

Populærvitenskapelig framstilling, 2011

Statistics for Innovation (sfi)<sup>2</sup> is one of the 14 Norwegian Centres for Research-based Innovation - Sentre for forskningsdrevet innovasjon (SFI). We invent new statistical methods and computational tools necessary to achieve innovation goals in four key sectors for the Norwegian society: petroleum, finance & insurance, marine and health. The acronym (sfi)<sup>2</sup> comes from the identical abbreviations of Statistics for Innovation and Sentre for Forskningsdrevet Innovasjon, both as sfi, with the square power indicating that we work for a more than linear growth and investment return! (sfi)<sup>2</sup> operates from 2007 and is part of Norsk Regnesentral (Norwegian Computing Center - NR) in Oslo, in partnership with the University of Oslo (UiO), the Norwegian University of Science and Technology (NTNU) and 11 partners: Biomex, DnBNOR, Gjensidige, Hydro, The Institute of Marine Research, Pubgene, Oslo University Hospital as, Sencel, Smerud Medical Research, Spermatech and Statoil. (sfi)<sup>2</sup> is funded by the Research Council of Norway and by the partners, with an annual budget of about 40 mill NOK, including in-kind contribution as research. About 100 researchers actively join the research projects of the centre. Statistics for Innovation participates to the UiO training programs and funds many PhD students, researchers and postdocs.

The quantity of data collected in science, business and industry is enormous, because we have the technology to measure and store data at reasonable costs. Data can be private and secret, but more often than ever, data are often free and public, easily available with a click from the internet. Up to maybe 15 years ago, the most precious thing a scientist had was her data. Often few numbers, representing years of work in the clinics or in the field, handwritten on a paper or on a simple excel file. The statistical methods, needed to read out knowledge from data, or to support a new hypothesis, were often rather standard. Today the situation is completely different: data are stored on giant disks, organised in efficient data bases, and freely available for all. The novelty is that there are now many different statistical methods available to make inference and predictions. Why? Because more data open the possibility to many competing models, each focusing on specific aspects of the problem. And because the questions to be answered, or the predictions to be made, are often about small margins, the 2-3% that allows understanding new mechanisms, makes better products, win over the competition. Traditional statements of statistics are of the type “this is true with an error margin of 5%”. But 5% is too much, today. And more: if we have to select the few genes which are responsible for a complex disease, like schizophrenia, and we are testing for this purpose the 30000 genes we have, a 5% of error rate means that automatically 1500 genes will be discovered by mistake! But also in business: if we are going to select among the 1 million clients the few ones who are committing frauds, a 5% error rate would not be useful! Statisticians have therefore invented new methods, which are explicitly tailored for multiple comparisons, like the ones in the examples. At (sfi)<sup>2</sup> we work on genomics, and we have in 2011 identified three new genes which probably are responsible for the fundamental human task of face recognition. The same statistical methodology was also used in selecting the customers, who have a car insurance, who are risky drivers, beyond the information the insurance company has about them. We test the residual risk (after taken care of car type, accidents in the past, age etc.) of thousands of insured drivers and deliver lists of those who are under-priced, with an error rate which is acceptable for marketing purposes. Both these results, are examples of the impact of basic research, here in statistics, on industry and science, in terms of creating value. At (sfi)<sup>2</sup> we have 11 projects at the interface between innovation and basic statistical science. With our partners, we are validating the new methods and ideas on real world assets and services, transferring new knowledge to production.