

MILTON FRIEDMAN, THE STATISTICAL METHODOLOGIST

David Teira

Dpto. de Lógica, Historia y Filosofía de la ciencia (UNED) and Urrutia Elejalde Foundation

ADDRESS:

Dpto. de Lógica, Historia y Filosofía de la ciencia
UNED. Edificio de Humanidades
Paseo de Senda del Rey 7
28040 Madrid (Spain)
E-mail: dteira@fsf.uned.es

ABSTRACT:

In this paper I study Milton Friedman's statistical education, paying special attention to the different methodological approaches (Fisher, Neyman and Savage) to which he was exposed. I contend that these statistical procedures involved different views as to the evaluation of statistical predictions. In this light, the thesis defended in Friedman's 1953 methodological essay appears substantially ungrounded.

ACKNOWLEDGEMENTS:

Thorough comments by D. Gillies, D. Hammond, K. Hoover, U. Mäki, S. Stigler, J. Urrutia and J. Zamora are gratefully acknowledged. This research was funded by the grant BF2003-04372 and the Urrutia Elejalde Foundation. The usual disclaimer on intellectual responsibility applies.

WORDCOUNT: 6749

MILTON FRIEDMAN, THE STATISTICAL METHODOLOGIST

1. Introduction

You asked me if I read methodology and philosophy, I've read a great deal in the field of statistical methodology and statistical philosophy [...]

(Milton Friedman to J. D. Hammond, 1988 [Hammond 1993, 224])

A theory should be tested by accuracy of its predictions independently of the realism of its assumptions. This probably was the most controversial thesis affirmed by Milton Friedman in his well-known 1953 methodological paper. Yet, the debate it elicited was focused more on what he denied (the need for realistic assumptions) than on what he affirmed, possibly because a response questioning the methodological relevance of accurate predictions in economics would have been out of place at the time the essay was published. At least, Friedman did not feel compelled to explain what an accurate prediction was and focused instead on the discussion of the realism issue, as if the former had not been an equally crucial subject for the case he was making. I will contend here that just the opposite happens.

Predictive success was probably an unambiguous concept in the positivist tradition which equated *prediction* to *explanation*, both being the result of deducing a particular conclusion from a law together with some initial conditions. Yet, the economic predictions that Friedman had in view while drafting his essay were of a quite different kind. Whereas the deductive predictions of the positivist philosopher could only be challenged by contesting the premises, Friedman dealt with statistical

inferences which required initial manipulation of the data, could be obtained with various estimation techniques and were later assessed from standpoints other than logic. In other words, economic predictions were debatable on many more grounds than the predictions theorised by the positivist philosopher. This contrasts sharply with the role that Friedman assigned to predictions, namely, the generation of consensus: first, within the profession on which theories were better; then among citizens when it came to estimate which policy was most desirable on the basis of its consequences (cf. Friedman 1953, §1). In this paper, I question why he could have expected economic predictions to provide consensus without a standard to decide which is the correct way to make statistical predictions.

The first point I will try to make in this paper is that these standards existed at the time Friedman published his methodological paper and he was well aware of them. In this respect, I will show in the next three sections how Friedman was successively exposed to the statistical approaches of R. Fisher (via Harold Hotelling), J. Neyman and L. J. Savage, each one of which arguing for a different assessment of the success of our predictions. Then I will discuss how it all affects his paper and what methodological moral might be inferred therefrom.

2. Fisher and Hotelling

As Stephen Stigler once put it, mathematical statistics started to exist as a discipline in the United States in 1933 (Stigler 1996, 244): the *Annals of Mathematical Statistics* had been founded three years before and the creation of the Institute of Mathematical Statistics laid only two years ahead. It was also in Fall 1933 when Milton Friedman, following the advice of Henry Schultz, moved from Chicago to Columbia for his second year of graduate study (Friedman and Friedman 1998, 43-59). There he was to

meet Harold Hotelling, then an intellectual partner of Schultz in his research on the interrelations of demand. Hotelling had been appointed at Columbia's Department of Economics two years before to replace Henry Ludwell Moore. Yet, his most outstanding contribution as a teacher seems to have been the introduction of updated statistical techniques in America. As Allen Wallis put it, by 1933 «Hotelling was practically the only person in the United States teaching statistics as we think of statistics today» (Olkin 1991, 133)¹. Whereas in Chicago emphasis was laid on general methods of collecting, tabulating, and analyzing data², Hotelling based his classes on the techniques introduced by Ronald Fisher, probably as they appeared in his *Statistical Methods for Research Workers*, whose first edition had appeared in 1925³. In Friedman's own words, the difference between these two approaches was that of «an integrated logical whole» in contrast with «a set of cook-book recipes». Yet, the mathematical sophistication of the former put it out of reach for a vast majority of the members of the centennial *American Statistical Association* —that is to say, for the average American statistician of that time (Hunter 1999)⁴. Hotelling and his students would thus lead the way in the reform of the profession, partly by settling a new paradigm of excellence in statistical research (Darnell 1988). Friedman himself became actively engaged in the creation of separate Department of statistics both at Wisconsin and Chicago⁵ according to the model advocated by Hotelling (Hotelling 1940), later assumed and extended by the commission of the Institute of Mathematical Statistics in a report coauthored by Friedman (Hotelling *et al.* 1948). Statistical predictions were now understood in a different light and their assessment considered tantamount to answering a very precise question: namely, how to justify our estimation of the value of a given parameter in a population by means of an analysis of a particular sample. To account for such inductive step, Fisher came to

claim that he had built a logic of inductive inference, no less cogent than our usual deductive reasoning. No other method of estimation ranked as high as maximum likelihood according to the standards of optimality established by Fisher himself (Efron 1998, 97-98). In this respect, he argued, maximum likelihood inferences would be equally cogent to all freely reasoning minds⁶.

When applied to the assessment of conventional statistical indexes such as the correlation coefficient, Fisher's approach was implemented with a particular technique, the analysis of variance (Fisher 1930, chap. VII, §40). This was sensitive to both the sample size and the grouping of the data and it often diverged from the traditional rule-of-thumb assessment of its standard error. Many empirical economists educated in the statistical school of Karl Pearson, such as Schultz (1929, 86), were certainly at pains to grasp the difference⁷. In 1933 Harold Hotelling's denunciation of the regression fallacies committed by one such economist, Horace Secrist, forced them to⁸.

All along his career, Friedman exploited the virtues of the analysis of variance⁹. In 1937 Friedman developed a statistical method to extend it to data originating in different populations without assuming normality, which may be read in turn as an expansion of several ideas advanced in 1936 by Hotelling and Margaret Pabst¹⁰. He also derived a number of economic applications —most significantly the distinction between temporary and permanent components of income¹¹. As a matter of fact, the use he made of statistics (most specifically, the assessment of regressions) in his economic analyses was generally Fisherian, both before and after his methodological paper. However, when the issue of predictive accuracy was raised in the 1953 essay this particular approach to the assessment of our inductive inferences was not

defended, not even mentioned. Part of the reason for this omission might well lie, as we will soon discover, in the alternative inferential approaches that he learned immediately after leaving Columbia in 1934 and the methodological puzzles they raised.

3. Neyman and Wald

As Harold Hotelling warned, early in the 1940s «the title “economist” covered many types of work, but much of it is largely statistical.» (Hotelling 1940, 458) Milton Friedman took advantage of this ambiguity on entering the labor market in 1935. In a time when the Roosevelt administration was hiring as many economists as possible in order to conduct the New Deal’s economic reform, Friedman joined the *National Resources Committee* as a research officer under the supervision of Hildergarde Kneeland (Friedman and Friedman 1998, 60-66). The task they were assigned was to plan and process two nation-wide statistical surveys on income and consumption, probably intended as technical reports on one of the most debated political issues at that time: whether the ongoing recession could be diagnosed as caused by an *underconsumption crisis*¹². However, even if Friedman’s statistical skills exceeded by far the administration’s standards, their assignment required an original sampling design, for which no theoretical account was then available¹³. In fact, sampling theory was almost at its very beginning. Being such essential need for the American administration, as soon as the great statistician Jerzy Neyman developed a new technique (Neyman 1934)¹⁴, the US Department of Agriculture Graduate School was eagerly invited to lecture on it¹⁵. This happened in 1937.

After completing the survey, Friedman attended these lectures and took the chance to pose a number of methodological questions related to their survey¹⁶. Neyman devoted

an entire paper to answering them (Neyman 1938a). Collecting the relevant data on income and consumption was an extremely expensive endeavor, given the size of the sample, so that reducing costs was imperative for the NRC team. In order to achieve this, they stratified the initial sample according to different characteristics supposed to be strongly correlated with the two variables at stake. A smaller sample was obtained by drawing at random from these strata, on which an in-depth study on income and consumption was cheaper to conduct. Yet, the size both of the strata and of the second sample were not determined by statistical considerations alone, but by introducing a number of different criteria, mainly practical in nature¹⁷. Most significantly, since the aim of the study was to assess the different factors affecting income and consumption by means of regression analysis, the sampling plan was defended on the basis that it produced regression coefficients with a smaller standard error¹⁸.

Neyman proposed instead to approach the case as a double sampling problem constrained by a cost function, where each sample was drawn at random up to a size ensuring that the estimate of the characteristic was unbiased and had the smallest possible variance. In other words, whereas the NRC team appraisal was probably inspired by the traditional *representative method*, Neyman's extended the alternative approach he had developed on an inferential basis by means of *confidence intervals*. Within Neyman's approach was embedded an epistemological thesis about induction—to which Friedman probably received his first exposure in Neyman's lectures.

As it happened with Fisher's *likelihood* (Hotelling 1931, 84-85), Neyman's concept of *confidence* could not be simply equated to conventional probabilities and required a

separate epistemological justification. On the occasion of presenting his method to the Royal Statistical Society, Neyman stated it as follows:

The term confidence coefficient is not synonymous to the term probability. It means an arbitrarily chosen value of the probability of our being right when applying a certain rule of behavior. (Neyman 1934, 623)

The validity of probability statements in the new form of the problem of estimation, which has been here so extensively discussed, depends on the permanent use of a system of confidence intervals. This system as a whole (not separate intervals) corresponds to a fixed probability that our predictions are correct [...]. (*ibid.*)

Neyman had come to this idea while working on hypothesis testing with Egon Pearson in the 1920s, skeptical as he was with the prospect of a unique solution for each statistical test, as intended by Fisher. According to Neyman, the confidence coefficient ϵ refers to the probability of making a mistake when a rule of behavior was put into practice —i.e., on accepting or rejecting a hypothesis—, not to the probability that ϵ corresponds to the true population value. Now, if our rule depends on the values of a random variable, i.e., the sampling data, the rule will be itself contingent and therefore no single instance may do to determine whether correct or not. It can only be said to minimize the error frequency when deciding on successive hypotheses, in so far as the agent holds to a consistent decision rule throughout the process. This is why Neyman replaces *inductive behavior* for Fisher's *inductive reasoning*:

We can *know* that the mathematical laws of large numbers applies in those cases that fall under the conditions under which the theorems are proven. We

can also know that the empirical law of large numbers was verified in certain experiments already carried out. But we can only believe that it will be verified again in future experiments (Neyman 1938b [1967], 353)

It is an act of the will to adjust our behavior to the results of the observation than no logic will rationalize (Neyman 1957, 12), granting universal assent to our choice.

Though Friedman seems to have a not particularly clear memory of Neyman's response (Friedman and Friedman 1998, 62), there are reasons to think that he must have been familiar with his approach as late as of the early 1940s. It is certain that he read the edition of his lectures which constituted one of the first American textbooks on the topic (Neyman 1952)¹⁹. Moreover, Neyman's techniques were incorporated into *Sampling Inspection*, another textbook that Friedman co-edited a few years after for the US Navy Statistical Research Group (Freeman *et al.* 1948, chapter 3), the mathematical task force based in Columbia under the direction of Harold Hotelling²⁰ which Friedman joined in 1942. Finally, Friedman cooperated with Abraham Wald, a devoted Neymanian, in the development of *sequential analysis* (Wald 1947)²¹, a technique allowing to stop sampling as soon as the evidence was strong enough to reach a decision on the hypothesis test. Friedman assumed Wald's approach²², even if avoiding committing himself to Neyman's epistemological stance —Wald himself did not either (Wald 1950, 28).

In view of all this, we may presume that an accomplished statistician like Friedman had had enough opportunities to grasp the difference between Neyman and Fisher as to the assessment of statistical inferences by the time he started drafting his methodological essay. Whereas Fisher's methodology was intended to yield a unique and universally acceptable estimation, for Neyman there were many possible

confidence intervals to ground each inference²³. i.e., Neyman put in focus the consensus generating role that Friedman would later assign to predictions. A discussion of which of these two approaches would render a better assesment of the predictive accuracy of a theory might have enormously complicated his 1953 piece. Besides, in the very process of writing his paper Friedman was introduced to still a third way of understanding statistics by one of his closest friends, Leonard Jimmie Savage.

4. Savage

Friedman met Savage at the SRG about 1943²⁴. He had been assistant to John von Neumann in Princeton in the year 1941-1942, where he learnt about the axiomatization of utility theory on a probabilistic basis (Lindley 1981, 38). At the SRG, Abraham Wald introduced him in all likelihood to Neyman's behavioristic approach²⁵, which Savage came to assume in full in the 1950s:

[I]t can be argued that all problems of statistics, including those of inference, are problems of action, for to utter or publish any statement is, after all, to take certain action (Savage 1951, 55)

Indeed, making choices under uncertainty so that the maximum loss that may be incurred was minimized, as advocated by Wald, was for Savage «the only rule of comparable generality proposed since Bayes' was published in 1763» (Savage 1951, 59). A full utilitarian approach to uncertain decision making may be thus applied even in those cases in which the standard *frequentist* approach failed to provide a probabilistic estimation. The agent may choose at random depending, on the one hand, on the utilitarian assessment of each alternative outcome and, on the other hand, on its respective probabilities estimated on a subjective basis.

At this point, a crucial philosophical insight was offered by an Italian mathematician, Bruno de Finetti. It was neither the cogency of each inductive judgment nor the number of failures of each inductive rule what was at stake in statistical testing: it was rather the *coherence of the process* throughout which an agent revises his own beliefs (De Finetti 1937). In his 1954 *Foundations of statistics*, Savage intended to deliver a rationale for this process in terms of a *personalistic approach* to probability.

This has also certain epistemological implications, since again statistics was to regain its role in the generation of public consensus. For Savage, a scientific community deciding whether to accept or reject a theory may provide a paradigm for statistical decision problems²⁶. The minimax solution would be «to act so that the greatest violence done to anyone's opinion shall be as small as possible» (Savage 1951, 62; see also Savage 1954, 177). Their respective beliefs on the state of the world may be appraised in terms of its *expected gain* for each individual:

Where effective experimentation is a component of some of the possible actions, practical agreement may well be reached in this way. For unless two opinions are originally utterly incompatible, enough relevant evidence will bring them close together, and even if they are utterly incompatible the holder of each will feel that that he has little to lose in agreeing to a sufficiently extensive fair trial. (Savage 1951, 62)

To sum up, even if the individual assessment of statistical predictions differed, as a result of our different probabilities and utilities, given certain mathematical constraints, it was still possible to make this assessments converge if additional evidence was provided. Again, we might presume that Friedman was exposed to this view at its early inception. Friedman and Savage cooperated in a number of research

reports at the SRG, some of them later published (Friedman and Savage 1947), and authored a couple of very influential papers on the expected utility hypothesis in economics (Friedman and Savage 1948, 1952). All this probably gave them the opportunity to debate the very foundations of their research:

Savage's work in economics was clearly influenced by his contact with Milton Friedman. It is not always recognized how much he owed to Friedman in other respects. For example, Friedman played an important role in developing his writing style^[27]. But most of all the two discussed the early ideas of personalistic probability and utility together, and the final form that found expression in his writings owes much to those debates. His respect for de Finetti was that of a kindred spirit; that for Friedman was of a stimulus, forever questioning and encouraging the new thoughts. (Lindley 1981,39)

Once more, Friedman had been a privileged witness of a new statistical breakthrough. But in this instance, unlike the previous occasions, Friedman explicitly acknowledged a methodological commitment to the Savagian approach, though almost forty years later:

A more important influence on the content of my article [«The Methodology of Positive Economics»] was my interest in statistics and my close friendship and collaboration with Jimmie Savage, at the time, he was in the process of writing *The Foundations of Statistics* (1954), a book that was to revolutionize the philosophical foundations of statistics. He regarded statistics as a method of reaching decisions, and replaced the concept of objective probability, which had been the key notion in the classical statistics that I had learned, with a personal probability, which is the key notion in what has come to be known as

“Bayesian” statistics. To oversimplify, Jimmie would say, “The role of statistics is not to discover truth. The role of statistics is to resolve disagreements among people. It’s to bring people closer together.”

Similarly, I argued in my essay that the role of positive economics was to contribute to resolving differences among economists. (Friedman and Friedman 1998, 216)²⁸

The paradox now is that we are forced to conclude that Friedman’s commitment with Savage as to the foundations of statistics must have been only in principle, since throughout his career he rarely made use of bayesian tools to predict or assess predictions (Pelloni 1996, 41). Moreover, he did not credit Savage for this idea neither in his 1953 methodological paper nor in any other publication he authored. Were it not for his correspondence and interviews with Frazer, Hammond and Pelloni, thirty years later, his Savagian faith would have remained utterly unknown.

5. Predictive success without rules

Having presented the three approaches to statistical inference, let us turn back to the methodological puzzles suggested in our first section. Each one of these, as we have seen, contained a different view as to the justification of our predictions and therefore potentially conflicting standards on their accuracy. For instance, a maximum likelihood estimation of a given parameter could be rejected on the basis of a particular confidence interval or of the particular prior on which we grounded our bayesian estimation. If Friedman wanted us to rank the scientific value of economic theories according to their predictive success, it would have been necessary to settle a particular standard to assess them. The main difficulty would have been, then and

now, to justify the particular epistemological position underlying our statistical technique of choice (Gillies 2000).

To consider just one of the questions to be addressed, we may focus on the Savagian stance which Friedman endorsed so many years after. Savage's approach to probability relied on a normative interpretation of the expected utility theory, according to which we should take as a maxim for *wise behavior*: once assumed, the individual must apply it to detect inconsistent choices and avoid them from then on (Jallais and Pradier, forthcoming). Friedman, in turn, wanted expected utility to be the core of a positive theory of economic behavior, i.e., justifiable on the basis of its predictive success alone. If Friedman had adhered to the probabilistic standpoint defended by Savage, it would have happened that his intendedly positive expected utility functions would have been judged on the basis of (Savagian) normative expected utility functions.

In this respect, we may retrospectively consider his paper incomplete, and this would account for the controversy around a significant number of predictions that Friedman delivered throughout his career. The mere fact of producing a figure that seemed accurate to him did not make his views more consensual, but usually the opposite (cf., e.g, Carl Noyes' comments in Friedman and Kuznets 1945, 405-410). I.e. economic predictions failed to accomplish the social task they were assigned in 1953, probably because the acceptance of a prediction requires previous agreement on the standard to assess them.

This may also cast some light on the particular motivations that Friedman could have entertained when he left our statistical question as to his predictive standards unanswered. We might conjecture that Friedman's concern with prediction was

mostly motivated not by statistics per se, but rather by his desire to account for the impartiality of the economist as a policy advisor. Let us recall that Friedman had decided to study economics shocked by his experience during the Depression and had plenty of opportunities immediately after graduating to serve as an advisor in Washington in multiple capacities, most of them related to his statistical expertise ([*]). Their different views on statistical methodology notwithstanding, Friedman expected his fellow economists not to take advantage of the multiple ways to assess a prediction to favor his own particular interests. He trusted their scientific uprightness and he wanted them to trust his, as Rose Friedman suggested.

I have always been impressed by the ability to predict an economist's positive views from my knowledge of his political orientation, and I have never been able to persuade myself that the political orientation was the consequence of the positive views. My husband continues to resist this conclusion, no doubt because of his unwillingness to believe that his own positive views can be so explained and his characteristic generosity in being unwilling to attribute different motives to others than to himself. (Friedman and Friedman 1998, 218-19)

I.e., probably Friedman expected that, independent of their statistical or political choices, disinterested predictors were bound to reach an agreement by virtue of their scientific ethics. However, one of the main philosophical challenges of the last three decades has been precisely to account for the fact that scientists may well act as economists would expect them to, i.e., self-interestedly (and not always in the public interest). The moral we can draw from this study is that if Friedman's methodological case has lost its strength, it is because we cannot take for granted the virtue of a single

scientific procedure (be it prediction or any other) separately from the rules and incentives that guide the agent who is applying it (Zamora Bonilla 2002).

6. References

Anderson, T. W., and M. Friedman. 1960. A Limitation of the Optimum Property of the Sequential Probability Ratio Test. In *Contributions to Probability and Statistics*, edited by I. Olkin et al., 57-69. Stanford: Stanford University Press.

Box Fisher, J. 1987. Guinness, Gosset, Fisher and Small Samples. *Statistical Science* 2.1: 45-52.

Brinkley, Alan. 1995. *The End of Reform. New Deal Liberalism in Recession and War*. N.York: Vintage Books.

Dalenius, T. 1991. Introduction to Neyman (1934) "On the Two Different Aspects of the Representative Method". In *Breakthroughs in Statistics. Vol. 2: Methodology and Distribution*, edited by Samuel Kotz and Norman L. Johnson, 115-22. N. York: Springer Verlag.

Darnell, A. 1988. Harold Hotelling 1895-1973. *Statistical Science* 3.1: 57-62.

David, H. A., and A. W. F. Edwards. 2001. The Origins of Confidence Limits. Comments on Fisher (1930). In *Annotated Readings in the History of Statistics*, edited by H. A. David and A. W. F. Edwards, 187-93. N. York: Springer.

De Finetti, B. 1937. La prévision, ses lois logiques, ses sources subjectives. *Annales de l'institut Henri Poincaré* 7: 1-68.

Duncan, J., and W. Shelton. 1992. U. S. Government Contributions to Probability Sampling and Statistical Analysis. *Statistical Science* 7.3: 320-38.

Efron, B. 1998. R. A. Fisher in the 21st Century. *Statistical Science* 13.2: 95-122.

Fisher, R. A. 1930. *Statistical Methods for Research Workers*. 3rd ed. London-Edinburgh: Oliver and Boyd.

Fisher, R. A. 1934. Discussion of “On the Two Different Aspects of the Representative Method of Stratified Sampling and the Method of Purposive Selection”, by J. Neyman. *Journal of the Royal Statistical Society* 97: 614-19.

Frankel, M., and L. King. 1996. A Conversation with Leslie Kish. *Statistical Science* 11.1: 66-87.

Frazer, W. 1988. *Power and Ideas. Milton Friedman and the Big U-Turn*. Gainesville (Flo.): Gulf/Atlantic Publishing Co.

Freeman, H. A., M. Friedman, F. Mosteller, D. H. Schwartz, and W. A. Wallis. 1948. *Sampling Inspection*. N. York: McGraw-Hill.

Friedman, M. 1937. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association* 32: 675-701.

———. 1940. A Comparison of Alternative Tests of Significance for the Problem of m Rankings. *Annals of Mathematical Statistics* 11: 86-92.

———. 1947a. Utilization of Limited Experimental Facilities When the Cost of Each Measurement Depends on Its Magnitude. In *Techniques of Statistical Analysis*, edited by C. Eisenhart, M. Hastay and W. A. Wallis, 319-28. N.York-London: McGraw-Hill.

———. 1947b. Planning an Experiment for Estimating the Mean and Standard Deviation of a Normal Distribution from Observations on the Cumulative Distribution. . In *Techniques of Statistical Analysis*, edited by C. Eisenhart, M. Hastay and W. A. Wallis, 339-52. N.York-London: McGraw-Hill.

———. 1953. The Methodology of Positive Economics. In M. Friedman, *Essays in Positive Economics*, 3-43. Chicago: University of Chicago Press.

———. 1992. Do Old Fallacies Ever Die?. *Journal of Economic Literature* 30: 2129-2132.

Friedman, M., and S. Kuznets. 1945. Income from Independent Professional Practice. N. York: NBER.

Friedman, M., and L. J. Savage. Planning Experiments Seeking Maxima. In *Techniques of Statistical Analysis*, edited by C. Eisenhart, M. Hastay and W. A. Wallis, 363-72. N.York-London: McGraw-Hill.

———. 1948. The Utility Analysis of Choices Involving Risk. *Journal of Political Economy* 56: 270-304.

———. 1952. The Expected Utility Hypothesis and the Measurability of Utility. *Journal of Political Economy* 60: 463-74.

Friedman, M., and R. D. Friedman. 1998. *Two Lucky People. Memoirs*. Chicago: The University of Chicago Press.

Geisser, S. 1991. Introduction to Fisher (1922) “On the Mathematical Foundations of Theoretical Statistics”. In *Breakthroughs in Statistics. Vol. 1: Foundations and Basic Theory*, edited by Samuel Kotz and Norman L. Johnson, 1-10. N. York: Springer Verlag.

Ghosh, B. K. 1991. A Brief History of Sequential Analysis. In *Handbook of Sequential Analysis*, edited by B. K. Ghosh and P. K. Sen, 1-19. N. York: M. Dekker.

Gillies, D. 2000. *Philosophical Theories of Probability*. London: Routledge

Hammond, J. D. 1993. An Interview with Milton Friedman on Methodology. In *The Philosophy and Methodology of Economics. Vol. 1*, edited by B. Caldwell, 216-38. Aldershot: Edward Elgar.

Hansen, M. H., T. Dalenius, and B. J. Tepping. 1985. The Development of Sample Surveys in Finite Populations. In *A Celebration of Statistics*, edited by A. C. Atkinson and S. E. Feinberg, 326-54. N. York: Springer Verlag.

Hotelling, H. Recent Improvements in Statistical Inference. *Journal of the American Statistical Association* 26: 137-75.

———. 1933. Review of H. Secrist, *The Triumph of Mediocrity in Business*. *Journal of the American Statistical Association* 28: 463-65.

———. 1939. The Work of Henry Schultz. *Econometrica* 7.2: 97-103

———. 1940. The Teaching of Statistics. *Annals of Mathematical Statistics* 11: 457-71.

———. 1951. The Impact of R. A. Fisher on Statistics. *Journal of the American Statistical Association* 46: 35-46.

Hotelling, H., and M. Pabst. 1936. Rank correlation and tests of significance involving no assumption of normality. *Annals of Mathematical Statistics* 7.29-43.

Hotelling, H., and H. Working. 1929. Applications of the Theory of Error to the Interpretation of Trends. *Journal of the American Statistical Association* 24.73-85.

Hotelling, H., W. Bartky, W. Edwards Deming, M. Friedman, and Paul Hoel. 1948. The Teaching of Statistics. A Report of the Institute of Mathematical Statistics Committee on the Teaching of Statistics. *Annals of Mathematical Statistics* 19: 95-115.

Hunter, W. 1999. An Unofficial Community: American Mathematical Statisticians before 1935. *Annals of Science* 56: 47-68.

Jallais, S., and P.-C. Pradier. Forthcoming. The Allais Paradox and its Immediate Consequences for Expected Utility. In *The Experiment in the History of Economics*, edited by P. Fontaine and R. Leonard. London: Routledge.

Jensen, E. R. 1985. Friedman's Chi-Square Test. In *Encyclopedia of statistical sciences*, edited by Samuel Kotz and Norman L. Johnson, 247-50. N. York: John Wiley and Sons.

Kneeland, H., E. Schoenberg, and M. Friedman. 1936. Plans for a Study of the Consumption of Goods and Services by American Families. *Journal of the American Statistical Association* 31: 135-40.

Lindley, D. 1981. L. J. Savage. His Works in Probability and Statistics. In *The Writings of Leonard Jimmie Savage; a Memorial Selection*, edited by L. J. Savage, 37-60. Washington: American Statistical Association and The Institute of Mathematical Statistics.

Mirowski, 2002. *Machine Dreams. Economics Becomes a Cyborg Science*. Cambridge: Cambridge University Press.

Mosteller, F. 1981. Memorial Service Tribute. In *The Writings of Leonard Jimmie Savage; a Memorial Selection*, 25-28. Washington: American Statistical Association and The Institute of Mathematical Statistics.

National Resources Committee. 1938. *Consumer Incomes in the United States*. Washington United States Government Printing Office.

———. 1939. *Consumer Expenditures in the United States*. Washington United States Government Printing Office.

Neyman, J. 1934. On the Two Different Aspects of the Representative Method of Stratified Sampling and the Method of Purposive Selection (with Discussion). *Journal of the Royal Statistical Society* 97: 558-625.

———. 1938a. Contribution to the Theory of Sampling Human Populations. *Journal of the American Statistical Association* 33: 101-16.

———. 1938b [1967]. L'estimation statistique traitée comme un problème classique de probabilité. In *A Selection of Early Statistical Papers of J. Neyman*, . Cambridge: Cambridge University Press.

———. 1952. *Lectures and Conferences on Mathematical Statistics and Probability*. Washington: Graduate School. U. S. Department of Agriculture.

———. 1957. "Inductive Behavior" as a Basic Concept of Philosophy of Science. *Review of the International Statistical Institute* 25.1-3: 7-22.

Olkin, I. 1991. A Conversation with W. Allen Wallis. *Statistical Science* 6.2: 121-40.

Pelloni, G. 1996. De Finetti, Friedman, and the Methodology of Positive Economics. *Journal of Econometrics* 75.1: 33-50.

Rao, R. 1992. R. A. Fisher: The Founder of Modern Statistics. *Statistical Science* 7.1: 34-48.

Reid, C. 1998. *Neyman*. N. York: Springer Verlag.

Samuelson, 1991. Statistical Flowers Caught in Amber. *Statistical Science* 6.4: 330-38.

Savage, L. J. 1951. The Theory of Statistical Decision. *Journal of the American Statistical Association* 46: 55-67.

———. 1954. *The Foundations of Statistics*. N. York: John Wiley.

———. 1976. On Rereading R. A. Fisher. *The Annals of Statistics* 4.3: 441-500.

———. 1981. *The Writings of Leonard Jimmie Savage; a Memorial Selection*. Washington: American Statistical Association and The Institute of Mathematical Statistics.

Schoenberg, E., and M. Parten. 1937. Methods and Problems of Sampling Presented by the U. *Journal of the American Statistical Association* 32: 311-22.

Schultz, H. 1929. Discussion. *Journal of the American Statistical Association* 24: 86-89.

Seng, Y. P. 1951. Historical Survey of the Development of Sampling Theory and Practice. *Journal of the Royal Statistical Society* 114: 214-31.

Stigler, S. 1996. The History of Statistics in 1933. *Statistical Science* 11.3: 244-52.

———. 1999. *Statistics on the table. The History of Statistical Concepts and Methods*. Cambridge (Mass.)-London: Harvard University Press.

Wald, A. 1947. *Sequential Analysis*. N. York: John Wiley.

———. 1950. *Statistical Decision Functions*. N. York: John Wiley.

Wallis, W. A. 1980. The Statistical Research Group, 1942-1945. *Journal of the American Statistical Association* 70: 320-30.

Wallis, W. A., and H. V. Roberts. 1956. *Statistics: a New Approach*. N. York: Free Press.

Zabell, S. L. 1992. R. A. Fisher and the Fiducial Argument. *Journal of the American Statistical Association* 7.3: 369-87.

———. 1994. A conversation with William Kruskal. *Statistical Science* 9.2: 285-303.

Zamora Bonilla, J. 2002. Scientific Inference and the Pursuit of Fame: a Contractarian Approach. *Philosophy of Science* 69: 300-323.

¹ Yet, cf. David 1998 on the activities at the statistical laboratory directed by George Snedecor at Iowa State College.

² Paul Samuelson, another fellow student of Friedman at Chicago, recalls being introduced to statistics by Aaron Director lecturing on Wesley Mitchell's *Business cycles* (1927) (Samuelson 1991, 334).

³ «My statistical mentors, Milton Friedman and W. Allen Wallis, held that Fisher's *Statistical Methods for Research Workers* was the serious man's introduction to statistics. They shared the idea with their own admired teacher, Harold Hotelling. They and some others, though of course not all, gave the same advice: «to become a statistician, practice statistics and mull Fisher over with patience, respect and skepticism.» (Savage 1976, 441-42) On the other hand, until 1945 there were not many other equivalent textbooks available (Rao 1992, 39). A general introduction to Fisher's life and works is provided by Box Fisher 1978.

⁴ «Editors ought perhaps not to accept volunteer reviews of books, but this was the way the reviews of the first editions of *Statistical Methods for Research Workers* and of *The Design of the Experiments* reached the Journal of the American Statistical Association. Apart from this contributed reviews I find no mention in any American book or journal of *Statistical Methods for Research Workers* during its five years, and only a very few allusions to Fisher.» (Hotelling 1951, 45)

⁵ On the Wisconsin experience, see Friedman and Friedman 1998, chapter 6 —a first draft of what later was transformed into the IMS report is mentioned in p. 100. Both Wallis (Olkin 1991, 125-127) and Kruskal (Zabell 1994, 293) attest to Friedman's instrumental role in the constitution of a Department of Statistics at Chicago.

⁶ The problem was that this likelihood was something different from mathematical probability. As Hotelling argued, it was rather a yardstick to measure our degrees of rational belief (Hotelling 1931, 84), but free from the arbitrariness introduced by bayesian priors (Geisser 1991, 3-7). Since we have no actual experience of the limit of any experimental frequency, as required by standard probabilities, likelihood will have to be defined in terms of *hypothetical frequencies* (ZABELL 1992, 374).

⁷ Schultz gradually learnt (Hotelling 1939, 99), but not fast enough as to meet the expectation of his most voracious disciples, such as Milton Friedman or Allen Wallis: cf. Friedman and Friedman 1998, 52, Olkin 1991, 123. This may partially account for their well-known antagonism in Chicago.

⁸ The details of this episode are magnifically discussed in Stigler 1996, epitomizing the statistical revolution that was taking place in the United States in the early 1930s. On the regression fallacy see also Stigler 1999, chap.9.

⁹ Neither did Wallis who was to provide his own Fisherian introduction to statistics (initially intended to be joint work with Friedman) soon afterwards: cf. Wallis and Roberts 1956.

¹⁰ Cf. Hotelling and Pabst 1936. Friedman's test deserved an entry in the *Encyclopedia of Statistical Sciences* (Jensen 1985).

¹¹ «As a student of Hotelling not long after his review had been published, I early became aware of the regression fallacy —or, perhaps better, trap.» (Friedman 1992, 2131) On the influence of this analysis on his own career, an explicit acknowledgment lies in Friedman and Friedman 1998, 74.

¹² Cf. NRC 1938 and NRC 1939, 1. Brinkley 1995, chap. 4 discusses how the idea of undeconsumption was shaped by different «worldly economists», who achieved great influence on the Roosevelt administration.

¹³ This was explicitly acknowledged by Friedman in Friedman and Friedman, 62. Duncan and Shelton 1992, 321 surveys the use of probability sampling by different government agencies in the USDA by 1933. On the development of sampling theory as such cf. Seng 1951, while a more recent one is Hansen, Dalenius and Tepping 1985.

¹⁴ See Dalenius 1991 for a retrospective assessment of Neyman's achievement. For an intellectual biography see Reid 1998.

¹⁵ Duncan and Shelton 1992, 323 comment on Neyman's influence on the American government statistics. Leslie Kish also provides evidence in this respect (Frankel and King 1996, 70).

¹⁶ «At a conference on Sampling Human Populations at the Department of Agriculture Graduate School in Washington held in April 1937, a problem was presented by Mr. Milton Friedman and Dr. Sydney Wilcox for which I could not offer a solution at that time.» (Neyman 1938a, 320). The NRC survey plan was exposed by Kneeland, Schoenberg and Friedman 1936 and further discussed in Schoenberg and Parten 1937.

¹⁷ Cf. Kneeland, Schoenberg and Friedman 1936, 137-139 and Schoenberg and Parten 1937, 312-313.

¹⁸ Cf. Kneeland, Schoenberg and Friedman 1936, 137 and Schoenberg and Parten 1937, 313.

¹⁹ Cf. the following passage from Neyman's biography: «[W. E. Deming] had worked very hard and very carefully to edit them [Neyman's lectures], and he flattered himself that he had been able "to bring your brilliant contributions into the open in a style that has pleased all who have seen the results". Among these he cited "Mr. Milton Friedman"» (Reid 1998, 148; see also pp. 137-138).

²⁰ The usual reference is Wallis 1980, now extended in Friedman and Friedman 1998, chap. 8. A controversial introduction lies in Mirowski 2002, 199-231.

²¹ For a brief historical account, cf. Ghosh 1991. Friedman provides quite a personal account of the episode: cf. in particular Friedman and Friedman 1998, 137

²² Anderson and Friedman 1960 extends a statistical result obtained by Friedman in a Waldian setting during his stay at the SRG.

²³ Zabell 1992, 374 comments on Fisher's criticism of Neyman's method (see Fisher 1934, 617-18).

²⁴ For a survey of Savage's achievements, cf. the set of memorial papers collected as an introduction to Savage 1981.

²⁵ «The minimax theory begins with some of the ideas with which the theory of personal probability, as developed in this book [*The Foundations of Statistics*, 1954], also begins. In particular, the notions of person, world, states of the world, events, consequences, acts, and decisions presented in §§ 2.2-5 apply as well to the minimax theory —*from which they were in fact derived*— as to the theory of personal probability» (Savage 1954, 158; emphasis added).

²⁶ «[T]he problem of statistics may often, if not always, be considered to be of this sort» (Savage 1951, 61)

²⁷ On this point, see also Wallis 1981, 19 and Mosteller 1981, 25-26.

²⁸ Cf. a previous statement in a 1984 interview with Frazer and Sawyer, transcribed in Frazer 1988, 135 and a more extensive one in Hammond 1993, 224-25.