sozialstatistisches Archiv, 17(3-4), 253–303

Proposed Quality Dimensions

Accuracy	„phenomenon is described correctly“
Robustness	„stable results despite small perturbations“
Explainability	„understand how results are generated“
Reproducibility	„reproduce identical results“
Timeliness & Punctuality	„deliver up-to-date results punctually“
Cost-effectiveness	„appropriate costs“

Predictions
↓
Model
↓
Infrastructure
↓
Processes

Our Approach

- What is needed to ensure compatibility of ML applications used in official statistics production with existing official statistics quality standards?
  1. **Quality dimensions**: What does it mean for ML to have „high quality“?
  2. **Quality guidelines**: How to implement quality along the above dimensions during development?
  3. **Quality indicators**: How to evaluate quality in development & production?
  4. **Quality documentation**: How to communicate quality of ML in an appropriate, standardised way?
- Work-in-progress: 1. completed, 2. in progress, 3. & 4. pending

Quality Guidelines – Some Examples

Dimension	Guideline
Accuracy	The quality measures relevant for the subject matter experts were determined, i.e. it was specified which quality measures were to be considered.
Robustness	Target variables that are to be the subject of robustness were defined.
Explainability	If the other quality dimensions were (approximately) the same, the more explainable model was used.
Reproducibility	It was ensured that the data is identical regardless of the time of access, i.e. archived unchanged at a known location.
Timeliness & Punctuality	Sufficient time was planned for the conceptualisation, selection and testing of possible ML solutions.
Cost-effectiveness	As part of a proof of concept, the existing database was analysed and the feasibility of the project assessed.

Fairness in Official Statistics

- Aim/Definition: to **avoid treating certain (sub-)groups unjustifiably differently** in a relevant way by or as a result of statistical procedures (like ML)
- In the context of official statistics, **effects are usually indirect**, e.g., through political decisions based on the published data
- **Example**: statistical aggregates are systematically over- or underestimated for certain sub-groups
  (e.g., economic sectors, types of households, regions, ...)
- **Connections** to accuracy (imbalanced data) and explainability

By the Way

- International conference
- Around 150 participants
- 19 countries
- 5 keynotes, 36 talks
- www.destatis.de/ml-conference

Statistisches Bundesamt

Thank you! Questions?

Florian Dumpert
Federal Statistical Office of Germany

florian.dumpert@destatis.de
+49 611 75-3887

www.destatis.de