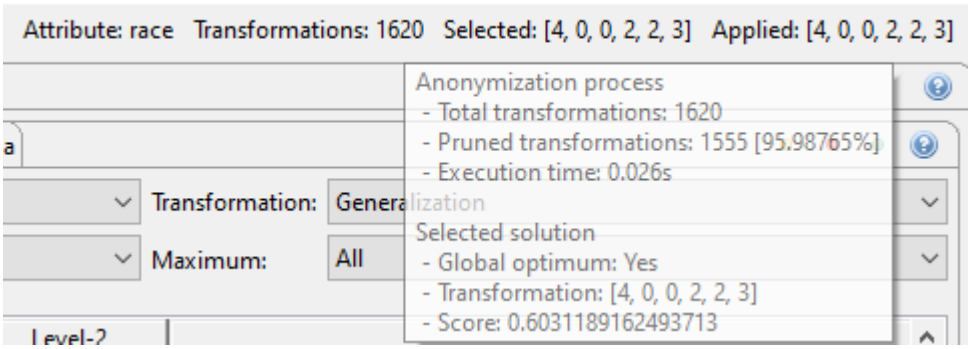
ARX Übung MIMIC Musterlösung

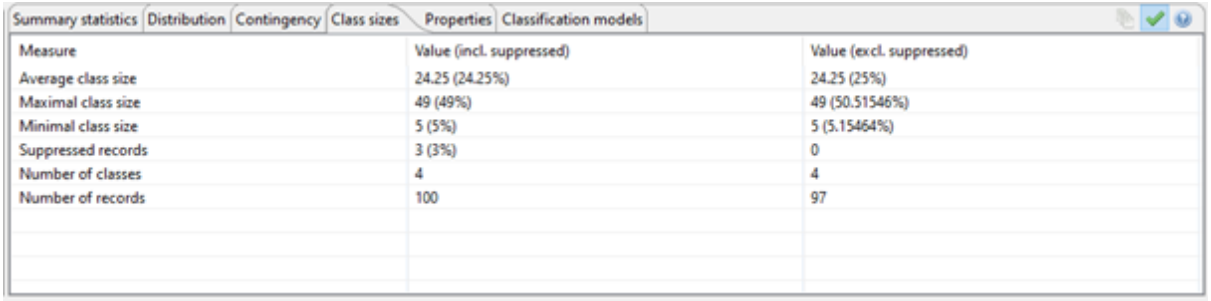
Aufgabe 1)
Import Zwischenlösung: mimic_ü2.deid



Optimale Lösung:

[4, 0, 0, 2, 2, 3]
1620 Transformationen
Informationsverlust (Score) = 60 %

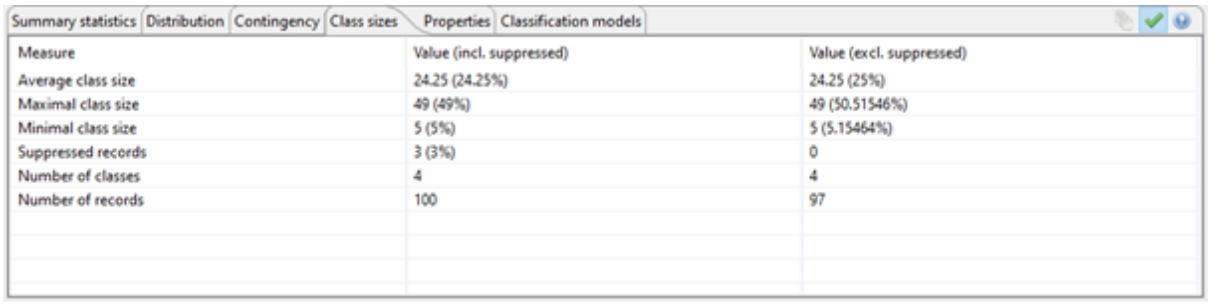
Aufgabe 2)
Import Zwischenlösung: mimic_ü2.deid



Measure	Value (incl. suppressed)	Value (excl. suppressed)
Average class size	24.25 (24.25%)	24.25 (25%)
Maximal class size	49 (49%)	49 (50.51546%)
Minimal class size	5 (5%)	5 (5.15464%)
Suppressed records	3 (3%)	0
Number of classes	4	4
Number of records	100	97

Optimale Transformation:

[4, 0, 0, 2, 2, 3]
Vollständig entfernte Attribute: age, martial-status, race, weight
Unterdrückte Einträge: 3 (3%)
Informationsverlust: 60 %
Average class size: 24 (24%)



Measure	Value (incl. suppressed)	Value (excl. suppressed)
Average class size	24.25 (24.25%)	24.25 (25%)
Maximal class size	49 (49%)	49 (50.51546%)
Minimal class size	5 (5%)	5 (5.15464%)
Suppressed records	3 (3%)	0
Number of classes	4	4
Number of records	100	97

Zweit beste Transformation:

[4, 0, 1, 2, 2, 3]
Vollständig entfernte Attribute: age, martial-status, race

Unterdrückte Einträge: 3 (3%)
Informationsverlust: 60 %
Average class size: 24 (24%)

Aufgabe 3)
Transformation: [4, 0, 0, 2, 3, 3] (Optimale Lösung)

Prosecutor Re-Identifikationsrisiko für den Eingabedatensatz

- **100 %** der Einträge sind einem Risiko von mehr als 20% ausgesetzt
- Die zu erwartende relative Anzahl an korrekt re-identifizierten Einträgen ist **99 %**
- Das niedrigste Risiko ist **50 %**, das höchste Risiko ist **100 %**
- Das höchste Risiko betrifft **98 %** der Datensätze

Prosecutor Re-Identifikationsrisiko für den Ausgabedatensatz

- **0 %** der Einträge sind einem Risiko von mehr als 20% ausgesetzt
- Die zu erwartende relative Anzahl an korrekt re-identifizierten Einträgen ist **4 %**
- Das niedrigste Risiko ist **2 %**, das höchste Risiko ist **20 %**
- Das höchste Risiko betrifft **20 %** der Datensätze

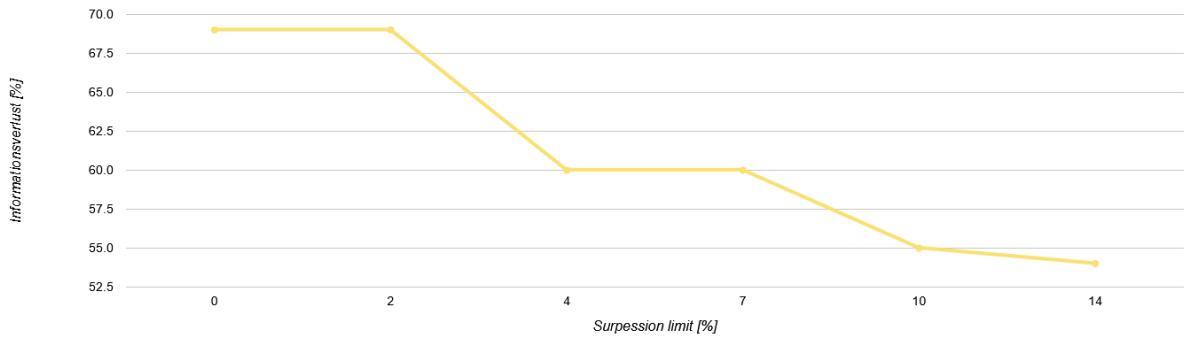
Feststellung 1: Im Worst Case können trotz 5-Anonymity der Einträge korrekt reidentifiziert werden

Feststellung 2: Im Schnitt liegt das Risiko für einzelne Einträge deutlich unter dem Grenzwert von 20%

Aufgabe 6)

Grenzwert ist: **14 %** (Anzahl surpressed Records bei 100 %)

14: 54 %, 10:55%, 7:60%, 4:60%, 2:69%, 0:69%

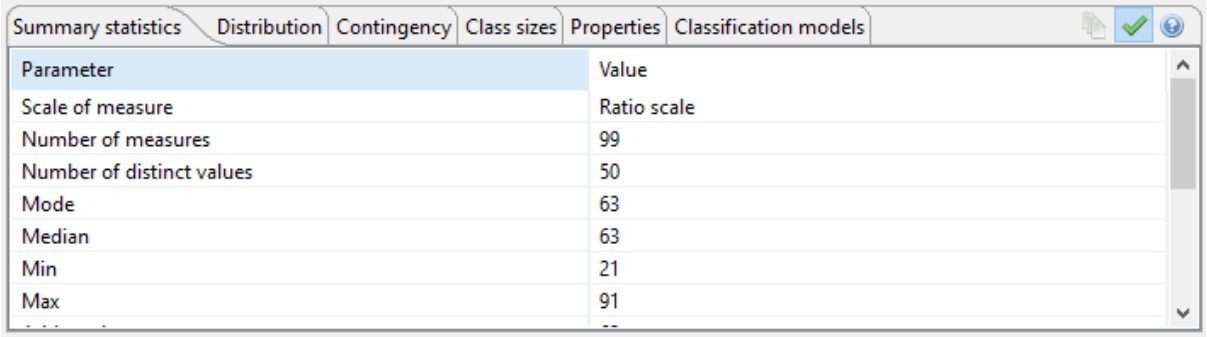


Feststellung 1: Starker Rückgang des Informationsverlustes
- Es ist meist ausreichend einen sehr kleinen Teil der Einträge zu entfernen

Feststellung 2: Keine starke Abflachung (nur ab 7 % Suppression limit)

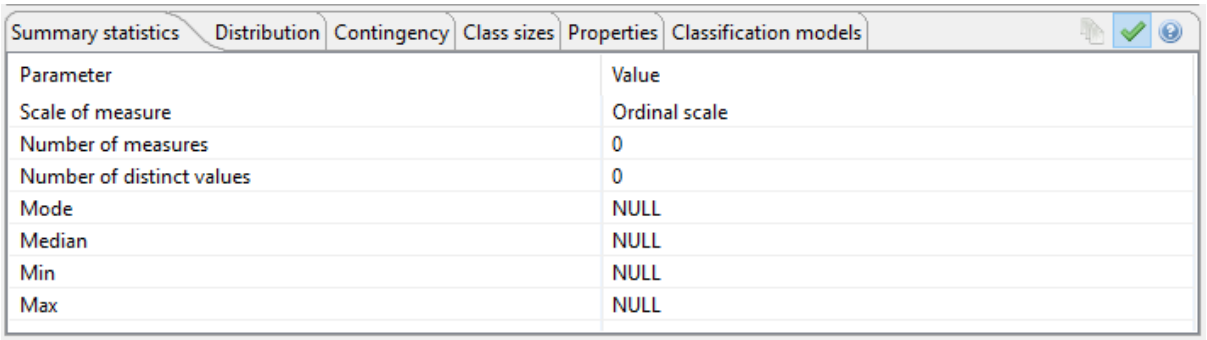
Weitere Feststellung: Ausführungszeiten von ARX steigen mit zunehmendem „Suppression limit“

Aufgabe 7)
Import Zwischenlösung: mimic_ü7.deid
Suppression Limit= 5%
Input:



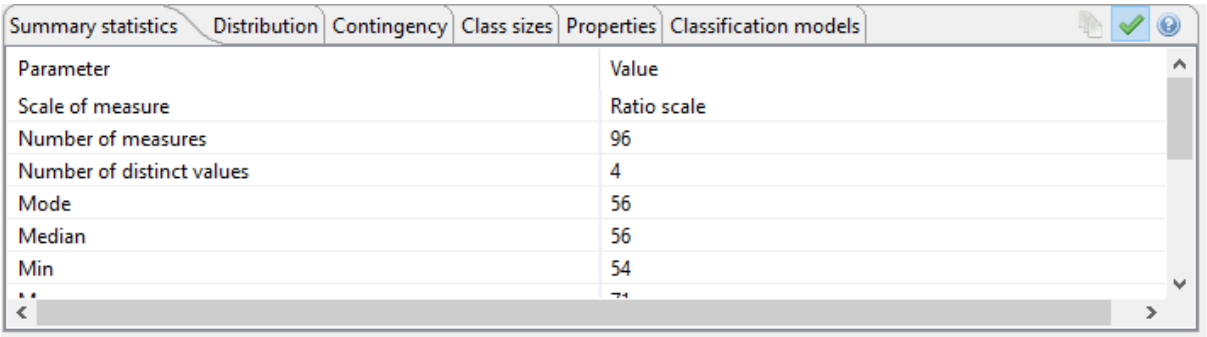
Parameter	Value
Scale of measure	Ratio scale
Number of measures	99
Number of distinct values	50
Mode	63
Median	63
Min	21
Max	91

Generalization:



Parameter	Value
Scale of measure	Ordinal scale
Number of measures	0
Number of distinct values	0
Mode	NULL
Median	NULL
Min	NULL
Max	NULL

Microaggregation:



Parameter	Value
Scale of measure	Ratio scale
Number of measures	96
Number of distinct values	4
Mode	56
Median	56
Min	54

	Input	Generalization	Microaggregation
Scale of measure	Ratio	Ordinal	Ratio
Mode	63	NULL	56
Median	63	NULL	56
Min	21	NULL	54
Max	91	NULL	71

Suppressed records	0	3 %	3 %

Feststellung 1: Eine generalisierte numerische Variable ist in ARX ordinalskaliert
Feststellung 2: Mikroaggregation erhält das Skalenniveau
Feststellung 3: Mikroaggregation ist nicht wahrheitserhaltend