Stockholms Universitet, Statistisk

Exam in: **Statistical databases and register**

Examiner: Mikael Möller and Bo Sundgren

Approved aids: Pen and pencil

Examination day: 110114

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)

- database
- statistical database
- register
- data warehouse
- statistical characteristic
- statistic

2: During the course you were taught how to make a graphical model, a so-called object 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: (10p)

A company produces different products. Each product is identified by means of a product identification number and belongs to a category and has a price per unit. The company sells products to customers. Each customer is identified by means of a customer number and belongs to a category and has a name and an address. When a customer wants to buy something from the company, the customer sends an order to the company via the Internet, telling which products the customer wants to buy, and how many units to be delivered of each product. The company then delivers the products to the customer together with an invoice stating what has been delivered, and how much the customer has to pay.

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

- Make a drawing of the relational data model, indicating relational tables, corresponding to object types (for example "product"), with columns corresponding to variables (for example "price") and rows corresponding to object instances (for example individual products). 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 products by category". Give examples of statistics that may be produced about the following objects:
  - customers
  - orders
  - deliveries

4: Give three advantages and three problems with producing statistics from registers rather than from traditional statistical surveys. (10p)

5: Give one or more methods for managing "missing data" (non-response) in registers. Which method is the best method according to Wallgren & Wallgren? (10p)

6: Give three arguments why statistical databases and registers must contain good metadata. (10p)

7: Describe the method of nearest neighbour. Use relevant pictures as part of your explanations. (10p)

8: Give an example of three clusters where a bad choice of starting points, in K-means, will not give three separate clusters as a solution. (10p)