## Exam: Introduction to information privacy

**Exercise 1.** Discuss how can clustering be used for data privacy. Discuss how disclosure risk and information loss can be controlled for this kind of approach.

**Exercise 2.** Let us consider the following four masking methods  $M_1$ ,  $M_2$ ,  $M_{id}$ ,  $M_r$ .

Methods  $M_1$  and  $M_2$  have been implemented and for a particular file under discussion we have the following.  $M_1$  achieves an information loss of 25 and a disclosure risk of 40. In contrast, the second one  $M_2$  achieves an information loss of 90 and a disclosure risk of 50.

Let us assume two other masking methods. X corresponds to the original file and Y to the protected file.

- 1. Method  $M_{id}$  is the identity function. That is, the protected file Y is made equal to the original file X.
- 2. Method  $M_r$  is a random function. The method replaces the original matrix by another with random numbers.

Discuss information loss and disclosure risk for these four methods. Discuss the effectiveness of these two methods, and compare with the results of the other ones represented in the figure.

## 

DR

**Exercise 3.** Consider the table below that represents a data file in a compact way. Rows represent the combinations of values for the four variables in the file, and the last row counts the number of records in the file with such combination. E.g., the first row informs us that there are 20 records with the values (10th, M, 40, 20Y0N).

Explain briefly k-anonymity.

Discuss whether this file satisfies k-anonymity for k = 2, k = 4 and k = 6, and if not, propose an updated file (with either supression, recoding) satisfying k-anonymity for these values of k.

Education	Sex	Work-Hrs	Class	Number of Records
10th	M	40	20Y0N	20
10th	M	30	0Y4N	4
9th	M	30	0Y2N	2
9th	F	30	0Y4N	4
9th	F	40	0Y6N	6
8th	F	30	0Y2N	2
8th	F	40	0Y2N	2

Exercise 4. Discuss the effect of rank swapping into correlation of variables.

Exercise 5. Discuss the effects of microaggregation on the mean and the variance of microaggregated variables.

**Exercise 6.** Discuss the information loss of a multidimensional algorithm for k-anonymization with respect to the information loss of a single-dimensional algorithm. Definition of single-dimensional and multidimensional recoding can be found in the following reference:

LeFevre, K., DeWitt, D. J., Ramakrishnan, R. () Mondrian Multidimensional K-Anonymity, Proc. ICDE 2006.

http://pages.cs.wisc.edu/lefevre/MultiDim.pdf