# Stockholms Universitet, Statistisk

Exam in: Statistical databases and register Examiner: Mikael Möller and Bo Sundgren Approved aids: Pen and pencil

## Examination day: 120113

#### 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.

Good luck!

1: On last page there is a presentation of some statistics from a Eurostat publication. (10p)

- a. The presentation contains both data and metadata. Which is which, that is, which parts of the presentation are data, and which are metadata?
- b. What is the purpose of the metadata in this presentation? Why are the metadata needed?
- c. Give some examples of other purposes that metadata may have in connection with statistical data, for example in connection with statistical data stored in a statistical database?

2: Let us assume that the statistics presented in Question 1 have been produced from (10p) a population database at Eurostat (the statistical office of the European Union), which contains microdata about all people living in the European Union, including data about important events for these people, such as births and deaths, marriages and divorces, migration events, etc.

- a. What is microdata as opposed to macrodata?
- b. As was shown during the course, the information contents of a statistical database may be expressed by means of a conceptual model, illustrated by an object graph. Outline an object graph illustrating the contents of the population database just mentioned. Show at least the object types, variables, and object relations that are needed to express the statistics in the presentation in the Question 1.

**3:** The Nordic countries have a long tradition of using administrative data from administrative register for producing official statistics, so-called register-based statistics production – as opposed to statistics produced by traditional statistical survey, where the data are collected directly for statistical purposes by statistical agencies. Many countries all over the world are now following the Nordic example.

#### Exam in: Statistical databases and register

- a. One argument in favour of register-based statistics production is that it is costeffective in comparison with traditional statistical surveys. A Dutch study, referred to by Wallgren & Wallgren in their book on register-based statistics, argues that a conventional census in Holland would be about x times as expensive as a census based on administrative data? Which is the figure x that the Dutch study mentions according to Wallgren & Wallgren? Is it 3, 5, 10, or 100?
- b. There are other important arguments than cost-effectiveness in favour of registerbased statistics. Give at least 3 such arguments!
- c. On the other hand, there are also arguments in favour of using traditional statistical surveys rather than administrative data for producing official statistics. Mention at least one such argument preferably the most important argument that you can think of!

### 4: Provide

(10p)

- a. a basic definition of a register in a strict and narrow sense;
- b. a definition of an extended register, a register in a broader sense;
- c. a definition of the difference between an administrative register and a statistical register;
- d. a definition of a register-based statistical system. Use the definitions presented to you on PowerPoint slides during the course.

**5:** Eurostat and several national statistical offices use a quality concept based on a (10p) number of quality components, typically something like 5 or 6 components.

- a. "Coherence" is one component in Eurostat's quality concept. What does it mean that statistical data are coherent?
- b. Name and define briefly at least 3 other quality components, which are often used by producers of official statistics in their quality declarations.

**6:** In statistical learning we have two main streams of algorithms - Supervised and (10p) unsupervised learning algorithms. Your task is to give

- a. a definition of supervised learning
- b. a definition of unsupervised learning
- c. three examples of algorithms for supervised learning
- d. three examples of algorithms for unsupervised learning

(forts)

# Exam in: Statistical databases and register

(forts)

7: Suppose we have a single neuron and we want it to learn the rules for  $x_1 \vee \neg x_2$  i.e (10p) we want it to learn the table

$$\begin{array}{cccccc} x_1 & \neg x_2 & D = x_1 \lor \neg x_2 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{array}$$

Start with the weights  $w_0 = w_1 = w_2 = 0$  and  $x_0 = 1$ . Let the activation function  $\phi$  be the treshold model. Calculate the final weights. Hint: You will need less than 8 iterations.

8: First give a written description of the 4 steps for the algorithm K-means. Second (10p) state the expression, in simplified form, that is minimized under the constraint p clusters. To answer the last question, it is more easy to give the sum of squares to minimize and from this sum deduce the answer.

#### AGE STRUCTURE

The impact of demographic ageing within the EU is likely to be of major significance in the coming decades. Consistently low birth rates and higher life expectancy will transform the shape of the EU-27's age pyramid; probably the most important change will be the marked transition towards a much older population and this trend is already becoming apparent in several Member States. As a result, the proportion of people of a working age in the EU-27 is shrinking at the same time as those who are considering retirement expands. The share of older persons in the total population will increase significantly from 2010 onwards, as the post-war baby-boom generation starts to reach retirement.

Low birth rates and rising life expectancy will likely result in a much older population structure, such that the ratio of the number of working-age people to those aged over 65 will be reduced from 4:1 in 2008 to less than 2:1 by 2060.

Figure 2.3: Age pyramid, EU-27, 2008 (% of total population)



4

10

Source: Eurostat (tps00010)

Table 2.6: Marriage and divorces (per 1 000 population)

|           | 0-14 | 15-24 | 25-49 | 50-64 | 65-79 | 80+ |
|-----------|------|-------|-------|-------|-------|-----|
| EU-27     | 15.7 | 12.5  | 36.2  | 18.6  | 12.7  | 4.3 |
| Euro area | 15.5 | 11.7  | 36.5  | 18.4  | 13.2  | 4.6 |
| BE        | 16.9 | 12.1  | 35.1  | 18.8  | 12.4  | 4.7 |
| BG        | 13.4 | 13.1  | 35.7  | 20.5  | 13.7  | 3.6 |
| CZ        | 14.2 | 13.0  | 37.0  | 21.2  | 11.2  | 3.4 |
| DK        | 18.4 | 11.7  | 34.4  | 19.9  | 11.5  | 4.1 |
| DE        | 13.7 | 11.6  | 36.0  | 18.6  | 15.3  | 4.6 |
| EE        | 14.8 | 15.2  | 34.8  | 18.0  | 13.5  | 3.7 |
| IE        | 20.6 | 14.1  | 38.9  | 15.5  | 8.2   | 2.7 |
| EL        | 14.3 | 11.2  | 37.6  | 18.3  | 14.6  | 4.1 |
| ES        | 14.6 | 11.2  | 40.6  | 16.9  | 12.0  | 4.6 |
| FR        | 18.5 | 12.8  | 33.7  | 18.6  | 11.5  | 4.9 |
| IT        | 14.0 | 10.2  | 37.2  | 18.6  | 14.6  | 5.5 |
| CY        | 17.4 | 15.4  | 37.4  | 17.3  | 9.7   | 2.8 |
| LV        | 13.8 | 15.6  | 35.7  | 17.8  | 13.7  | 3.5 |
| LT        | 15.4 | 15.9  | 36.1  | 16.8  | 12.6  | 3.3 |
| LU        | 18.2 | 11.8  | 38.7  | 17.3  | 10.6  | 3.4 |
| HU        | 15.0 | 12.7  | 35.7  | 20.4  | 12.5  | 3.7 |
| MT        | 16.2 | 14.1  | 34.5  | 21.3  | 10.7  | 2.8 |
| NL        | 17.9 | 12.1  | 35.6  | 19.7  | 11.0  | 3.8 |
| AT        | 15.4 | 12.3  | 37.4  | 17.8  | 12.6  | 4.6 |
| PL        | 15.5 | 15.5  | 36.0  | 19.6  | 10.5  | 3.0 |
| PT        | 15.3 | 11.6  | 37.4  | 18.2  | 13.2  | 2.5 |
| RO        | 15.2 | 14.6  | 37.0  | 18.2  | 12.1  | 2.8 |
| SI        | 13.9 | 12.2  | 37.6  | 19.9  | 12.7  | 3.6 |
| SK        | 15.8 | 15.3  | 38.1  | 18.8  | 9.4   | 2.6 |
| FI        | 16.9 | 12.4  | 32.7  | 21.5  | 12.2  | 4.3 |
| SE        | 16.8 | 13.0  | 33.0  | 19.6  | 12.2  | 5.3 |
| UK        | 17.6 | 13.4  | 34.9  | 18.0  | 11.6  | 4.5 |
| HR        | 15.4 | 12.6  | 35.0  | 19.4  | 13.9  | 3.2 |
| МК        | 18.5 | 15.9  | 37.0  | 17.3  | 9.7   | 1.7 |
| TR        | 26.4 | 17.6  | 37.0  | 11.9  | 5.9   | 1.1 |
| IS        | 20.9 | 14.7  | 36.4  | 16.5  | 8.4   | 3.2 |
| LI        | 16.8 | 12.2  | 38.6  | 20.1  | 9.3   | 3.1 |
| NO        | 19.2 | 12.7  | 34.9  | 18.6  | 10.0  | 4.6 |
| СН        | 15.5 | 11.9  | 37.2  | 19.0  | 11.7  | 4.7 |

#### Table 2.3: Population by age class, 2008 (% of total population)

Source: Eurostat (demo\_pjan)

#### MARRIAGES AND DIVORCES

A marriage is the act, ceremony or process by which the legal relationship of husband and wife is constituted. Divorce is defined as the final legal dissolution of a marriage, that is, a separation of husband and wife which confers on the parties the right to remarry under civil, religious and/or other provisions. Divorce is possible in all of the Member States, except Malta.

Crude marriage and divorce rates measure the number of marriages/divorces in relation to the average population; these rates are expressed per 1 000 inhabitants.





Source: Eurostat (tps00012)

|           | N    | Marriages (1) |      |      | Divorces (2) |  |  |
|-----------|------|---------------|------|------|--------------|--|--|
|           | 1998 | 2003          | 2008 | 1998 | 2003         |  |  |
| EU-27     | 5.1  | 4.9           | 4.9  | 1.8  | 2.0          |  |  |
| Euro area | :    | 4.7           | 4.5  | :    | 1.8          |  |  |
| BE        | 4.4  | 4.0           | 4.4  | 2.6  | 3.0          |  |  |
| BG        | 4.3  | 3.9           | 3.6  | 1.3  | 1.5          |  |  |
| CZ        | 5.4  | 4.8           | 5.0  | 3.1  | 3.8          |  |  |
| DK        | 6.6  | 6.5           | 6.8  | 2.5  | 2.9          |  |  |
| DE        | 5.1  | 4.6           | 4.6  | 2.3  | 2.6          |  |  |
| EE        | 3.9  | 4.2           | 4.6  | 3.2  | 2.9          |  |  |
| IE        | 4.5  | 5.1           | 5.2  | 0.4  | 0.7          |  |  |
| EL        | 5.1  | 5.5           | 4.6  | 0.7  | 1.1          |  |  |
| ES        | 5.2  | 5.1           | 4.5  | 0.9  | 1.1          |  |  |
| FR        | 4.6  | 4.6           | 4.3  | 2.0  | 2.1          |  |  |
| IT        | 4.9  | 4.5           | 4.1  | 0.6  | 0.8          |  |  |
| CY        | 11.4 | 7.7           | 7.5  | 1.3  | 2.0          |  |  |
| LV        | 4.0  | 4.3           | 5.7  | 2.6  | 2.1          |  |  |
| LT        | 5.2  | 4.9           | 7.2  | 3.3  | 3.1          |  |  |

2008

1.9

| DK | 6.6  | 6.5 | 6.8 | 2.5 | 2.9 | 2.7 |
|----|------|-----|-----|-----|-----|-----|
| DE | 5.1  | 4.6 | 4.6 | 2.3 | 2.6 | 2.3 |
| EE | 3.9  | 4.2 | 4.6 | 3.2 | 2.9 | 2.6 |
| IE | 4.5  | 5.1 | 5.2 | 0.4 | 0.7 | 0.8 |
| EL | 5.1  | 5.5 | 4.6 | 0.7 | 1.1 | 1.2 |
| ES | 5.2  | 5.1 | 4.5 | 0.9 | 1.1 | 2.8 |
| FR | 4.6  | 4.6 | 4.3 | 2.0 | 2.1 | 2.2 |
| IT | 4.9  | 4.5 | 4.1 | 0.6 | 0.8 | 0.9 |
| CY | 11.4 | 7.7 | 7.5 | 1.3 | 2.0 | 2.1 |
| LV | 4.0  | 4.3 | 5.7 | 2.6 | 2.1 | 2.7 |
| LT | 5.2  | 4.9 | 7.2 | 3.3 | 3.1 | 3.1 |
| LU | 4.8  | 4.4 | 3.9 | 2.4 | 2.3 | 2.0 |
| HU | 4.4  | 4.5 | 4.0 | 2.5 | 2.5 | 2.5 |
| MT | 6.5  | 5.9 | 6.0 | -   | -   | -   |
| NL | 5.5  | 4.9 | 4.6 | 2.1 | 1.9 | 2.0 |
| AT | 4.9  | 4.6 | 4.2 | 2.2 | 2.3 | 2.5 |
| PL | 5.4  | 5.1 | 6.8 | 1.2 | 1.3 | 1.7 |
| PT | 6.6  | 5.2 | 4.1 | 1.5 | 2.2 | 2.4 |
| RO | 6.5  | 6.2 | 7.0 | 1.8 | 1.5 | 1.7 |
| SI | 3.8  | 3.4 | 3.1 | 1.0 | 1.2 | 1.1 |
| SK | 5.1  | 4.8 | 5.2 | 1.7 | 2.0 | 2.3 |
| FI | 4.7  | 5.0 | 5.8 | 2.7 | 2.6 | 2.5 |
| SE | 3.6  | 4.4 | 5.5 | 2.3 | 2.4 | 2.3 |
| UK | 5.2  | 5.1 | 4.4 | 2.7 | 2.8 | 2.4 |
| HR | 5.4  | 5.0 | 5.3 | 0.9 | 1.1 | 1.1 |
| MK | 7.0  | 7.1 | 7.2 | 0.5 | 0.7 | 0.6 |
| TR | :    | 6.8 | 9.0 | :   | 0.7 | 1.4 |
| IS | 5.6  | 5.3 | 5.1 | 1.8 | 1.8 | 1.7 |
| LI | :    | 4.4 | 5.8 | :   | 2.5 | 2.8 |
| NO | 5.3  | 4.9 | 5.3 | 2.1 | 2.4 | 2.1 |
| СН | 54   | 5.5 | 54  | 25  | 23  | 26  |

(?) EU-27, euro area, Ireland, Spain, Cyprus and the United Kingdom, 2007 instead of 2008; Cyprus, break in series, 2003.
(?) Ireland, Greece, Spain, Cyprus, Austria, Portugal and the United Kingdom, 2007 instead of 2008; France, 2006 instead of 2008.

Source: Eurostat (tps00012 and tps00013)