PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 15 (1995), pp. 47–60

RECONSIDERING NEYMAN ON EXPERIMENTATION AND SAMPLING: CONTROVERSIES AND FUNDAMENTAL CONTRIBUTIONS*

BY

STEPHEN E. FIENBERG (PITTSBURGH, PENNSYLVANIA)
AND
JUDITH M. TANUR (STONY BROOK, NEW YORK)

Introduction. What seems especially fascinating to us, having taken the occasion of the 100th anniversary of Neyman's birth to reread many of the early papers and writings in the areas of sampling and experimentation, is the pivotal role of Neyman's pair of 1923 papers in Polish. These appear in retrospect to have served as the wellspring of many of the ideas that he pursued for the next 13 years.

We began by rereading the excerpt of the 1923 paper on experimentation that was recently translated and published in *Statistical Science*, and we were especially struck by the importance that repeated random sampling played in Neyman's thinking. This seemed to us to foreshadow, at the very least, the use of randomization in experimentation. Reid ([32], p. 44) quotes Neyman considerably later as denying his priority here:

> ... I treated theoretically an unrestrictedly randomized agricultural experiment and the randomization was considered as a prerequisite to probabilistic treatment of the results. This is not the same as the recognition that without randomization an experiment has little value irrespective of the subsequent treatment. The latter point is due to Fisher and I consider it as one of the most valuable of Fisher's achievements.

Since we see one of the major purposes of experimental randomization as the necessary precondition for probabilistic inference from the results, we would join Rubin ([33], p. 477) in saying that had Neyman later claimed priority rather than denying it, we would have had no reason to quarrel with that claim. Rubin [33] also reminds that the use of randomization was "in the air" in the early 1920's, citing Student ([37], pp. 281–282) and Fisher and MacKenzie [13], and further reminds us that "[t]his situation, with its juxtaposition of implicit suggestion and explicit contrary attribution from the

* Presented at the International Conference on Statistics to Commemorate the 100th Anniversary of Jerzy Neyman's Birth, Jachranka, Poland, November 25–26, 1994.

same author, emphasizes ... the dangers of over interpreting, with ebullient and embellished hindsight, early writings of great men."

While this priority issue has an "after you, Alphonse" quality, other disagreements about priority are far more acrimonious, in particular one regarding some of the basic results in sampling theory. In this paper, we describe that controversy and attempt to set Neyman's important contributions in a broader perspective. We remain convinced that despite the many contributions of others to the technical development of sampling in the period from 1900 to 1925, Neyman's 1934 paper [27] played a pivotal role in turning the dry mathematics of expectations into real sampling plans for actual randomly selected large scale surveys. We shall also touch on the issue of optimal allocation and note that here Neyman failed to acknowledge Tchouproff's priority in advancing that idea.

In addition to examining this issue of priority, we attempt to assess Neyman's contributions to sampling and experimental design, and we identify three intertwined themes: (i) Neyman's pioneering contributions to the application of statistical theory to agricultural experimentation, (ii) Neyman's fundamental contributions to the theory and practice of sample surveys, (iii) the interrelated nature of Neyman's work on experiments and surveys and the pivotal role that randomization played in his thinking in both areas.

Parallels between surveys and experiments. For a number of years, we have pursued the parallels and linkages between surveys and experiments (e.g., see, Fienberg and Tanur [7]–[9]) and, in particular, Neyman's role as a progenitor of ideas in both areas. Although in fact surveys and experiments had developed very long and independent traditions by the start of the 20th century (e.g., see Cochran [5] and Deming [6]), it was only with the rise of ideas associated with mathematical statistics in the 1920's that the tools for major progress in these areas became available. The key intellectual idea was the role of randomization or random selection, both in experimentation and in sampling, and both Neyman and Fisher utilized this idea, although in different ways. Over the next decade or so, both men made major contributions to these developments, but because of the bitterness that grew out of the dispute between Fisher on the one hand and Neyman and Pearson on the other, first over tests of hypotheses and then later over confidence intervals and experimental design (e.g., see Neyman [30]), they were never able to bring their ideas together and benefit from the fruitful interaction that would likely have occurred had they done so. And in the aftermath, Neyman staked out intellectual responsibility for sampling while Fisher did the same for experimentation.

It was in part because of this rift between Fisher and Neyman that the fields of sample surveys and experimentation drifted apart. But it remains true that one can adapt the ideas from one field to the other and that one can profitably link them, as in embedded surveys in experiments. Neyman's first 1923 paper [34] on experimental design embodies this sort of adaptation. He conceptualizes the assignment of treatments to units in an experiment as the drawing without replacement of balls from urns, one urn for each treatment. These urns have the special property that the removal of a ball (representing the outcome of an experimental unit) from one urn causes it to disappear from the other urns as well. Thus Neyman shows that when there is a finite pool of experimental units that need to be assigned to treatments, the random assignment of units to treatments is exactly parallel to the random selection of a sample from a finite population. Hence, when the number of units used in an experiment is a large fraction of the units in the population, a finite population correction must be used in an experiment, just as it is in a sample survey. The parallel between experiments and sampling is particularly close in this case. Nevertheless, all but a few modern investigators have lost sight of the parallel and fail to take advantage of insights offered in the parallel literature.

Neyman's initial contributions to the theory of sampling in the 1920's. In 1927, Greenwood and Isserlis published [18] a complaint in the *Journal of the Royal Statistical Society* that Neyman had failed to acknowledge the published papers of the recently deceased Russian statistician, Alexander Alexandrovitch Tchouproff. They accused Neyman of intellectual dishonesty — at worst plagiarism, at best failure to acknowledge Tchouproff's earlier work.

In a 1925 reprinting in *Biometrika* of some of the results he originally published in Polish in 1923 (see [34]), Neyman gave the higher moments of the means and variances of samples from finite populations. The paper is succinct and to the point. He begins by deriving formulas for moments of the sampling distribution of the mean of samples of size n from a finite population of size m, where the sample members are drawn without replacement, and he relates these to the moments of the underlying finite population. Then he focuses on the variance, skewness, and kurtosis, of the sampling distribution, i.e., M_2, $B_1 = M_3^2/M_2^3$, and $B_2 = M_4/M_2^2$, where M_2, M_3, and M_4 are the central moments of the sampling distribution, and he expresses these in terms of μ_2, $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$, where μ_2, μ_3, and μ_4 are the central moments of the finite population of m elements. He explains how these moment quantities of the sampling distribution tend to the usual ones for sampling with replacement when we let $m \to \infty$. He then turns to the first and second moments of the sample variance, expressing them in terms of μ_2 and β_2, and the correlation between the square of the sample mean and the sample variance as well as the correlation between the sample mean and the sample variance, expressing them in terms of β_2 and β_1 and β_2, respectively. In the final paragraph of the paper, by letting the finite population size $m \to \infty$, Neyman uses the correlation formulas to argue that the independence of the sample mean and the sample variance holds only for normal distributions.

The Polish paper was one of those that Neyman had shipped to Karl Pearson before his arrival in London, and the suggestion to republish some part of it in English originated with Pearson himself. But Pearson believed that 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. Neyman recounts to Reid (see [32]) Pearson's difficulty as a confusion between independence and lack of correlation. When Neyman tried to explain the difficulty (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." (See [32], 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 settling 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 Pearson, who finally convinced his father that Neyman was not mistaken. Thus the *Biometrika* version does contain this observation from the Polish version, as we noted above.

Greenwood and Isserlis ([18], p. 348) quote Neyman as writing of the formula for the second moment of the variance "[t]his result, being a generalization of formulae given by other authors, is, I believe, novel and of considerable importance." Indeed, after that statement on page 477 of his 1925 paper [34], Neyman did actually cite one of Tchouproff's(1) 1918 *Biometrika* papers [38], [39] as well as a 1925 paper by Church [3]. Greenwood and Isserlis nevertheless take him to task for failing to cite works by Karl Pearson, Isserlis, and Edgeworth besides those of Tchouproff. In their view, Neyman's felony was compounded by the work of Church, especially a 1926 paper [4] published in *Biometrika*, which cited Neyman for some of the formulas for moments rather than citing Tchouproff. There is a little question but that Neyman's results can be found at least in some form in the sea of formulas in Tchouproff's 1923 papers [40], [41], although there remains some issue as to the correctness of all of the results in [38], [39] (Church [3] was written to correct one of these). What appears to have been happening in these and other papers is the alternative algebraic derivation and expression of a variety of moment expressions, most of which were quite complex. One of the nice features of Neyman [34] is the relatively clean derivations and succinct formulas.

By the time this critique was published, Neyman had returned to Poland. In obvious distress at the accusation, he wrote to Egon Pearson soliciting

(1) This was the transliteration of Tchouproff's name, as it appeared in the 1918 *Biometrika* articles. We use the spelling Tchouproff throughout, even though when he published in *Metron*, the editor used Tschuprow, and elsewhere he is referred to as Chuprov and Chouprow.

advice. Reid ([32], p. 75) quotes his letter:

> I have got the reprint of Isserlis's and Greenwood's paper. It is indeed very ugly! What is [striking] is that if dr Greenwood wanted really to defend Tchouproff it would be the easiest way to tell me or write me two words – he [Greenwood] was a teacher of mine and I attended to his lectures – of course it would be no what French people call tapage [an uproar]. I think I shall write some thing of that sort in J.R.S.S.? Would you? Will you correct my paper? I am afrayed that as I do not know quite the English customs and circumstances I can make some stupide step – my temperament will not help me in not doing it.

The reply, it turned out, was handled by Karl Pearson, who editorialized in *Biometrika* in Neyman's defense (as well as Church's). He argued that Neyman's original publication in Polish in 1923 (see [34]) was certainly contemporaneous with Tchouproff's pair of 1923 articles [40], [41] in *Metron*, and perhaps actually predated those Tchouproff publications because delays in the publication of *Metron* caused it to appear later than its cover date. But it seems to us that this argument from Pearson, while it may well be true, is irrelevant to the controversy – Greenwood and Isserlis were accusing Neyman of ignoring not these specific papers of Tchouproff but a whole body of literature that originated some 20 years earlier with work by Pearson himself. We might speculate that Pearson was in some sense defending himself as editor of *Biometrika* for his laxness in failing to urge Neyman, the young foreigner, to update his work for an English-reading audience with citations to the literature in English. That Neyman could have profited by such urging seems clear from another facet of Pearson's editorial defense – he cites a letter from Dr. K. Bessalik, Professor of the University of Warsaw, who in 1922 was Director of the State Institute of Agricultural Research in Bydgoszcz(²). In that letter, Bessalik certifies that Neyman's paper was written in Bydgoszcz in 1922 and that "no English Journals" were accessible to him. According to Reid [32], Neyman did have access to *Biometrika* at the Central Statistical Office by the time he was located in Warsaw in 1925 and in London during the summer of the same year, but that was long after the 1923 paper had been written. Had Pearson urged Neyman to place the republication of his results in the historical context about which he had presumably been ignorant during the time he was writing the original paper in Poland, we believe that Neyman would undoubtedly have agreed. He had, after all, come to London expressly to study with Pearson, whose *Grammar of Science* had served as an inspiration to Neyman since he discovered it in 1916.

Since Neyman's 1923 papers [34], [35] did not represent a conceptual breakthrough in the theory of sampling from finite populations, we are left with two questions of historical interest. The first is to ask how the isolated scholar arrived at what may well to be an independent derivation of some key results in

(²) Pearson in his editorial spelled the name of the city as Bydgoner.

finite sampling and the second is to ask why Neyman's results, in particular, seem to have had a more profound influence than did similar one obtained by others.

Perhaps the answer to the first question is that, while surely out of the mainstream of early twentieth century statistics, neither in Russia nor in Poland was Neyman completely isolated. Neyman studied probability theory in Kharkov under the direction of the Russian mathematician S. N. Bernstein as early as 1915 or 1916. Although these activities had been forgotten by Neyman by the time he was describing his early career to Constance Reid in 1978 (and perhaps this forgetfulness in his later life contributes to our image of Neyman's early years as isolated), in the mid 1920's Neyman described himself as having continued his studies under Bernstein through 1921. He also notes that he worked in 1920 under Bernstein to apply the theory of probability to experimentation in agriculture ([32], p. 30). We know that before he left Poland he worked with a Professor Isseroff on statistical analysis of agricultural experiments and even lectured to the Agricultural Department at the University of Kharkov on the application of probability theory to experimental problems in agriculture. These activities were clearly precursors to the papers dealing with experimental design in agriculture and finite sampling theory which appeared in *Biometrika* in 1925.

Just as Fisher was stimulated by the challenges of real experimentation at Rothamsted to gather together ideas that were "in the air" to make his masterly synthesis of the design of experiments and the analysis of variance, so Neyman could well have been stimulated by the challenges of experimentation at the Institute in Poland to synthesize ideas of sampling from finite populations that were "in the air." And that would lead rather naturally, given the variable citation practices at the time, to the kind of papers Neyman wrote in 1923, papers that had little reference to the work of others. The lack of references to work in English is particularly understandable, since as his advisor certified to Pearson, Neyman had no access to English journals while he was in Poland. Tchouproff's influence is evident not only from Neyman's actual citation of Tchouproff's work but from the actual similarities that Greenwood and Isserlis noted. While Neyman's formulas are more succinct that Tchouproff's (whose equations often ran on for several pages), they are of a similar style, and deal with some (but not all) of the same problems that Tchouproff did. That the ideas were "in the air" and that parallel publications occurred is demonstrated by a trio of papers by Isserlis and Tchouproff. Although the brief exposition in Isserlis [19] was in large part "scooped" by the two major papers by Tchouproff [38], [39], Karl Pearson published all three papers in the same volume of *Biometrika*. Similarly, in 1923, both Tchouproff and Neyman published a careful mathematical exposition of moments and how they are used in sampling from a finite population, although Tchouproff does cite a number of earlier papers including his own.

To answer the second question, that of Neyman's profound influence, we need to look not at the 1923 paper (or its 1925 republication – see [34]), but at Neyman's watershed 1934 paper [27], in which he was able to capture the essential ingredients of the problem of sampling, synthesize his own contributions and those of others, and effectively demolish the idea of purposive sampling.

Bowley, Jensen, Gini, and the ISI discussions. Before Neyman returned to the problems of sampling from a finite population that he had addressed in 1923, there were major developments surrounding presentations made at the 1925 and 1927 sessions of the International Statistical Institute (ISI), and published in 1926 and 1928, respectively. In this section, we review these briefly since they serve as the backdrop for Neyman's 1934 paper. For an excellent and somewhat more detailed discussion see Kruskal and Mosteller [23].

The 1925 ISI Meeting was held in Rome and a substantial proportion of the time was taken us with discussions of the "method of representative sampling." In 1924, the Bureau of the ISI appointed a Commission, consisting of Arthur Bowley, Corrado Gini, Adolph Jensen, Lucien March, Verijn Stuart, and Frantz Zizek, to study the representative method. Jensen served as rapporteur and leader of the discussion at the meeting. The Commission Report (Jensen [20]) contains a description of the 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.

Four of the Commission members wrote separate Annexes to the Commission Report. In his Annex, Jensen [21] described at length the practice of the representative method, including the random selection of groups (cluster sampling), and provided selected references. In their Annexes, Stuart [36] and March [25] made the link to earlier ISI discussions of the representative method, and March provided about a page of theory. Finally, Bowley [2] provided a lengthy theoretical development, but he failed to describe fully how the statistical theory of purposive sampling works. What is remarkable to us in this Report and its Annexes, especially in light of the controversy that Isserlis and Greenwood were to ignite only then next year, is the singular lack of references to the theoretical work on sampling from finite populations by Isserlis, Neyman, and Tchouproff. [Jensen's [21] reference list did include a single Tchouproff reference, to his 1910 thesis, but not any references to any of his published articles, e.g., those in *Biometrika* and *Metron*.]

The formal discussion of the Report and its Annexes (pp. 58–69) included statements by Jensen, March, Bowley, and Tchouproff. Bowley was particularly pointed in the criteria he demanded for statistical practice and in a polite way seemed to make a strong argument against the purposive method.

[Neyman ([27], p. 607) had chided Bowley for failing to choose between the two methods and Bowley replied: "Certainly I thought I damned it with faint praise at the end of the summary of my report."] Jensen, however, was clearly more supportive of the purposive method and he published a later note [22] in the *Journal of the Royal Statistical Society* arguing that the use of controls in the method can yield far greater gains than Bowley had suggested.

The ISI adopted the recommendations of approval of the representative method contained in the Commission Report (Jensen [20], pp. 377–380) and Kruskal and Mosteller [23] summarize the situation as follows:

> Thus at the 1925 meeting the discussion is not whether to do sampling, but how to do it. Progress has appeared in stratified and even cluster sampling. The idea of purposive sampling continues to live and is endorsed, although it remains vague.

Yates, in reviewing these proceedings, notes ([42], p. 13) the "lack of any clear conception of the possibility, except by the selection of units wholly at random, or by the inadequate procedure of sub-dividing the sample into two or more parts, of so designing sampling inquiries that the sampling errors should be capable of exact estimation from the results of the inquiry itself." He goes on to describe the developments in random sampling that took place in England linked to agricultural experimentation between 1925 and 1935, stimulated largely by the elements of randomization in experimentation and the analysis of variance, both of which he attributes to Fisher. We see this as another indication of the lack of impact of Neyman's early contributions in these two interrelated fields.

At the 1927 ISI meeting, Corrado Gini presented a paper on the application of the purposive method to the sampling of records from the 1921 Italian census (see Gini [14] and Gini and Galvani [15]), which was to play a pivotal role in the later work by Neyman. Their problem was simply stated. They needed to discard most of the records of the 1921 census before taking the next one, and they proposed to retain a sample for future analyses and reference. To make their sample representative, they chose to retain all of the data from 29 out of 214 large administrative districts into which Italy was divided. They applied the purposive method in order to select the 29 through a process which attempted to match the averages on seven important characteristics with those for the country as a whole. When they did this, they discovered that there were substantial deviations between the sample and the entire country for other characteristics. This, they claimed, called into question the accuracy of sampling. It remained for Neyman to take up the challenge of this claim.

The 1934 JRSS paper. Neyman took up the challenge in his classic 1934 paper [27] presented before the Royal Statistical Society, *On the two different aspects of the representative method.* We review the paper's elements and emphasize how Neyman rescued clustering from the clutches of purposive sampling and gave it a rightful place in the foundations of random sampling methodology. These ideas were later to be adapted and developed by many people.

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 [32], p. 105). Published in 1933, the original Polish version of the paper [26] 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 gives enormous credit to Bowley for his 1926 synthesis of the theory underpinning sampling methods, but also notes the hundreds of contributions in the area on which his comparison between purposive and random sampling builds. There is, however, the curious presentation of optimal allocation with no reference back to Tchouproff [40], [41]. On the basis of our recent rereading of all of these papers, the kind interpretation is that this was merely an oversight.

The goal of the paper is very clearly stated in terms of the comparison of purposive and random sampling. But the elements of a synthesis are prominent as well. Neyman describes stratified sampling, noting that Bowley considers only proportionate stratified sampling but stating that such restriction is not necessary. He gives (pp. 568–569) a crisp description of cluster sampling: "Suppose that the population Π of M' individuals is grouped into M_0 groups. Instead of considering the population Π we may now consider another population, say π, having for its elements the M_0 groups of individuals, into which the population Π is divided... If there are enormous difficulties in sampling individuals at random, these difficulties may be greatly diminished when we adopt groups as the elements of sampling." This is a new synthesis — earlier conceptualizations of clustering had coupled it with purposive sampling. Indeed, Neyman quotes Bowley (p. 570) as maintaining that "in purposive selection the unit is an aggregate, such as a whole district, and the sample is an aggregate of these aggregates, while in random selection the unit is a person or thing, which may or may not possess an attribute, or with which some measurable quantity is associated." Neyman goes on to explicitly uncouple clustering and purposive sampling, saying (p. 571), "In fact the circumstance that the elements of sampling are not human individuals, but groups of these individuals, does not necessarily involve a negation of the randomness of the sampling." He calls this procedure "random sampling by groups" and points out that, although Bowley did not consider it theoretically, he used it in practice in London, as did O. Anderson in Bulgaria.

Neyman also speaks of combining stratification with clustering to form "random stratified sampling by groups." [Bowley, in his role as the lead discussant of Neyman's paper, notes the innovativeness of Neyman's suggestion of random stratified sampling of groups and acknowledges that in fact it was what he had been driven to use in his work even though he has not fully recognized the implications of what he had done.] Then Neyman refers to the full theory for best linear unbiased estimation(3) of a population average developed in his 1933 Polish publication [26], and he gives the now familiar formula for the variance of the "natural" weighted average estimator in stratified cluster sampling.

In order to compare Gini and Galvani's method of purposive selection with his own method of random stratified sampling by groups, Neyman imbues the purposive method with a structure that allows one to treat it *as if* it were based on a special form of random selection, even though this was clearly not the way in which Gini and Galvani actually selected or conceived of their groups (districts)(4). In doing so, Neyman constructs a sturdy coffin in which to bury the method of purposive selection, and then he proceeds to slam the lid of the coffin tight and nail it shut, by presenting one statistical argument after another. In the first, Neyman considers the conditions under which the purposive method, which utilizes the regression of group means on a control variate y, produces unbiased estimates. He then notes, on the basis of calculations from Gini and Galvani's own data, that the conditions appear not to be satisfied in practice. Neyman goes further, however, and points out that, even if there were not a problem of bias, Gini and Galvani make the implicit assumption that the groups (clusters) are themselves random samples from the population, something that is decidedly false. Neyman's own method of random stratified sampling by groups does not have these deficiencies. Neyman carries the argument one final step by exploring, using his variance formula, whether it is preferable to sample a small number of large groups (in effect, a variant on Gini and Galvani) or a large number of smaller units. The latter turns out to be the clear choice and this provides Neyman with the final nail to

(3) There was an implicit assumption in Neyman's work regarding the uniqueness of the best linear unbiased estimate in sampling from finite populations. This assumption turns out to be false. Godambe and Thompson ([17], pp. 386–387) observe that "Neyman proposed the use of the Gauss–Markoff technique to prove the optimality (UMV-ness) of the sample mean and other similar estimators. Indeed, during the discussion, Fisher concurred in this method; and it appears that the Gauss–Markoff technique, now known to be of doubtful validity in sampling theory (Godambe [16]), was one of the very few points on which Neyman and Fisher were in agreement."

(4) Neyman envisions a universe of districts divided up into strata according to the values of one or more control variates, and then each stratum is subdivided into substrata according to the number of units in the district. Then within each substratum he speaks of the random selection of a preassigned number of districts. Of course, because of the large size of the districts in Gini and Galvani's situation, most strata would contain zero or one district. This leads to the sampling bias which Neyman then illustrates.

hammer into the lid of the coffin of the purposive method. Shortly later, the coffin was buried, although the ghost of the purposive method continues to rise until this day in the form of quota sampling.

The immediate effect of Neyman's paper was to restore the primacy of the method of stratified random sampling over the method of purposive selection, something that was left in doubt by the 1925 presentations by Jensen and Bowley to the meeting of the International Statistical Institute. But the paper's longer-term importance for sampling was the consequence of Neyman's wisdom in rescuing clustering from those who were the advocates of purposive sampling and integrating it with stratification in a synthesis that laid the groundwork for modern-day multistage probability sampling.

The heart of the 1934 paper, however, in terms of the amount of space Neyman devoted to exposition and in terms of emphasis in the discussion by others at the presentation before the Royal Statistical Society, is the material on confidence intervals. There was much confusion as to whether this was just Fisher's fiducial method presented in a slightly different fashion or a new inferential approach. Neyman introduced the idea of coverage in repeated samples, but the discussion did not pick up on its originality. It is somewhat ironic that in his discussion of Neyman [27], Leon Isserlis, who only seven years ealier had condemned Neyman for not giving enough credit to Tchouproff, completely overlooked the fact that Neyman failed to credit Tchouproff in this paper, this time for proposing the notion of optimal allocation in stratified sampling. We speculate that Isserlis' failure to question this point arose largely from the trouble he was having understanding Neyman's concept of confidence intervals, the latter being the sole topic of Isserlis' discussion.

Fisher's discussion of the Neyman paper is also worthy of note in the present context, not so much for his challenge that Neyman's confidence intervals were really fiducial intervals but under a different name, but rather for his observations on the importance of the parallelism between sampling in economic research and the role sampling of finite populations in agricultural experimentation. The major distinction, he was reported as saying, was that (p. 616): "In a well-designed experiment, however, the mathematics were simplified, and all anxiety was avoided in respect to different systems of weighting."

Our question of why the Neyman paper had such a profound influence compared to the earlier work of Tchouproff and others, was also raised by Bellhouse [1] and Kruskal and Mosteller [23]. We believe that while Tchouproff had clearly derived a number of the technical results a decade earlier, and had broken out of the mold of constant probabilities of selection, his papers were abstract and formal in nature, and his results were far removed from real-world application. Neyman more clearly laid the groundwork for statistical practice by his innovative integration of clustering and stratification, and his clear and convincing exposition of the inferiority of the purposive method. Neyman provided the recipe for others to follow and he continued to explain its use in convincing detail to those who were eager to make random sampling a standard diet for practical consumption (e.g., see Neyman [29] for a description based on his 1937 lecture on the topic at the U.S. Department of Agriculture Graduate School, and Neyman [28] for a two-phase sampling approach to optimal allocation).

REFERENCES

[1] D. R. Bellhouse, *A brief history of random sampling methods*, in: *Handbook of Statistics*, Vol. 6, P. R. Krishnaiah and C. R. Rao (Eds.), North-Holland, Amsterdam 1988, pp. 1–14.

[2] A. L. Bowley, *Measurement of the precision attained in sampling (Annex* A *to the Report by Jensen)*, Bull. Internat. Statist. Inst. 22, Supplement to Liv. 1 (1926), pp. 1–62.

[3] A. E. R. Church, *On the moments of the distribution of squared standard deviations for samples of N drawn from indefinitely large population*, Biometrika 17 (1925), pp. 79–83.

[4] – *On the moments of the distribution of squared standard deviations of small samples from any population*, Biometrika 18 (1926), pp. 321–394.

[5] W. G. Cochran, *Early development of techniques in comparative experimentation*, in: *On the History of Probability and Statistics*, D. Owen (Ed.), M. Dekker, New York 1976, pp. 1–25.

[6] W. E. Deming, *Sample surveys: The field*, in: *International Encyclopedia of Statistics*, W. H. Kruskal and J. M. Tanur (Eds.), Macmillan and Free Press, New York 1978, pp. 867–885.

[7] S. E. Fienberg and J. M. Tanur, *Experimental and sampling structures: parallels diverging and meeting*, Internat. Statist. Rev. 55 (1987), pp. 75–96.

[8] – *From the inside out and the outside in: combining experimental and sampling structures*, Canad. J. Statist. 16 (1988), pp. 135–151.

[9] – *Combining cognitive and statistical approaches to survey design*, Science 243 (1989), pp. 1017–1022.

[10] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh 1925.

[11] – *The arrangement of field experiments*, Journal of the Ministry of Agriculture 33 (1926), pp. 503–513.

[12] – *The Design of Experiments*, Oliver and Boyd, Edinburgh 1935.

[13] – and W. A. MacKenzie, *The manurial response of different potato varieties*, Journal of Agricultural Science 13 (1923), pp. 311–320.

[14] C. Gini, *Une application de la méthode représentative aux matériaux du dernier recensement de la population italienne (1er décembre 1921)*, Bull. Internat. Statist. Inst. 23, Liv. 2 (1928), pp. 198–215.

[15] – and L. Galvani, *Di una applicazione del metodo rappresentative all'ultimo censimento Italiano della popolazione (1° decembri, 1921)*, Annali di Statistica, Series 6, 4 (1929), pp. 1–107.

[16] V. P. Godambe, *A unified theory of sampling from finite population*, J. Roy. Statist. Soc., Ser. B, 17 (1955), pp. 269–278.

[17] – and M. E. Thompson, *Bayes, fiducial and frequency aspects of statistical inference in regression analysis in survey sampling (with discussion)*, ibidem 17 (1971), pp. 361–390.

[18] M. Greenwood and L. Isserlis, *A historical note on the problem of small samples*, ibidem 90 (1927), pp. 347–352.

[19] L. Isserlis, *Formulae for determining the mean values of products of deviations of mixed*

moment coefficients in two to eight variables in samples taken from a limited population, Biometrika 12 (1918), pp. 183–184.

[20] A. Jensen, *Report on the representative method in statistics*, Bull. Internat. Statist. Inst. 22, Liv. 1 (1926), pp. 359–380.

[21] – *The representative method in practice (Annex* B *to the Report by Jensen)*, ibidem 22, Liv. 1 (1926), pp. 381–439.

[22] – *Purposive selection*, J. Roy. Statist. Soc. 91 (1928), pp. 541–547.

[23] W. H. Kruskal and F. Mosteller, *Representative sampling, IV: The history of the concept in statistics, 1895–1939*, Internat. Statist. Rev. 48 (1980), pp. 169–195.

[24] E. L. Lehmann and C. Reid, *In Memoriam, Jerzy Neyman 1894–1981*, Amer. Statist. 36 (1982), pp. 161–162.

[25] L. March, *Observations sur la méthode représentative et sur la projet de rapport relatif à cette méthode (Annex* D *to the Report by Jensen)*, Bull. Internat. Statist. Inst. 22, Liv. 1 (1926), pp. 444–451.

[26] J. Neyman, *Zarys teorii i praktyki badania struktury ludności metodą reprezentacyjną (An Outline of the Theory and Practice of Representative Method Applied in Social Research)*, Institute for Social Problems, Warsaw 1933 (in Polish with an English summary).

[27] – *On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection (with discussion)*, J. Roy. Statist. Soc. 97 (1934), pp. 558–625.

[28] – *Contributions to the theory of sampling human populations*, J. Amer. Statist. Assoc. 33 (1938), pp. 101–116.

[29] – *Lectures and Conferences on Mathematical Statistics and Probability*, U.S. Department of Agriculture, Washington, DC, 1952. (The 1952 edition is an expanded and revised version of the original 1938 mimeographed edition.)

[30] – (with the cooperation of K. Iwaszkiewicz and S. Kołodziejczyk), *Statistical problems in agricultural experimentation (with discussion)*, Supplement to J. Roy. Statist. Soc. 2 (1935), pp. 107–180.

[31] K. Pearson, *Another "historical note on the problem of small samples"* (Editorial), Biometrika 19 (1927), pp. 207–210.

[32] C. Reid, *Neyman from Life*, Springer, New York 1982.

[33] D. B. Rubin, *Comment: Neyman (1923) and causal inference in experiments and observational studies*, Statist. Sci. 5 (1990), pp. 472–480.

[34] J. Spława-Neyman, *Contributions of the theory of small samples drawn from a finite population*, Biometrika 17 (1925), pp. 472–479. (The note on this republication reads "These results with others were originally published in: La Revue Mensuelle de Statistique, publ. par l'Office Central de Statistique de la République Polonaise, Tom vi (1923), pp. 1–29".)

[35] – *On the application of probability theory to agricultural experiments. Essay on principles (Section 9)*, Statist. Sci. 5 (1990), pp. 465–472. Translated and edited by D. M. Dąbrowska and T. P. Speed from the Polish original, which appeared in: Roczniki Nauk Rolniczych (Annals of Agricultural Sciences), Tom X (1923), pp. 1–51.

[36] A. C. V. Stuart, *Note sur l'application de la méthode représentative (Annex* B *to the Report by Jensen)*, Bull. Internat. Statist. Inst. 22, Liv. 1 (1926), pp. 440–443.

[37] 'Student', *On testing varieties of cereals*, Biometrika 15 (1923), pp. 271–293.

[38] A. A. Tchouproff, *On the mathematical expectation of the moments of frequency distributions* (Chapters 1 and 2), ibidem 12 (1918), pp. 140–169.

[39] – *On the mathematical expectation of the moments of frequency distributions* (Chapters 3 and 4), ibidem 12 (1918), pp. 185–210.

[40] A. A. Tschuprow, *On the mathematical expectation of the moments of frequency distributions in the case of correlated observations* (Chapters 1–2), Metron 2 (1923), pp. 461–493.

[41] — *On the mathematical expectation of the moments of frequency distributions in the case of correlated observations* (Chapters 4–6), ibidem 2 (1923), pp. 646–683.

[42] F. Yates, *A review of recent developments in sampling and sample surveys (with discussion)*, J. Roy. Statist. Soc. 109 (1946), pp. 12–42.

Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213-3890, U.S.A.

Department of Sociology
State University of New York at Stony Brook
Stony Brook, NY 11794-4356, U.S.A.

Received on 15.12.1994