Lithuanian Journal of Statistics *2011, vol. 50, No 1, pp. 114-122*
Lietuvos statistikos darbai *2011, 50 t., Nr. 1, 114-122 p.*
www.statisticsjournal.lt

PROFESSOR JERZY NEYMAN – SOME REFLECTIONS

Jan Kordos

Warsaw School of Economics
Aleja Niepodległości 162, 02-554 Warszawa, Poland
E-mail: jan1kor2@aster.pl

Received: April 2011 Published: November 2011

Abstract. The author cooperated with Professor Jerzy Neyman in Poland in 1958 on the subject of sampling surveys. He begins with Neyman's ancestors to explain his connections with Poland, and Neyman's primary, secondary and university education. Next the author focuses on Neyman's period in Poland in the years 1921-1934, paying special attention to Neyman's involvements in the development of theory and practice of sampling surveys.

He begin with major developments in sampling at the 1925 and 1927 sessions of the International Statistical Institute. This is the background for discussion of Neyman's contribution to theory and practice of sampling. Next Neyman's initial contribution to the theory of sampling in the 1920s and his cooperation with Karl Pearson is emphasized. Before presenting Neyman's famous 1934 paper, which synthesizes his own contributions and those of others, and effectively demolishes the idea of purposive sampling, the author stresses that main ideas of this paper are presented in Neyman's booklet prepared in 1932 in Polish and published in 1933. Summary of this booklet is given in Annex 1.

Finally, special attention is paid to the co-operation of Professor Neyman, after the Second World War, with Polish sampling statisticians, and mainly with the Central Statistical Office of Poland. At the end, some reflections of the author are presented.

Key words*:* Confidence interval, International Statistical Institute, Jerzy Neyman, Optimal sample allocation, Purposive sampling, Representative sampling, Representative method, Sampling method.

1. Introduction

I cooperated with Professor Jerzy Neyman in 1958 when he gave six weeks consultation for the Central Statistical Office of Poland (GUS) on sampling method (Zasępa, 1958). Day by day we discussed different aspects of application of sampling method carried out or planned by GUS. We talked also about Professor's connection with Poland in the past. So my reflections concern not only sampling surveys but also connections of Professor Neyman with Poland and his Polish ancestors. I have discussed these problems during different conferences, I gave also a talk on the J. Neyman's contributions to theory and practice of sampling methods at the International Conference in Memory of the Hundredth Anniversary of the Birth of Professor Jerzy Neyman, Jachranka, near Warsaw, Poland, 25-26 November, 1994 (Kordos, 1994), and also a talk on the impact of J. Neyman and W.E. Deming on sample surveys practice in Poland, at European Survey Research Association, ESRA 2009 Warsaw Conference (Kordos, 2009), but have not published it yet.

There have already appeared numerous papers on Neyman's revolutionary contributions to statistics and on his many activities in the area of scientific planning and organization. There is a book (Reid, 1982) describing in details the life of Jerzy Neyman and biographical papers from a variety of sources, e.g. Bartoszynski (1978), Bartoszynski & Klonecki (1977), Fienberg (1992), Kendall et al (1982), Klonecki, (1995), Lehmann & Reid (1982) and Rao (1981). Fienberg & Tanur's paper (1996) reconsiders the fundamental contributions of Fisher and Neyman on experimentation and sampling.

I begin with the birth of Jerzy Neyman, his ancestors and connections with Poland. Next I present Neyman's contributions to theory and practice of sampling surveys, focusing on the development of "representative method", discussions at the International Statistical Institute in 1920s, Neyman's booklet in Polish (Neyman, 1933), and its connection with his famous paper (Neyman, 1934), which is a powerful rejection of purposive sampling and also sets the basis for

estimation from cluster samples. At the end, cooperation with the Central Statistical Office of Poland in 1950 and 1958 is described and some of my reflections presented.

2. Neyman's ancestors

We know that Neyman's ancestors settled in the western part of Poland (Poznańskie) in 17th century. In 1775 Stanislaw August, the King of Poland granted Mateusz Neyman, great-grand father of Jerzy Neyman, a coat of arms, given to noble men for special achievements, and an additional name "Splawa" (which is a sign of nobility). So Jerzy Neyman published his earliest scientific articles under the name of Jerzy Splawa-Neyman. Latter, at the age of 30, he dropped Splawa and used the name Jerzy Neyman or J. Neyman. Jerzy Neyman's father, Czesław Splawa-Neyman, with his mother and grand-mother, was exiled as a twelve-year-old to the Crimea, after the 1863 Polish uprising against the Russians. Czeslaw's father was killed and his brothers, who participated in the insurrection of Polish patriots against Russian occupation, were sent to Siberia (Klonecki, 1995).

3. Neyman's education

Jerzy Neyman was born on the 16th of April 1894 in a small town Bendery, which lies on the river Dniester about 100 km away from Odessa. Neyman's father was a judge in Bessarabia. He was also a historian, a member of the Kraków Academy of Sciences. Jerzy was born into a Roman Catholic family which considered itself to be Polish and was certainly Polish speaking.

As a young boy, Jerzy lived in several different towns: first Bendery, then Kherson, then Melitopol in the Crimea, and by the time he was eight years old the family was living in Simferopol also in the Crimea. Up to the age of ten he was taught at home by a governess and then he entered the local highschool. Remarkably he could speak five languages by that time, Polish, Ukrainian, Russian, French and German. In 1906, however, his father died of a heart attack and his mother, now having little money to bring up her son, moved to Kharkov where she had relatives. Neyman excelled at the secondary school at Kharkov and he decided, probably because he had an outstanding mathematics teacher, that he would study mathematics at university.

Neyman began his studies at Kharkov University in the autumn of 1912. He was interested both in physics and in mathematics and at first it was practical physics experiments which he enjoyed most. Many students left for military service when World War I started in 1914 but Neyman failed the eyesight test so remained at the University. In the academic year 1915-16 A. Bernstein lectured to him on probability; he strongly influenced Neyman and encouraged him to read Karl Pearson's *The Grammar of Science.*

In September 1917, having completed his undergraduate studies, Neyman remained at Kharkov University preparing for an academic career. He lectured and assisted with tutorial sessions and began to take an interest in statistical ideas. However the last year of the war, the Russian Revolution, and the civil war, totally disrupted the academic life of the University. There was great hardship and not surprisingly Neyman's health began to deteriorate. The doctors diagnosed tuberculosis and he was told to go south to recover.

After recovery, in the fall of 1920, Neyman passed the examinations for his master's degree at the University of Kharkov and remained at the university to teach a variety of mathematical courses. However, by the end of his first year of teaching, Neyman received word that he was about to be arrested again. As a result, he moved with his wife, mother, and grandmother out of Russia to Poland. During the period of the Russian Revolution Neyman was arrested as an enemy. After the treaty signed in Riga in 1921 that ended the war between Poland and the Soviet Russia, Poles living in the territory of Soviet Russia were allowed to move to Poland. Neyman's family took this opportunity, and settled in various parts of Poland. Jerzy Neyman settled in Warsaw to be close to W. Sierpinski, a famous Polish mathematician, whose outstanding research aroused Neyman's interests in pure mathematics (Reid, 1982).

4. Neyman in Poland

In Poland, he accepted a position as the senior statistical assistant for the National Agricultural Institute in Bydgoszcz, located about 200 km west from Warsaw. During the two years he spent there, Neyman began working on numerous papers that applied mathematical probability to agricultural problems.

While in Poland he lectured on statistics at the University of Warsaw, at Jagiellonian University in Krakow and at the School of Agriculture in Warsaw. He cooperated with the Central Statistical Office of Poland (GUS) and with various agricultural institutes. He developed an especially lively activity as the head of the Statistical Laboratory of the M. Nencki Institute of Experimental Biology in Warsaw.

Neyman gained high recognition in Poland. He was elected an honorary member of the Polish Mathematical Society and an ordinary member of the Polish Statistical Association, Neyman received a honorary doctoral degree of the University of Warsaw (1974), and was elected a Foreign Member of the Polish Academy of Sciences. In 1974 an International Symposium to honour Jerzy Neyman was organized in Warsaw and a special volume of the presented papers was published (Klonecki, 1995).

5. Neyman's Initial Contributions to the Theory of Sampling in the 1920s

In 1923 Neyman accepted an offer to direct the Biometric Laboratory of the Nencki Institute for Experimental Biology in Warsaw and began lecturing at the Central College of Agriculture in Warsaw. He also continued his advanced studies and graduated with his doctoral degree in mathematics from the University of Warsaw in 1924. Continuing his interest in statistical probability, Neyman submitted his thesis in two papers that addressed issues of probability related to agricultural experimentation. Written in Polish and summarized in French, the papers were titled "*On the Applications of the Theory of Probability to Agricultural Experiments*" and *"Memoir on the Application of Mathematical Statistics on Resolving Some Agricultural Problems*" (Splawa-Neyman, 1990/1923a) The papers provided a careful analysis of problems in estimation, including a preliminary discussion of "*true value*."

In the fall of 1925, with the help of a postdoctoral fellowship from the University of Warsaw, Neyman traveled to London to study at the Biometric Laboratory at University College. The laboratory was run by Karl Pearson, the author of the influential mathematical work *Grammar of Science* and commonly considered the founder of modern statistics.

Before his arrival in London, Neyman sent to Karl Pearson several of his statistical publications, including the two 1923 papers on agricultural experimentation, and Pearson suggested that Neyman republish part of the second paper in English in *Biometrica* (Spława-Neyman, 1925 [1923b]). But Pearson believed Neyman's statement at the end of the paper, that it is only in sampling from a normal population that the sample mean and sample variance are independent, to be mistaken. When Neyman tried to explain to Pearson his confusion between independence and lack of correlation (in halting English and in front of several other Pearson students), Pearson interrupted: "*That may be true in Poland, Mr. Neyman, but it is not true here"* (Reid, 1982, p. 57). Dismayed at having offended Pearson and at perhaps having lost a chance to publish in English – his Polish mentors had sent Neyman to England as a kind of test to see if his ideas were worth anything, and a publication in English would go a long way towards setting that matter – Neyman searched for a way to communicate his explanation to Pearson. He finally offered his explanation to J.O. Irwin, who communicated it to Egon Person, who finally convinced his father that Neyman was not mistaken. Thus the *Biometrica* version does contain this observation from the Polish version.

The 1925 paper might have passed unnoticed, but for the fact that, in 1927, Major Greenwood and Leon Isserlis published a complain about it in the *Journal of the Royal Statistical Society,* pointing out that Neyman had failed to acknowledge the published papers of the recently deceased Russian statistician, Alexander A. Tchouproff (1923).

In 1926 Neyman continued his postdoctoral studies, spending the academic year in Paris with the support of a Rockefeller Fellowship. The next year he returned to Poland and assumed his position as faculty docent at the University of Warsaw. Although his fame as a mathematician and statistician eventually placed him among the elite of the academic world, he was never offered a full professorship in Poland.

6. The 1925 and 1927 ISI Discussions on Sampling

Before Neyman returned to the problems of sampling from a finite population that he had addressed in 1923, major developments in sampling were presented at the 1925 and 1927 Sessions of the International Statistical Institute (ISI), and published in 1926 and 1928, respectively. Kruskal and Mostller (1980) give a related and somewhat more detailed discussion.

A substantial portion of the 1925 ISI Meeting was taken up with discussions of the *"method of representative sampling*". In 1924, the Bureau of the ISI appointed a Commission, consisting of *A. Bowley, C. Gini, A. Jensen,*

L. March, V. Stuart, and F. Zizek, to study the representative method. Jensen served as a reporter and leader of the discussion at the meeting. The Commission Report (Jensen, 1926a) contains a description of two methods: random sampling (with all elements of the population having the same probability of selection), and purposive selection of large groups of units (in modern terminology clusters) chosen to match the population on selected control variates. The report does not really attempt to choose between the methods.

At the 1927 ISI meeting, C. Gini presented a paper on the application of the purposive method to the sampling of records from the 1921 Italian census (Gini, 1928, and Gini and Galvani, 1929), which was to play an essential role in the later work by Neyman.

7. Neyman's Contributions to the Theory and Practice of Sampling in the 1930s

J. Neyman published several articles and scientific contributions in Polish in 1920s and 1930s (Neyman (1921, 1923, 1930, 1933). However, the most important booklet in Polish which contributed to theory and practice in sampling methods significantly is:

Zarys teorii i praktyki badania struktury ludnosci metoda reprezentacyjna (*An Outline of the Theory and Practice of Representative Method Applied in Social Research*) published in Warsaw in 1933.

This booklet was written in connection with an inquiry carried out by the Institute for Social Problems (*Instytut Spraw Społecznych*) in Warsaw. The scope of the inquiry consisted in the construction of certain tables for actuarial purposes, presenting the distribution of the working class in Poland according to different characteristics, such as occupation, sex, age, number of dependents, etc. The investigation was based upon a sample of about 400 000 individuals drawn from the census data of the 9th of December 1931. As the author stressed in English summary of this book, the theory has been built up from a point-view slightly different from that of the previous publications of the other authors, concerned with the same problems. In this Polish publication there are some points which were further developed in his English papers published in 1934 and later, such as advocacy for probability sampling approach, and criticism of the purposive sample selection**;** optimum allocation of sample to strata, and confidence intervals. At the end of the summary of his Polish publication he stressed: "*It is hoped that the new results contained in the present publication will be soon published in an international language*".

Fienberg & Tanur's paper (1996) discusses relations between these two papers: "Neyman originally prepared the paper in 1932 in Polish (with an English summary) as a booklet growing out of his practical experience. He had been working for the Institute for Social Problems on a project involving sampling from the Polish census to obtain data to describe the structure of the working class in Poland. As he wrote to Egon Pearson, he used the opportunity and "pushed a little the theory" (Reid, 1982, p. 105). Published in 1933, the original Polish version of the paper traces a good deal of history (Neyman had clearly learned his lesson about citation practices and was now unquestionably familiar with the literature); the 1934 version published in the Journal of the Royal Statistical Society includes even more history and carefully cites many of the people who were to be in the room for the presentation before the Society."

Neyman (1952a) acknowledging Tschouproff's priority in discovering optimal sample allocation mentioned also on his 1933 Polish publication: "I am obliged to Dr. Donovan J. Thompson of the Statistical Laboratory, Iowa State College, Ames, Iowa, for calling my attention to the article of A.A. Tschuprow, "*On the mathematical expectation of the moments of frequency distributions in the case of correlated observations*" published in Metron, Vol. 2, No. 4 (1923), pp. 646-680, which contains some results refound by me and published, without reference to Tschuprow, in 1933. The results in question are the general formula for the variance of the estimate of a mean in stratified sampling and the formula determining the optimum stratification of the sample. These formulae appeared first in a Polish booklet *An Outline of the Theory and Practice of Representative Method*, Applied in Social Research published in 1933 by the Warsaw Institute of Social Problems. Later on they were republished in English in the Journal of the Royal Statistical Society, Vol. 97 (1934), pp. 558-625. Finally, the same formulae, again without a reference to Professor Tschuprow, were given in the second edition of my book, *Lectures and Conferences on Mathematical Statistics and Probability*, Washington, D.C., 1952. The purpose of this note is, then, to recognize the priority of Professor Tschuprow, to express my regret for overlooking his results and to thank Dr. Thompson for calling my attention to the oversight."

8. Cooperation with the Central Statistical Office of Poland after the World War II

During his visit in Poland Professor J. Neyman twice consulted some problems connected with the application of sampling methods in the Central Statistical Office of Poland (GUS). Results of these consultations were published in Polish by Fisz (1950) and Zasepa (1958).

The GUS Mathematical Commission which was established in 1949 was responsible for preparing methodology of sampling surveys. The Commission was headed in the consecutive years by outstanding Polish statisticians, such as Professors: S. Szulc, M. Fisz, W. Welfe, Z. Pawłowski and R. Zasepa. The Commission consisted of about 15 members from various universities in Poland and the GUS staff.

Although the Commission played mainly the advisory and opinion-making role, its particular members dealt with practical matters such as the preparation of the sampling design, determination of the sample size, working out of the estimation methods and estimation of sampling errors. For each new sample survey a member of the Commission was responsible as a sampling statistician, and prepared a paper together with subject matter-expert from GUS unit. The paper was discussed at the plenary session of the Commission, assessed by two referees. A considerable number of those papers were then published and some of the compilations were issued in the form of monographs. The Commission also organised training in the sampling methods for the GUS employees. It took part not only in designing of sample surveys, but also in the implementation of the new mathematical methods, such as time series, taxonomic methods, econometric modelling and forecasting methods. The activities of the Commission were terminated in 1993.

8.1. Consultation in 1950

The General National Census was carried out in December 1950. In order to accelerate the acquirement of basic results, the census was processed by the sampling method. The sampling design, the sample size, and estimation method, were consulted with J. Neyman. Professor consulted also (Fisz, 1950):

a. application of sampling method to accelerate processing of age structure of students of primarily schools;

b. problem of letters representation;

c. „*Tables of Random Numbers*“ prepared by E. Vielrose; and

d. post-enumeration agricultural survey (area sown under different crops).

8.2. Consultation in 1958

During Prof. Neyman's subsequent visit to Poland in 1958, the application of the sampling method in various statistical surveys was further discussed (Zasepa, 1958). First, Professor met the GUS Mathematical Commission at plenary session, presenting a talk on new development in sampling theory and practice, and answering questions. Next detailed programme of consultancy was established. In particular, the following topics were consulted:

1) the sample-based processing of the national population census in 1960,

2) comparing the sample-based results of the 1950 census with the complete results obtained therein,

3) the household budget survey – current practice and future plans,

4) the sample estimates of the agricultural crops for individual counties and voivodship,

5) introducing the sampling method in agricultural censuses held in June each year.

I participated in all the meetings with Professor Neyman in GUS and was responsible for organizing particular meetings with the GUS subject-matter experts and the Commission sampling statisticians with Professor, and preparing notes from each meeting.

Opinion of Professor Neyman on Household Budget Surveys

Out of discussed topics, I selected the household budget survey for which I was responsible as a sampling statistician, and that topic was the most controversial during the consultation with Prof. Neyman. We presented currently applied sampling method in household budget survey, and our plan for changing it in the future.

However, Professor stressed that for household budget survey it is impossible to apply sampling method, since non-sampling errors are very high. Nonresponse rates are rather high, and methods of data collection are usually biased

(interview and record keeping). For me such an opinion was a surprise and had some disturbing effect on our plan in the future.

Anyway, we have continued some experiments with household budget survey and changed methodology which is described in my paper (Kordos, 1996). Now I understand the opinion of Professor Neyman: from the point of view of confidence intervals, constructed by Professor, if methods of data collection are seriously biased, it is impossible to construct confidence intervals which cover true value of the estimated parameter with a given level of confidence.

In social surveys, such as household budget surveys, LFS and EU-SILC collected data are biased for different reasons (e.g. non-response, measurement errors, response errors, etc.). In such a case, it is possible to estimate the precision of the estimated parameters, but not confidence intervals, if additional efforts are not undertaken to improve accuracy of the collected data. For quality assessment, it seems reasonable to use only precision or relative standard error (CV). If a confidence interval must be used, then its interpretation should be changed, avoiding words "*to cover the true value*....". Unfortunately, very often we forget about it.

9. Concluding remarks

Neyman made fundamental contributions to the science of statistics. He put forward the theory of confidence intervals, the importance of which in statistical theory and analysis of data cannot be overemphasised. His paper on sampling stratified populations paved the way for statistical theory which, among other things, gave us modern sampling surveys. Neyman's work has led to the whole new field of experimentation so much used in agriculture, biology, medicine, and physical sciences.

References

1. Bartoszynski, R, (1978), Some thoughts about Jerzy Neyman, in: *Proceedings, Six International Conference, Wisla (Poland)***,** Lecture Notes in Statist. 2, Springer, 1978, pp. *ix-xvi*.
2. Bartoszynski, R, and W. Klonecki, W. (1977), Some thoughts about the contribution of Jerzy Neyman to statistics, in: *Proceedings of the Symposium to Honour Jerzy Neyman,* PWN-Polish Scientific Publishers, Warszawa 1977, pp. 9-15.
3. Fisz, M. (1950), *Consultation with Prof. Neyman and conclusions*, (In Polish).. *Studia i Prace Statystyczne*, nr 3-4.
4. Fienberg, S.E (1992), A brief history of statistics in three and one-half chapters: a review essay, *Statist. Sc*i. 7, pp. 208-225.
5. Fienberg, S.E. and Tanur, J.M. (1996), Reconsidering the Fundamental Contributions of Fisher and Neyman on Experimentation and Sampling, *International Statistical Review*, vol. 64, pp. 237-253.
6. Greenwood, M. and Isserlis, L. (1927). A historical note on the problem of small samples. *Journal of the Royal Statistical Society*, 90, 347-352.
7. Jensen, A. (1926a). Report on the representative method in statistics. Bulletin of the International Statistical Institute, 22, Liv. 1, 359-380.
8. Jensen, A. (1926b). The representative method in practice. (Annex B to the Report by Jensen). Bulletin of the International Statistical Institute, 22, Liv. 1, 381-439.
9. Kendall, D. G., .Bartlett, M. S. and PAGE, T.*L.* (1982), *Jerzy Neyman* 1894-1981. In: *Biographical Memoirs of Fellows of the Royal Society,* Vol. 28, November 1982, pp. 37-112.
10. Kiaer, A.N. (1895), Observations et experiences concernant des denombrements representatives. *Bulletin of the International Statistical Institute,* vol. 9, Book 2, pp. 313-330.
11. Klonecki, W. and Urbanik, K. (1982), Jerzy Neyman (1894-1981), *Probab. Math. Statist.* **2** (2) (1982), i-iii.
12. Klonecki, W. (1995), Jerzy Neyman (1894-1981), Dedicated to the memory of Jerzy Neyman, *Probab. Math. Statist.* **15** (1995), 7-14.
13. Kordos, J. (1994), *J. Neyman's contributions to theory and practice of sampling methods - the Polish connections",* The International Conference in Memory of the Hundredth Anniversary of the Birth of Professor Jerzy Neyman, Jachranka, near Warsaw, Poland, 25-26 November 1994.
14. Kordos, J. (1996), Forty Years of the Household Budget Surveys in Poland. *Statistics in Transition*, Vol. 2, No. 7, pp. 1119-1138.
15. Kordos, J. (2009), Impact of J. Neyman and W. E. Deming on Sample Survey Practice in Poland, *ESRA– Warsaw Conference – 2009- New Challenges in Sampling.*
16. Kruskal, W.H. and Mosteller, f. (1980), Representative sampling, IV. The History of the Concepts in Statistics, 1895-1939, *International Statistical Review*, Vol. 48, pp. 169-195.
17. Lehmann, E., L. and C. Reid, C. (1982), In memoriam, Jerzy Neyman 1894-1981, *American Statistician.* 36, pp. 161-162.
18. Neyman, J. (1921), Issue of Organization of International Statistics in League of Nations (in Polish), "*Miesiecznik Statystyczny*", tom IV, pp. 95-98.

19. Neyman, J. (1925), Contributions to the Theory of Small Samples Drawn from a Finite Population**,** "*Biometrika*", nr 17.
20. Neyman, J. (1930), *Beginnings of Probability Theory and Mathematical Statistics* (in Polish)**.** "*Kwartalnik Statystyczny*", tom VII, z. 3, 115 pages.
21. Neyman, J. (1933). Zarys teorii i praktyki badania struktury ludności metoda reprezentacyjna (An Outline of the Theory and Practice of Representative Method, Applied in Social Research). Institute for Social Problems, Warsaw. (in Polish with an English summary), 123 pages.
22. Neyman, J. (1934), On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection (with discussion), "*Journal of the Royal Statistical Society*", nr 97, pp. 558-625.
23. Neyman, J. (1935), On the Problem of Confidence Intervals**,** "*The Annals of Mathematical Statistics*".
24. Neyman, J. (1938), Contribution to the Theory of Sampling Human Population**,** "*Journal of American Statistical Association*", nr 33, pp. 101-116.
25. Neyman, J. (1974). (Editor), *The Heritage of Copernicus: Theories "Pleasing to the Mind":* The *MIT* Press, Cambridge, Massachusetts, and London, England, 1974.
26. Neyman, J. (1981). Introduction to the Correspondence Between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics, (Kh. O. Ondar, Ed.), Springer-Verlag, New York.
27. Rao, C. R. (1981), Professor Jerzy Neyman (1894-1981), *Sankhya Ser. A* 43 (3), 247-2.
28. Reid, C. (1982), *Neyman - from Life,* Springer, New York 1982.
29. Splawa-Neyman, J. (1925[1923b]) Contributions of the theory of small samples drawn from a finite population. Biometrika, 17, 472-479. (The note on this republication reads "These results with others were originally published in La Revue Mensuelle de Statistique, publ. parl'Office Central de Statistique de la Republique Polonaise, tom. vi. pp. 1-29, 1923").
30. Splawa-Neyman, J. (1990 [1923a]). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Statistical Science, 5, 465-472. Translated and edited by D.M. Dabrowska & T.P. Speed from the Polish original, which appeared in Roczniki Nauk Rolniczych, Tom X (1923): 1-51 (Annals of Agricultural Sciences).
31. Tschuprow, A. (1923), On the mathematical expectation of the moments frequency distributions in the case of correlated observations. *Metron* 2, pp. 646-680.
32. Zasepa, R. (1958), Problems of Sampling Surveys of CSO in the light of Consultation with Prof. J. Neyman, (in Polish), "*Wiadomosci Statystyczne*", nr 6, pp. 7-12.

ANNEX I

Jerzy NEYMAN, **Zarys teorii i praktyki badania struktury ludnosci metoda reprezentacyjna** (in Polish with summary in English) **(*An Outline of the Theory and Practice of Representative Method Applied in Social Research*)**, Warszawa, 123 pages.

SUMMARY

The present memoir has been written in connection with an inquiry carried out by the Institute for Social Problems. The scope of the inquiry consisted in the construction of certain tables for actuarial purposes, presenting the distribution of the working class in Poland according to different characters, such as occupation, sex, age number of dependents etc. The investigation was based upon a sample of about 400 000 individuals drawn from the census data of the 9-th of December 1931.

Accordingly the theory has been build up from a view-point slightly different from that of the previous publications of other authors, concerned with the same problems.

1. I do not trust much the method of purposive selection. This is based on some far-reaching hypotheses, concerning the population under consideration, which we are not able to test. Furthermore it requires some, say a priori, knowledge of characters of different parts of the population which might be "purposely selected", the which characters ought to be correlated with the investigated ones. I have the impression that the chance of getting accurate results by applying the method of purposive selection is based rather on good luck, than on the law of big numbers, which alone may be considered as the safe ground of statistical work. Therefore the publication is concerned only with the accuracy of the results which can be obtained by the method of stratified sampling.

2. The census data are kept in bundles containing information concerning the inhabitants of several neighboring houses, which form the so-called statistical district. The process of forming a stratified sample from the population of Poland consisted in (a) dividing the whole territory into 26 partial populations, (b) in subdividing the partial populations into several "sub-populations" and (c) in drawing at random several statistical districts from each sub-population.

A certain proportion of the material thus obtained concerned the working-class. This has been selected, classified and tabulated.

The problem considered in the present publication consists in estimating the accuracy of tables obtained by the above method. As the tables are to be used for actuarial purposes, the criterion of their accuracy is *chosen* as follows.

Denote by

$$p_1, p_2, \ldots, p_x, \ldots, p_w, \qquad \sum_x p_x = 1, \tag{1}$$

the known true distribution of workers, for instance according to their age. Were these numbers known, they should be used by the Institutions of Social Insurance to calculate the sum of the form

$$Z = \sum_x \alpha_x \cdot p_x , \tag{2}$$

where the coefficient α_x, depending upon the death or sickness rates etc. are to be considered as known.

Denote by n_x the total numbers of workers at age *x* as estimated from the sample. Then the estimate of p_x will be $\frac{n_x}{n_o}$, where

$$n_o = \sum_x n_x \tag{3}$$

is the estimate of the whole number of workers in Poland. The estimate of *Z* will be

$$Z_1 = \frac{1}{n_o} \sum_x \alpha_x n_x \tag{4}$$

It is convenient to consider the accuracy of the estimate *Z* by means of the ratio Z_1 as the measure of accuracy of the table, on which the calculation of Z_1 is based.

3. The accuracy of Z_1 depends upon the accuracy of the numerator and the denominator in (4). This depends upon the following factors: (a) The size of the sample (= the number of statistical districts selected to form the sample). (b) The method of sampling (stratified or unrestricted). (c) The degree of variability within the sub-populations and the numbers of statistical districts selected from different sub-population.

The points (a) and (b) are already sufficiently discussed, for instance in the remarkable publication of Bowley*. The point (c) requires some explanation. The sum in the numerator of (4) may be presented as a linear function of analogous sums corresponding to single statistical districts selected from the sample. Let x_{ij} denote such a sum corresponding to the *j*-th district selected from the *i*-th subpopulation. M_i and m_i - the total number of districts and the number of districts selected from the *i*-th subpopulation. Then

$$\sum \alpha_x n_x = \sum_i \frac{M_i}{m_i} (x_{i1} + x_{i2} + \ldots + x_{im_i}) = \sum_i M_i x_i \qquad \text{(say)} \tag{5}$$

Denote further by y_{ij} the number of workers within the *j*-th district selected from the *i*-th subpopulation. Then

$$n_o = \sum_i \frac{M_i}{m_i} (y_{i1} + y_{i2} + \ldots + y_{im_i}) = \sum_i M_i y_i \qquad \text{(say)} \tag{6}$$

We may now say that the accuracy of the numerator of (4) depends upon the interrelationship between the variances $\sigma_i^2(x)$ of x_{ij} within single subpopulations, the sizes of these subpopulations and the numbers of elements selected from the same. It was possible to show that the maximum precision is obtained when the m_i are proportional to $M_i \sigma_i(x)$. An analogous conclusion applies to the precision of the denominator in (4).

This is the question which - as far as I know - has not been discussed by other authors. As the numbers x_{ij} and y_{ij} are highly correlated (both being linear functions of the same variables, namely the numbers of workers within single statistical districts at different ages), if we could select a system of numbers m_i, assuring the maximum accuracy of the numerator in (4), the same system will be rather satisfactory from the view-point of the accuracy in the denominator

* Artur L. Bowley: "*Measuring of Precision Attained in Sampling*", Cambridge, 1925.

of (4) and vice versa. The numbers M_i being generally known, we may construct a "good" system of numbers of statistical districts to be randomly selected from each sub-population, on condition that by means of a preliminary inquiry we estimate the variances $\sigma_i^2(x)$ or $\sigma_i^2(y)$.

4. Another question connected with the precision attained in sampling, which seems to be new, consists in estimation of the frequency distribution of Z_1. This is done as follows. The numerator and the denominator of (4) are linear functions of variables $x_1, x_2, \ldots, x_k$ and $y_1, y_2, \ldots, y_k$, where *k* means the (considerable) number of sub-populations. As the variate x_i is correlated only with y_i but is entirely independent of x_i and y_i, when $i \neq j$, we can apply the very important result of S. Bernstein[*] and conclude that the mutual frequency distribution of the enumerator and the denominator in Z_1 may be fairly accurately represented by the normal surface. Then we apply the results of Geary[**] and deduce the frequency distribution of Z_1. The preliminary inquiry carried out to estimate the variation within the sub-populations allows to estimate the constants of the frequency distributions of Z_1.

It is hoped that the new results contained in the present publication will be soon published in an international language.

PROFESORIUS JERZY NEYMAN – KAI KURIE APMĄSTYMAI

Jan Kordos

Santrauka. Imčių tyrimų srityje autorius bendradarbiavo su profesoriumi Jerzy Neimanu 1958 m. Lenkijoje. Pradžioje, siekiant paaiškinti Neimano ryšį su Lenkija, pateikiama žinių apie jo kilmę, išsilavinimą. Toliau aptariama Neimano veikla 1921–1934 m. Lenkijoje, pabrėžiant jo indėlį vystant imčių tyrimų teoriją ir taikymus. Tai buvo pradžia sietina su svarbiais imčių teorijos pasiekimais, aptartais 1925 ir 1927 m. Tarptautinio Statistikos Instituto (ISI) sesijose. Atsižvelgiant į šiuos pasiekimus galima pagrįstai įvertinti Neimano indėlį į imčių teoriją bei praktiką. Toliau kalbama apie 1920-ųjų pirmuosius Neimano darbus, pažymint jo bendradarbiavimą su Karlu Pirsonu.

Prieš pristatant įžymųjį 1934 m. Neimano straipsnį, kuriame apjungiamas jo ir kitų statistikų indėlis nuvainikuojant tikslinio ėmimo idėją, yra pabrėžiama, kad pagrindinės šio Neimano straipsnio idėjos buvo pateiktos 1932 metais lenkų kalba parašytoje ir 1933 m. atspausdintoje brošiūroje. Šios brošiūros santrauka pateikiama 1 priede.

Paminėtas Neimano bendradarbiavimas su Lenkijos statistikais po Antrojo pasaulinio karo, daugiausia su Centrine Lenkijos statistikos tarnyba. Pabaigoje pateikiama autoriaus apmąstymų.

Reikšminiai žodžiai: pasikliautinasis intervalas, Tarptautinis Statistikos Institutas, Jerzy Neyman, optimalusis imties paskirstymas, tikslinis ėmimas, reprezentatyvusis ėmimas, reprezentatyvusis metodas, imčių metodas.

[*] S. Bernstein: *Sur l'extension du théoréme limite du calcul des probabilités aux sommes de quantités dépendantes.* Math. Ann., Bd. 97, pp. 1-59.

[**] R. C. Geary: *The Frequency Distribution of the Quotient of Two Normal Variates.* J. R. S. S.,Vol. XCIII, pp. 442-446, 1930.