

PREFACE

These notes in statistical genetics are the result of auditing a course in 1970 taught by Dr. C. Clark Cockerham at North Carolina State University and of teaching a course myself since 1972 at Purdue University. The outline of the course was patterned after Cockerham's course, and many, many ideas of his are embodied throughout the notes. I have also drawn freely upon some mimeographed material which Clark handed out in his class. These notes were initially written with the idea that the notes might eventually evolve into a book on the subject coauthored jointly with Clark. I had always encouraged Clark Cockerham himself to write a book on the subject, but I could never interest him in undertaking such a project. With the publication of the excellent book by Michael Lynch and Bruce Walsh, entitled *Genetics and Analysis of Quantitative Traits*, published in 1998, by Sinauer Associates, and a second volume forthcoming by Walsh and Lynch, there is no need to duplicate their effort and attempt to publish a book on the subject. Other books have also been published particularly in the last decade or more.

The central objective underlying these notes was to provide the theoretical framework for designing and interpreting quantitative genetic studies, particularly in plant and animal breeding. These notes were written for the advanced animal-breeding and plant-breeding students in the 1970s and 1980s. The notes were written for plant and animal breeders--not mathematicians or statisticians. For these researchers who do not deal with mathematical ideas on a daily basis, it is important to lead the readers step-by-step to further their understanding of the subject. The notes were written to be self-contained, so most basic elementary formulas which one would want to refer to later are presented within the notes. All equations were numbered for completeness. The notes attempt to give the derivation of all equations and to cross-reference them to previous equations. The repeated referencing of equation numbers may almost strike one as being a bit boring, but the writer tried to be careful to avoid pulling equations out of "thin air". The writer also tended to overuse "where" statements as part of numbered equations so that all symbols were clearly defined and could be easily referenced by equation numbers. Derivations are presented for all equations; lengthy derivations are given in boxes. When the writer was writing on a topic, he did not hesitate to develop a particular topic in considerable depth. He did not concern himself at the time with balance of topics discussed.

With respect to mathematical background, college algebra is assumed. In addition some elementary knowledge of matrix operations is assumed (see Lynch and Walsh, 1998, pp. 182-192). Some differential calculus is used in obtaining normal equations in connection with least-squares theory in statistics (see Lynch and Walsh, 1998, pp. 39-43, Steel, Torrie, and Dickey, 1997, McGraw-Hill, Chapter 10). With respect to statistics itself a good rigorous one-year sequence in statistical methods is assumed. A rather complete familiarity with the analysis of variance and expected mean squares is assumed (see Steel, Torrie, and Dickey, 1997, Chapters 1 to 18). A somewhat more extensive knowledge of probability is assumed than is presented in most statistical methods courses. Set notation is introduced because it is so convenient and appropriate (see Freund, 1962, Prentice-Hall, Chapters 1-3). A thorough knowledge of the mean and variance of linear functions is assumed (see Steel, Torrie, and Dickey, 1997, Section 5.10).

The notes are incomplete. Chapter 2 was rewritten and expanded considerably in 1975 and modified again in 1983. Chapters 3, 4, 5, and 6 were rewritten on sabbatical leave in Edinburgh in 1976-77. Chapter 7 was rewritten during a three-month sabbatical leave at Iowa State University in 1983 (June to September), and Chapters 8 and 9 were rewritten during a three-month sabbatical leave at North Carolina State University in 1983 (September to December). The last section of Chapter 8, Section 8.4, is an add-on and would eventually be transferred to a later chapter. These chapters, 2 through 9, are in particularly good order. Another 500 pages needs to be written particularly on mating designs for noninbred relatives (Chapter 10), quantitative genetics models for inbred populations, covariance of inbred relatives, and mating designs for inbred relatives (Chapter 11). The present Chapter 12 on selection needs to be greatly expanded into several chapters. The present Chapters 1, 10, and 12 are largely unchanged from their original version written in 1972. The present Chapter 13 is a throw-in on a very specialized topic, written while I was coauthoring an article published in *Biometrics* (39:607-619, 1983).

The approach used in many situations has been somewhat different or atypical. For example, in considering the model for quantitative variation in Chapter 8, a cross between two populations with multiple alleles per locus was the starting point. (After all breeders commonly deal with a cross between two populations.) A single random-mating population was a subset of that more general situation. The two-allele case was also regarded as a subset of the multiple-allele case.

A comment about the word "Assumptions" might be in order. That word follows the title of major sections within a chapter. The original idea was that commonly stated assumptions in statistical genetics would be enumerated and briefly discussed or listed somewhere, possibly in an appendix. Then the relevant assumptions for the section would simply be listed by number. The idea was maybe overly ambitious and was never developed.

With the advent of molecular markers, statistical genetics has taken a significant turn. The absence of any discussion of molecular markers and quantitative trait loci (QTL) in these notes makes them out-of-date. On the other hand, there are many original ideas and topics not discussed or presented in detail any place in the literature to the knowledge of the writer. It is for these reasons that the notes are presented on the world wide web. The objective was simply to make them available to the profession for whatever merit they may have.

The notes are presently in an old obsolete word-processing program known as Microsoft Word, Version 4. It was a version in use in the early eighties. The notes were originally entered into electronic memory in one of the first available, non-Microsoft, word-processing program, called Word 11. Then they were converted to Microsoft Word, Version 4. The notes were last revised in 1990, mostly correcting known errors. To convert the notes to a present day, modern word-processing program such as Microsoft Word with the Equation Editor or MathType or TEX proved to be a task requiring several months of full-time work. Fortunately technology has developed only recently whereby one can scan a document and digitize the images to create a portable document format (pdf) file which can then be read onto a world-wide-web site and downloaded by any one who has a connection to the world wide web.

The option of simply scanning a document and producing a pdf file directly was chosen rather than converting a document to a modern word-processing program prior to the creation of the pdf file. The limitation of that option was that a pdf file is a noneditable file. If an up-to-date editable file had been available, some editing might have been undertaken. For example, renumbering of the equations would have been desirable, because in some chapters the numbering became very awkward. I was not free to completely renumber the equations, because the original numbers had to be retained in that they were referenced in later chapters which were not being revised at the time. Second, when these notes were initially written, I made the decision to list the male first and female second. I now regret that decision. This pattern of listing the male parent first was commonly used by breeders of large farm animals which historically dominated statistical genetics. However, plant workers always list the female parent first and many animal workers also list the female first (see my parenthetical note in Chapter 2, p. 18). This change itself would require a lot of editorial revision. (However, following Lynch and Walsh, p. 48, who let m = mother, instead of male, and f = father, instead of female, would lessen the work involved.) Third, I designated the generation of the sporophyte and the gametes produced by that sporophytic generation to be the same. However, the alternative way of designating the gametes and the sporophytes produced therefrom as being in the same generation has much to commend it (see another parenthetical note in Chapter 3, p. 43, and my Remark section on pp. 40-42 of that same chapter which emphasized the necessity of the *next* generation). Again changing the labeling of the gametic and sporophytic generations would entail considerable revision. Fourth, there are many other possible changes too numerous to mention here. Finally, there are a few known typographical errors which would be desirable to correct. There are also, no doubt, many unknown errors, both typographical and conceptual. Identification of errors and/or comments are welcomed.

I gladly acknowledge the many contributions to these notes by the late Dr. C. Clark Cockerham, Department of Statistics, North Carolina State University. He has been my mentor throughout my career and an enormous source of inspiration. I also want to thank my many students who have help me think through problems and have identified errors of one kind or another in the various drafts. I also want to thank the Department of Statistics, Purdue University, for the use of their top-of-the-line scanning equipment and the creation of the high quality pdf files from scanning at 600 dots per inch. File sizes became quite large, but file size is no longer a real issue for most workers who would be interested in downloading my notes from the world wide web and have access to high-speed cable modems. I would be remiss if I did not acknowledge Bill Muir who first suggested putting these notes on the web a few years ago. Finally I am very grateful to Bruce Walsh who volunteered to create a web site for these notes, has helped me prepare a suitable file structure for these notes, and has cross-linked them with his web site. For all of this I could ask no more.

Wyman E. Nyquist, Professor Emeritus
Department of Agronomy
Purdue University
1150 Lilly Hall
West Lafayette, IN 47907-1150
March 2003