Stockholms Universitet, Statistisk

Exam in: **Statistical databases and register**

Examiner: Mikael Möller and Bo Sundgren

Approved aids: Pen and pencil

Examination day: 110214

Examination time: 5 hours

All assumptions and notations should be explained and defined (also those that have been used during the course). All answers, reasoning and explanations should be easy to follow. Answers and arguments which cannot be understood give 0 p. The questions do not come in an order of difficulty so read the whole exam before you start with your solutions. I will visit you during the second hour.

Good luck!

1: Give short definitions and descriptions of the following concepts: (10p)

- microdata

- macrodata

- metadata

- statistical characteristic

- statistic

2: During the course you were taught how to make a graphical model, a so-called object (10p) graph, of a piece of reality, using different symbols for objects, variables of objects, and relations between objects. The film renting activities of a videoshop was used as an example, and many more examples were given from the world of official statistics. Now you are asked to **make an object graph** for the following piece of reality:
A university gives courses to students, who attend the courses. Teachers teach the courses. Students, courses, and teachers are identified by student numbers, course numbers, and person numbers, respectively. Students and teachers are characterised by name, address, sex, age, and citizenship. Courses are characterised by topic, name, and points.

3: The object graph you drew in the previous task is used as a basis for the design of (10p) a database, implemented by means of a relational database management system. Solve the following tasks:

- Make a drawing of the relational data model, indicating relational tables, corresponding to object types (for example "course"), with columns corresponding to variables (for example "points") and rows corresponding to object instances (for example individual courses). Also indicate how different relational tables refer to each other, representing relations between object types.

– The database can be used for producing statistics about the company. An example of a statistic that can be produced is "number of teachers by sex, age, and citizenship". Give examples of statistics that may be produced about the following objects:

  – courses

  – teachers

  – course attendances (by students; note that a student may attend several courses in parallel)

4: Give at least two examples of how registers can be used for improving the quality of traditional statistical surveys, and at least two examples of how surveys can be used for improving the quality of registers. (10p)

5: Eurostat and other international organisations and national statistical agencies use a quality concept based upon a number of quality dimensions. Describe five important quality dimensions and discuss them briefly with respect to (a) traditional statistical surveys; (b) register-based statistics. (10p)

6: Describe the method of linear classification and derive the classification line. (10p)

7: Describe Hebbian learning and give an example for the logical OR table. Start with the weights $w_0 = w_1 = w_2 = 0$. (10p)

8: Decribe the basic ideas behind hierarchical clustering. (10p)