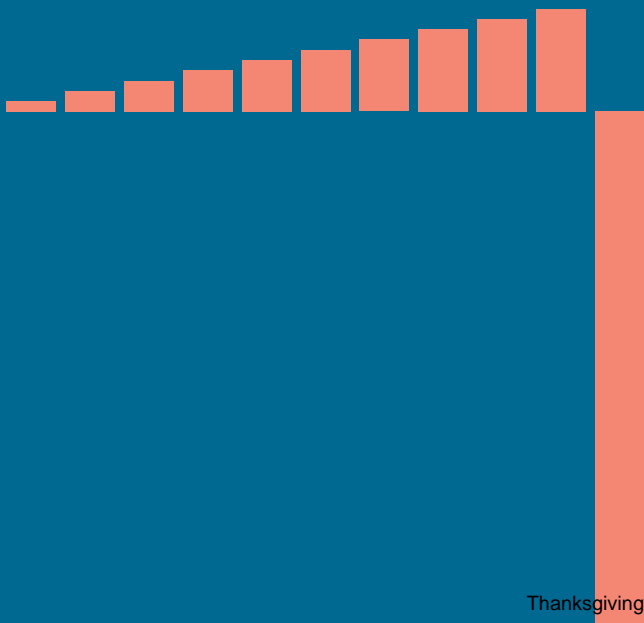


SWISS STATISTICAL SOCIETY

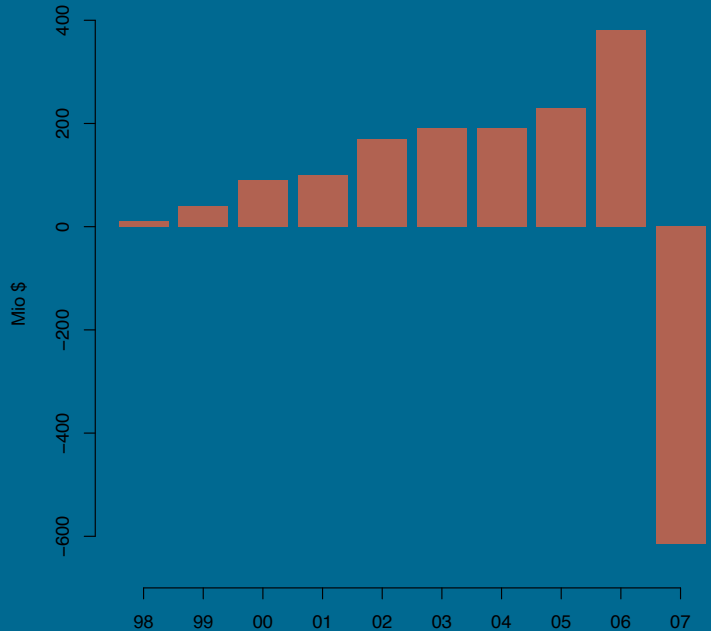
www.stat.ch

Weight Gain of a Turkey



Thanksgiving

IndyMac's (a US Bank)
Annual Net Income



Graphics inspired by
Nassim Nicholas Taleb:
"The Forth Quadrant"
(www.edge.org)

Opening for a Full / Adjunct Professor in Econometrics

The Department of Econometrics, Faculty of Economics and Social Sciences, University of Geneva invites applications for a position of a Full Professor or an Adjunct Professor in Econometrics starting August 1, 2009 or at a convenient date. Candidates should hold a relevant Ph.D., must have a proven track record of research and publications in theoretical econometrics, in particular in microeconomics, in top journals. Experience in applied econometric research, in various areas of economics including environmental economics, is desirable.

The candidate must also have experience in teaching and doctoral supervision. The position involves teaching courses at the undergraduate and graduate levels. Teaching at the undergraduate level is in French (with a period of adjustment) and teaching at the graduate level is in English. The successful candidate will also be required to supervise Masters-level and doctoral research work and fulfill administrative tasks related to the position.

The application deadline is November 21, 2008.

Applications, including a cover letter, a detailed CV with the list of publications and three references (two copies), should be sent to:

The Office of the Dean, Faculty of Economics and Social Sciences (SES), University of Geneva, Uni-Mail, Boulevard du Pont-d'Arve 40, CH-1211 Genève 4, Switzerland.

For further information contact: Prof. Jaya Krishnakumar, Head of Department (jaya.krishnakumar@unige.ch)

With a view towards increasing the proportion of female professors, the University of Geneva encourages female candidates to apply.



marcel.baumgartner@nestle.com

Liebe Statistikerinnen und Statistiker,

Wir durchlaufen turbulente Zeiten: die Wirtschaftskrise in den USA und ihre Auswirkungen bei uns werden uns wohl leider noch viel Diskussionsstoff liefern. Wie die Titelseite illustriert, ist es nicht einfach die sehr seltenen Ereignisse, welche aber sehr grosse Auswirkungen haben, vorauszusagen. Nassim Nicholas Taleb benennt solche Ereignisse „Black Swans“, und seine Arbeit wird durch die heutige Krise in ein ganz neues Licht gesetzt.

Als Statistiker kennen wir Variabilität, Unsicherheit und Risiko, und wir haben Modelle, um sie zu quantifizieren. Wir wissen auch, wann unsere Modelle nicht mehr gültig sind. Doch dieses Wissen scheint bei den vielen Statistikbenutzern nicht genügend verankert zu sein. Wir müssen daher unsere Bestrebungen vervielfachen, damit diese Risikoschätzungen extremer Ereignisse präziser werden. Taleb ermutigt uns, „zu wissen, was wir nicht wissen“!

Dieses Bulletin beinhaltet den dritten Artikel von Professor O'Hagan über Bayesianische Statistik. Die SSS ist sehr stolz Ihnen einen **Weiterbildungskurs** mit Anthony O'Hagan anzubieten, im Juni 2009 in Kandersteg: nutzen Sie diese einmalige Gelegenheit damit Sie vom Bayesianischen Ansatz in Ihrer täglichen Arbeit profitieren können.

Marcel Baumgartner

Chères statisticiennes et chers statisticiens,

Nous vivons une période turbulente: la crise économique aux USA et ses répercussions chez nous générera malheureusement encore beaucoup de discussions. Comme la page de garde l'illustre, il n'est pas facile de prédire les événements rares, mais qui ont un impact important. Nassim Nicholas Taleb les appelle «Black Swans», et son travail sera d'autant plus apprécié suite à cette crise.

En tant que statisticien, nous connaissons la variabilité, l'incertitude et le risque, et nous avons des modèles pour les quantifier. Nous savons aussi quand nos modèles ne sont plus valables, mais il semble que cette connaissance n'est pas appliquée par les nombreux utilisateurs de la statistique. Nous devons donc multiplier nos efforts pour que ces estimations du risque deviennent plus précises. Taleb nous encourage de «savoir ce que nous ne savons pas» !

*Ce bulletin contient le 3ème article du Professeur O'Hagan sur la statistique bayésienne. La SSS est très fière de pouvoir vous offrir un **cours** avec Anthony O'Hagan qui aura lieu en juin 2009 à Kandersteg: profitez de cette opportunité exceptionnelle, pour que vous puissiez appliquer l'approche bayésienne dans vos travaux quotidiens.*

Marcel Baumgartner

BAYES FACTORS

Bayes factors are somewhat essential to Bayesian statistics. **Tony O'Hagan** explains their basics.

Bayes factors are a fundamental part of the Bayesian approach to testing hypotheses. In frequentist statistics, hypothesis testing is a matter of choosing a test statistic and calculating the p -value. The p -value is regarded as a measure of the evidence in favour of (or against) a null hypothesis. In general, the smaller its value, the stronger the evidence against the null hypothesis, but more precise interpretation of p -values is a matter of convention. For instance, $p < 0.05$ is widely regarded as a yard-stick: any larger value of p is considered to be insufficiently strong evidence against the null hypothesis for it to be rejected.

In interpreting p -values, and in communicating their meaning to others, researchers frequently fall into the trap of saying that $p < 0.05$ means that there is a less than 5% chance that the null hypothesis is true. This is wrong. Unfortunately, p -values have no such direct interpretation as measures of evidence.

Bayesian hypothesis testing

It is only within the Bayesian framework that we can discuss the probability that a hypothesis is true. Indeed, the essence of a Bayesian hypothesis test is, quite simply, to compute and report this probability.

Given data x and a hypothesis H , we require the posterior probability $P(H | x)$ that H is true, and this is obtained using Bayes' theorem as:

$$P(H|x) = \frac{P(H)P(x|H)}{P(x)}$$

In this formula, $P(x | H)$ is the probability of observing the data x if H is true, while $P(x)$ is the unconditional probability of observing x . This is usually expressed as:

$$P(x) = P(H)P(x|H) + P(H^c) P(x|H^c)$$

where H^c denotes the complement of H (i.e., the hypothesis that is true if H is false). In the frequentist terminology, we could call H the null hypothesis, and H^c the alternative hypothesis.

In both these formulae, we have the prior probability $P(H)$ that H is true, and we also have $P(H^c) = 1 - P(H)$.

Thus, the posterior probability $P(H | x)$ depends on three things: $P(x | H)$, $P(x | H^c)$ and $P(H)$, and the two formulae together seem to imply that it depends on these three things in complicated ways. The Bayes factor arises from a different way of writing these expressions, which clarifies how these three quantities influence the posterior probability.

Consider the posterior odds in favour of H , which we define to be

$$\frac{P(H|x)}{P(H^c|x)} = \frac{P(H|x)}{1 - P(H|x)}$$

As $P(H | x)$ increases, it is clear that this ratio increases (because the denominator decreases), so the odds value is another way of representing the strength of evidence in favour of H . In particular, if the posterior probability of H is $P(H | x) = 0.5$, then the odds in favour of H is 1, so odds of 1 corresponds to a situation where the evidence favours neither H nor H^c . Odds greater than 1 corresponds to the weight of posterior evidence being in favour of H , while odds less than 1 corresponds to the weight of evidence being against H . In simple terms, odds of z means that H is z times as likely as H^c .

There is a corresponding prior odds, $P(H)/P(H^c)$, and the key formula is:

$$\frac{P(H|x)}{P(H^c|x)} = \frac{P(H)}{P(H^c)} \times \frac{P(x|H)}{P(x|H^c)} \quad (1)$$

The final factor on the right is the *Bayes factor*, $B_{\pi}(x)$. In words, this formula says that the posterior odds is equal to the prior odds multiplied by the Bayes factor.

If the Bayes factor is greater than 1, then the posterior odds will be larger than the prior odds, and so the posterior probability of H will be larger than its prior probability. Conversely, if $B_{\pi}(x) < 1$, then $P(H | x) < P(H)$. So, the Bayes factor says how the evidence in the data modifies the prior probability. The formula distinguishes two separate influences on the posterior probability of H : first, there is the prior probability, based on prior evidence; second, there is the Bayes factor, which measures the strength of the new evidence in the data, x .

Interpreting Bayes factors

The Bayes factor has a very clear interpretation as a measure of evidence in favour of the (null) hypothesis H . If $B_{r,H}(x) < 0.05$, then the posterior odds in favour of H will be less than a twentieth of the prior odds. This represents appreciable evidence against H . For instance, if you initially had negligible prior evidence for or against H , then the prior odds would be 1, and given a Bayes factor less than 0.05, the posterior odds will also be less than 0.05. We can easily convert odds back to probability by using the formula:

$$\text{Probability} = \frac{\text{Odds}}{1 + \text{Odds}}$$

Thus, posterior odds less than 0.05 converts to a posterior probability less than

$$\frac{0.05}{1.05} = 0.0476$$

With a Bayes factor below 0.05, then, there is only a small posterior probability that H is true, unless there was prior evidence in its favour.

To take another example, suppose that prior information suggested that H was twice as likely to be true as false, in which case the prior odds would be 2. Now suppose that we obtain data x such that the Bayes factor is $B_{r,H}(x) = 0.01$, which is clearly strong evidence against H . The posterior odds would now be only 0.02, and H is 50 times more likely to be false than to be true. This is the effect of the strong Bayes factor; the posterior probability is

$$\frac{0.02}{1.02}$$

which is also close to 0.02.

There are two points to be aware of in interpreting Bayes factors. The first is that the Bayes factor only measures evidence in the data, and must be interpreted relative to the prior evidence. A Bayes factor of, say, 0.01 does indeed represent strong evidence against H , but if the prior evidence was strongly in its favour,

The Bayes factor only measures evidence in the data, and must be interpreted relative to the prior evidence

for instance, by prior odds of 1000, then the posterior odds will still be in favour of H . This might be the case, for instance, if H were the null hypothesis that the subject in an experiment does not have extra-sensory perception (ESP). In this case, most people would demand extremely strong evidence in the data to overcome a strong prior belief that people do not have ESP.

The second point is that in practice the Bayes factor still makes use of prior information. To see why this is the case, we need to consider the distinction between simple and compound (or composite) hypotheses.

Deriving the Bayes factor

To calculate the Bayes factor, we need to find the probability of the data x under the hypothesis H and under its complement H^c . Consider the example of ESP. Suppose that the data consist of asking the subject to guess the colour (red or black) of a random sequence of 20 playing cards. The subject correctly guesses 15. The null hypothesis H of no ESP means that the probability of guessing correctly is 0.5 for each card. Hence, $P(H | x)$ is the probability that $x = 15$ when x has the binomial distribution $\text{Bi}(20, 0.5)$, which is

$$\binom{20}{15} 0.5^{20} = 0.0418.$$

In frequentist theory, the p -value is the probability of getting 15 or more correct guesses, which is 0.0207 and is small enough, by convention, to imply appreciable evidence against H .

We have $P(x | H) = 0.0148$, which is the numerator of the Bayes factor, but what about the denominator, $P(x | H^c)$? To derive this, we need to know the probability of a correct guess under the alternative hypothesis, H^c , which is the hypothesis that the subject does have ESP. But that probability depends on how strong the ability is. If the subject had the power to correctly identify every card, then the probability of a correct guess would be 1. Any weaker ability would give a probability less than 1,

but greater than 0.5. If we denote that probability by θ , then we can find

$$P(x|q) = \binom{20}{15} q^{15} (1 - q)^5$$

for any θ . But the denominator of the Bayes factor is not $P(x | \theta)$ for some specified, it is $P(x | H^c)$, where, in effect, we are only specifying that $\theta > 0.5$.

The denominator, $P(x | H^c)$, is an average of the values $P(x | \theta)$, averaged over the prior distribution of θ . Thus, to compute the Bayes factor we have to say how likely it would be, a priori, that θ would take any given value between 0.5 and 1, assuming that the subject does indeed have the power of ESP.

This is the second point mentioned above about Bayes factors. If H is a simple hypothesis, then it completely specifies the probability of x because it specifies the values of all the unknown parameters in the statistical model. However, in practice it is extremely rare for both H and H^c to be simple hypotheses. (In the example, H is simple but H^c is not.) When one or both of the hypotheses is compound (i.e., not simple), then the Bayes factor depends on the prior distributions for the unknown parameters.

The fact that Bayes factors depend on prior information is an important distinction between the Bayesian and frequentist approaches

Weak prior information

The fact that Bayes factors depend on prior information is an important distinction between the Bayesian and frequentist approaches to hypothesis testing, because p -values do not have this property. This is a standard feature of Bayesian statistics, and the requirement for prior distributions to be specified has been at the core of the debate between the two approaches for decades. One response by Bayesians has been to make use of prior distributions that represent, in some formal sense, very weak prior information. For instance, in the context of the ESP question, it can be argued that a uniform prior distribution (giving equal prior density to all values of θ between 0.5 and 1) represents an absence of prior information about how strong the subject's powers might be if H is false.

See the box “Extra-sensory perception?” for the result of using the prior distribution.

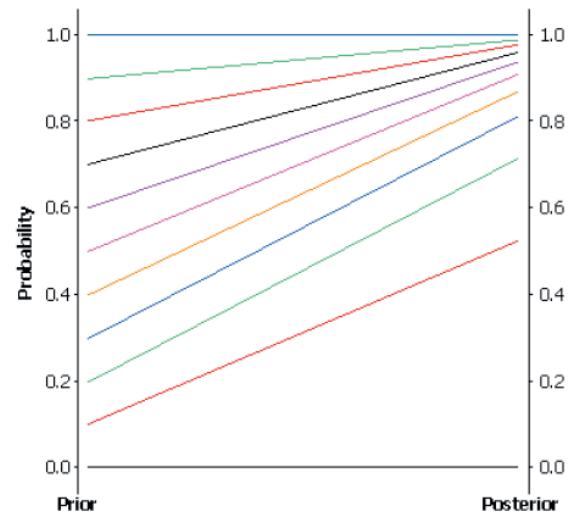


Figure 1. The effect of a Bayes factor of 10 in favour of H . Even with a prior probability of only 0.1, the posterior probability of H is better than 0.5

There is some arbitrariness about how we try to represent weak information, but in most applications of Bayesian statistics, given enough data, the posterior conclusions will not be sensitive to which representation we use. Unfortunately, one area in which this argument does not work is in testing hypotheses. The posterior probability of H may remain sensitive to the choice of weak prior (and may even be formally undefined for most choices), regardless of how much data we have. A variety of more or less *ad hoc* variations on the Bayes factor have been

One advantage of the Bayesian approach is that it is possible to compare more than two hypotheses

proposed to address this difficulty, including the “intrinsic Bayes factor” and the “fractional Bayes factor”. These are supposed to have the same interpretation as the full Bayes factor, but can be applied in situations of weak prior information; however, this is an area of on-going discussion within Bayesian statistics.

Many hypotheses

The preceding discussion has been based entirely within the context of comparing two hypotheses, which we can

characterise as a null hypothesis and an alternative hypothesis in line with frequentist terminology. However, one advantage of the Bayesian approach is that it is possible to compare more than two hypotheses. The key formula (1) generalises to

$$\frac{P(H_1|x)}{P(H_2|x)} = \frac{P(H_1)}{P(H_2)} \frac{P(x|H_1)}{P(x|H_2)}$$

where where H_1 and H_2 are any two hypotheses. We then call the second factor on the right-hand side of this formula, $P(x | H_1) = P(x | H_2)$ the Bayes factor for H_1 versus H_2 and denote it by $B_{12}(x)$. No matter what other hypotheses are under consideration, the ratio of posterior probabilities for H_1 and H_2 is their prior ratio multiplied by the Bayes factor $B_{12}(x)$.

Bayes factors are an important part of the applied Bayesian statistician's toolkit. Although it is not possible in an account such as this to go into technical detail about how they are derived or computed in practice, I have tried to explain what they mean, issues around their use and how they relate to frequentist p -values. As Bayesian methods are increasingly used in statistical applications, a basic understanding of how to interpret them should be a part of every scientist's toolkit.



Bayes factor

The Bayes factor in favour of H (and against its complement H^c), based on data x is

$$B_H(x) = \frac{P(x|H)}{P(x|H^c)}$$

Extra-sensory perception?

A subject is asked to guess the colour (red or black) of each of 20 randomly selected playing cards. She correctly guesses 15 colours. Does this level of success indicate that she has extra-sensory perception (ESP)?

Under the null hypothesis H that she does not have ESP, the probability of correctly identifying 15 out of 20 colours is $P(x | H) = 0.0148$. If she does have ESP (the alternative hypothesis H^c), then we suppose that she has a probability $\theta > 0.5$ of correctly guessing each card. Under the assumption that all values of θ from 0.5 to 1 are equally likely, the probability that she correctly identifies 15, given that she does have ESP, is $P(x | H^c) = 0.0940$. The Bayes factor in favour of the null hypothesis is therefore $B_H(x) = 0.0148/0.0940 = 0.16$. This is clearly not strong evidence against H , and would not be enough to persuade most people to believe in ESP. For instance, if you start with a prior belief that the subject does not have ESP is 100 times as likely as her having ESP, your prior odds is 100. Multiplying this by the Bayes factor of 0.16 gives a posterior odds of 16. So you still believe it is 16 times more likely that the subject does not have ESP than that she does.

If we had a different prior distribution for θ under the alternative hypothesis, we would get a different posterior odds. However, it can be shown that, based on the data of 15 correct answers out of 20 and the prior odds of 100 against ESP, then the posterior odds will always be at least 7.

Originally published in *Significance*, December 2006.
Reprinted with permission from the
Royal Society of Statistics.

BAYESIAN STATISTICS, JUNE 22 – 24, 2009

Prof. Anthony O'Hagan, University of Sheffield
Waldhotel Doldenhorn, Kandersteg



Lecturer

Prof. Tony O'Hagan is a leading researcher in Bayesian statistics, with extensive practical experience in applying Bayesian methods in a variety of fields.

Tony obtained his PhD in 1974 from University College, London, and has taught at the Universities of Dundee, Warwick, Nottingham and Sheffield. He formally retired from the University of Sheffield in February 2008, but continues to work for them parttime in consultancy and research.

Aims and objectives

This course is intended to provide a basic grounding in the theory and concepts of Bayesian inference. Whilst many participants will have encountered the Bayesian approach to some extent, no specific background knowledge will be assumed.

At the end of the course, participants should:

- understand the fundamentals of Bayesian inference, and how it differs from frequentist inference;
- appreciate the role of the prior distribution, how it may be constructed, and its basis in the personal interpretation of probability;
- have intuitive understanding of how Bayes' theorem synthesises prior information and data;
- be able to derive the posterior distribution and appropriate summaries/inferences in simple situations;
- be aware of how more complex problems are tackled, through hierarchical modelling and modern computational tools;
- understand the benefits of Bayesian analysis, the extra demands that it makes, and the kinds of problems where it is proving to be most powerful.

Synopsis

The course will be structured as 8 lectures plus 3 exercise/practical sessions. Lecture topics will be as follows.

1. The Bayesian paradigm. Bayes' theorem, prior, likelihood and posterior. Simple one-parameter models.
2. Some standard models. A detailed analysis of the cases of binomial and normal samples.
3. Inference. Bayesian inferences, and how they differ from frequentist inferences. Formal and informal inference.
4. Prior. Personal probability. Elicitation. 'Default' prior distributions. Interpretation of inferences.
5. Bayes theory. Likelihood, sufficiency and ancillarity. Asymptotics. Preposterior analysis.
6. Structuring prior information. Independence and exchangeability. Hierarchical models. Shrinkage and smoothing.
7. Tackling real problems. Bayesian modelling and computation. Outline of a real example to show how Bayesian analysis works in complex problems.
8. For and against. What do we get from a Bayesian analysis, and at what cost? In what situations does the Bayesian approach give the greatest benefit?

Practicals

The first practical/tutorial session will be a practical exercise in specifying beliefs about exchangeable sequences. The second will provide experience with theoretical manipulations. The third will include a roleplay exercise and group work.

Location

The course will be held in the Waldhotel Doldenhorn in Kandersteg; see www.doldenhorn-ruedihus.ch. The hotel offers a free shuttle service to the train station.

Date and hour

Monday, June 22 till Wednesday, June 24, 2009. Course starts at 09.00 and ends at 17.00.

COURSE

Course fee

- CHF 1'350.– for members of the Swiss Statistical Society, other applicants CHF 1'550.–.
- Reduced course fee CHF 950.– for students. Please send a copy of the certification. A limited number of course places are reserved for students.

The course fee includes all meals and accommodation in single rooms. Arrival on Sunday evening, June 21, 2009, and departure on Wednesday, June 24, 2009, 17.00.

The number of participants is limited to 20, with a minimum of 12.

Registration deadline:

April 15, 2009.

Registration and further information

Swiss Statistical Society
Sabine Probst
Bergacher 8
CH-3253 Schnottwil
+41 (0)32 353 70 94
sabine.probst@stat.ch

The Swiss Statistical Society (SSS) reserves the right to cancel a course up to 14 days prior to the course due to insufficient enrollment. The SSS is not liable for any participants' expenses incurred prior to cancelled courses. Payment of the course registration fee is required prior to the start of the course. Cancellations received in writing more than 30 days before the start of the course will be refunded 100% of the course fee. Cancellations received between 30 and 14 days prior to the course will be refunded 50% of the course fee. The SSS regrets that no refunds are allowed for cancellations received within 14 days of the course start date.

DIE VOLKSZÄHLUNG 2010: AUF DEM WEG ZU EINEM UMFASSENDEN SYSTEM DER HAUSHALTS- UND PERSONENSTATISTIKEN

Markus Schwyn, MSc, MAS. Leiter der Abteilung Bevölkerungsstudien und Haushaltssurveys, Bundesamt für Statistik, Neuchâtel.

Das einzig Konstante ist der Wandel, im Alltag genauso wie in der Politik. Im Zuge des rasanten wirtschaftlichen und gesellschaftlichen Veränderungsprozesses sind auch die Ansprüche an die öffentliche Statistik gestiegen. Sie muss häufiger und schneller Informationen liefern, aktuelle Fragen und Themen flexibel aufgreifen und beantworten können. Die neue Volkszählung ist Teil der Antwort auf diese neuen Herausforderungen. Sie will in einer ganzheitlichen Betrachtungsweise die vielfältigen Bedürfnisse abdecken: Sie bietet Grundlageninformationen für die Öffentlichkeit, dient als Input für die wissenschaftliche Analyse oder als Planungs- und Entscheidungsgrundlage für die öffentliche Hand und Private.

Am 22. Juni 2007 hat das Parlament die Totalrevision des Bundesgesetzes über die eidgenössische Volkszählung verabschiedet. Das Gesetz ist am 1. Januar 2008 in Kraft getreten. Mit dem neuen Gesetz über die eidgenössische Volkszählung haben Bundesrat und Parlament eine Modernisierung der amtlichen Statistik beschlossen. Mit der neuen Volkszählung wird ein umfassender Systemwechsel vollzogen: die Vollerhebung alle zehn Jahre wird abgelöst durch ein integriertes statistisches System. Das System kombiniert die Verwendung bestehender Personenregister mit Stichprobenerhebungen, die im Einjahresrhythmus durchgeführt und ausgewertet werden.

Kernelemente der Volkszählung

Das Volkszählungsgesetz gibt den Auftrag vor: jährlich sind Daten über die Bevölkerungsstruktur und die gesellschaftliche Entwicklung in der Schweiz zu erheben. Das Volkszählungsgesetz beschreibt in Artikel 1 die geforderten Themenfelder. Diese thematische Breite wird mit

vier Erhebungsgefässen abgedeckt, die jährlich durchgeführt werden: die Registererhebung, die Strukturhebung, eine thematische Erhebung sowie der CH-Omnibus (Abbildung 1).

Die jährliche Registererhebung basiert auf den Einwohnerregistern von Gemeinden und Kantonen, den wichtigsten Bundespersonenregistern sowie dem eidgenössischen Gebäude- und Wohnungsregister. Damit sind jährlich Basisinformationen über die Bevölkerung sowie zu Gebäuden und Wohnungen in räumlich kleinster Auflösung verfügbar. Mit dem Gesetz über die Harmonisierung der amtlichen Personenregister, welches seit dem 1. Januar 2008 in Kraft ist, wurde eine zentrale Voraussetzung für eine effektive und vereinfachte Nutzung der Registerdaten geschaffen. Das Gesetz bestimmt die Identifikatoren und die Merkmale, die in den Registern geführt sein müssen, formuliert die inhaltlichen und formalen Anforderungen und regelt den Datenaustausch zwischen den verschiedenen Registern.

Die Strukturhebung ist als jährliche Stichprobenerhebung bei 200'000 Personen konzipiert. Als Personenerhebung liefert sie wichtige, in den Registern heute nicht verfügbare Merkmale. Befragt werden Personen in Privathaushalten, die 15 Jahre sind oder älter. Die befragte Person gibt dabei Auskunft über sich und über den Haushalt. Eine jährliche Stichprobenerhebung bei 200'000 Personen erlaubt statistische Auswertungen für alle Kantone bzw. für Gruppen von 15'000 Personen. Nach fünf Jahren sind durch die Methode des Datenpoolings Aussagen zu Gruppen von 3'000 Personen möglich. Innerhalb dieser Gruppen können Einheiten von 140 nach einem Jahr, resp. 28 Personen nach 5 Jahren entdeckt werden. Kantone können die Erhebung auf ihrem Gebiet auf eigene Kosten aufstocken und damit eine Resultatsverbesserung erzielen.

Die thematischen Stichprobenerhebungen werden ebenfalls jährlich mit einer Stichprobengrösse von 10'000 bis 40'000 Personen durchgeführt. Die Themen alternieren in einem Fünfjahresrhythmus mit folgenden Schwerpunk-

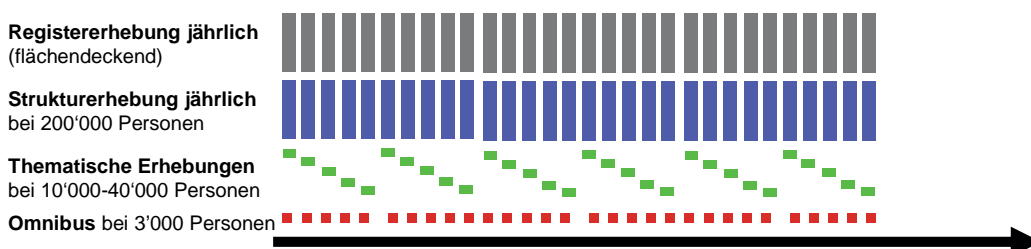


Abbildung 1: Erhebungen im Zeitablauf

ten: „Mobilität“, „Bildung“, „Gesundheit“, „Familien“ sowie „Sprache, Religion und Kultur“. Die heute bestehende Gesundheitsbefragung und der Mikrozensus „Verkehr“ werden in dieses System integriert. Aufgrund der Stichprobengrösse sind bei den thematischen Erhebungen aussagekräftige Ergebnisse für die gesamte Schweiz sowie die sieben Grossregionen zu erwarten. Der Mikrozensus „Mobilität und Verkehr“ wird mit einer Stichprobengrösse von 40'000 Personen zudem Resultate für die Agglomerationsgebiete bereitstellen. Auch diese Erhebungen können die Kantone auf eigene Kosten aufstocken.

Mit dem als CH-Omnibus bezeichneten Erhebungsinstrument wird neu ein flexibles Gefäss für die rasche Beantwortung von aktuellen Fragestellungen bereitgestellt. Die jährlich durchgeführte Stichprobenerhebung bei rund 3'000 Personen bietet interessierten Kreisen die Möglichkeit, mit spezifischen Fragen bei der Befragung „aufzuspringen“. Die Erhebung liefert Resultate für die Schweiz, die schnell aufgearbeitet und publiziert werden können.

Die Volkszählung als Teil eines Gesamtsystems

Mit ihrer Neuorientierung wird die Volkszählung zum Rückgrat eines neuen Gesamtsystems für Haushalts- und Personenstatistiken (SHAPE). Sie ist künftig Teil eines Systems, das die systematische Verwendung bestehender Verwaltungsregister und Stichprobenerhebungen bei Personen und Haushalten miteinander kombiniert. Die verschiedenen Erhebungen werden inhaltlich, methodisch und organisatorisch miteinander verbunden und aufeinander abgestimmt. Gerade in einem längeren Betrachtungshorizont ergibt das Gesamtsystem einen Mehrwert, der über die Summe seiner Einzelteile hinausreicht. Die verschiedenen Gefässe der Volkszählung ergänzen vorhandene Erhebungen und vervollständigen sie zu einem umfassenden Bild sozio-ökonomischer Statistiken (Abbildung 2). In den Themenfeldern „Arbeit“, „Einkommen, Konsum und Lebensbedingungen“ sind bereits drei jährliche Erhebungen etabliert.

- Die Schweizerische Arbeitskräfteerhebung (SAKE) liefert Daten zum Arbeitsmarkt und zum Erwerbsleben im Allgemeinen. Mit der Anpassung an europäische Statistiken wird sie künftig vierteljährliche Ergebnisse liefern.
- Die Haushaltsbudgeterhebung (HABE) stellt regelmässige Informationen über die Konsumgewohnheiten und

die Einkommenssituation der privaten Haushalte in der Schweiz bereit.

- Die 2007 neu geschaffene Erhebung der Einkommen und Lebensbedingungen (SILC: *Statistics on Income and Living Conditions*) deckt einen breiten Bereich ab, wie z.B. die Einkommen, die Ausbildung, die Arbeit, die Kinderbetreuung, die Haushaltszusammensetzung, die Wohnsituation oder die Gesundheit.

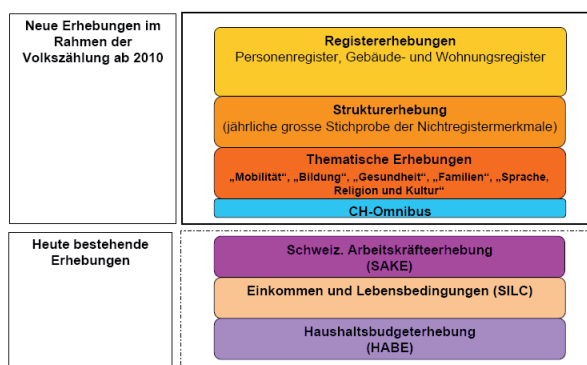


Abbildung 2: Gesamtsystem für Personen- und Haushaltsstatistiken (SHAPE)

Statistiken und Erhebungen

Im Zentrum des neuen statistischen Informationssystems steht die Wohnbevölkerung der Schweiz: die Personen und ihre Haushalte. Über die Person wird die Verbindung zu den Wohnungen und Wohngebäuden hergestellt. Das neue Volkszählungssystem führt vier Erhebungsgefässe und die daraus resultierenden Daten zusammen. Als Ergebnis dieses Gesamtsystems werden künftig vier Arten von Statistiken vorliegen, die unterschiedliche thematische und räumliche Tiefenschärfen aufweisen.

- Jährliche Basisstatistiken zu Bevölkerung, Haushalten und Wohnen auf Basis der flächendeckenden Registererhebungen.
- Jährliche Strukturstatistiken auf Basis der Strukturerhebung und der Registererhebungen. Darin enthalten sind auch die «traditionellen» Themen der Volkszählung. Sie liefern ergänzende Informationen zu den Basisstatistiken und Grundlageninformationen für die Analyse der Themenfelder, jeweils auf Stufe der Kantone.
- Jährlich eine vertiefende Statistik zu den Themenfeldern „Arbeit“ und „Einkommen, Konsum und Lebensbedingungen“ auf Basis der thematischen Erhebungen und der Registererhebungen.

- Jährlich eine vertiefende Statistik zu einem der übrigen Themenfelder auf Basis der entsprechenden thematischen Erhebung und der Registererhebungen.
- Jährliche Statistiken zu aktuellen Fragestellungen auf Basis des CH-Omnibus und der Registererhebungen.

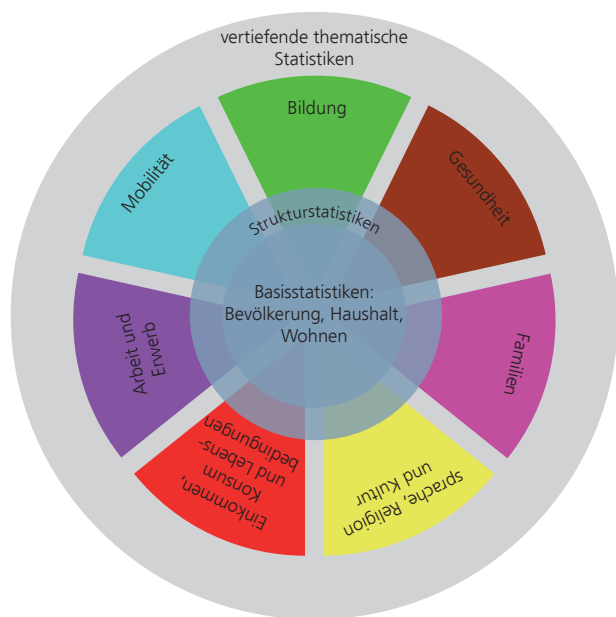


Abbildung 3: Gesamtsystem Statistiken und Themen

Das Gesamtsystem integriert von Basisstatistiken über die Strukturstatistiken bis zu den thematisch vertiefenden Statistiken alle Informationen über Personen, Haushalte und Wohnen (Abbildung 3). Durch diese transversale Perspektive und die Integration der Resultate der verschiedenen

Erhebungen können die sieben Themenfelder umfassend abgedeckt werden. Die Prinzipien der Integration können dabei vereinfacht folgendermassen dargestellt werden:

Die Stichproben können ab 2010 auf Basis der Informationen der Einwohnerregister und des GWR gezogen werden. Die Bevölkerungsstatistik übernimmt eine zentrale Funktion als Lieferantin von Referenzgrössen für die Stichprobenerhebungen. Zur Planung, Gewichtung und Hochrechnung von effizienten Stichprobenerhebungen sind die Informationen über die Gesamtbevölkerung und deren Bevölkerungsgruppen sowie über Wohngebäude und Wohnungen sehr wichtig.

Die neue Volkszählung kann ihr volles Potenzial erst als funktionierendes Gesamtsystem ausschöpfen. Sie ist mehr als nur die Summe verschiedener einzelner Statistiken. Um diesen Systemeffekt zu erzielen, werden Integrationselemente benötigt, welche die Erhebungen mit ihren unterschiedlichen Datenquellen zusammenführen können. Diese Funktion übernehmen vier Integrationselemente:

- Die gemeinsam verwendeten Grundgesamtheiten: Die Zusammenführung der verschiedenen Erhebungen benötigt eine gemeinsame Basis. Notwendig sind gemeinsam verwendete Grundgesamtheiten, die einheitlich definiert und harmonisiert sind. Sie stecken den Rahmen der Personen-, Haushalts- und Wohnstatistiken ab mit den zu beobachtenden statistischen Einheiten.
- Die neuen AHV-Nummern, die als Personenidentifikator die eindeutige Identifikation einer Person in unterschiedlichen Datenquellen ermöglicht.

Datenquelle	Inhalt	Geografie	Thema. Detaillierung
Omnibus und thematische Erhebungen	Thematische Vertiefung	CH	hoch
Struktur-erhebung	Basis-	Gross- regionen Kantone	mittel
Register-erhebung	Informationen	Gemeinden Meterkoordinaten	tief

Abbildung 4: Integration der Statistiken

- c. Die Gebäude- und Wohnungsidentifikatoren, die die Haushaltsbildung und die Zuordnung der Personen und der Haushalte zu den Gebäuden und den Wohnungen erlauben.
- d. Die Schlüsselmerkmale, die die Vergleichbarkeit der Resultate sowie die gemeinsame Definition der Bevölkerungsgruppen gewährleisten. Die Schlüsselmerkmale sollen in Zukunft harmonisiert in allen Erhebungen enthalten sein. Sie erlauben die einheitliche Abgrenzung und Identifikation von Bevölkerungsgruppen. Damit wird ein minimaler gemeinsamer Nenner geschaffen, der die Vergleichbarkeit der Resultate aus den verschiedenen Statistiken und das statistische Monitoring bestimmter Bevölkerungsgruppen gewährleistet. Mit dem Konzept der harmonisierten Schlüsselmerkmale können zu den Merkmalsträgern in verschiedenen thematischen Bereichen Aussagen gemacht werden. So können beispielsweise Aussagen gemacht werden von Gruppen gemäss ihrer „höchsten abgeschlossen Ausbildung“ zum Mobilitäts- und Verkehrsverhalten, zur Gesundheitsprävention, zur Verwendung von Kulturangeboten, zu ihren Sprachkompetenzen oder zur religiösen Praxis und zu anderem mehr. Damit können abgrenzbare Personengruppen (z.B. Personen mit Universitätsabschluss) im Kontext der Themenfelder beschrieben und analysiert werden.

Fazit

Das neue System bringt vielfältige Vorteile mit sich. Die Volkszählung gewinnt an Aktualität, Flexibilität und räumlicher Genauigkeit. Informationen sind häufiger, thematisch breiter und innert kürzester Zeit vorhanden. Mit der Verfügbarkeit jährlich aktueller Daten wird die regelmässige systematische Beobachtung wichtiger und politikrelevanter soziokultureller Themenkreise verbessert. Das neue System kann laufend aktualisiert und weiterentwickelt werden. Und nicht zuletzt weist es auch ein optimales Kosten-Nutzen-Verhältnis auf. Eine verbesserte Koordination und verstärkte Synergieeffekte führen zu einem deutlich geringeren finanziellen und administrativen Aufwand und es bringt eine deutliche Entlastung bei den Befragten und den Gemeinden.

Weitere, laufend aktualisierte Informationen zum Gesamtsystem finden Sie unter:

- Volkszählung 2010: www.VZ2010.bfs.admin.ch
- SHAPE: www.bfs.admin.ch / Modernisierungsprojekte / SHAPE
- Registerharmonisierung: www.register-stat.admin.ch

Auf dem Internet publizierte Grundlageninformationen:

- Die Volkszählung 2010: Auf dem Weg zu einem umfassenden System der Haushalts- und Personenstatistik. Informationsgrundlagen für die Kantone, Bundesämter und weitere interessierte Kreise
- Strukturhebung. Aufstockungsmodelle für die Kantone
- SHAPE: Das neue System der Haushalts- und Personenstatistiken

Rechtsgrundlagen (Auswahl):

- Bundesstatistikgesetz (BStatG) vom 9. Oktober 1992 (SR 431.01)
- Verordnung vom 30. Juni 1993 über die Durchführung von statistischen Erhebungen des Bundes (SR 431.012.1)
- Bundesgesetz vom 22. Juni 2007 über die eidgenössische Volkszählung (Volkszählungsgesetz) (SR 431.112)
- Verordnung zum Volkszählungsgesetz (in Arbeit)
- Bundesgesetz über die Harmonisierung der Einwohnerregister und anderer amtlicher Personenregister (RHG) (SR 431.02)
- Verordnung vom 21. November zum Registerharmonisierungsgesetz (Registerharmonisierungsverordnung) (SR 431.021)

TÄTIGKEITSBERICHT DES ETHIKRATES (ER) 2008

Der Ethikrat stützt sich bei seiner Tätigkeit auf die Charta der öffentlichen Statistik der Schweiz und sein Reglement. Seine Entscheide fällt er politisch unabhängig und in der Sache nach bestem Wissen und Gewissen.

Mit den Statistiktagen 2007 in Luzern ging die erste Periode des Ethikrates zu Ende. Nach längeren Verhandlungen zwischen den Trägerschaften Bundesamt für Statistik (BFS) und Konferenz der regionalen Statistischen Ämtern (KORSTAT) wurden die künftigen Rahmenbedingungen für den Ethikrat festgelegt. Zu Beginn der zweiten Amtsperiode konnten die nötigen Weichenstellungen für die Zukunft des Ethikrates umgesetzt werden.

An der Generalversammlung der SSS-O 2007 wurden die Mitglieder des Ethikrates für die zweite Periode 2008-2011 gewählt.

Personelle Zusammensetzung des Ethikrates 2008

Präsident

Peter Tschopp, Honorarprofessor an der Universität Genf und ehemaliger Nationalrat

Experten

Regula Stämpfli, Wissenschaft und Politik

Christoph Menzel, FEDESTAT

Sekretär

Felix Bosshard, Statistisches Amt des Kantons Zürich

Kantone

Simone Nuber, Statistik Stadt Zürich

Hervé Montfort, Office cantonal de la statistique Genève (OCSTAT)

Bund

Philippe Eichenberger, Bundesamt für Statistik

Markus Schwyn, Bundesamt für Statistik

Weichenstellung für die Zukunft des Ethikrates

Gemäss Beschluss der Generalversammlung wird in den ersten zwei Jahren der neuen Periode das Sekretariat durch ein regionales Statistisches Amt geführt. Das Statistische Amt des Kantons Zürich war bereit mit Dr. Felix Bosshard das Sekretariat zu stellen. Das Bundesamt für Statistik bietet weiterhin die technische Infrastruktur insbesondere den Weiterbetrieb der Internetseite und die Übersetzungen an.

Der Ethikrat hat ein Budget von maximal 20'000 Franken, das je zur Hälfte vom Bundesamt für Statistik (BFS) und von der Konferenz der regionalen Statistischen Ämtern (KORSTAT) gedeckt wird. Zusätzliche ausserordentliche Ausgaben für unvorhergesehene Projekte würden ebenfalls paritätisch finanziert. Dafür muss allerdings zuerst die Zustimmung von BFS und KORSTAT eingeholt werden. Damit sind zu Beginn der zweiten Amtsdauer die Rahmenbedingungen für den Ethikrat klar geregelt.

Seit der Neubesetzung des Ethikrates an den Statistiktagen 2007 fanden drei ordentliche Sitzungen statt.

Peer Review 2007 der Bundesstatistik und Vorbehalt zu Artikel 10 der Charta

Die Bundesstatistik wurde 2007 einer von der EU vorgesehenen Peer Review unterzogen. Der Ethikrat ist befremdet, dass er im Rahmen der Peer Review nicht konsultiert worden ist, wie er dies ausdrücklich gewünscht hatte. Im nächsten Geschäftsjahr will er sich vertieft mit den Ergebnissen der Peer Review befassen.

Das BFS hat bei der Unterzeichnung der Charta einen Vorbehalt angebracht. Der Ethikrat kann den Vorbehalt des BFS zu Artikel 10 nicht nachvollziehen. Er bedauert diesen Schritt, insbesondere auch deshalb, weil das Bundesamt zu den Promotoren der Charta und zu den ersten Unterzeichnern gehört. Der Ethikrat wird sich mit der Auswirkung des Vorbehalts und dem Verhältnis der Charta zum europäischen Verhaltenskodex weiter auseinandersetzen.

Umfrage über die Tätigkeit des Ethikrates

Die Charta der öffentlichen Statistik legt Wert auf einen transparenten, nachvollziehbaren und redlichen Umgang mit öffentlichen Daten. Der Ethikrat der öffentlichen Statistik der Schweiz musste in den letzten Jahren mehrfach feststellen, dass statistische Daten, insbesondere für politische Zwecke, missbraucht wurden.

Anlässlich seines fünfjährigen Bestehens hat der Ethikrat beschlossen, im Hinblick auf die Statistiktage 2008 im September in Davos eine Umfrage über die präventive Wirkung der Charta der öffentlichen Statistik zu lancieren.

Bei dieser Umfrage geht es darum, die Wirksamkeit der Charta zu evaluieren. Die bescheidene Anzahl von Anfragen und Beschwerden, die an den Ethikrat gelangen, stellen diesen vor die Frage, ob allenfalls die Charta eine präventive Wirkung hat. Es gibt Anzeichen, die dafür sprechen. Aus der Umfrage bei Ämtern der öffentlichen Statistik erhofft sich der Ethikrat Antworten auf diese Frage.

Überprüfung der Einhaltung der Charta

Im vergangenen Jahr hat sich der Ethikrat vor allem inhaltlich mit der Lohnstatistik und einer Studie der Credit Suisse (CS) im Auftrag des Seco auseinandergesetzt.

Lohnstatistik

Der Ethikrat befasste sich mit verschiedenen Fragen zur Lohnstatistik und lud Herrn Froidevaux (Verantwortlicher beim BFS) zu einer Präsentation ein. Der Ethikrat zog daraus folgende Schlüsse:

- Im Gegensatz zu verschiedenen Branchenstatistiken erfüllt die Lohnstatistik die Ansprüche bezüglich Qualität und Transparenz.
- Die Lohnstrukturerhebung (LSE) weist Lücken auf. Obwohl die 2-jährige Periodizität der LSE die EU-Anforderungen (alle 4 Jahre) übertrifft, ist sie für die Schweizer anscheinend ungenügend. Im Weiteren fehlen in der LSE die Einkommen von Selbständigerwerbenden und diejenigen aus dem 1. Sektor. Diese Lücken können teilweise durch die Schweizerische Arbeitskräfte-Erhebung (SAKE) behoben werden. Die Vergleichbarkeit der Löhne der Bauern mit anderen Löhnen ist schwierig. Zudem werden die Bonuszahlungen, die insbesondere im Finanzbereich eine wesentliche Lohnkomponente darstellen, nicht vollständig erfasst.

Studie der CS im Auftrag des Seco zu „Wirtschaftliche Erneuerungsgebiete“

Mit zwei Vertretern der CS führte der Ethikrat ein Hearing zu der im Auftrag vom Seco erstellten Studie „Wirtschaftliche Erneuerungsgebiete“ durch. Grund dafür war die Anfrage einer öffentlichen Stelle. Obwohl diese Stelle das vom Ethikrat vorgesehene Fact Sheet nicht eingereicht hat, befasste sich der Ethikrat mit diesem Thema.

Aus der Sitzung ergaben sich folgende Erkenntnisse:

- Die Studie verwendet einen Indikator für die regionale Bruttowertschöpfung. Für diesen Indikator fehlen regionale Daten, weshalb Hilfsgrössen wie die Branchenstruktur verwendet werden. Der verwendete Indikator weist auf eine empfindliche Lücke der öffentlichen Statistik hin. Das BFS berechnet keine regionale Wertschöpfung, da aus seiner Sicht die Datenlage dafür nicht ausreicht. Durch die grosse Nachfrage bzw. „Wichtigkeit“ werden solche Indikatoren deshalb von Privaten berechnet und als Produkt vermarktet.
- Die Öffentliche Statistik verwendet hier mangels Alternativen oft wenig transparente Indikatoren von Privaten. Die Transparenz der einzelnen Berechnungen fehlt häufig bzw. wird aus Sicht der Firmen als USP (*unique selling proposition*) gehortet.

Ein wichtiges Fazit aus der Besprechung mit der CS sind für den Ethikrat die aufgezeigten Grenzen der Statistik bei Daten zur regionalen Wertschöpfung. Die qualitativen Anforderungen an die Primärdaten sind hoch und können nicht immer erfüllt werden. Deshalb berechnet das BFS keine solchen Werte. Gerade bei solchen Indikatoren sind aber Transparenz über die Berechnung und eine sorgfältige Qualitätskontrolle von grosser Bedeutung. Die Auslagerungen von Untersuchungen öffentlicher Institutionen an private Unternehmungen nehmen zu. Solange Transparenz und Nachvollziehbarkeit gewährt sind, stellt dies kein Problem dar. Es ist aber auch zu beachten, dass mit den Auslagerungen Know-how ausserhalb der Verwaltung entsteht, was auf die Dauer nicht unproblematisch ist.

Der Ethikrat hält fest:

- Für den Leser der veröffentlichten Studie bleibt die Zusammenarbeit zwischen CS und dem Seco undurchsichtig. Insbesondere sind auch die Vorgaben, die vom Seco gemacht worden sind unklar. Leider können auch die Resultate der Studie nicht in allen Bereichen nachvollzogen werden.
- Der Ethikrat weist darauf hin, dass bei der Verwendung nicht offizieller Daten grösste Sorgfalt an den Tag gelegt werden muss. Insbesondere gilt dies für Studien, die als politische Entscheidungshilfen verwendet werden.
- Bei öffentlichen Aufträgen ist die Nachvollziehbarkeit wichtig. Im vorliegenden Fall hätte das Seco zusätzlich zum Bericht einen Methodenbericht in Auftrag geben

sollen. Für den Auftraggeber sollte es keine vertraulichen Daten und Methoden geben. Der Ethikrat versteht, dass solche Informationen nicht veröffentlicht werden, sie müssten aber in einem Methodenbericht zumindest für den Auftraggeber transparent sein.

Beiträge von Mitgliedern des Ethikrates

Der Start in eine neue Periode veranlasste den Ethikrat, diesen Jahresbericht mit einigen Gedanken, die über das Tagesgeschäft hinausgehen, anzureichern. Im Anhang zu diesem Tätigkeitsbericht sind deshalb Beiträge der Experten des Ethikrates, Frau Regula Stämpfli und Christoph Menzel, zu finden. Zahlen sind für jede Statistik eine unabdingbare Grundlage. In ihrem Beitrag „Die Politik der Zahlen: Allgemeine Überlegungen zur Arbeit des Ethikrats“ geht Regula Stämpfli auf die Kraft von Zahlen ein. Mit dem Bilateralen Statistikabkommen zwischen der Schweiz und der Europäischen Gemeinschaft wurde der Verhaltenskodex der europäischen Statistik für die Schweiz relevant. In seinem Beitrag „Peer Review 2007 der Bundesstatistik“ stellt Christoph Menzel das Ergebnis der Peer Review dar. Mit der Peer Review wird sich der Ethikrat im kommenden Geschäftsjahr auseinandersetzen.

Dr. Felix Bosshard
08.08.2008

Anhänge

Die Politik der Zahlen: Allgemeine Überlegungen zur Arbeit des Ethikrats

von Regula Stämpfli, Dr. phil., Politikwissenschaftlerin/Dozentin/Autorin

„Mit Zahlen lässt sich trefflich streiten, mit Zahlen ein System bereiten“ liess schon Goethe im „Faust“ verlauten. Die numerisch starke Gegenwart übertrifft die Literatur jedoch in grossem Masse. Gerade die Schweizer sind ein Volk von Zahlenspezialisten, Listen-Lesern und Komparatisten, neudeutsch „Fans von Rankings“. Aus dem Sport stammend überfluten die Ranglisten nun auch die Politik. Es gibt Städte-Regionen- und Subventionsrankings, die nicht in geringem Masse über Ansehen, Macht und Geld bestimmen. Deshalb ist es nicht unwichtig, zu jedem Ran-

king eine klare, transparente, nachvollziehbare Datenbasis zu haben. Dies ist jedoch nicht immer der Fall.

Solange die Rankings als Medienspektakel die Zeitungsspalten, Radio- und Fernsehminuten füllen, ist dies eigentlich kein Problem – obwohl punkto Meinungsumfragen im Vorfeld der Abstimmungen und Wahlen durchaus Fragen angebracht wären, doch dies soll uns hier nicht kümmern. Sobald die nicht nachvollziehbaren, methodisch undurchsichtigen und punkto Datenerhebung zufällig zusammengesetzten Rankings jedoch Politik werden, lohnt es sich hinzuschauen. Das ist mitunter die Aufgabe des Ethikrats für öffentliche Statistik.

Anhand mehrerer Beispiele stellte der Ethikrat fest, dass in der Schweiz ein Grossteil der Datenerhebungen, die Politikgrundlage darstellen, immer noch vorwiegend in der Hand privater Interessenorganisationen oder in den Händen von den Daten selber betroffenen Institutionen liegen. Die Santésuisse erhebt Daten, die für die Krankenkassenansätze bestimmend sind, der Bauernverband erhebt Daten, welche über die Subventionen im Agrarbereich entscheiden, der Schweizerische Nationalfonds erhebt Daten, die die künftige Forschungspolitik finanzieren und bestimmen, die Gewerkschaften liefern Daten über die künftige Arbeitslohnpolitik und und und.

Diese Verknüpfung zwischen staatlichen und wirtschaftlichen Interessen hat in der Schweiz lange Tradition, ist sinnvoll, weil viele Politiken so auch gesellschaftlich gestützt sind, ist aber dann in Frage zu stellen, wenn zwischen Interessenpolitik, Datenlage und Subventionspraxis eigentlich nicht mehr unterschieden werden kann.

Damit die Erhebung dieser Daten, deren Nachvollziehbarkeit, Transparenz und Qualität gesichert werden kann, braucht es den Ethikrat für öffentliche Statistik. Denn schon im 19. Jahrhundert stellten unsere liberalen Verfassungsväter fest, dass es zwischen Eigen- und Allgemeininteresse so zu unterscheiden galt, dass öffentliche Institutionen bestehen müssen, um Normen, Recht und Übersicht zu garantieren.

Dieser feine Unterschied scheint jedoch in der politischen Gegenwart manchmal etwas verloren gegangen zu sein. Da reagieren sämtliche private und öffentliche Institutionen pikiert, wenn der Ethikrat es wagt, bei gewissen Statistiken, Beständen, Listen, ja, eben Rankings genauer nachzufragen, woher die Daten stammen, wie sie erhoben wurden, welche Fehlerquoten ausgewiesen werden und wie es um die Transparenz und Nachvollziehbarkeit steht. Denn genau letztere sind gerade bei dieser manch-

mal seltsam anmutenden Aufteilung privat-öffentlich, die in der Schweiz gleichzeitig für eine gewisse Verankerung zwischen Verwaltung und Bevölkerung sorgt, aber gleichzeitig auch den Beigeschmack von Filz nie wirklich los wird, von entscheidendem öffentlichen, politischen, gesellschaftlichen und wirtschaftlichen Interesse.

Zahlen und Statistiken per se sind neutral, gleichgültig. Ihr Wert ergibt sich nur durch Interpretation, durch den politischen Diskurs und die politische Legitimation. Nur eindimensional denkende PositivistInnen meinen, mit Zahlen jeden Bereich der Welt erklären zu können. Alle anderen denkenden Menschen wissen, dass nicht Zahlen, sondern die Menschen die Politik bestimmen. Also ergeben sich für den Ethikrat mehr und mehr neue Aufgaben. Denn nicht nur die Frage der Transparenz, Nachvollziehbarkeit und Gültigkeit der präsentierten Statistiken, die oft aus privater Hand entstammen, muss überprüft werden, sondern auch die Frage, ob tatsächlich die erhobenen Daten genügend Transparenz, Nachvollziehbarkeit und Legitimation für weitgreifende Politiken bieten.

Der Ethikrat für öffentliche Statistik heisst eben nicht zuletzt deshalb Ethikrat und nicht Numerologenrat. Entscheidend bleibt, ob Daten wirklich in einem Umfeld der Redlichkeit, der Objektivität und der Basis erhoben werden. Transparenz ist nicht nur das Zauberwort der Good Governance, sondern auch des Ethikrats.

Zu einer Demokratie gehört, das System, das hinter den Zahlen steckt, transparent, öffentlich und nachvollziehbar für die Stimmbürger und Stimmbürgerinnen zu machen. Dazu gehört auch, private und allgemeine Interessen zu unterscheiden. Wer sich dagegen wehrt, versucht wohl nur eines: Bisher erworbene Zahlen-Macht-Position eifersüchtig zu behalten und ohne Legitimation weiterzuwursteln.

Vielleicht sollte dieser Blick für Zusammenhänge statt der gebannten Blick auf schöne Tabellen wieder einmal eingeübt werden und auch von der Politik her öfters gefragt werden.

Peer Review 2007 der Bundesstatistik

von Christoph Menzel

Der Ethikrat kann bezüglich der Anwendung der Grundprinzipien der Charta Feststellungen und Empfehlungen formulieren sowie Informationen verbreiten. Der Ethikrat ist unabhängig und überwacht die Einhaltung der Grund-

prinzipien, die Anwendung der Prinzipien und unterstützt deren Umsetzung.

Die Überprüfung der Einhaltung des Verhaltenskodex 2007 durch die EU

Vorgeschichte: Am 31. Oktober 2007 nahm der Bundesrat Kenntnis vom europäischen Verhaltenskodex, der Eurostat und die nationalen statistischen Ämter und damit auch die Schweiz verpflichtet, die ethischen Grundsätze des europäischen Kodex zu befolgen.

Die Einhaltung des Kodex wird überprüft. Im Anschluss an eine Selbstbeurteilung fand in den Vorjahren eine Beurteilung der nationalen statistischen Ämter in Form von Peer Reviews statt. Vom 21. – 23. November 07 führten zwei Vertreter/innen nationaler statistischer Stellen (Norwegen, Italien) zusammen mit einem Vertreter von Eurostat eine solche Review in der Schweiz durch. Die Prüfung beschränkte sich auf das Bundesamt für Statistik und seine Rolle als Koordinator. Geprüft wurde die Einhaltung von 7 von den insgesamt 15 Grundsätzen des europäischen Kodex, nämlich:

1. Fachliche Unabhängigkeit; 2. Mandat zur Datenbeschaffung; 3. Ressourcen zur Erfüllung des Auftrages; 4. Verpflichtung zur Qualität; 5. Statistische Geheimhaltung; 6. Unparteilichkeit und Objektivität; 7. Zugänglichkeit und Klarheit.

Ergebnisse: Im Folgenden werden die für den Ethikrat wichtigsten Ergebnisse kurz zusammengefasst.

Die Prüfung erfolgte aufgrund von schriftlichen Unterlagen, die den Reviewern im Voraus zur Verfügung standen und Interviews. Befragt wurden neben dem BFS und dem GS des EDI, der Präsident der Kommission für Bundesstatistik (Prof. H. W. Brachinger), Benutzer aus der Wissenschaft, Korstat (G.-A. Paravicini), die SNB, die Oberzolldirektion, das Seco, die KOF und die Medien. Hingegen wurde der Ethikrat nicht begrüsst.

Der Ethikrat hat das Ergebnis der Peer Review zur Kenntnis genommen. Drei Gebiete gaben den Prüfenden Anlass zu insgesamt 7 Empfehlungen. Es sind dies Koordinationsprobleme, unpräzise rechtliche Grundlagen und Qualitätsmängel. Für den Ethikrat unerwartet stellten die Prüfer ins Gewicht fallende Mängel im Qualitätsmanagement des BFS fest. Sie empfehlen dem BFS deshalb, ein umfassendes Qualitätsmanagement sowie klare Prozesse zur Qualitätskontrolle und Dokumentation in allen Statistikbereichen einzuführen (Peer Review, S. 25). Nach dem Urteil des Prüferenteams befindet sich das BFS in einem

sehr frühen Stadium des Qualitätsmanagements und der regelmässigen Qualitätskontrollen in allen Produktionsprozessen. Das BFS plant, ein umfassendes Qualitätsmanagement bis 2010 einzuführen (S. 10). Das Review-Team anerkennt, dass das BFS über eine klare Strategie verfügt, um die Mängel zu beheben (S. 3).

Mit Genugtuung nimmt der Ethikrat zur Kenntnis, dass die Prüfer die Charta als einen Meilenstein für das schweizerische Statistiksystem beurteilen und dass eine weitere Verbreitung der Charta und deren Anwendung („...further extension and implementation...“) wichtig sei, um die Qualität der öffentlichen Statistik zu verbessern. Das Team stellt fest, dass nicht alle Statistikproduzenten, die auch am Mehrjahresprogramm der Bundesstatistik teilnehmen, die Charta unterschrieben haben.

Das Prüferteam bemängelt ebenfalls die verhältnismässig schwache gesetzliche Grundlage zur Koordination der

Statistikstätigkeit. Als Folge des Beitritts zum europäischen statistischen System könnte sich ein zusätzlicher Ressourcenbedarf ergeben und es wird empfohlen, insbesondere innerhalb der amtsinternen Wirtschaftsstatistik Aufgaben und Verantwortung neu zu verteilen sowie die Koordination zu verbessern (S. 24).

Zusätzlich empfiehlt die EU, eine transparentere und klarere Politik bezüglich der vorzeitigen Veröffentlichung von Statistiken anzustreben.

In den Bereich der Zugänglichkeit von Statistiken fällt die Empfehlung, analytische Beiträge zur verwendeten Methodologie zu verfassen. Dies würde den Benutzern den Zugang erleichtern und den Nutzen der einzelnen Statistik für verschiedene Zwecke erhöhen (S. 19).

Die Peer Review findet sich auf dem Portal des BFS/Institutionen/öffentliche Statistik/Ethische Prinzipien.

RAPPORT D'ACTIVITÉ DU CONSEIL D'ÉTHIQUE (CE) 2008

Le Conseil d'éthique s'appuie dans ses activités sur la Charte de la statistique publique de la Suisse et sur son règlement. Il prend ses décisions indépendamment de toute influence politique et en pleine connaissance de cause.

La première période administrative du Conseil d'éthique a pris fin lors des Journées statistiques 2007 à Lucerne. Les nouvelles conditions-cadre du CE ont été fixées après de longues négociations entre l'Office fédéral de la statistique (OFS) et la Conférence suisse des offices régionaux de statistique (CORSTAT), les deux institutions responsables du CE. Au début de la deuxième période administrative, les lignes directrices nécessaires pour assurer l'avenir de la CE a pu être mises en œuvre.

A l'occasion de l'assemblée générale de la SSS-O 2007, les membres du Conseil d'éthique ont été élus pour la période 2008-2011.

Composition du Conseil d'éthique 2008

Präsident

Peter Tschopp, Honorarprofessor an der Universität Genf und ehemaliger Nationalrat

Experten

Regula Stämpfli, Wissenschaft und Politik

Christoph Menzel, FEDESTAT

Sekretär

Felix Bosshard, Statistisches Amt des Kantons Zürich

Kantone

Simone Nuber, Statistik Stadt Zürich

Hervé Montfort, Office cantonal de la statistique Genève (OCSTAT)

Bund

Philippe Eichenberger, Bundesamt für Statistik

Markus Schwyn, Bundesamt für Statistik

Lignes directrices pour l'avenir du Conseil d'éthique

En vertu de la décision de l'assemblée générale, le secrétariat est assuré pour un office statistique régional

durant les deux premières années de la nouvelle période administrative. Il a été confié à Felix Bosshard, de l'Office statistique du canton de Zurich. L'Office fédéral de la statistique continue à mettre son infrastructure technique à la disposition du CE, en assurant notamment l'exploitation du site Internet et les traductions.

Le Conseil d'éthique dispose d'un budget de 20'000 francs au maximum, couvert à parts égales par l'OFS et la CORSTAT. Ces deux institutions prendraient aussi en charge de manière paritaire des dépenses extraordinaires qui seraient occasionnées par des projets imprévus, à condition qu'elles donnent leur accord au préalable. Les conditions-cadre du CE sont ainsi clairement définies au début de cette deuxième période administrative.

Trois séances ordinaires ont eu lieu depuis l'entrée en fonction du CE dans sa nouvelle composition.

Peer Review 2007 de la statistique fédérale et réserve concernant l'article 10 de la Charte

La statistique fédérale a été soumise en 2007 à une des Peer Review prévues par l'UE. Le Conseil d'éthique s'étonne de ne pas avoir été consulté dans le cadre de la Peer Review, alors qu'il l'avait expressément souhaité. Lors du prochain exercice, il traitera de manière approfondie des résultats de cette Peer Review.

L'OFS a apporté une réserve lors de la signature de la Charte. Le Conseil d'éthique ne peut comprendre la réserve de l'OFS concernant l'article 10. Il regrette cette dernière, notamment aussi parce que l'Office fédéral de la statistique fait partie des promoteurs de la Charte et qu'il a été l'un des premiers à la signer. Le Conseil d'éthique va continuer d'évaluer l'effet de cette réserve ainsi que la relation de la Charte avec le Code de bonnes pratiques de la statistique européenne.

Enquête concernant l'activité du Conseil d'éthique

La Charte de la statistique publique de la Suisse vise une utilisation transparente, fondée et loyale des données publiques. Ces dernières années, le Conseil d'éthique de la statistique publique de la Suisse a constaté à plusieurs reprises que des données statistiques ont été utilisées de manière abusive, en particulier à des fins politiques.

Pour marquer ses cinq ans d'existence, le CE a décidé de lancer une enquête sur l'effet préventif de la Charte de la statistique publique dans la perspective des Journées statistiques 2008 qui auront lieu à Davos en septembre. Cette enquête est destinée à évaluer l'efficacité de la Charte. Le nombre modeste de demandes et de plaintes qui parviennent au Conseil d'éthique amène ce dernier à se demander si la Charte a un effet préventif. Il y a des indices qu'il en soit effectivement ainsi. Par le biais de cette enquête menée auprès des offices de la statistique publique, le CE espère recueillir des réponses à ce sujet.

Examen de cas pour savoir si la Charte a été respectée

L'année dernière, le Conseil d'éthique a notamment examiné deux cas, l'un concernant la statistique des salaires et l'autre une étude du Credit Suisse (CS) mandatée par le Seco.

Statistique des salaires

Le Conseil d'éthique a traité diverses questions relatives à la statistique des salaires et a invité Monsieur Froidevaux (chef de la section LOHN à l'OFS) à une présentation. Au terme de cette dernière, le Conseil d'éthique est ensuite arrivé aux conclusions suivantes:

- Contrairement à différentes statistiques de la branche économique, la statistique des salaires satisfait aux exigences posées sur les plans de la qualité et de la transparence.
- L'enquête sur la structure des salaires (ESS) présente des lacunes. Même si l'ESS est réalisée tous les deux ans, ce qui va au-delà des exigences de l'UE (tous les 4 ans), cette périodicité n'est apparemment pas suffisante pour les Suisses. Par ailleurs, l'ESS ne considère pas les revenus des indépendants ni les revenus dans le secteur primaire. Ces lacunes peuvent être comblées en partie par l'Enquête suisse sur la population active (ESPA). Il est difficile de comparer les salaires des paysans avec d'autres salaires. En outre, les bonis, qui représentent une composante du salaire essentielle dans le domaine financier en particulier, ne sont pas saisis de manière exhaustive.

Etude du CS consacrée aux «zones économiques de redéploiement», réalisée sur mandat du Seco

Le Conseil d'éthique a mis sur pied une audition avec deux représentants du CS concernant l'étude «Zones économiques de redéploiement» réalisée sur mandat du Seco. Cette audition a été organisée suite à une demande adressée à ce propos par un service public. Même si ce service n'a pas adressé la note explicative prévue par le Conseil d'éthique, ce dernier s'est penché sur ce cas.

L'audition a abouti aux résultats suivants:

- L'étude utilise un indicateur pour la valeur ajoutée brute régionale. Il manque des données régionales pour établir cet indicateur, raison pour laquelle on a recours à des grandeurs d'appoint comme la structure des branches. L'indicateur utilisé met en lumière une lacune sensible de la statistique publique. L'OFS ne calcule pas de valeur ajoutée régionale, car il estime que la base de données nécessaire n'est pas suffisante. En raison d'une forte demande et de leur «importance», de tels indicateurs sont établis par des entreprises privées et commercialisés comme produit.
- Faute de solutions de remplacement, la statistique publique utilise ici des indicateurs privés qui sont souvent peu transparents. La transparence de ces calculs fait souvent défaut ou elle est mise en avant comme USP (*unique selling proposition*) par les entreprises qui les établissent.

L'audition menée avec les représentants du CS a permis au CE de prendre conscience des limites de la statistique quant aux données relatives à la valeur ajoutée régionale. Les exigences posées aux données primaires sur le plan de la qualité sont élevées et ne peuvent pas toujours être satisfaites. C'est pourquoi l'OFS ne calcule pas de telles valeurs. Pour ces indicateurs justement, il est très important d'assurer une bonne transparence dans le calcul et un contrôle minutieux de la qualité. Les enquêtes d'institutions publiques sont de plus en plus souvent confiées à des entreprises privées. Cette délégation dans la réalisation des enquêtes ne représente pas un problème aussi longtemps que la transparence et la vérifiabilité sont garanties. Mais il convient de relever que le savoir-faire est ainsi forgé à l'extérieur de l'administration, ce qui ne va pas sans poser de problèmes à terme.

Le Conseil d'éthique émet les remarques suivantes:

- La collaboration entre le CS et le Seco reste opaque pour le lecteur de l'étude publiée. Les critères posés par le Seco pour cette étude notamment ne sont pas connus. Les résultats de cette dernière ne peuvent malheureusement pas être vérifiés dans tous les domaines.
- Le Conseil d'éthique relève qu'il convient de faire preuve d'une grande circonspection dans l'utilisation de données non officielles. Cela vaut en particulier pour des études qui sont utilisées comme aide à la prise de décisions politiques.
- La vérifiabilité est importante dans les mandats publics. Dans le cas présent, le Seco aurait dû commander un rapport méthodologique en sus du rapport d'étude. Pour le mandant, il ne devrait pas y avoir de données et de méthodes confidentielles. Le Conseil d'éthique comprend que de telles informations ne soient pas publiées, mais celles-ci devraient être au moins portées à la connaissance du mandant par le biais d'un rapport méthodologique.

Contributions de membres du Conseil d'éthique

L'entrée dans une nouvelle période administrative a incité le CE à enrichir le rapport d'activité annuel de quelques réflexions allant au-delà des affaires courantes. En annexe au présent rapport d'activité, vous trouverez des contributions de Madame Regula Stämpfli et de Monsieur Christoph Menzel, experts du Conseil d'éthique. Dans sa contribution « La politique des chiffres: réflexions générales sur le travail du Conseil d'éthique », Regula Stämpfli s'intéresse au pouvoir des chiffres. Avec l'entrée en vigueur de l'accord bilatéral sur la statistique conclu entre la Suisse et l'UE, le Code de bonnes pratiques de la statistique européenne est désormais aussi valable pour la Suisse. Dans sa contribution « Peer Review 2007 de la statistique fédérale », Christoph Menzel évoque les résultats de cette Peer Review. Le CE traitera de ces derniers au cours de l'exercice à venir.

Felix Bosshard

08.08.2008

Ndlr: Veuillez noter que les annexes de Regula Stämpfli et Christoph Menzel, disponibles dans la version allemande (pages 14 à 16 dans ce bulletin) du rapport d'activité, n'ont pas été traduites en français.

LE 'R PACKAGE SAMPLING'

Yves Tillé et Alina Matei

*Institut de statistique, Université de Neuchâtel
Pierre à Mazel 7, 2000, Neuchâtel
yves.tille@unine.ch, alina.matei@unine.ch*

Résumé

R est un logiciel libre pour le calcul statistique et la création de graphiques. Le logiciel de base est accompagné d'un ensemble de plus de 1500 modules complémentaires (packages) dans tous les domaines de la statistique. Le 'R package sampling' est un module d'échantillonnage permettant de sélectionner des unités statistiques dans des populations finies au moyen de plans complexes, à probabilités égales ou inégales, de construire des estimateurs calés en choisissant la fonction de calage et en imposant éventuellement des bornes sur les poids, de réaliser des poststratifications et de traiter les non-réponses.

Mots clés : statistique d'enquête, sondages, R.

1. Introduction

R¹ est un logiciel libre pour le calcul statistique et la création de graphiques. Le logiciel de base est accompagné d'un ensemble de plus de 1500 modules complémentaires (packages) dans tous les domaines de la statistique. Il peut être téléchargé librement de l'adresse : www.r-project.org.

Le 'R package sampling' (Tillé et Matei, 2008) est un module d'échantillonnage développé à l'Institut de statistique de l'Université de Neuchâtel. Il permet de sélectionner des unités statistiques dans des populations au moyen de plans complexes : plans stratifiés, par grappes, équilibrés, à plusieurs degrés. Il est possible de tirer des échantillons à probabilités égales ou inégales. Pour les plans à probabilités inégales, de nombreux algorithmes sont implémentés. Le module permet également de calculer des poids pour un estimateur calé et son estimateur de variance, de réaliser des poststratifications et de traiter les non-réponses. De plus, trois bases de données sont incluses dans le module : la base 'MU284' issue du livre de Särndal, Swenson et Wretman (1992), et des données sur les communes

belges et suisses provenant des offices de statistique belge et suisse. De nombreux exemples accompagnent chaque fonction.

Le module 'sampling' répond à deux objectifs principaux :
– c'est un logiciel libre pour le traitement d'une enquête,
– il offre des fonctions semblables à celles qu'offrent les logiciels payants.

Le module a vu le jour en 2005 afin de servir d'outil pédagogique pour des cours avancés sur les méthodes d'échantillonnage, organisés par l'Office Fédéral de la Statistique sous l'égide d'Eurostat et de l'Association Européenne de Libre Échange (AELE). La version courante est la version 2.0 et le module est en développement continu.

Pour installer le module (version 2.0) à partir d'une session R, on utilise la commande :

```
install.packages("sampling")
```

Pour charger le module on utilise :

```
library(sampling)
```

La liste de toutes les fonctions disponibles est obtenue en utilisant la commande :

```
help(package=sampling)
```

Les noms des fonctions incluses dans le module sont donnés dans le texte en italique. Par exemple, *UPpoisson* est la fonction pour tirer un échantillon en utilisant le plan de Poisson.

Pour obtenir l'aide sur une fonction on utilise son nom, comme dans l'exemple suivant :

```
help(UPpoisson)
```

2. Fonctions principales

2.1 Plans à probabilités égales et inégales

Le module contient des fonctions pour tirer des échantillons avec ou sans remise, et à probabilités égales ou inégales. Un plan simple (à probabilités égales) est implémenté à l'aide de deux fonctions (*srswor*, *srswor1* pour un plan sans remise, et *srswr* pour un plan avec remise). Les plans à pro-

¹ Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

babilités inégales implémentés sont les suivants : le plan de Brewer (1963), le plan à entropie maximale (Hájek, 1964), le plan de Midzuno (1952), le plan multinomial (Hansen et Hurwitz, 1943), la méthode du pivot (Deville et Tillé, 1998), le plan ordonné (Rosén, 1997), le plan de Poisson (Hájek, 1958), le plan au support minimal (Deville et Tillé, 1998), le plan systématique (Madow, 1949), le plan de Sampford (1967) et le plan de Tillé (1996). Les fonctions pour tirer des échantillons à probabilités inégales sont précédées du préfixe UP (“unequal probability”). La liste de fonctions de type UP est : *UPbrewer*, *UPmaxentropy*, *UPmidzuno*, *UPmultinomial*, *UPpivotal*, *UPpips*, *UPrandompivotal*, *UPpoisson*, *UPminimalsupport*, *UPSampford*, *UPsystematic*, *UPrandomsystematic*, *UPtille*.

Les probabilités d'inclusion d'ordre un et deux sont calculées pour les plans suivants : à entropie maximale, Midzuno, systématique et Tillé.

2.2 Plans complexes

Des plans de sondages complexes sont implémentés dans notre module. Ainsi, il est possible d'avoir :

- un plan stratifié (*strata*), avec un tirage à probabilités égales ou inégales,
- un plan par grappes (*cluster*) avec un tirage à probabilités égales ou inégales,
- un plan à plusieurs degrés (*mstage*), en utilisant un plan stratifié, un plan par grappes ou un plan simple à chaque degré, avec un tirage à probabilités égales ou inégales.

2.3 La méthode du cube

On dit qu'un échantillon est équilibré s'il permet de restituer avec exactitude les moyennes de la population pour toutes les variables connues. La méthode du cube² (Deville et Tillé, 2004 ; Tillé, 2006) est une méthode générale pour sélectionner des échantillons équilibrés à probabilités égales ou inégales (*samplecube*). Les deux phases de la méthode peuvent être exécutées séparément (*fastflightcube*, *landingcube*). Des fonctions additionnelles sont utilisées pour des plans équilibrés stratifiés, par grappes et

à deux phases (*balancedstratification*, *balancedcluster*, *balancedtwostage*).

2.4 Estimations

Dans la pratique des sondages on utilise souvent un estimateur qui a la propriété d'être sans biais, nommé l'estimateur d'Horvitz-Thompson. Il est implémenté en utilisant les fonctions *HTestimator* et *HTstrata*.

Estimation par calage

Le calage (voir Deville et Särndal, 1992) est une technique qui utilise des informations auxiliaires dans les pondérations des estimateurs. Cette technique assure l'amélioration de l'estimateur d'Horvitz-Thompson : on maintient presque parfaitement son caractère sans biais, et on diminue la largeur de l'intervalle de confiance (la fourchette) en fonction de l'information (le pouvoir prédictif) apportée par des variables auxiliaires, dont les totaux sont connus dans la population.

La fonction *calib* permet de construire des poids (“g-weights”) pour estimateurs calés en choisissant la fonction de calage et en imposant éventuellement des bornes sur les poids. Quatre méthodes sont implémentées (“linear”, “raking”, “truncated”, “logit”). La fonction *checkcalibration* vérifie si le calage est possible. L'estimateur calé et sa variance estimée sont calculés ensuite à l'aide de la fonction *calest*.

Poststratification

Comme le calage, la poststratification permet d'améliorer la précision des estimations. Supposons que l'on connaisse la répartition par sexe et âge de la population, mais cette information n'a pu être utilisée au stade de l'échantillonnage (par stratification), car l'information nécessaire ne figure pas dans la base de sondage. On utilise cette information a posteriori en modifiant les poids de façon à retrouver les bons effectifs de la population. La poststratification est im-

² Notons que l'Institut national de la statistique et des études économiques (INSEE) vient de publier les chiffres du dernier recensement en France. La qualité des statistiques résultant de cette nouvelle procédure de recensement a été rendue possible grâce à l'application de la méthode du cube.

plémentée à l'aide de la fonction *poststrata*, et l'estimateur poststratifié par la fonction *postest*.

D'autres estimateurs utilisant l'information auxiliaire sont disponibles :

- l'estimateur par le quotient (*ratioest*, *ratioest_strata*);
- l'estimateur par la régression généralisée (*regest*, *regest_strata*), avec le calcul des coefficients de régression, l'erreur standard, la valeur du test de Student et sa p-valeur, la matrice de variance-covariance des coefficients, des informations qui ne sont pas données dans la fonction *calib*.

2.5 La non-réponse

Quelle que soit la qualité de l'échantillonnage et du processus de collecte, l'échantillon sera déformé par le fait que certaines unités ne répondront pas, parce qu'elles sont injoignables, refusent explicitement de répondre, ou à cause d'autres motifs. C'est ce que l'on appelle la non-réponse totale. La non-réponse partielle est la situation où certaines parties du questionnaire n'ont pas été remplies. Généralement ces données manquantes sont estimées par des modèles statistiques de prévision (imputation) à partir des données disponibles. Pour corriger la non-réponse totale on a implémenté deux approches :

- la création des groupes de réponses homogènes (les répondants ont une réponse homogène dans une classe ; voir *rhg*, *rhg_strata*)
- l'utilisation des modèles statistiques pour re-pondération (on prédit les probabilités de réponses à l'aide d'une régression logistique; voir *rmodel*).

Les deux approches peuvent être utilisées dans la fonction de calage, *calib*.

2.6 Autres fonctions

Des fonctions additionnelles sont présentes pour :

- calculer les probabilités d'inclusion pour un plan avec probabilité proportionnelle à la taille (*inclusionprobabilities*);
- calculer les probabilités d'inclusion pour un plan stratifié (*inclusionprobastrata*);
- extraire une information d'une base de données à partir d'un échantillon tiré (*getdata*);

- lister tout les échantillons possible de taille fixe précisée (*writesample*);
- effacer les strates vides (*cleanstrata*);
- créer une codification à partir d'une variable qualitative (*disjunctive*).

3. Un exemple

Pour exemplifier nos fonctions on suit les étapes suivantes:

- on génère des données artificielles, un jeu de données contenant 235 observations et 4 variables : "state" (l'Etat, une variable à deux catégories A/B), "region" (la région, une variable à trois catégories), "income" (le revenu, une variable continue), "sex" (le sexe, une variable à deux catégories, homme - 1/femme - 2); ce jeu de données sera notre population finie;
- on calcule les probabilités d'inclusion en utilisant la variable revenu et un bruit aléatoire distribué selon une loi Uniforme (0, 1);
- on tire un échantillon systématique, de taille 75, en utilisant la fonction *Upsystematic* à partir de la population générée;
- on ajoute la variable «status» qui indique si une unité tirée a répondu ou pas;
- on crée des groupes homogènes de réponses selon la région et l'Etat (*rhg*);
- on applique la technique de calage en utilisant les totaux connus dans la population pour homme/femme;
- on calcule l'estimateur calé pour la variable «income» et son estimateur de la variance, et on donne un intervalle de confiance à 95% pour le total des revenus dans la population.

Le code R est donné ci-dessous. On utilise le symbole '#' pour faire des commentaires.

```
# on génère des données artificielles
data = rbind(matrix(rep("A", 165), 165, 1, byrow = TRUE), matrix(rep("B", 70), 70, 1, byrow = TRUE))
data = cbind.data.frame(data, c(rep(1, 100), rep(2, 50), rep(3, 15), rep(1, 30), rep(2, 40)), 1000 * runif(235))
sex = runif(nrow(data))
for (i in 1:length(sex)) if (sex[i] < 0.3) sex[i] = 1 else sex[i] = 2
data = cbind.data.frame(data, sex)
```

```

names(data) = c("state", "region", "income", "sex")

# on calcule les probabilités d'inclusion
pik=inclusionprobabilities(data$income+runif(1),75)

# on ajoute les probabilités à la base de données
data=cbind.data.frame(data,Prob=pik)

# on tire un échantillon aléatoire de type systématique
s=s1=UPsystematic(pik)

# on extrait l'information de la base de données selon l'échantillon tiré
s=getdata(data, s)

# la variable 'status' est générée aléatoirement en utilisant la loi Uniforme(0,1)
# 1 – signifie qu'une unité a répondu, 0 sinon
# on ajoute cette variable à la base de données
status = runif(nrow(s))
for (i in 1:length(status))
  if (status[i] < 0.3) status[i] = 0 else status[i] = 1
s = cbind.data.frame(s, status)

# on calcule les groupes de réponses homogènes
# en utilisant les variables 'region' et 'state'
s = rhg(s, selection = c("region", "state"))

# on sélectionne l'échantillon des répondants
sr = s[s$status == 1,]

# on crée une matrice X contenant les informations concernant le sexe pour toute la population; on va caler sur cette information auxiliaire (le sexe), qui doit être connue pour toutes les unités de la population
X = matrix(0, nrow = nrow(data), ncol = 2)
for (i in 1:nrow(data)) {
  if (data$sex[i] == 1)
    X[i, 1] = 1
  if (data$sex[i] == 2)
    X[i, 2] = 1
}

# on calcule les totaux de la population pour chaque modalité de la variable 'sex'
total = c(t(rep(1, nrow(data))) %*% X)

# on crée une matrice Xs contenant l'information concernant le sexe,
# mais uniquement pour l'échantillon tiré
Xs = matrix(0, nrow = nrow(sr), ncol = 2)
for (i in 1:nrow(sr)) {
  if (sr$sex[i] == 1)
    Xs[i, 1] = 1
  if (sr$sex[i] == 2)
    Xs[i, 2] = 1
}

# les totaux de la population sont
total
# les totaux de l'échantillon sont
colSums(Xs)

# on calcule les poids initiaux en utilisant les probabilités d'inclusion et les probabilités de réponse
d = 1/(sr$Prob * sr$prob_resp)

# on calcule les poids (g-weights) pour le calage en utilisant la méthode 'raking'
g=calib(Xs, d, total, method = «raking»,description=TRUE)

# on vérifie si le calage a pu être réalisé
checkcalibration(Xs, d, total, g)

# on calcule les probabilités d'inclusion d'ordre deux
pik1=UPsystematicpi2(pik)

#on extrait les probabilités d'ordre deux pour les unités répondantes pikls=pikl[sr$ID_unit,sr$ID_unit]

# la variable d'intérêt est le revenu ('income')
Ys=sr$income

# on calcule l'estimateur calé et sa variance estimée
est=calibev(Ys,Xs,total,pikls,d,g,with=FALSE,EP S=1e-6)
est

# on calcule également un intervalle de confiance à 95%
cat("[",est$calest-1.96*sqrt(est$sevar),",",est$calest+1.96*sqrt(est$sevar),"]\n")

```


Remerciements : Les auteurs remercient l'Office Fédéral de la Statistique qui soutient ce projet.

Bibliographie

Brewer, K. R. W. (1963), *A model of systematic sampling with unequal probabilities*, Australian Journal of Statistics, vol. 5, p. 5-13.

Deville, J.-C. (2005), *Peut-on croire aux sondages ?*, Document de travail, ENSAI, Rennes, France.

Deville, J.-C. et Tillé, Y. (1998), Unequal probability sampling without replacement through a splitting method, Biometrika, vol. 85, p. 89-101.

Deville, J.-C. et Tillé, Y. (2004), *Efficient balanced sampling: the cube method*, Biometrika, 91, p. 893-912.

Deville, J.-C. et Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382.

Hájek, J. (1958), *Some contributions to the theory of probability sampling*, Bulletin of the International Statistical Institute: Proceedings of the 30th session (Stockholm), ISI, The Hague, vol. 36, book 3, p.127-134.

Hájek, J. (1964), *Asymptotic theory of rejective sampling with varying probabilities from a finite population*, Annals of Mathematical Statistics, vol. 35, p. 1491-1523.

Hansen, M.H. et Hurwitz, W.N. (1943), *On the theory of sampling from finite populations*, Annals of Mathematical Statistics, vol. 14, p. 333-362.

Madow, W.G. (1949), *On the theory of systematic sampling, II*, Annals of Mathematical Statistics, vol. 20, p. 333-354.

Midzuno, H. (1952), *On the sampling system with probability proportional to sum of size*, Annals of the Institute of Statistical Mathematics, vol. 3, p. 99-107.

Rosén, B. (1997), *Asymptotic theory for order sampling*, Journal of Statistical Planning and Inference, 62:135-158.

Sampford, M.R. (1967), *On sampling without replacement with unequal probabilities of selection*, Biometrika, vol. 54, p. 499-513.

Särndal, C.-E. and Swensson, B. et Wretman, J.H. (1992), *Model Assisted Survey Sampling*, Springer, New York.

Tillé, Y. (1996), *An elimination procedure of unequal probability sampling without replacement*, Biometrika, vol. 83, p. 238-241.

Tillé, Y. (2006), *Sampling Algorithms*, Springer Series in Statistics, Springer, New York.

Tillé, Y. et Matei, A. (2008), *sampling: Survey Sampling*. R package version 2.0. <http://cran.r-project.org/src/contrib/Descriptions/sampling.html>.

SOCIÉTÉ SUISSE DE STATISTIQUE - SECTION STATISTIQUE PUBLIQUE RAPPORT D'ACTIVITÉ 2007-2008

Organisation du Comité

Le Comité de la SSS-O a pris acte de trois démissions lors de l'Assemblée générale (AG) du 17 novembre 2007: Franziska Spaeti (LU), Yves-Alain Gerber (OFS) et Ernst Matti (OFS) ont quitté le Comité; nous les remercions très chaleureusement pour leur importante contribution aux activités de la Section. Pour les remplacer, l'AG a élu les trois nouveaux membres du Comité suivants : Dieter Koch (OFS), André de Montmollin (OFS) et Simon Villiger (ville de ZH). Les autres membres du Comité ont été réélus : Sophie Rossillion (GE), Martial Clément (FR), Peter Moser (ZH), Alexandre Oettli (VD, Président) et Yves Tillé (Uni NE). Alexandre Oettli et Dieter Koch ont également pour mission de représenter la Section au sein du Comité de la SSS.

Pour assumer le secrétariat de la Section, le Comité salue l'arrivée de Zoubeida Zenati, de l'OFS, que nous profitons pour remercier de son engagement au cours de cette première année.

Les principaux dossiers suivis par le Comité ont été répartis de la façon suivante :

- Organisation des Journées suisses de la statistique publique 2008 (JSSO) : Sophie Rossillion et Dieter Koch, lequel assume également la présidence du Comité d'organisation ;
- Organisation d'ateliers et de séminaires (workshop) : Yves Tillé et André de Montmollin
- Forum et flyer de présentation : Alexandre Oettli
- Bulletin de la SSS : Martial Clément et Peter Moser
- Relations avec le Conseil d'éthique : Simon Villiger
- Représentation de la SSS au sein de l'Académie suisse des sciences humaines et sociales (ASSH/SAGW) : Peter Moser.

Durant l'exercice, le Comité de la SSS-O s'est réuni à trois reprises : en janvier, mai et août.

Activités du Comité

Un séminaire sur les « small area estimation », donné par Pascal Ardilly (INSEE), initialement prévu au printemps 2008, se déroulera le 26 septembre 2008 à l'OFS.

Un bilan mitigé a été fait concernant le forum Internet de discussion. Au vu de l'absence de coûts engendrés, décision a été prise de le conserver. Il pourrait être réactivé au besoin.

Afin de faire connaître les activités de la Section auprès des nouveaux venus dans le monde de la statistique publique, le Comité a diffusé un flyer de présentation de la Section auprès des offices membres de la CORSTAT et de FEDESTAT. Ce flyer a également été inséré dans les serviettes des Journées suisses de la statistique 2007 et 2008.

Finalement, une activité importante du Comité de la SSS-O a été la préparation des Journées suisses de la statistique publique 2008 ainsi que le lancement des opérations des Journées suisses de la statistique 2009.

En cette fin d'exercice, nous devons malheureusement annoncer le départ du Comité de Peter Moser (ZH). Qu'il soit ici sincèrement remercié de son engagement durant ces quatre années au sein du Comité.

Nous remercions de même très chaleureusement les autres membres du Comité de la SSS-O pour leur travail tout au long de cet exercice.

*Pour le comité : Alexandre Oettli, Président
Lausanne, le 8 septembre 2008*

Offizielles Organ der Schweizerischen Gesellschaft für Statistik (Swiss Statistical Society)
Organe de publication officiel de la Société Suisse de Statistique (Swiss Statistical Society)

Erscheint 3 Mal jährlich
 März, Juni, Oktober
Parution: 3 fois par an (mars, juin, octobre)

Redaktion / Rédaction
 Vertrieb / Distribution
 Marcel Baumgartner
 marcel.baumgartner@nestle.com
 Sabine Probst
 sabine.probst@stat.ch

Konzept, Gestaltung
Mise en page, graphisme
 Peter Gaffuri AG
 Kornhausplatz 7, 3011 Bern
 www.gaffuri.ch

Inserate / Publicités
 Sabine Probst
 sabine.probst@stat.ch
s/w / n/b 1/4 S. / p. Fr. 100.–
s/w / n/b 1/2 S. / p. Fr. 200.–
s/w / n/b 1/1 S. / p. Fr. 300.–
 Umschlag innen 1/1 S.
Couverture intérieure 1/1 p. Fr. 400.–
 Umschlag aussen 1/1 S.
Couverture extérieure 1/1 p. Fr. 500.–

Auflage: 460 Exemplare
Tirage: 460 exemplaires

Vorstand / Comité
Präsident / Président
 Prof. Andreas Ruckstuhl
 ZHAW Zürcher Hochschule
 für angewandte Wissenschaften
 IDP Institut für Datenanalyse
 und Prozessdesign
 Rosenstrasse 3
 Postfach
 8401 Winterthur
 Tel. +41 (0)58 934 78 12
 andreas.ruckstuhl@zhaw.ch

Vizepräsident / Vice-président
Präsident / Président SSS-BI
 Dr. Diego Kuonen
 Statoo Consulting
 Morgenstrasse 129
 3018 Berne
 Tel. +41 (0)31 998 45 90
 kuonen@statoo.com

Kassier / Trésorier
 Dr. Anne Renaud
 Office fédéral de la statistique
 Espace de l'Europe 10
 2010 Neuchâtel
 Tel. +41 (0)32 713 62 65
 anne.renaud@bfs.admin.ch

Aktuar / Secrétaire
 Marcel Baumgartner
 Nestec Ltd.
 Av. Nestlé 55
 1800 Vevey
 Tel. +41 (0)21 924 18 03
 marcel.baumgartner@nestle.com

Geschäftsführerin / Administratrice
 Sabine Probst
 Bergacher 8
 3253 Schnottwil
 Tel. +41 32 353 70 94
 sabine.probst@stat.ch

Präsident / Président SSS-O
 Alexandre Oettli
 SCRIS
 Rue de la Paix 6
 1014 Lausanne
 Tel. +41 (0)21 316 29 89
 alexandre.oettli@vd.ch

Präsident / Président SSS-ER
 Dr. Eva Cantoni
 Dept. of Econometrics
 University of Geneva
 Bd du Pont d'Arve 40
 Tel. +41 22 379 82 40
 eva.cantoni@metri.unige.ch

Mitglied / Membre
 Dieter Koch
 Bundesamt für Statistik
 Espace de l'Europe 10
 2010 Neuchâtel
 Tel. +41 32 713 68 43
 dieter.koch@bfs.admin.ch

© 2008, SWISS STATISTICAL SOCIETY

Die Schweizerische Gesellschaft für Statistik dankt der Schweizerischen Akademie der Naturwissenschaften SANW für die finanzielle Unterstützung des Bulletins.

Die Schweizerische Gesellschaft für Statistik, gegründet im Jahr 1988, fördert die Anwendung und die Entwicklung der Statistik in der Schweiz, vertritt die Interessen der auf diesem Gebiet tätigen Personen in Praxis, Forschung und Lehre und trägt zur Anerkennung der Statistik als eigenständige wissenschaftliche Disziplin bei. Das Bulletin wird jedem Mitglied der Gesellschaft zugeschickt. Der Jahresbeitrag beträgt Fr. 70.– für natürliche Personen und Fr. 350.– für juristische Personen. Bitte senden Sie Ihren Beitrittsantrag oder allfällige Adressänderungen direkt an die Geschäftsführerin. Artikel, Kommentare und Informationen für das Bulletin sind sehr willkommen. Bitte schicken Sie Ihre Beiträge der Redaktion.

La Société Suisse de Statistique (SSS) a été fondée en 1988. Ses objectifs consistent à promouvoir le développement et l'application de méthodes statistiques en Suisse, de représenter les intérêts de toutes celles et de tous ceux qui s'occupent de la méthodologie statistique dans l'industrie, dans l'administration, dans la recherche et l'enseignement, et de contribuer activement à la reconnaissance de la statistique en tant que branche scientifique indépendante. Le bulletin SSS est envoyé à tous les membres de la société. La cotisation annuelle s'élève à 70 francs suisses par personne et à 350 francs pour une société. Nous vous serions reconnaissants de faire parvenir votre inscription à la société ou de communiquer d'éventuels changements d'adresses à l'administratrice. Articles, commentaires et informations sur le bulletin SSS sont les bienvenus. Vous êtes priés de les envoyer directement à la rédaction.

SWISS STATISTICAL SOCIETY

- Ich möchte der Schweizerischen Gesellschaft für Statistik beitreten
- Je désire être membre de la Société Suisse de Statistique*

Ich möchte zusätzlich folgender
 Sektion beitreten:
*Je désire également faire partie
 de la Section :*

- Öffentliche Statistik
Statistique publique
- Business & Industry
- Education & Research

Name / Vorname • Nom / Prénom

Adresse

PLZ / Ort • NPA / Lieu

Tel.

E-Mail

Schicken Sie die Beitrittserklärung an die Schweizerische Gesellschaft für Statistik, Sabine Probst, Bergacher 8, 3253 Schnottwil
Formulaire à envoyer à La Société Suisse de Statistique, Sabine Probst, Bergacher 8, 3253 Schnottwil



Thrive in an environment of innovation.
Experience exceptional career opportunities.
Think what's possible.

A global healthcare leader, Novartis has one of the most exciting product pipelines in the industry today. A pipeline of innovative medicines brought to life by diverse, talented, performance-driven people. All of which makes us one of the most rewarding employers in our field.

Statisticians and Senior
Statisticians for International
Clinical Development
Pharmaceuticals
Basel, Switzerland and
Horsham, UK

Working as part of the Biostatistics group of the Clinical Information Sciences function of Novartis Pharmaceuticals Development, you will provide input to clinical development plans, in co-operation with the Global Clinical Development Team. As a team member you will bring statistical expertise to the design, analysis and reporting of clinical studies from phases I to IV and continuously strive for innovative solutions for our studies.

You will have access to an international team of statisticians in our Global Biostatistics group and world-class Statistical Methodology Consultants. You will ensure high scientific standards for studies, explore data for product profiling, hypothesis generation, new research topics and publications. You will co-ordinate project and/or study tasks with other statisticians and statistical programmers. We are organized in a truly global and team-focused environment.

It all calls for experience in the field of clinical trials, a working knowledge of methodology and regulations, and a relevant PhD or MSc. level. Strong communication, interpersonal and negotiation skills, as well as a highly flexible and creative approach, will also be essential. Knowledge of SAS and ideally some alternative statistical software will be required.

In return, we offer a competitive salary package, excellent benefits and state-of-the-art facilities.

To apply, please visit our career website www.novartis.com/careers using the Job ID 41670BR for Switzerland, or 38768BR for the UK. If you have questions, please contact us at friedhelm.hornig@novartis.com or call us on +41 61 324 85 14.