



UNIVERSITÄT  
LEIPZIG

Medizinische Fakultät

Informationsmanagement in der klinischen Forschung

# MI-Lab Übung: ARX als Anonymisierungswerkzeug

Leipzig, Datum

Mona Perbix

**imise.**

## DISCLAIMER

- Die Übung basiert zu wesentlichen Teilen auf einer Präsentation der Funktionalitäten von ARX von (damals) Dr . Fabian Prasser vom Institut für medizinische Statistik und Epidemiologie, TU München
- Die Rechte an den Inhalten liegt bei der TMF Technologie und Methodenplattform für die vernetzte medizinische Forschung e.V. Berlin
- Keine Weiterverbreitung ohne Genehmigung der obigen Stellen!

# INHALT

## Vorbereitung

- Einführung und Installation

## Grundlegende Funktionalitäten

- Aufgabe 1: Schutz vor Re Identifikation mit  $k$  Anonymität
- Aufgabe 2: Aufzeigen der optimalen Transformation und Finden der zweitbesten Lösung
- Aufgabe 3: Vergleich der Identifikationsrisiken vorher/nach Transformation
- Aufgabe 4: Erstellen einer eigenen Generalisierungshierarchie
- Aufgabe 5: Auswirkung der Generalisierungshierarchie auf die Distribution einzelner Merkmale

# INHALT

## Transformationsmodelle

- Aufgabe 6: Auswirkung von Suppression
- Aufgabe 7: Auswirkung von Mirkoaggregation

# EINFÜHRUNG

## Aufgabenstellung

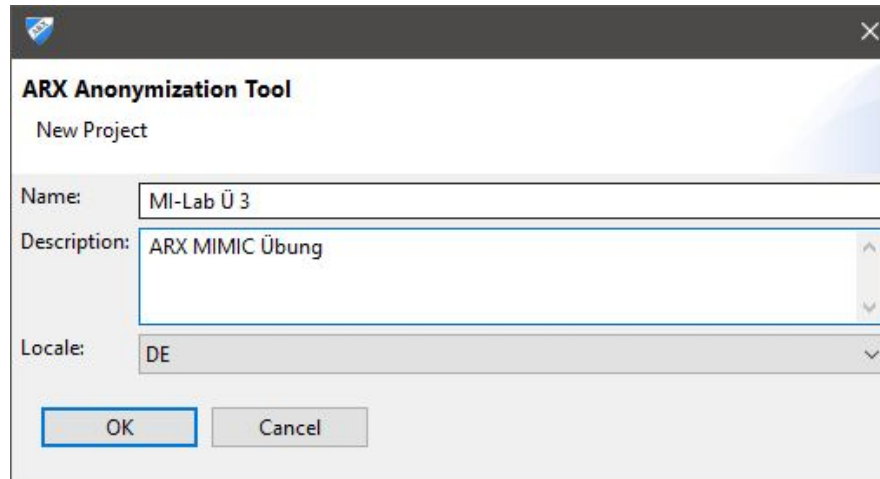
- Erstellen Sie eine ARX Projektdaten für einen gegebenen Datensatz

## Dateien

- Eingabedaten
  - MIMIC Ü3 full.csv
- Hierarchien
  - Age.csv, insurance.csv, language.csv, marital-status.csv, race.csv, weight.csv

## PROJEKT ANLEGEN

- Öffnen Sie ARX und wählen Sie aus dem Anwendungsmenü „File“ → „New project...“
- Vergeben Sie einen Projektnamen und klicken Sie “OK”



The screenshot shows the 'ARX Anonymization Tool' window with the 'New Project' tab selected. It contains three input fields: 'Name' with the text 'MI-Lab Ü 3', 'Description' with the text 'ARX MIMIC Übung', and 'Locale' with the dropdown menu set to 'DE'. At the bottom are 'OK' and 'Cancel' buttons.

ARX Anonymization Tool

New Project

Name: MI-Lab Ü 3

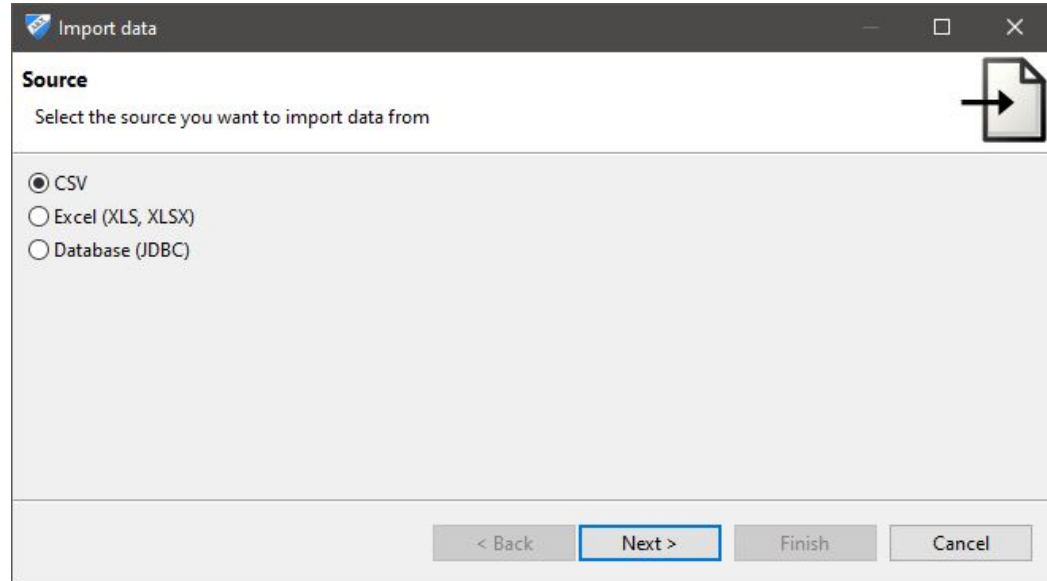
Description: ARX MIMIC Übung

Locale: DE

OK Cancel

## DATEN IMPORTIEREN (1)

- Wählen Sie aus dem Anwendungsmenü „File“ → „Import data“
- Wählen Sie "CSV" und klicken Sie „Next“
- Wählen Sie "Browse...", und selektieren Sie die Datei „MIMIC Ü3.csv“



## DATEN IMPORTIEREN (2)

- Die Syntax der Datei wird automatisch erkannt

Import data

CSV

Please provide the information requested below

Location: D:\IMISE-Arbeit\NFDI4Health\Uebung-3-ARX\MIMIC IV Demo full Ü3.csv Browse...

Charset: WINDOWS-1252 (System default)

Delimiter: ,

Quote: "

Escape: "

Linebreak: Windows

☒ First row contains column names

subject_id	age	gender	admission_type	insurance	language	marital_status	race	vor
10014729	21	F	SURGICAL SA...	Other	ENGLISH	SINGLE	WHITE - OTHE...	
10003400	72	F	EW EMER.	Medicare	ENGLISH	MARRIED	BLACK/AFRIC...	I48
10002428	80	F	OBSERVATION...	Medicare	ENGLISH	WIDOWED	WHITE	
10032725	38	F	EW EMER.	Other	ENGLISH	SINGLE	BLACK/AFRIC...	
10027445	48	F	EW EMER.	Other	ENGLISH	WIDOWED	WHITE	
10037928	78	F	OBSERVATION...	Medicare	?	WIDOWED	HISPANIC/LAT...	
10001725	46	F	EW EMER.	Other	ENGLISH	MARRIED	WHITE	
10040025	64	F	AMBULATORY...	Other	ENGLISH	DIVORCED	WHITE	I48
10008454	26	F	EW EMER.	Other	ENGLISH	SINGLE	WHITE	
10020640	91	F	EW EMER.	Other	ENGLISH	WIDOWED	WHITE	
10014078	60	F	EW EMER.	Medicaid	ENGLISH		UNABLE TO O...	

< Back **Next >** Finish Cancel



## EINGABEDATEI KONFIGURIEREN

- Im folgenden Dialog können Sie die Eingabedatei konfigurieren

Import data

Columns

Click right to edit

Selected	Name	Data type	Format
✓	subject_id	Integer	
✓	age	Integer	
✓	gender	String	
✓	admission_type	String	
✓	insurance	String	
✓	language	String	
✓	marital_status	String	
✓	race	String	
✓	vorhofflimmern	String	
✓	hyperlipidaemie	String	
✓	creatinine	String	
✓	hematocrit	Decimal	#,##0.###
✓	infusion	String	
✓	ventilation	String	
✓	ibuprofen	String	
✓	fentanyl	String	
✓	weight	String	

Move up Move down

☒ Perform data cleansing

< Back Next > Finish Cancel

# VORANSICHT

- Es erscheint eine Übersicht über den Datensatz

Import data

**Preview**  
Please check whether everything is right

subject_id	age	gender	admission_type	insurance	language	marital_status	race	vor
10014729	21	F	SURGICAL SA...	Other	ENGLISH	SINGLE	WHITE - OTHE...	
10003400	72	F	EW EMER.	Medicare	ENGLISH	MARRIED	BLACK/AFRIC...	148
10002428	80	F	OBSERVATION...	Medicare	ENGLISH	WIDOWED	WHITE	
10032725	38	F	EW EMER.	Other	ENGLISH	SINGLE	BLACK/AFRIC...	
10027445	48	F	EW EMER.	Other	ENGLISH	WIDOWED	WHITE	
10037928	78	F	OBSERVATION...	Medicare	?	WIDOWED	HISPANIC/LAT...	
10001725	46	F	EW EMER.	Other	ENGLISH	MARRIED	WHITE	
10040025	64	F	AMBULATORY...	Other	ENGLISH	DIVORCED	WHITE	148
10008454	26	F	EW EMER.	Other	ENGLISH	SINGLE	WHITE	
10020640	91	F	EW EMER.	Other	ENGLISH	WIDOWED	WHITE	
10014078	60	F	EW EMER.	Medicaid	ENGLISH		UNABLE TO O...	
10007795	53	F	DIRECT OBSER...	Medicare	ENGLISH	SINGLE	WHITE	
10001217	55	F	DIRECT EMER.	Other	?	MARRIED	WHITE	
10019172	63	F	ELECTIVE	Other	ENGLISH		UNABLE TO O...	148
10039831	57	F	SURGICAL SA...	Other	ENGLISH	MARRIED	UNABLE TO O...	
10007928	59	F	EW EMER.	Other	ENGLISH	DIVORCED	WHITE	
10019003	65	F	OBSERVATION...	Other	ENGLISH	WIDOWED	WHITE	
10002930	48	F	EU OBSERVATI...	Medicare	ENGLISH	SINGLE	BLACK/AFRIC...	
10010471	89	F	OBSERVATION...	Medicare	ENGLISH	WIDOWED	WHITE	
10027602	71	F	SURGICAL SA...	Other	ENGLISH	SINGLE	WHITE	

< >

< Back Next > Finish Cancel

## ARX GUI

- Klicken Sie auf die Attribute (Spalten) und ändern Sie den Typ

The screenshot displays the ARX Anonymization Tool interface. The 'Input data' table on the left contains the following columns: subject\_id, age, gender, admission\_type, and insurance. The table lists 26 rows of data. The 'Attribute metadata' panel on the right shows the 'Data transformation' tab, where the 'Type' dropdown is set to 'Quasi-identifying' and the 'Transformation' is set to 'Generalization'. The 'Minimum' and 'Maximum' dropdowns are set to 'Sensitive' and 'All' respectively. The 'Sample extraction' panel at the bottom shows 'Size: 100 / 100 = 100%' and 'Selection mode: None'.

	subject_id	age	gender	admission_type	insurance
1	10014729	21	F	SURGICAL SAME...	Other
2	10003400	72	F	EW EMER.	Medicare
3	10002428	80	F	OBSERVATION A...	Medicare
4	10032725	38	F	EW EMER.	Other
5	10027445	48	F	EW EMER.	Other
6	10037928	78	F	OBSERVATION A...	Medicare
7	10001725	46	F	EW EMER.	Other
8	10040025	64	F	AMBULATORY O...	Other
9	10008454	26	F	EW EMER.	Other
10	10020640	91	F	EW EMER.	Other
11	10014078	60	F	EW EMER.	Medicaid
12	10007795	53	F	DIRECT OBSERV...	Medicare
13	10001217	55	F	DIRECT EMER.	Other
14	10019172	63	F	ELECTIVE	Other
15	10039831	57	F	SURGICAL SAME...	Other
16	10007928	59	F	EW EMER.	Other
17	10019003	65	F	OBSERVATION A...	Other
18	10002930	48	F	EU OBSERVATION	Medicare
19	10010471	89	F	OBSERVATION A...	Medicare
20	10027602	71	F	SURGICAL SAME...	Other
21	10031757	67	F	DIRECT EMER.	Other
22	10016742	58	F	OBSERVATION A...	Medicaid
23	10039997	67	F	SURGICAL SAME...	Medicare
24	10012853	91	F	DIRECT OBSERV...	Medicare
25	10000032	52	F	EW EMER.	Medicaid
26	10031404	82	F	URGENT	Other

- 
- The screenshot displays the ARX Anonymization Tool interface. The main window is titled "ARX Anonymization Tool - MI-Lab U3". It features a menu bar (File, Edit, View, Help) and a toolbar with various icons for file operations and analysis. Below the toolbar, there are tabs for "Configure transformation", "Explore results", "Analyze utility", and "Analyze risk". The central area is divided into two panes. The left pane, labeled "Input data", shows a table with columns: subject\_id, age, gender, admission\_type, and insurance. It contains 26 rows of data, each with a checkbox in the first column. The right pane, labeled "Data transformation", shows settings for "Attribute metadata". Under "Type", it's set to "Quasi-identifying". Under "Transformation", it's set to "Generalization". There are also "Minimum" and "Maximum" dropdowns both set to "All". Below these settings is a table showing generalization levels (Level-0 to Level-4) for different attributes. At the bottom, there are tabs for "Privacy models", "Population", "Costs and benefits", "General settings", "Utility measure", "Coding model", and "Attribute weights". The "General settings" tab is active, showing "Suppression limit" at 0%, "Approximate" checked, and "Precomputation" unchecked. A "Sample extraction" section at the very bottom shows "Size: 100 / 100 = 100%" and "Selection mode: None".



UNIVERSITÄT  
LEIPZIG

Medizinische Fakultät

# VIELEN DANK!

**Mona Perbix**

Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)

[www.imise.uni-leipzig.de](http://www.imise.uni-leipzig.de)