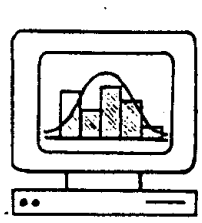
Information Bulletin



Czech Statistical Society English issue, December, 1992

A few words on the affinity of the dissimilars

Stanislav Komenda

Not so many topics exist in which peoples' opinions coincides in such a complete way. Two of them are coming to me - namely that (1) meteorologists are mistaken and (2) statisticians are telling lies. And what aggravates them - their lies are very tedious.

A few words can be said about the last point.

In many versions and modifications the characteristics like those (cited by H. Swoboda: Modern Statistics): " Three kinds of lies are known - lies, proscribed lies and statistics" - circulate in the information stream of mankind. Since this discipline of statistics was not able to deduce from its bad reputation any adequate conclusion and lasts on its right to exist in the neighbourhood of other sciences with a honest reputation (let me introduce otorhinolaryngology or sexuology as an example), the question arises about the reason for it. Were not the mendacious statisticians to able also falsify the statistics demonstrating the increasing use of the statistics within the biomedical sciences frequently?

The author of these rows is a statistician and thus a man not independent and prejudiced in any case. The truthfulness of his viewpoints, statements and opinions decreases due to it. Well, let us do a gentlemen's agreement - by which he will not express himself entirely as a statistician, and the reader will believe (suppose there will be anyone ready to read these rows) at least partially to his words. Say - one third seems to be enough. And we will see later.

The world we live in is full of absurdities. By the common opinion a statistical table is a top ennui. We are ready to recommend psychiatric treatment for anybody studying mortality and morbidity tables in bed before going to sleep - and the fact that no such treatment has been applied till now we explain by the hypothesis that the nearest medical institution is completely occupied.

The ennuing man is going to play. Or to look at a play, at least. To play football or to watch people playing football is a reliable way for a lot of people to amuse themselves. Play or games seem to be in opposition to statistics.

But the statistician and the player or gambler as well - both refer to the same person. They refer to what is being denoted as - Chance. Mortality tables and the score indices

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

are being influenced by the same chance, undoubtedly. In this sense, Statistics - tedious and dismal gammer, and Play - merry girl ready to amuse and distract, are closely related. Maybe sisters, sisters connected just by this common chromosome of Chance.

By the way - maybe you have read one of the best plays of George Voskovec and Jan Werich - The Donkey and its Shadow. Perhaps you have an opportunity to take it from the shelf of your library. If it is the case, do it. It's a play with a great "P". On the introductory page the dedication can be found: To our Muse, Karel Schönbaum, thankfully - V + W. In the Memoires of Jan Werich (Jan Werich remembering ..., namely Potlach, edited by Melantrich, Praha 1982) the explaining text can be found on page 46: "... this play The Donkey and its Shadow has been written by Voskovec and me as a paraphrase of the anecdote by Lukianos, some 2 000 years old. The initiative was that of our friend and good fellow and an excellent man Charlie Schönbaum, correctly JUDr.Karel Schönbaum, to whom the play was dedicated ... Charlie Schönbaum was a solicitor, but more strongly attracted to mathematics. There is a detective story on my bedside table - and there was a thick book entitled Equations of Higher Mathematics laying on his, and he has been reading in it every evening as in a detective story. The idea of The Donkey and its Shadow of Lukianos was his, and all German occurring in the play, various German verses, originated with him".

It was Jan Werich.

JUDr. Karel Schönbaum, solicitor in love with mathematics, has been murdered by Nazis in a concentration camp. The chromosome of Chance connecting Statistics with Play also demonstrated itself in this family in other ways. Karel Schönbaum had a brother, University Professor Dr. Emil Schönbaum, prominent specialist in the statistics of the period between the wars, Czechoslovakia, engaged in the insurance mathematics, and the co-author of the project of social insurance policy (I am grateful to Dr. Marcel Josífko for this information). Thus, Statistics and Play are able to tolerate each other.

Certainly, it can be objected that - just as a single swallow does not bring spring - a single statistician does not make statistics for recreation. It is right. The World Association of Statisticians does not make statistics amusing. However it does not matter at all. What's important is that statistics deals with the events and circumstances of our everyday life and is considering them in the language of mathematics, in the language of figures and categories, necessarily rigorous and of economy. This means cool and unbiased.

The difficulty arises from the fact that these cool and unbiased statistical tables are being taken into our hands, in the capillaries of which circulate not only plasma and erythrocytes, but also favours and wishes to see the matters occur in a particular light, usually tendentious, wanting the matters go in just this way and not the other, wanting the results of a survey to advocate this and to reject that.

While matrix theory, differential geometry, topology and other branches of mathematics address the mathematicians, statistics speaks to the more general public, such as medical statistics, to offer an example. Statistics differ from other mathematics just in this point. Statistics do not tell lies - but their interpretations can and do. The theory of statistics is a mathematical theory and the domain of specialists as such. On the other side, interpretation of statistical results is open to anybody. Thus also to laymen who want these results confirm

the desire of them. And because they consider their desires noble-minded, they do not see any reason why to set any limits of their nobility.

Living organisms possess the senses which enable them orientation in the environment, which is knowledge and understanding sui generis. A human being is the only biological species learning his/her environment consciously, in such a degree, that learning has been elimited as an autonomous activity. The reason is that just this ability to learn systematically and actively was recognized by him as his most imporant power. The process of learning about the environment and about himself, too, has its own organization. A very complicated organization - Science and Research - formed into the network of deeply and widely founded disciplines, among them biomedical sciences. The basic idea of the Science and Research is that of the recognizability of our World and our ability to grasp it through the system of general laws, accessible by means of particular observations and measurements derived from controlled experiments, through the laws concerning relations and predictions among the events and variables. The fundamental instruments in the biomedical sciences are - observation, measurement and experimentation, the same principle being applied in the case of weighing a newborn or to an assessment of the computer tomography data. In contrast to the physical, chemical and technical measurements, those carried out in biology and medicine have a very significant specificity - consisting in the relatively high complexity of the objects and subjects under consideration. Due to it the statements being made about them are also relatively complicated. In other words, the results of biomedical measurements are - as a rule - under the influence of many factors, a great deal of which our control is not able to take into account - the effect of them being significant and far from to be considered negligible.

The history of science and the methodology of science proved the ability of statistical concepts to apprehend operating of these factors when the laws of Nature are to be formulated. The reason is that these factors behave, under certain circumstances, like the chance - this chance being in relation to the statistics. The way in which these factors operate in the area of biomedical sciences we are used to denoting by the term "biological variability". Some kind of uncertainty and ambiguity of measurement manifests itself through it, and thus also of the statements deduced from these measurements. There are no mysteries in it - biological variability is an ingredient of our everyday life with the self-evidence similar to that of respiration and digestion.

In the set of pregnancies having finished by the deliveries of healthy children the length of these pregnancies varies within some weeks. Besides it, body weight of these newborns varies within some hundreds of grams and of their length at delivery within some centimeters. Certainly, there exist factors like heredity, age of the mother and others, which make prediction of the respective values of these variables more precise. Nevertheless, such a forecast can never be of absolute certainty - which guarrantees the field where statistics might help. Telling lies through statistics does not give reason for a condemnation of statistics, but is the reason to condemn the liar - and, of course, to stimulate better understanding of the instrument by means of which a good deal of our world could be understood.

Discrete Gaussian distribution

Jiří Anděl

Seeing the title of this paper, the reader must come to the idea that a new nonsense like "round square" has been made. It has been known for more than some 200 years that the Gaussian distribution (which is also called normal distribution) is continous. Of course, I mean a regular Gaussian distribution!

Let us start at the beginning. A few weeks ago I made a business trip to Bratislava. Just before my departure back to Prague I was asked by Professor MUDr. M. Mikulecký, DrSc., who is a well known doctor of medicine and friend of statistics, to consult an interesting problem. He told me that he used a pocket calculator and added the values of the density of the $N(0,1)$ distribution

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

at the points $x = 0, \pm 1, \pm 2, \ldots$ and that the sum was equal to 1 for many decimals. He asked me a question, if it is really true that

$$\sum_{k=-\infty}^{\infty} \varphi(k) = 1$$

or if his result holds only approximately. I answered that I thought that the sum is not exactly one. I motivated my decision by the fact that the difference between the integral (which is, of course, exactly one) and the analyzed sum would contain some complicated members of the Euler–Maclaurin formula, which would be so complicated that they could not cancel to give zero. Mr. Mikulecký was very pleased. First of all, my answer was also negative as the answers of the other two matematicians who had been asked before me (Professor Komorník and Professor Huťa), and second, because I introduced promptly two names of well known mathematicians Euler and Maclaurin. It looked like a scientific diagnosis in medicine.

Having returned to Prague I tried to solve the problem, but it seemed to be rather complicated. Of course, I started with a computer and checked if the sum is really so near to one. But Mr. Mikulecký was right, the result was 1 point and some eight zeros and only after that there was a non-zero figure.

Now, I should take a little break, say 1 month or so, to leave the readers a chance to solve this problem. Whoever has good nerves can try it. Who does not can continue reading.

I also entrusted this interesting problem to my colleagues. I must tell you that many of them devoted much time and effort to its solution. It was not a question how to make numerical procedures more exact. If the sum were exactly one (and it seemed quite possible, because there were so many zeros after 1), no computation could prove it. And, moreover, we are proud of our mathematical skills, aren't we?

The answer was given practically immediately by Professor Dr. Břetislav Novák, DrSc., who is a specialist in the number theory and so he knows perfectly many useful properties of theta function. One of them implies that for every positive s we have

$$\sum_{k=-\infty}^{\infty} e^{-k^2\pi s} = \frac{1}{\sqrt{s}} \sum_{k=-\infty}^{\infty} e^{-k^2\pi/s}. \tag{1}$$

To my great surprize this formula can also be found in the most popular texbook of mathematical analysis written by Jarník (Integral Calculus II, Chap. XIII, formula (127)). I was quite near to it when I was looking for the solution. It was clear to me that the series under consideration could have some relation with the integral

$$\int_{-\infty}^{\infty} e^{-k^2\pi s} \cos 2k\pi x \, dx = \frac{1}{\sqrt{s}} e^{-k^2\pi/s},$$

but I looked only into the textbook by Fichtengolc and there were no furter details there. Professor Jarník, who was also a well known specialist in the number theory explained such consequences with great pleasure in his book . By the way, formula (1) is a corolary of the Poisson formula from the area of the Fourier series. Now, it is very easy. It suffices to put $s = 2\pi$ in (1) and one gets

$$\frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} e^{-k^2/2} = \sum_{k=-\infty}^{\infty} e^{-2k^2\pi^2},$$

which is equivalent to

$$\sum_{k=-\infty}^{\infty} \varphi(k) = 1 + 2\sum_{k=1}^{\infty} e^{-2k^2\pi^2}.$$

From here it follows that the sum equals one plus a series with positive terms. That's why the sum must be larger than one. But not much larger. The series on the right hand side converges fantastically fast. It is easy to calculate that one is exceeded only by 5.4×10^{-9}.

Sorry, the discrete Gaussian distribution really does not exist. But it was not very far from existing, was it?

Don't be obstinate

Jiří Anděl

In the American journal *Parade Magazine* the following question was posed in 1990:

"Suppose you are on a game show and given a choice of three doors. Behind one is a car; behind the others are goats. You pick door 1, and the host, who knows what's behind them, opens door 3, which has a goat. He then asks if you want to pick door 2. Should you switch?"

Ms. Marylin von Savant, the columninst, replied: " Yes you should switch. The first door has a 1/3 chance of winning, but the second door has a 2/3 chance." She gave some reasons and a proof of her assertion but the professional statisticians called it dubious and false. Soon after publishing Ms. von Savant's answer letters from three Ph. D.'s appeared. All of them unanimously claimed that the correct probability of winning with either remaining door is 1/2. Ms. von Savant reported: " I'm receiving thousands of letters, nearly all insisting I'm wrong." If you want to know it exactly, then 92% of the letters from the general public were against her answer and 65% of letters from universities were against her answer.

In the paper Morgan et al. (1991) from where I have the information given above, the following possible solutions are proposed.

1. If the host gives no opportunity to switch, the probability of winning would be 1/3. Hence the probability that the player wins if he (or she) does switch is 2/3.

2. Denote briefly the car by A (it means auto) and the goat by G. The sample space is then AGG, GAG, GGA and each element of it has probability 1/3. The player choosing door 1 will win in two of these cases if he (or she) switches. Hence the probability of winning is again 2/3 in this case.

3. Play the game many times. Use three cards, one of which represents the car. You will verify that the probability of winning is 2/3 if the player switches.

4. Consider the same sample space as in the second solution. The door 3 has been shown to contain a goat and so GGA is no longer possible. The remaining two events must have equal probability. Hence the probability of winning is 1/2 regardless of switching or not.

5. The probability that a player is shown a goat is 1. Therefore conditioning this event cannot change the probability of 1/3 that the car is behind the door 1. So the player has a 2/3 probability if he (or she) switches.

6. The sample space is $AGG2$, $AGG3$, $GAG3$, $GGA2$ where the number indicates the door opened by the host. The corresponding probabilities of the mentioned elementary events are 1/6, 1/6, 1/3, 1/3. Let S be the event "the player switches and wins" and let D3 be the event "the host opens the door 3 and shows a goat". Calculations give

$$P(S|D3) = \frac{P(S \text{ and } D3)}{P(D3)} = \frac{P(GAG3)}{P(AGG3 \text{ or } GAG3)} = \frac{\frac{1}{3}}{\frac{1}{6}+\frac{1}{3}} = \frac{2}{3}.$$

Have you already chosen which solution is most attractive? Or have you even found another one giving an answer different from 1/3 or 1/2?

Morgan et al. namely claim that all these solutions are generally not correct and, especially, the solution of Ms. von Savant is also not correct. I believe that most of letters presenting disagreement are clearly caused by the fact that people make often errors when they use conditional probabilities. But in the opinion of Morgan et al. it is not possible to

add any other assumptions to those formulated in the original question. For example they claim that with a positive probability the host can open the door 3 also in the case when the car is behind it. I cannot imagine a host who would do it and who would ask then if the player wants to switch. I think that this would be the last opportunity for the host to act on a game show because a medical care in a psychiatric clinic would be inevitable. At the same time Morgan et al. assume that the car is behind door i with probability 1/3 for $i = 1, 2, 3$. But this is also an assumption which is not given in the question. By the way, I do not think that it is reasonably fulfilled in similar games.

I propose to write down the assumptions under which the problem is solved. If a reader disagrees with some of them, he/she can replace them by different ones and derive another solution. But it will be a solution of a different problem. We accept, for example, all the three solutions of the well known Bertrand paradox as correct although they give three different answers. We consider them as solutions of three different problems, the formulations of which are nearly identical.

Well, we shall definitely assume:

(i) Probability that the car is behind the door i, is p_i. The sum $p_1 + p_2 + p_3$ is 1.
(ii) The host never opens the door chosen by the player.
(iii) The host never opens the door behind which is the car.
(iv) In the case AGG the host opens door 3 whith probability p and the door 2 with probability $1 - p$.

Our assumptions imply

$$\begin{aligned} P(AGG) = p_1, \qquad & P(GAG) = p_2, \qquad & P(GGA) = p_3, \\ P(D3|AGG) = p, \qquad & P(D3|GAG) = 1, \qquad & P(D3|GGA) = 0. \end{aligned}$$

The Bayes theorem yields

$$P(GAG|D3) = \frac{p_2 \times 1}{p_1 \times p + p_2 \times 1 + p_3 \times 0} = \frac{p_2}{p_1 \times p + p_2}.$$

Generally, we cannot say anything else. If $p_2 = 0$, $p_1 > 0$ and $p > 0$ then $P(GAG|D3) = 0$ and so switching is harmful. In the case $p_2 > 0$ and $p_1 = 0$ we get $P(GAG|D3) = 1$ and switching is the winning strategy. The car should have to be always behind the door 2. But both the situations are so simple that no reader would ask the question. We get an interesting result if we also assume that

$$p_1 = p_2 = p_3 = \frac{1}{3}$$

which most readers would obviously do and which also Morgan et al. did. Then we obtain

$$P(GAG|D3) = \frac{1}{1 + p}.$$

And this probability is at least 1/2 for every $p \in [0,1]$. Then it is advantageous not to be obstinate, not to insist on door 1 but to switch and pick door 2. Let me introduce that this result already follows from the solution derived by Morgen et al. if some of their parameters are zero. In spite of it some people claim that they wouldn't switch even if they know that the probability of winning is larger after switching. They would have pangs of conscience when it appears later that their original decision was correct.

It is clear that Ms. von Savant based her solution on the assumption that $p = 1/3$. Then, of course, $P(GAG|D3) = \frac{2}{3}$.

At the end I would pose another question. As far as I know, Ms. von Savant has the meaning Miss von Savant as well as Mrs. von Savant. What do you think is correct? Perhaps I should call her simply Marylin, it sounds so lovely.

References

Morgan J. P., Chaganty N. R., Dahyia R. C., Doviak M. J. (1991): Let's make a deal: The player's dilemma. *Amer. Statist.* **45**, No. 4, 284 –287.

I wouldn't be obstinate if ...

A comment on the article *Don't be obstinate* by J. Anděl

Jan Klaschka

I would take the author's advice and wouldn't be obstinate, if the equality $P(D3|GAG) = 1$ could be relied upon. However, this follows from assumptions (i) – (iv) only if the host *has to* open some door. (Whatever I do, I can only see that he has opened one.)

Provided there is also a possibility that the host does not open any door, the solution presented requires a revision. Assumption (iv) should be modified – the host opens door 2 in the case AGG with probability q which is not greater than $1-p$. In addition, $P(D3|GAG)$ may be assumed to take on some value r. The Bayes theorem yields this time

$$P(GAG|D3) = \frac{p_2 \times r}{p_1 \times p + p_2 \times r}\,,$$

so that on assumption $p_1 = p_2 = p_3 = \frac{1}{3}$

$$P(GAG|D3) = \frac{r}{p+r},$$

which is less than one half for $r < p$. thus a player willing to win a car with probability of at latest 50% regardless of p and r values obviously cannot be advised better than to toss a coin.

Organizers of such game shows perhaps rarely make every effort to minimize prizes. Nevertheless if I still played the part of a stingy host and was competent to do so I would choose $r = 0$ and $q = 1 - p$ (i.e. the strategy to "examine" a player just when he/she is picking the correct door). A reader of the Bulletin following the recommendation in the title of the article commented on would then become my easiest victim.

Statisticians Anonymous ?

Jan Klaschka

Almost everybody has heard about the international (and even intercontinental) movement Alcoholics Anonymous, whose members, mostly abstaining alcoholics, help each other in maintaining abstinence. The article "Fundamentalists and the others" by Pavel Hartl (Lidové noviny, 8 August 1991, Sunday supplement, p. 2) informs that a number of similar self-help movements or groups act in the civilized countries: nearly 200 kinds of problems concerned by these groups or movements have been counted by American experts. There are Gamblers Anonymous, Overeaters Anonymous, Lovers Anonymous (associating women vexed by compulsion to enter repeatedly similar hopeless relationships, not to be mistaken), Fundamentalists Anonymous... Reading the article, an idea is coming to me: And what about Statisticians Anonymous?

Does this make sense? Let me explain. I have spent most of my professional career until now (10 years out of 12) as a practicing statistician in biomedical research, and this work is the permanent source of my dissatisfaction. An attempt to indicate some of the reasons for these feelings is as follows.

(1) As a member of a statistical minority in the institution where I work, I have to face the danger that (to utilize contemporary political slang) I might lose my identity (i. e. contact with my own discipline) and end up occupying the role of an inferior alien worker.
(2) Making oneself understood with users about fundamental statistical issues is as much impossible as attempting it is inevitable. A permanent feeling of vanity is thus induced with high effectivity. (However, let's not oversimplify: There are three faces at the stage altogether - a user, me and the methodological establishment.)
(3) No application of statistics is ultimately correct; no one may be more frustrated by that than a mathematician.

The objections listed above are of such general nature that I would wonder if they would not resonate with the feelings or thoughts of some other unfortunates. If they do, it may be concluded that the profession of a practicing statistician is not only my personal problem, but a problem in itself. And such a fact might be a sufficient reason for establishing the Statisticians Anonymous club.

What should this club be occupied with? Primarily with occasional discussions of the problems of both statistics and statisticians. Improving the professionalism of participants might be a desirable side-effect, but the principal aim should lie in helping statisticians to adapt to their profession. The discussions should by no means be similar to scientific seminars, where only a successful solution of a problem qualifies one to speak. On the contrary, in my opinion even the problems that cannot be hoped to be factually solved within short (see the objection number 3 above, for instance) should not be avoided because, quoting the article by Pavel Hartl, "one of the key processes in self-help groups is offering an opportunity to the members to recognize common features of the problem".

As for the abstaining alcoholics' movement, the anonymity means the form of contact of needy people with local groups. With regard to the proposed statisticians' club, the word has another sense: I suppose that the idea should address especially those statisticians, who are not very successful (mainly in the theoretical work) and hence are anonymous. (Hopefully not only the stupid and lazy ones will be attracted).

Well, I have to confess: The intention to establish the Statisticians Anonymous club is half serious and half a joke. Which one of these bearings becomes prevalent is left to the readers. Self-help movements are said to be initiated commonly with a newspaper advertisement. Let this article be an advertisement too. Do you find the idea interesting? Don't then hesitate to contact me.

Author's address: RNDr. Jan Klaschka, Psychiatric Center, Ústavní 91, 181 03 Prague 8 - Bohnice, Czechoslovakia. E-mail: pcpoam@csearn.bitnet.

Mr. Jan K l a s c h k a
Psychaitric Center,
Ústavní 91,
181 03 Prague 8 - Bohnice

Anonymous Alcoholics or A letter of a Statistician to the Statistician

Dear Colleague,

let me answer on your paper and the ten-year experience of you as presented in the Inf. Bull. of the Czech Statistical Society (November 1991), this response being that of a statistician destined to join his life with the biomedical research more than 30 years ago. This long series of years are not meant as any credit or qualification - they are considered as an objective biosocial event, doubtlessly in a correspondence with the topic.

In the last paragraph of your paper you admit your opinion (foundation of KASTA or KANASTA) to be half a joke and half a serious matter. This point of view I consider sympathetic - as a proof that the complications you have met on your path of life as an applying statistician did not break you in spirit and are proof of your ability not only to see the problems but also to see them from a broad-minded distance.

Sometimes in the middle of the Sixties when I was a biometrician for the same period as you are now, a fixed idea persecuted me to write just the same paper as you did. This idea never found its fulfilment due to my ignorance of the Alcoholics Anonymous movement and of the non-existence of a platform where to publish it, as well. Nevertheless, if I did hate cleaning less than I do, some meditations would be able to find in a bottom drawer of my table considering topic of a "loneliness of a field-applying statistician". In comparison with yours I had one reason more for such a meditation : the distance between the Hanakian City of Olomouc and the Prague Center or its quarter of Karlín is much longer than that of the Prague quarter of Bohnice.

But, let's get down to details.

An applying statistician convinced of the meaningfulness and vitality of his/her mission, who wants to be the statistician among physicians or biologists, should not compare his position with that of a virgin kidnapped by the Turks into the sultan's harem. Might he be able to communicate with his clients, it is necessary to find common speech - which needs him to observe with his eyes open and to listen to, to receive and accept the paradigms of experimental research and to offer - avoiding any abstract intellectual provocations - suitable statistical support. Doing it, his statistical individuality should be maintained, without any tendencies to play the role of "one among the physicians". I never held it as a loss of my statistical honor to do what was required to be done. In my opinion the identity could be saved by having some effective statistical principles in the repertoire and by applying and transferring them to the collaborators who are able to accept and to comprehend. The presence of a statistician in the collective of experimenters (of an institution or a faculty) should be distinguishable (by the way of editing results and conclusions formulation) and should be known. Of course, besides the objective factors some subjective parameters also play their part - such as the statistician's social integrity, assertion, effectivity.

The role of an applying statistician is complicated by the absence of his everyday activities in the list of disciplines officially accredited. This problem is one of those the statisticians have to solve themselves.

The key question to answer at the very beginning by the statistician taking his/her role seriously, is that the reality of the biomedical research offers only exceptionally an occasion for the mathematical "Let's have ... " What has been done when applying, is almost always conditioned and in some way relative. The categories applied have the features of something fuzzy, their contours being unclear. The statistician should defend against false fictions to stress objectivity, to criticize, to fight against logical inconsistencies, to measure, estimate, evaluate.

It is the statistician's problem to confront the demands of "hard" conclusions from the "soft" observations with the mirror of reality - and to put it through that also in the information area and that of knowledge no perpetuum mobile is possible to construct.

Each applying statistician knows the role of Pontius Pilatus washing his hands over the innocent Jesus Christ committed by the Jewish power - in the hope to relieve himself of his own responsibility. It is not permitted to avoid treatment of questionable data - not to tarnish a statistician's reputation - because the respective client might treat them by himself committing thus greater loss in comparison with the possible gain of a cooperation of both of them - client and statistician.

A statistician's experience on the way of biometrical applications allows him to recognize there some periods - maybe as follows:

(1) Period of amazement, due to that "all is going other way" - what has been written in the statistical textbooks and presented in the lectures does not offer ready, well-fitted rules how to search for responses to the questions the experimenters put to a statistician.

It is not easy to recognize : are the questions bad or the rules offering the response? By the way, this version is the better one; in the worse one the experimenter asks the statistician not only to respond but also to formulate questions.

(2) Period of disillusion, linking up continuously to the amazement period and in the same way passing into a

(3) period of scepticism, sometimes taking on the form of the philosophical agnosticism and behavioral cynism codified by the sentence "Anything is worthless", this period leading - in the worse case - into a

(4a) Period of a compromise, resignation and searching for a compensatory raison d'etre, or a

(4b) Period of understanding that searching for a both-side acceptable basis is inevitable; period of a creative adaptation; period of the comprehension that "risk" is a notion denoting entities actual in a similar way as the notions of "hardness", "temperature" or intelligence"; a period of recognition that the greater amount of jokes addressed to statistics in comparison with those addressed to algebra or topology does not prove statistics to be worse than algebra or topology but to be more provocative towards real life.

Nevertheless, it's the topic for another paper.

Dear Colleague, these remarks of mine should be a letter. So it is polite to conclude as I would in a letter.

I am convinced that I did not write anything unknown to you. It does not matter - to give information seems not to be so important than the possibility to let the addressee known that his/her deal is being shared among more human beings related by the same lot. For life as such and also for the statistician's life no exhaustive instructions could be given-but it's always possible to hold a speech.

As I indicated at the beginning, your attitude towards applications is sympathetic to me. A certain amount of a reasonable scepticism can not be harmful for your marriage with the lady Biometrics - on the contrary it may become an efficient prevention against divorce.

Sincerely yours

Stanislav Komenda

1991 census of population, houses and homes

Problems and solutions

Lubomír Balík, Pavel Čtrnáct, Štěpánka Morávková, Hana Vorlová

The midnight of March 2 to 3, 1991 was the decisive moment of the seventh population, houses and homes census in the history of Czechoslovakia. The authors would like to acquaint the reader in two installments with the methodological and organizational conditions under which the census was effected, and also to give him a brief survey of the preliminary results that have been published only recently. They deem it fitting that mention should be made on this occasion of the basic figures of the preceding censuses, as well as the fact that the census of December 31, 1869 is considered to have been the first of its type in Czechoslovak territory. The history of this most extensive statistical undertaking is thus longer than 120 years.[1)]

Date of census	Czech and Slovak Federal Republic	Czech Republic	Slovak Republic
	Estimated numbers of inhabitants[2)]		
4th - 6th century	500 000	350 000	150 000
round 1000	1 500 000	1 100 000	400 000
end of 12th cent.	1 500 000	1 100 000	400 000
begin. of 15th cent.	3 000 000	2 300 000	700 000
begin. of 17th cent.	4 000 000	3 000 000	1 000 000
mid–17th cent.	3 000 000	2 000 000	1 000 000
1705 - 1720	3 400 000	2 400 000	1 000 000
1754	4 820 000	3 360 000	1 460 000
1787	6 300 000	4 355 000	1 945 000
1840	8 724 000	6 369 000	2 355 000
1850	9 230 000	6 792 000	2 411 000
1860	9 656 000	7 256 000	2 400 000
	Results of population censuses[2)]		
31.12.1869	10 099 041	7 617 230	2 481 811
31.12.1880	10 699 534	8 222 013	2 477 521
31.12.1890	11 260 601	8 665 421	2 595 180
31.12.1900	12 155 139	9 372 214	2 782 925
31.12.1910	12 995 294	10 078 637	2 916 657
15. 2.1921	13 003 446	10 009 587	2 993 859
1.12.1330	13 998 497	10 674 386	3 324 111
1. 3.1950	12 338 450	8 896 133	3 422 317
1. 3.1961	13 745 577	9 571 531	4 174 046
1.12.1970	14 344 987	9 807 697	4 537 260
1.12 1980	15 283 095	10 291 927	4 991 168
3. 3.1991	15 567 666	10 298 731	5 268 935

[1)] Comp. the papers by J. Růžičková in Statistika No.7,8 and 9/1989.

The preparation and the census

The statistical public will remember the really unbelievable concurrence of adverse circumstances during the periods of the preparation of the census and the census itself. After the November revolution the Czechoslovak statistical organs started revising some of the material and methodological problems of the census the solution of which had until then been blocked. This applies to questions concerning religious affiliation, which are quite currently asked the respondents in census in a number of advanced countries (in Czechoslovakia these questions were last posed in 1950), the methods of ascertaining the economic activity of the population (the classification having been extended so as to incorporate job seeking individuals), social groups (the group of employers added, as well as a special group of unpaid persons helping in family households), and nationality (separate surveys of the Ruthenians and Ukrainians: the demand for equal classification of the Romanies having been successfully pushed through before), as well as some minor changes in the formulation of the individual questions.

These amendments were aimed at improving the international comparability of the results of the census, as well as a closer approximation to the changing social situation. The amendments adopted however involved revision of roughly eighty per cent of the higher to completed project documentation. Some more important differences remained only in the classification of education and employment where such rapid changes in the existing national system of classification would have involved great risks of failure.

The following breakneck development of the situation led to local elections being called at the end of November 1990. This called forth the necessity to change the original date of the census (Dec.1, 1990) and postpone it till March 3, 1991. the delay may have seemed quite welcome because it enable cutting of the arrears in the preparation of the project, but it also brought three serious complications: collision with the term of spring vacations in some regions. budgetary problems linked with the rapid growth of prices and wages: and, primarily, major changes in both federal and local administration close before the census. Apart from these facts, various nationalistic tendencies were gaining intensity, which of course also had its effects on the preparation of the census.

The system of national committees, which played an important role in the preparation of the census went factually to pieces long before the local elections. This resulted in a delay of the preparatory operations particularly as far as the demarcation of the census districts and their descriptions are concerned. In some of the districts the work scheduled for August 1990 had not been completed until February 1991, viz. close before the census. This work was of course participated in by the regional organs of the Czech Statistical Office and the Slovak Statistical Office: in some of the districts statisticians practically substituted fully for the old disestablished national committees. The newly established organs of local government were often unexperienced as far as that kind of work is concerned and they

[2)] Converted to today s territory. From 1869 till 1950 the population present: since 1961 resident population.

were surprised and literally caught unawares by the extent and the exactions of the task that was to be fulfilled. In that situation the organs of the regional statistical offices were actually of decisive importance and it was thanks to them that the territorial preparation of the census and the distribution of the printed forms were successfully completed in time despite various complications.

Another adverse circumstance was the selection of the census-administering organs and their financial remuneration based on the Labour Code, the amendment of which, effective since February 1, 1991, was the cause of the greatest problems. When that amendment was being debated the issues raised by the Federal Statistical Office were not taken any account of and the work of the census-administering organs was classified with the so-called publicly useful work for which the employee should be released by his employer, though without any right to compensation for the loss of his regular pay. Since the budgetary resources for that pay refund could not be garnered in time, the situation resulted in a large number of economically active commissioners rejecting any offers to participate in census administration. This also explains the unusually large share of retired people and students in the census staff, the other members being entirely dependent upon the benevolence of their employers: some of them had to discharge their duties of census organs out of their working time or office hours. This of course had a negative effect and resulted in a time squeeze under which the commissioners and the inspectors were assembling the materials and checking them, and summarizing the data for preliminary results. In a number of municipalities as many as fifty per cent of commissioners had to be released from their duties of census organs and replaced by completely new staff close before the census, and even after it was launched.

Indicator	Czech and Slovak Federal Republic	Czech Republic	Slovak Republic
municipalities	8 602	5 768	2 834
parts of munisipalities	13 437	11 627	1 810
basic residemtal units[3)]	29 396	21 990	7 406
of which: urban districts[3)]	9 587	8 007	2 580
census districts[4)]	69 847	49 306	20 542

Tab.2. Territorial units in the 1991 census

Since the beginning of this year the census has been the target of a wave of attacks from various directions. In the Federal Assembly the members of the Movement for Democratic Self–Government – the Society for Moravia and Silesia, as well as other members of the Federal Assembly, launched a campaign for the "recognition" of Moravian and silesian

[3)] Incl. parts of the basic residential units and urban districts

[4)] Except census districts in the armed forces, the organs of corrective training and abroad.

nationalities. Claims have been made and demands voiced in the form of official letters addressed to the President of the Federal Statistical Office and to the members of the law-making bodies, various petitions and demonstrations of the squares of Moravian towns and villages.

In this context it should be emphasized that the formulations of the questions in the census sheet concerning nationality, as well as the formulation of the explanatory notes, are in full accord with the declarations of the Charter of Basic Human Rights and Liberties adopted on January 9, 1991. and that free option of every citizen as far as nationality is concerned has never been doubted. The statistical organs never claimed the right to give a definition of the terms "nation" or "nationality", even though the respective legislation has been lacing. It should however be admitted that so far as the processing of the results is concerned coding of Moravian and silesian nationalities was actually not reckoned with in the list of numerical characters. Under the pressure of the vents and facts mentioned above, as well as the real threat of the census being boycotted in some of the regions of Moravia and silesia, the respective committee of the Czech National Council and the Central Commission in charge of the census recommended the methods of ascertainment of nationality to be modified so as to satisfy the demands raised. An unprecedented situation was thus given rise to where radical intervention in the methods and the project of census-data processing had to be undertake and the population as well as the census commissioners informed through the press and on television.

Another polemic was engaged in concerning the anonymity of the census forms (which of course would prevent any possibility of the completeness of the census being checked on and the missing data supplied for the citizens who were absent) and directed against the questions about the outfit of the respondent s flat or home, his recreational possibilities and facilities and his ownership of a car. Even mentioning the employer (essential for coding the branches of industry) and the address of the workplace (used to process data concerning commuting) was in some cases considered to be inadmissible because it "violates the privacy" of the respondent. In a number of mass media these views were unfortunately given greater publicity than the points of view of expert statisticians, which fostered distrust among the population concerning the census. The situation culminated close before the census so that boycott by parts of the population and foiling of the census loomed threateningly in the same way as in the Federal Republic of Germany in 1981.

Petitions doubting any surveys of religious affiliation were being sent to a number of addresses incl. the Supreme Court, the General Attorney s Office, the Federal Statistical Office, the Czech Statistical Office and the Federal Assembly. In the end the problem was even put on the agenda of the constitutional and legal committees of the Federal Assembly, which called on the Government of the Federal Republic (three days before the census) to look into the formulation of the respective questions and find guarantees against the data being misused. The Government of the Czech and Slovak Federal Republic dealt with these problems at its session on March 1, and in a press conference the Government s spokesman assured the population that the doubts raised by the critics were unfounded. The same point of view was also voiced on television by the prime minister of the Czech

and Slovak Federal Republic and the president of the Federal Statistical Office the same day. That was an important moment that has saved the census from failure.

The authors can declare with all their responsibility that the whole preparations of the census, both its material content and the method of collecting data and processing them, are in perfect harmony with the respective Czechoslovak legislation and that any doubts that may have been raised can be easily and successfully refuted. The methodology, the material content and the type of the census do in no way differ from the practice that is quite usual in the majority of advanced countries. The authors themselves are surprised to realize in how many features and details of these features, incl. for instance the negative reactions and the type of argumentation of the opponents, the Czechoslovak census is fully comparable with the censuses eg. in France. Reading the reports by our Japanese colleagues gave us the impression that they must have been in Czechoslovakia and that they were actually describing our own specific problems ... And we cannot but dream of the high prestige of the censuses in Canada, Australia, Austria and other countries, e. g. the United States of America, where Census Day is regarded as something like a national holiday. And mention should also be made of the material conditions under which censuses are organized in the advanced countries, starting with the quality colour-printed census sheets and informative brochures and ending with optical scanning of the inputs to be processed.

The quality of the data and the preparation for processing them

The employees working for the respective departments of the Federal Statistical Office, the Czech Statistical Office and the Slovak Statistical Office have undertaken a number of checks on the completeness and the quality of the census material handed over by the inspectors at the district branches of the Czech and Slovak Statistical Offices. They acquainted themselves in detail with the quality of the census material as well as the input data for the preliminary results. The quality of the census forms filled in can be briefly characterized as follows;

– the quality may be a little lower than that of the 1980 census, but positively better than it could be expected in view of the signals in the press and the reception shown by the public before the census. Generally speaking the material enables the results to be processed to the extent originally planned without any risks as far as the reliability of the data is concerned;

– the resultant quality of the census material is closely linked with the level of the preparation of the census in a given territorial unit, the quality of work of the commissioners and inspectors of the consistency of the checks on the census materials at their acceptance by the district branches of the Czech Statistical Office and the Slovak Statistical Office;

– the quality fluctuates relative to the largeness of the census district (the countryside districts faring better than the municipal districts), the level of education and the social status of the persons censused. It is paradoxical that with higher educated persons the census sheets should be very often filled in incompletely (which must thus be regarded as

intentional);

– there were a few cases where the census commissioner failed to adhere to the basic methodological instructions. This applies particularly to the collection of the filled in census forms before the decisive moment of the census and the negligence concerning the check on the correct filling out of the forms on the occasion of the second visit of the commissioner in the person's home. Several cases were reported of the commissioners returning the forms to the respondents to be "corrected" (viz. Czech nationality to Moravian nationality) land even manipulating the respective entries (which was however of no qualitative importance);

– the situations where the respective data were unavailable due to a strict refusal of the censused person to cooperate were very rare. It can safely be claimed that contrary to all expectations the number of these cases was really most insignificant;

– a certain distorsion of the data concerning the equipment of households had been expected (viz. a greater share of unavailable responses and intentional withholding of information as a result of misinformation published in the press, as well as various unfounded rumours). The comparison of the preliminary results with the data of the 1989 Microcensus has however not corroborated these predictions.

From the information gained in field–work, from the preliminary results of the census and the process of marking the census material for final processing it follows quite unequivocally that the census can be relied on to fulfil its purpose as far as the quality of the data is concerned.

Processing preliminary results

Decentralized processing of the preliminary results with the help of personal computers in the individual branches of the Czech and the Slovak Statistical Offices has been applied for the first time in this year s census. The Federal Statistical Office worked out the project and supplied the technical design. Computer Technology Ltd. Prague, The Prachatics division, is the author of the respective software, which is perfect and has proved to be absolutely functional and reliable. The inputs were supplied by the individual district branches, which prepared them from the material handed over to them by the census inspectors. Clean files on diskettes were sent to the Federal Statistical Office for central processing.

In spite of the fact that the basic material from the district branches was available with a week s delay (viz. not until June 22 evening), The Federal Statistical Office was able to hand over the processed results on diskettes to the Statistical Offices of the two republics within four to five days (on June 26 and 27). Only eight working days after the delivery of the material from the district branches of the Statistical Offices of the two republics the publication containing the preliminary results of the census could be printed, bound and prepared for sale in czech, Slovak and English versions in the Czech and the Slovak Republics. The graphical layout was in the charge of the Geographical Institute of the Czechoslovak Academy of Sciences in Brno and some of the copies (particularly the English version) had colour graphs run off on the printing press loaned by Mannesmann

Co. For the users the publication has diskettes with the preliminary results attached to it.

Commented preliminary results are now being published e. g. in the journals Statistika, Demografie and S'91 (National Administration for Local Government). The Federal Statistical Office has published A Brief Analysis of the Preliminary Results of the 1991 Population, Houses and Homes Census.

A field of paradoxes and paradoxes of the field

J. Tvrdík

Preface to English version:

This short essay was written for the St. Nicklaus Party of our Faculty of Sciences, University of Ostrava. It was considered as a joke for colleagues and as a kind of advertisment for our newly established department. The Czech version was written as a balanced mixture of experience and demagogy. English is not my native language, so I am afraid that the mixture may not be balanced. Auxiliary explanation:

1) *The term for "Computer Science" sounds something like "Informatics" in Czech*
2) *Our Faculty of Science consists of six departments: Geography, Physics, Chemistry, Biology, Mathematics and Informatics*

If you ask somebody what informatics is, you can hear a vague answer like

a) the field of a great future (it is said about all fields including archeology)
b) the field in which we are very backward compared with western countries (there are many such fields)
c) the base of our happiness in the information age (who knows what happiness is ?)
d) something with computers (and you are sure that it is misty)

Informatics is, in our vocabulary, without any good definition. It seems to me that nobody cares about it. The term "Computer Science" is preferred to "Informatics" in English-speaking countries, but it is also not self-explanatory.

The main concept of the field is neither information nor computer as one can infer from the names, but the algorithm. What is an algorithm? Almost everyone feels it, but it does not have an acceptable definition.

We expect naively that informatics could help us to process data on the real world. But informatics itself (maybe herself, informatics is probably female) produces so much new data on the artificial world, that the tools created by informatics fail sometimes in practical processing of that information. Moreover, computer scientists are not able to see the reality over the stacks of their books, journals, manuals and displays.

Most experts agree with the statments "No programming for PCs is needed yet" and "Good programs cannot be produced outside the U.S.A." In spite of those statments the same experts teach the programming for PCs as one of the main topics in courses of informatics over the world. We cannot imagine any computer science courses without programming. Programming is even the most attractive thing for students and they like best the unimportant part of programming - playing with displays.

Perhaps no other skills become old as quickly as the skills in informatics. The things taught in school will be overcome at the time of the end of study. Is it important, what we teach our students? Moreover, teachers are experts in the past only, because the true experts of the present time have no time to teach, they are in a hurry to create the future.

In spite of these dilemmas, informatics has many useful results. One of those are commercial application programs, other very strange things. They are goods that are often sold very well. But we are not able to weigh, count or measure them like meat, rolls or carpets. If someone steals your program, you do not mention it, because nothing is missing form your disk. But stolen programs can be recognized and the thief can be punished.

There are many things in informatics which do not allow our mind to rest in peace. Informatics is an adventure that you do not find in other fields. Is it comparable with physics as defined by Newton? Small corrections added by nuclear physicists are several decades old. Cosmology is focused on negligible seconds from the furthest past. What about geography, for which such minor events like volcano eruptions or migration movment in Africa are phenomena of great significance? Such conservative fields like mathematics, chemistry or biology cannot take part in competition. The only challenge for really brave students is informatics!

Cech Statistical Society, the annual report.

Gejza Dohnal, Hana Řezanková

The Society Committee has been meeting approximately twice every three months. In what follows, I will give a summary of the main problems which were discussed in the committee meetings in the last year.

1) **The compilation of the choice from the ROBUST proceedings.** We decided to choose 16 of the most successful papers from the current published proceedings of the favorite conference about robust statistics (and not only about it) ROBUST to be included in the new publication. The colleague Íváek was responsible for the edition and the problems connected with the preparation for publication. His effort resulted in the soft-cover publication "ROBUST 1980-90" which is currently offered for sale at the University of Economics. Unfortunately, Dr. Íváek could not influence the relatively high price of this book (350,-Ks) due to the casualty which kept him in the hospital for a long time.

2) Two following problems arose along with the unexpected and tragic **death of our colleague, Tomáš Havránek**, above all: to prepare Tomáš's book for publication and to issue the special Information Bulletin which would be dedicated to Tomáš's memory. In the case of the book, there were economical reasons for stopping the editorial activity. Our aim is to renew this activity so that the book will be published. Several people were looking for some solution to this problem very intensively, for example Prof. Katětov, on behalf of the Collegium of the mathematicians and Prof. Anděl, who wrote the letter to the President of the Academy of Sciences, looking for private publishers too. In accordance with the latest information it appears that the book will be published.

3) Professor Jílek took on the role of **foreign relations** and correspondence after Tomáš Havránek died. Prof. Jílek took part in the meeting of the Statistical Societies in Cairo (a brief description was in one issue of IB). Prof. Anděl sent the Information about the activities of our Society to the ISI. The similar materials with information about several societies from all over the world are at the members' disposal. We will be bringing it in this Bulletin successively.

4) The members of the committee decided to prepare for publication one **English issue of the Information Bulletin** of the Czech Statistical Society per year. The first one was issued in 1991 and it was distributed to the addresses of societies which established contact with us. The English issue of IB will not be sent to all members of the Czech Statistical Society automatically, but it is possible to obtain it after sending a request to the editor.

5) Professor Likeš asked our Society for co-operation in the preparation of the memorial

action in occasion of **the 100th anniversary of professor Janko's birth** (in December 1993). Two main actions are being planned: the organization of a one day meeting and issue of a memorial publication. For these purposes, the preparation committee will be constituted. This committee will consist of approximately 10 members who represent the University of Economics, Charles University and the Czech Statistical Society. Three following sections will be involve in the committee: publishing (memorial publication), organizing and programming (memorial meeting). Our Society will participate in this project.

6) Ing. Roth put forward **the foundation of the section on biometrics**, which could become a collective member of the Biometrics Society and form a regional group (possibly Czech-Polish), as the case may be. The membership in the Biometrics Society will be determined by the amount of the participation fee. Further, Ing. Roth suggested close co-operation with the Medical Society, for example by means of common seminars.

7) **The economy of Czech Statistical Society** is very simple. The only receipts are the members' dues in the amount of 10568,- Ks in the year 1991. The main expenses were the charges connected with the publication and distribution of the Information Bulletin (1380,- Ks, approx.). The other big expenses were spent on the issue of "Proceeding's choice", the wreath to Tomáš Havránek and the remuneration for editorial and economic services. At the end of 1991 we had 14820,- Ks in our account.

8) **The second annual meeting** was held at January 21, 1991, at the University of Economics in Prague. The members' interest corresponded with our prediction – 60 participants, approximately. The meeting was opened by Prof. Anděl, who summarized activities of the Society during last year. Prof. Anděl commemorated the two persons – Tomáš Havránek and Jiří Žváček, who could not attend this meeting. After that, the economy and editorial reports were delivered. Lectures on the following topics were held:

- Ing. Fišer spoke about the role complicated position of governmental statistics under the new intricate circumstances
- Prof. Anděl offered us a small excursion into the history of the testing of statistical hypotheses
- Dr. Antoch emphasized several valuable, but often neglected areas of interest, such as how to introduce the computational statistics into our work
- Ing. Roth, who represented medical statisticians and statisticians interested in biometrics, suggested the foundation of a section of biometrics.

The state of membership of the Czech Statistical Society up to December 15, 1991

266 statisticians were registered in our Society up to the present time. Two of them died and one colleague changed her employment so that the connection with her was broken. Thus the present state is 263 members. From the point of view of fields of interest, the structure of the Czech Statistical Society is shown in the following table:

field of interest	number of members	field of interest	number of members
mathematical statistics	112	technical sciences	45
computational statistics	121	agriculture	30
economical statistics	84	econometrics	27
governement statistics	78	business statistics	22
social sciences	71	biometrics	11
medicine	58	biology	7
demography	46		

Contents:

[0]**Information Bulletin** of the Czech Statistical Society is issued five times per year in czech and one time in english. President of the Society: Prof. Dr. Jiří Anděl, DrSc., MFF UK, Sokolovská 83, 180 00 Praha 8, ČSFR. Editorial adress: Dr. Gejza Dohnal, Jeronýmova 7, 130 00 Praha 3, ČSFR. E-mail: dohnal@cspuni12.bitnet