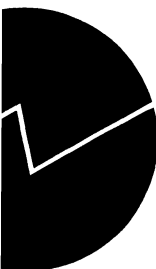


Statistics Norway
Department for Coordination and Development

Hans Viggo Sæbø and Lars Rogstad

**Dissemination of Statistics
on Maps**

Documents



1. Introduction and summary¹

The needs for better presentation and interpretation of statistics, together with improved and more available tools for geographical analyses and statistical mapping, have put geographical information systems (GIS) on the agenda in most national statistical institutes. The widespread use of such tools has contributed to increased demand for regional and georeferenced statistics.

Statistics Norway has taken several steps to meet and to take advantage of this development. This paper presents a CD product ("Statistics on maps") which combines regional statistics with digitised maps and a simple GIS tool for presenting statistics on maps.

There are several issues and purposes behind this project:

- Co-operation between the national statistical institute, the national mapping authority and a GIS company is of mutual benefit to all parties, and such a product would have been difficult or impossible to achieve without such co-operation. For Statistics Norway, the co-operation provides access to digital maps and expertise within GIS and system development.
- Even if the product is interactive in the sense that the user can define the statistics and maps to be presented, emphasis has been put on simplicity of use and on avoiding misuse in the presentation of data. This has been done by including only statistics that are suitable for presenting on maps, and the types of maps the users can make are adapted to the different statistics.
- The present CD-ROM can be regarded as a pilot for later similar products with other statistics and smaller regional units. In particular, this is relevant in connection with the dissemination of data from the year 2000 Population and Housing Census. Statistics for small regional units is a main issue in this census since it will mainly be based on administrative registers.

The project should be seen in connection with current Statistics Norway dissemination strategy of increasing electronic publication in various media on the basis of common databases containing official statistics. Special emphasis is put on regional statistics in this strategy.

The paper gives a presentation of the project, considerations on the issues mentioned above and some examples.

2. Project and product description

The work on "Statistics on maps" started several years ago, but the first step was fulfilled one year ago with a demonstration CD containing statistics and maps for 19 Norwegian counties. This CD was free, and it was distributed to a large number of schools, libraries and some other public institutions, and it is available on Internet for downloading. The idea was to evaluate the market and to get feedback to improve the final product. This new version will be released this autumn and contains data and maps for all 435 Norwegian municipalities.

A product that integrates statistics, maps and mapping tools provides new and powerful possibilities for not only experts within management and research, but for media, education and the public at large. "Statistics on maps" has been designed for the educational sector in addition to the more traditional users within local government administration, and the challenge is to combine flexibility with the necessary simplicity.

¹ A short version of this paper has been presented at the seminar on New Techniques and Technologies for Statistics in Sorrento, Italy, 4 – 6 November 1998 (NTTS 98).

The opening screen shows a menu with statistics to be presented and a map of Norway as an illustration. The menu is organised according to Statistics Norway's standard subject matter grouping. There are about 50 statistical tables with about 200 single data series included in the current version of the CD. The data cover most subject matter areas from environment statistics to industrial activities and public sector statistics. When statistics or data are selected, the regional level and coverage must be selected. Possible regional levels are counties, municipalities and in some cases parts of towns. Choices for coverage are the whole country, parts of the country, one county, the surroundings of one particular municipality or one of the four largest urban municipalities in Norway. With the selection of statistics, regional level and coverage (figure 1), the program has enough input to provide a default map (figure 2). This will be a choropleth map in the case where data represent relative figures (such as income per capita) and a symbol map if data represent total figures (such as the total population in the region). It is possible to change the map foremost by changing the number of classes and the class limits in a choropleth map. To support these choices the data distribution can be displayed (values by region in an ordered diagram). There is also a default choice of palette (from yellow to red for displaying choropleth maps) which can be changed (to for example grey scales). On the map the user can select to show features such as administrative regional borders (counties and municipalities), administrative centres, lakes and rivers. Names of regions and centres may also be shown. The map can be edited by changing objects to be shown, adding texts, legend and changing position on the page before printing or exporting to image formats or as clip art to other programs.

The CD also contains some pre produced maps from the Mapping Authority (in bitmap format), which can be displayed and printed out. These include topographical maps, land use maps, meteorological maps and satellite images.

The product is based on three principles:

- Interactive use
- Simplicity
- Avoid misuse

The application is *interactive* because of the number of possible user selections described above. In particular the construction of classes for choropleth maps should be mentioned. The default choice is four classes each comprising the same number of regions. It is possible to select equal ranges or intervals for each class or self defined intervals. As mentioned, study of the data distribution is useful when selecting intervals.

The application is *simple* compared to GIS software in general and also compared to programs for statistical thematic mapping, since choices are more limited and since the application is integrated with the most relevant statistical series and maps.

Misuse of statistical thematic mapping is an important issue that is briefly discussed in chapter 4. Here it should be pointed out that limitations in the program and in the selections that can be made by the users ensure that misuse is avoided to a large extent.

The application has been developed by using MapObjects from ESRI (also known for the world wide GIS tools ArcInfo and ArcView). Delphi has also been used to develop some of the more complex dialogue boxes. The program runs on a standard Pentium PC with WinNT/95 and minimum 16 MB RAM.

"Statistics on maps" is marketed and sold by a publisher which has the education sector as its special field (Gyldendal Undervisning). A user's guide including installation instructions will follow the product.

Ideally, there are several components that should be reflected in the price:

- Program licences
- Digital maps
- Statistics
- Development and marketing costs

However, CD products have a market price which represent an upper price limit. This price has been reduced during last years, and we have decided to price the product according to a suitable market price compared to other CD-products, and not calculate a price based on the more sophisticated considerations mentioned. The price for a one user licence has been set to NOK 495 (about USD 65), but special prices for several users (i.e. several users at one school) are being considered.

The Norwegian National Centre for Educational Resources has supported the project, and the parties think it is more important to spread the product than necessarily to make a profit on it as such. This is in accordance with price policies of a national statistical institute. As for the private GIS company, the education sector is a minor user of its products today. If the product could contribute to more interest in GIS among students it could turn out to be a profitable investment for such a company in the long run.

3. Co-operation

Statistics Norway has for many years had a co-operation with the Norwegian Mapping Authority. This co-operation has been formalised and based on a general agreement including several fields of co-operation, such as:

- Exchange of experience
- Co-ordination of standards
- Co-ordination of land use statistics
- Statistical thematic mapping in general
- The CD-ROM with statistics on maps

In addition, co-operation with the Norwegian Mapping Authority will be central in connection with the year 2000 Population and Housing Census.

Exchange of experience has been done by allowing employees from the two institutions to participate in each others' courses. One common course in statistical thematic mapping has been set up, and run twice with a total of 25 participants from each institution.

Co-operation on standards comprises standards of georeferencing and exchange formats for data and digital maps.

The national mapping authority has in many ways the same role in mapping as the national statistical institute (NSI) has in official statistics. Information technology has increased the possibilities for bordering or overlapping activities in both institutions, and this alone calls for a close co-operation. Traditionally, digital maps have been expensive and may be not easily available for many NSIs, and co-operation with the mapping authorities ensures availability for statistical purposes at the same time as it may lead to products which are beneficial to both parties.

In the same way, co-operation with private GIS companies provides experience in this field and in IT in general, experience that is scarce in many NSIs. And a product like "Statistics on maps" depends on a suitable GIS tool which depends on agreement with a company. In this case we have co-operated with GEODATA AS which is the Norwegian representative and dealer of ESRI products. A national

statistical institute is independent and emphasises equal conditions for both companies and the public. However, this does not prevent co-operation with private companies on specific products. We can co-operate with different companies in different fields and on different products, but no company gets exclusive rights to our data.

4. Misuse of statistics on maps

Statistical thematic mapping represents a simple and powerful way of presenting statistics, but it can easily be misused. A map reflects geographical patterns, which can be misinterpreted because it most often displays only one dimension (marginal) of a statistical table.

Frequent reasons for misuse of statistics on maps are:

- Irrelevant selection of statistics to be displayed
- Incorrect selection of map type, for example choropleth maps for presenting totals, not only ratios (such as the number of something per inhabitant or per area unit)
- Incorrect or biased groupings used in choropleth maps
- Wrong proportionality factors used in symbol maps

Only statistics where the geographical variation is relevant should be displayed on maps. In the case where statistics are based on rare events (such as the incidents of a seldom illness), a map may indicate a geographical correlation which is not significant. In this project we have tried to avoid statistics which can be misused in this way.

A widespread mistake when displaying statistics on maps is to present totals (such as the number of something or total income in an area) on choropleth maps. This is misleading, since the larger an area is the higher the total will tend to be (merging of two regions will give a higher total). In the project this mistake has been avoided by only allowing choropleth maps to be used for presenting relative figures. For totals a symbol map is the only possibility.

Choropleth maps are based on groupings of the statistical variables. The selection of groups will affect the result. There is no simple answer to what is right or wrong groupings, but one has to be careful to ensure that the map pattern is not biased due to the selected grouping. The possibility to display the data distribution (statistics by region) is important in this context.

In our product the only symbols available are circles, and the areas of these are proportional to the statistical number they represent, thus avoiding the common mistake which consists of using a linear proportionality factor. In fact, it is believed that the most "correct" proportionality factor lies in between linear and square due to the way the human eye conceives the symbols. Anyway, area proportionality is better than linear proportionality. No 3D-effects are possible in the product. Such effects can emphasise a message, but often they lead to exaggerations and misinterpretations.

There are other sources of misusing statistical thematic maps which cannot be completely avoided in any application for presenting statistics on maps. In a choropleth map statistics are visually given weights proportional to the area of the regions. These areas are inhomogenous and in particular the population or economic significance of the regions will vary. The larger cities do not occupy large areas, and their significance will be visually underestimated. A technique that is often used is to present two maps: One choropleth map displaying a relative figure and one symbol map showing one of the corresponding totals (for example income per capita and total population).

A symbol map can have several dimension (combining symbols and choropleth/colours), but such possibilities are not included in the present product. On the other hand, the possibility of making the map too complicated and thus blurring the message is avoided.

5. Future prospects

There will always be a wish to include new features in a possible next version of a product like this. This will depend on the market response to the first version, but some possible extensions could already be mentioned. In addition to more statistical data and more maps (see below) this concerns extended functionality like possibilities to select different symbols and better possibilities for editing the final maps. Possibilities for importing own data and doing calculations on the data should also be considered.

Statistics for small geographical areas is one of the main ideas behind the next Norwegian Population and Housing Census which will be carried out for the year 2000. The provision of statistics for small areas will be facilitated by linking different registers with the register for Ground properties, Addresses and Buildings (GAB) which is georeferenced down to geographical co-ordinates. The population part of the census will be exclusively based on administrative registers. A product like "Statistics on maps" is one of several products to be considered for the dissemination of statistics. In this case maps and statistics will refer to smaller areas than municipalities such as the (statistical) basic units. Norway is divided in about 15 000 basic units compared to 435 municipalities. This product could be aimed at the municipality sector.

Another possible development is to base the updating of later issues on Internet. Subscription could include updating of existing data or even new data sets and maps. The whole content of the CD-ROM could of course be distributed by downloading from Internet as well.

Finally, it will be considered to run the complete application on Internet. Statistics Norway distribute a lot of data on Internet which has become our main channel for dissemination with more than 800 000 hits on our Web pages in October 1998. Data can be downloaded in different formats, but we have no real interactive services on Web so far. However, tests are being made of interactive access to a regional database where users can specify multidimensional statistical tables. This is based on a three level architecture, with data on one server, the application on another and the interface on the local clients. A similar structure will be convenient for a future statistics on maps application on Internet.

6. Examples

The application is in Norwegian, but the figures should still give some ideas of the user interface and possibilities described in chapter 2.

In this case, the following selections are made in figure 1:

- Statistics: Percentage of men per 01.01.1998
- Geographical level: Municipality
- Area: Oslo with neighbouring municipalities

Figure 2 shows the resulting default map with groupings and data distribution in separate windows. In the middle of the map the municipality of Oslo has a lower share of men compared to the surrounding municipalities, as also the data distribution clearly shows.

Figure 3 shows a completed choropleth map of the population density in Norway.

Figure 1. Selection of statistics, geographical level and area

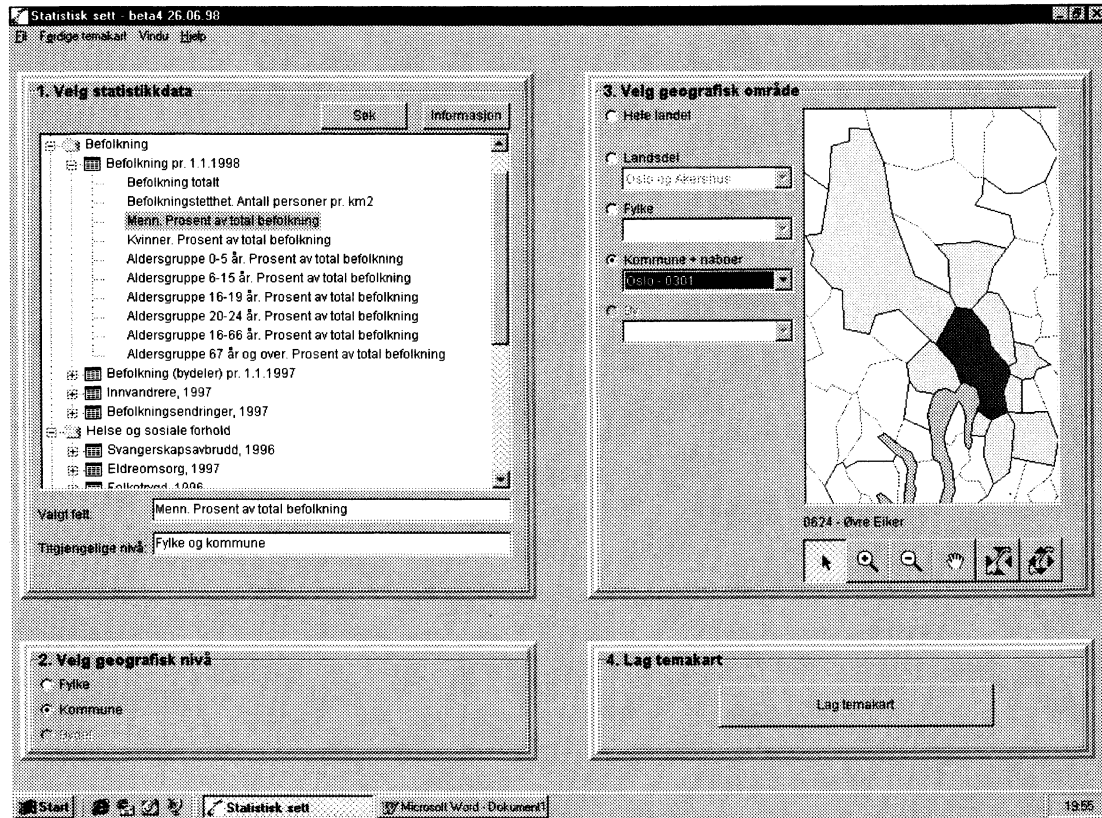


Figure 2. Choropleth map with selection of classes and data distribution

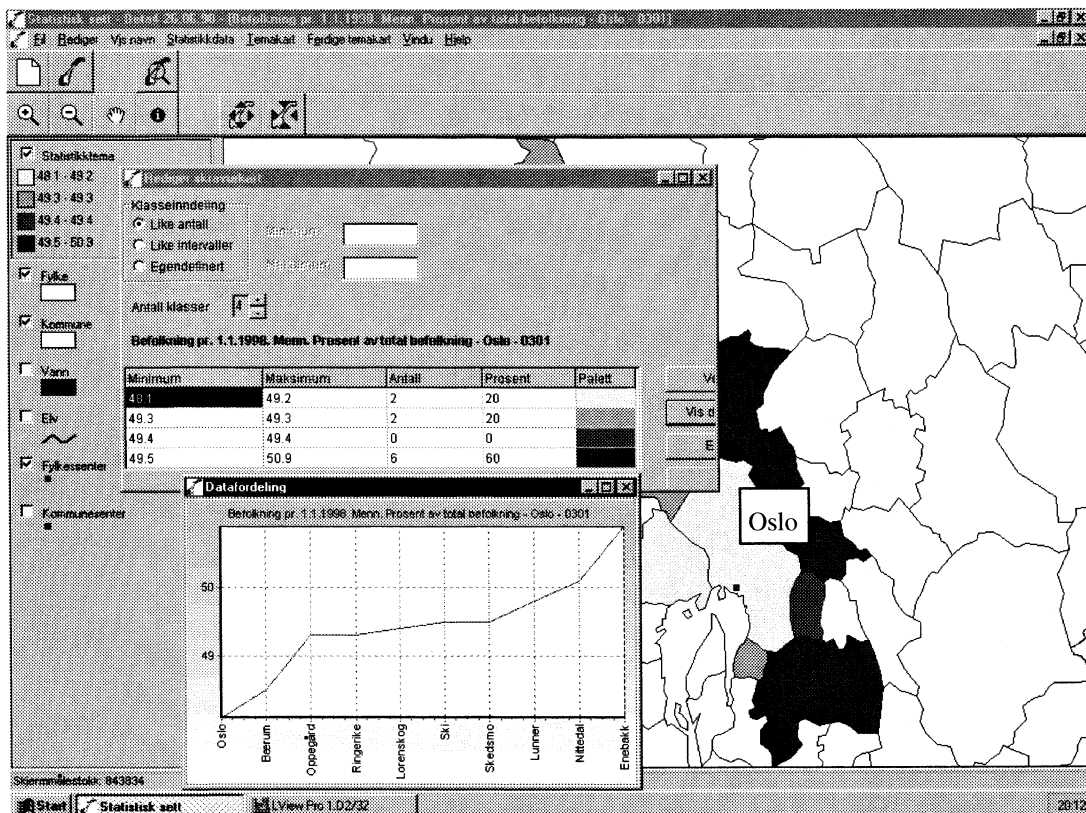
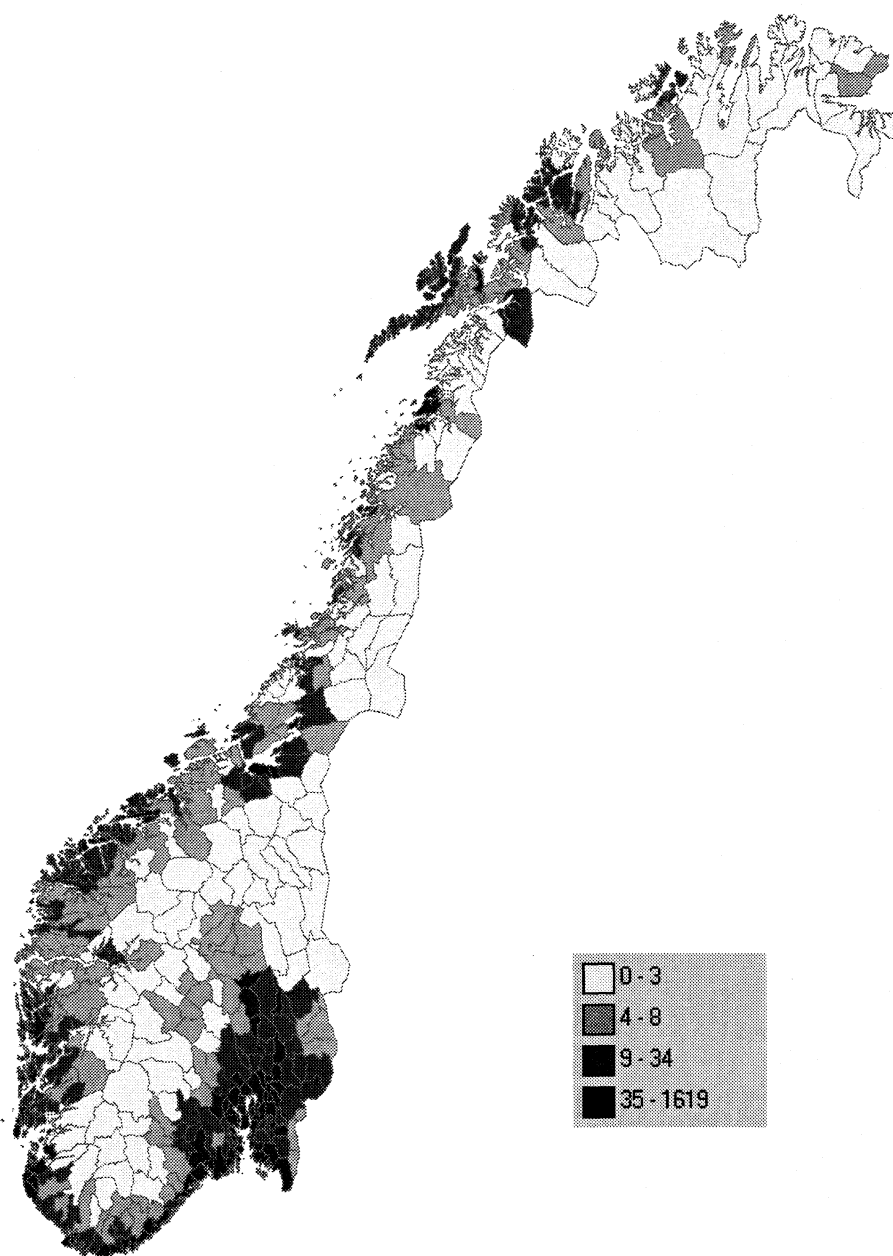


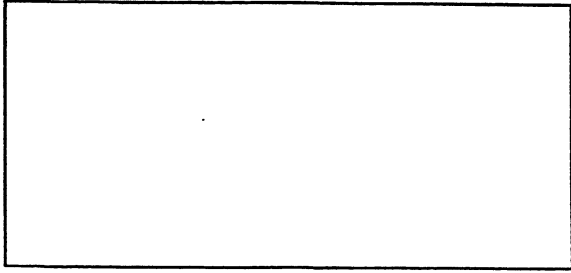
Figure 3. Population density in Norway 01.01.1998. Number of people per square kilometre



Recent publications in the series Documents

- 97/1 T.C. Mykkelbost and K. Rypdal: Material Flow Analysis of Cadmium and Di-2-ethylhexylphthalate (DEHP) in Norway
- 97/2 S. Grepperud: The Impact of Policy on Farm Conservation Incentives in Developing countries: What can be Learned from Theory
- 97/3 M. Rolland: Military Expenditure in Norway's Main Partner Countries for Development Assistance. Revised and Expanded Version
- 97/4 N. Keilman: The Accuracy of the United Nation's World Population Projections
- 97/5 H.V. Sæbø: Managerial Issues of Information Technology in Statistics Norway
- 97/6 E.J. Fløttum, F. Foyn, T.J. Klette, P.Ø. Kolbjørnsen, S. Longva and J.E. Lystad: What Do the Statisticians Know about the Information Society and the Emerging User Needs for New Statistics?
- 97/7 A. Bråten: Technical Assistance on the Jordanian Consumer Price Index
- 97/8 H. Brunborg and E. Aurbakken: Evaluation of Systems for Registration and Identification of Persons in Mozambique
- 97/9 H. Berby and Y. Bergstrøm: Development of a Demonstration Data Base for Business Register Management. An Example of a Statistical Business Register According to the Regulation and Recommendations of the European Union
- 97/10 E. Holmøy: Is there Something Rotten in this State of Benchmark? A Note on the Ability of Numerical Models to Capture Welfare Effects due to Existing Tax Wedges
- 97/11 S. Blom: Residential Concentration among Immigrants in Oslo
- 97/12 Ø. Hagen and H.K. Østereng: Inter-Baltic Working Group Meeting in Bodø 3-6 August 1997 Foreign Trade Statistics
- 97/13 B. Bye and E. Holmøy: Household Behaviour in the MSG-6 Model
- 97/14 E. Berg, E. Canon and Y. Smeers: Modelling Strategic Investment in the European Natural Gas Market
- 97/15 A. Bråten: Data Editing with Artificial Neural Networks
- 98/1 A. Laihonon, I. Thomsen, E. Vassenden and B. Laberg: Final Report from the Development Project in the EEA: Reducing Costs of Censuses through use of Administrative Registers
- 98/2 F. Brunvoll: A Review of the Report "Environment Statistics in China"
- 98/3: S. Holtskog: Residential Consumption of Bioenergy in China. A Literature Study
- 98/4 B.K. Wold: Supply Response in a Gender-Perspective, The Case of Structural Adjustments in Zambia. Technical Appendices
- 98/5 J. Epland: Towards a register-based income statistics. The construction of the Norwegian Income Register
- 98/6 R. Chodhury: The Selection Model of Saudi Arabia. Revised Version 1998
- 98/7 A.B. Dahle, J. Thomasen and H.K. Østereng (eds.): The Mirror Statistics Exercise between the Nordic Countries 1995
- 98/8 H. Berby: A Demonstration Data Base for Business Register Management. A data base covering Statistical Units according to the Regulation of the European Union and Units of Administrative Registers
- 98/9 R. Kjeldstad: Single Parents in the Norwegian Labour Market. A changing Scene?
- 98/10 H. Brüngger and S. Longva: International Principles Governing Official Statistics at the National Level: are they Relevant for the Statistical Work of International Organisations as well?
- 98/11 H.V. Sæbø and S. Longva: Guidelines for Statistical Metadata on the Internet
- 98/12 M. Rønsen: Fertility and Public Policies - Evidence from Norway and Finland
- 98/13 A. Bråten and T. L. Andersen: The Consumer Price Index of Mozambique. An analysis of current methodology – proposals for a new one. A short-term mission 16 April - 7 May 1998
- 98/14 S. Holtskog: Energy Use and Emmissions to Air in China: A Comparative Literature Study
- 98/15 J.K. Dagsvik: Probabilistic Models for Qualitative Choice Behavior: An introduction
- 98/16 H.M. Edvardsen: Norwegian Regional Accounts 1993: Results and methods
- 98/17 S. Glomsrød: Integrated Environmental-Economic Model of China: A paper for initial discussion
- 98/18 H.V. Sæbø and L. Rogstad: Dissemination of Statistics on Maps

Documents



Tillatelse nr.
159 000/502

B *Returadresse:*
Statistisk sentralbyrå
Postboks 8131 Dep.
N-0033 Oslo

Statistics Norway
P.O.B. 8131 Dep.
N-0033 Oslo

Tel: +47-22 86 45 00
Fax: +47-22 86 49 73

ISSN 0805-9411



Statistisk sentralbyrå
Statistics Norway