PREFACE TO FIRST EDITION (1995)

A prospective reader is surely entitled to know what kind of material is contained in a book and whether and for what reason he or she should consider reading it. I accept that this statement is true but I nonetheless think that the prospective reader in the present instance might be given a better foundation for this decision if I tell how I came to write this book and what my expectations are concerning it.

As a very young man I encountered, and was amazed by, Karl Pearson's analysis of a male mortality table. (This analysis is the centerpiece of the last Section, 142, of this book.) I wanted to go and do likewise. To this end I diligently studied the Pearsonian system and the relatively small number of theoretical discrete distributions then available. In those long-ago days I dreamed a typically inflated dream of the very young. I thought that every observed distribution must have, so to speak, a signature – a reason why it is as it is and no other way. On the one hand, it turns out that this is certainly a horrendous over-simplification but, on the other hand, it was absolutely impossible to find out because of the enormous difficulties in making the necessary kinds of calculations. Maximum likelihood and minimum Chi-square estimation were simply not achievable with hand calculations.

My interest in the attempt never flagged. For thirty years I taught a graduate level course with heavy emphasis on the advantages of fitting observed distributions and what it might tell us. With the advent of computers and Basic programming it became possible to do some of the calculations, although with difficulty. My collection of examples increased at an exorbitant rate, unfortunately not matched by a comparable growth in the number of analyses.

Then came the giant step that transformed the whole subject: Mathematica. With Mathematica I could finally make, with relative ease, all of the innumerable analyses that I had always wanted to make. As far as I am concerned, Mathematica is the greatest intellectual tool ever developed and it is not possible for me to express the tremendous pleasure I take in the fact that I lived long enough to see it. That is the reason this book is dedicated to Mathematica.

What assistance was available to me when I began my long search to find the meaning hidden within an observed frequency distribution? When I started, immediately after World War II, there was one book available: W. Palin Elderton's Frequency Curves and Correlation; first edition in 1906, second edition in 1927, and third edition, the one I used, in 1938. As I end my efforts, almost fifty years later, what assistance is available? The same book, now titled Systems of Frequency Curves, and with a co-author, Norman L. Johnson, published in 1969. As far as I am aware, there is no periodical literature on the subject that is worthy of the name.

One would imagine from the lack of any literature that there are no conceptual or practical problems worthy of mention in the attempt to fit an observed distribution. Nothing could be farther from the truth. There is, indeed, an almost total lack of information on, or apparently interest in, the subject. Perhaps that is the explanation for the astonishing ignorance prominently displayed regarding it. Here are a few examples: many published examples of incorrectly fitted distributions including ones by well-known statisticians; innumerable articles developing multi-modal theoretical distributions when it is a fact that multimodality is invariably a sign of heterogeneity, as all of the famous older generation of statisticians knew; no awareness of the major problems surrounding the 0-class in discrete distributions, so prevalent that the whole idea of modified distributions was developed to deal incorrectly with this difficulty; no understanding of the frequent need to dissect observed distributions nor of the methods for so doing; no awareness of the problems of

convergence that arise in the Carver system and extensions thereof; and so forth - and these are simply the elementary examples.

This is the situation that I have attempted to address. This book has 69 theoretical distributions with 632 fits to almost 200 observed, real-world distributions. Many other distributions are discussed. The book presents carefully and with numerous examples every difficulty that I have seen arise in fitting observed distributions in almost 50 years of such analyses. I wrote this book for two reasons. First, I wanted to see if it could be done. Second, I would have been overjoyed had a book such as this one been available to me at any time during those 50 years of trying. The second reason seems to me to be the ultimate reason for writing such a book.

This is how I came to write this book. What are my expectations for it? On the one hand I would be extremely surprised if there were any interest in this book among statisticians. First, the mere scarcity of any literature on the subject, already pointed out, is a strong hint that there is little interest. Surely, if there had been potential demand then books would have been written to meet that demand. Second, the predominant opinion among statisticians is probably best reflected in a statement by Kendall and Stuart: "The fitting of mathematical curves to observational data has a certain intrinsic interest which is apt to outrun its statistical usefulness." (Advanced Theory of Statistics; vol. 1; p.173). Third, the current interest of statisticians seems definitely to be in the direction of what is called "data analysis" and not at all in the classical approach of fitting theoretical distributions. In short, I think this book is out of the mainstream of statistics and will be so treated.

On the other hand, I believe that some practicing scientists might find this book to be of considerable interest. When a scientist devotes the time and effort that are required in order to determine an observed distribution he or she surely wants to wring out from it every drop of meaning. This book probably would be helpful in that regard. I have made a determined effort to separate those theoretical distributions that may be useful to the practicing scientist from the very great number that definitely will not be. Included among the successful analyses are quite a few theoretical distributions that even an expert is unlikely ever to have encountered before. The practicing scientist may very well find many useful features in the book.

At least I have tried and in trying I have always had in mind the needs of the practitioner. If a few practicing scientists discover this book and find it useful then my efforts will have achieved their purpose.

PREFACE TO SECOND EDITION (1998)

The first edition of Fitting Frequency Distributions; Philosophy and Practice; vol. I : Discrete Distributions; and vol. II: Continuous Distributions appeared in 1995. The books were privately published in a durable format. This method of publication was used so that the cost of the books could be kept to an amount that would be affordable for the persons to whom the book was directed. There have been quite a few email communications that suggest that this strategy was successful. Indeed, a significant number of persons stated their opinions that the price could have been - perhaps should have been - quite a bit more than the $65 charged for the two volumes. I am pleased that apparently my strategy was successful and I intend to continue it. However, because of inflationary pressures it is necessary to increase the price for the first two volumes to $80.

One of the disadvantages of my strategy was that I did not have the help of an editor. The result of this was perhaps more typographical and other errors than would have occurred if I had had editorial help. The contents of these books - many mathematical expressions and equations and very many quantitative tables - were particularly prone to errors of transposition, faulty typing, skipped lines, and such. Most of the errors that escaped detection in the first edition were relatively trivial ones that probably would not have occasioned anyone very much difficulty - things like finding slightly different theoretical frequencies than the ones given in the text or slightly different summary values of Chi-square. Unfortunately, however, Murphy's Law operates here as well as elsewhere and one error was enough to make an angel weep. This occurred in the procedure presented on pp. 484-5 for fitting a Johnson S_U distribution. Step 8 has a dropped √. It should read: "The new guess for ω is

$$\omega = \sqrt{-1 + \sqrt{1 - C}}$$

With this value, go to Step 3." This omission makes the carefully planned procedure fail to converge. Other errors are sometimes unsightly - for example, not including the whole numerator in the top expression for r_1 on p. 522 - but do not cause any major problems.

Apart from my strong desire to correct any errors, by re-analysis of various examples given in the text I sometimes found markedly better fits from distributions other than the ones reported in the first edition. An example is the dissection of the U.S. income distribution for 1972, discussed in Section 141. In the first edition this was dissected into a Pearson Type III (gamma distribution) for incomes ≥ $15,000 and an Ord distribution for incomes from $4,000 through $15,000. I have since discovered that the Pearson Type I (beta distribution) gives a much nicer fit than does the Ord. Therefore, in this second edition I have given this better fit. There are a reasonable number of such improvements in fits that are now in this second edition.

Correction of errors and improvement in fits requires only a second edition of vols. I and II, now available. The price for these two volumes is now $75 and the volumes can certainly be obtained from Amazon.com. However, persons who already own the first edition of these two volumes should read on in this Preface.

Since my goal throughout has been to keep the price of these volumes as low as possible, I certainly do not want now to initiate a procedure that would require owners of the first edition to buy the second edition in order to correct errors. I decided that I could avoid this if I added a vol. III to the second edition. Therefore, I must immediately state that there were three major new things that I wanted to include in these volumes. First, I included hardly any examples of distributions from the stock market and from commodity

exchanges in the first edition. I have a great many such. Second, I wanted to present the essentials of the probability of ruin argument that is such a natural adjunct to analyses of stock market distributions in order to control the size of investments. Third, and in response to numerous requests, I wanted to make available a significant number of Mathematica programs for fitting the various theoretical distributions. All three of these things will be a part of vol. III, now in slow preparation. In addition, I am adding to vol. III a list of errors that are corrected in the second edition plus a summary of changes in the theoretical fits that are incorporated in the second edition. Therefore, persons who own the first edition need merely purchase vol. III, when it is available, and they will have all the new material plus a summary of the changes made in the second edition.

I am anxious to call attention to these plans since quite a number of emails made it clear that many of the purchasers of the two volumes of the first edition very much wanted some additional material. Much of this material will be in the forthcoming volume III. But as Shakespeare puts it, “How brief the life of man Runs his erring pilgrimage, That the stretching of a span Buckles in his sum of age.” Or, more briefly, “Art is long, life is short”. It remains to be seen when and if volume III sees light of day.

PREFACE TO THIRD EDITION (2005)

I want to compare the situation with regard to errors in the text, as described in the Preface to the Second Edition, with the current situation. I am happy to report that subsequent printings of the set have enabled me to eliminate most of the errors that have been discovered. In addition, re-analysis of the various examples given in the text has sometimes disclosed markedly better fits from distributions other than the ones reported in the first edition. A good example of this is the dissection of the U. S. income distribution for 1972, analyzed in Section 141 and mentioned in the Preface to the Second Edition. In this third edition I have given this better analysis. There are a reasonable number of such improvements in fits that are now in this third edition.

The possibility of a vol. III, discussed in the Preface to the Second Edition, still exists but it is certainly not yet a reality. There are many messages to be found in the somewhat old-fashioned approaches described and illustrated in vols. I and II and that would continue in vol. III. Thus, as an example, we find that a very large proportion of the distributions of commodity prices, stock prices, and stock indices are Pearson Type IV. Directly from this fact we can deduce that the moments >4 are infinite. This is the reality - not the common claim that the variances are infinite. There is quite a long history of attempts to summarize the characteristics of stock price distributions. Nonetheless, as far as I know there has never been any analysis of such distributions in terms of the Pearson system. It seems likely that some very interesting truths are ready to be found by taking this approach.

The good news is that with this edition we will finally have book covers that are both apropriate and handsome. This is due entirely to my daughter, Rachael L. Miller, who is a professional designer and who volunteered to do the work necessary to find and use some artistic quantitative presentations as background to our titles. Thank you, Rachael!