

Preface

This book was written to serve students and researchers of the animal sciences, with the primary purpose of helping them to learn about and apply appropriate experimental designs and statistical methods. Statistical methods applied to biological sciences are known as biostatistics or biometrics, and they have their origins in agricultural research. The characteristic that distinguishes biometrics within statistics is the fact that biological measurements are variable, not only because of measurement error, but also from their natural variability from genetic and environmental sources. These sources of variability must be taken into account when making inferences about biological material. Accounting for these sources of variation has led to the development of experimental designs that incorporate blocking, covariates and repeated measures. Appropriate techniques for analysis of data from these designs and others are covered in the book.

Early in the book, readers are presented basic principles of statistics so they will be able to follow subsequent applications with familiarity and understanding, and without having to switch to another book of introductory statistics. Later chapters cover statistical methods most frequently used in the animal sciences for analysis of continuous and categorical variables. Each chapter begins by introducing a problem with practical questions, followed with a brief theoretical background and short proofs. The text is augmented with examples, mostly from animal sciences and related fields, with the purpose of making applications of the statistical methods familiar. Some examples are very simple and are presented in order to provide basic understanding and the logic behind calculations. These examples can be solved using a pocket calculator. Some examples are more complex, especially those in the later chapters. Most examples are also solved using SAS statistical software. Both sample SAS programs and SAS listings are given with brief explanations. Further, the solutions are often given with sufficient decimal digits, more than is practically necessary, so that readers can compare results to verify calculation technique.

The first five chapters of the book are: 1) Presenting and Summarizing Data; 2) Probability; 3) Random Variables and Their Distributions; 4) Population and Sample; and 5) Estimation of Parameters. These chapters provide a basic introduction to biostatistics including definitions of terms, coverage of descriptive statistics and graphical presentation of data, the basic rules of probability, methods of parameter estimation, and descriptions of distributions including the Bernoulli, binomial, hypergeometric, Poisson, multinomial, uniform, normal, chi-square, t , and F distributions. Chapter 6 describes hypothesis testing and includes explanations of the null and alternate hypotheses, use of probability or density functions, critical values, critical region and P values. Hypothesis tests for many specific cases are shown such as population means and proportions, expected and empirical frequency, and test of variances. Also, the use of confidence intervals in hypothesis testing is shown. The difference between statistical and practical significance, types of errors in making conclusions, power of test, and sample size are discussed.

Chapters 7 to 10 present the topics of correlation and regression. The coverage begins with simple linear regression and describes the model, its parameters and assumptions. Least squares and maximum likelihood methods of parameter estimation are shown. The concept of partitioning the total variance to explained and unexplained sources in the analysis of variance table is introduced. In chapter 8 the general meaning and definition of

the correlation coefficient, and the estimation of the correlation coefficient from samples and testing of hypothesis are shown. In chapters 9 and 10 multiple and curvilinear regressions are described. Important facts are explained using matrices in the same order of argument as for the simple regression. Model building is introduced including the definitions of partial and sequential sum of squares, test of model adequacy using a likelihood function, and Conceptual Predictive and Akaike criteria. Some common problems of regression analysis like outliers and multicollinearity are described, and their detection and possible remedies are explained. Polynomial, nonlinear and segmented regressions are introduced. Some examples are shown including estimating growth curves and functions with a plateau such as for determining nutrient requirements.

One-way analysis of variance is introduced in chapter 11. In this chapter a one-way analysis of variance model is used to define hypotheses, partition sums of squares in order to use an F test, and estimate means and effects. Post-test comparison of means, including least significant difference, Tukey test and contrasts are shown. Fixed and random effects models are compared, and fixed and random effects are also shown using matrices.

Chapters 12 to 21 focus on specific experimental designs and their analyses. Specific topics include: general concepts of design, blocking, change-over designs, factorials, nested designs, double blocking, split-plots, analysis of covariance, repeated measures and analysis of numerical treatment levels. Examples with sample SAS programs are provided for each topic.

The final chapter covers the special topic of discrete dependent variables. Logit and probit models for binary and binomial dependent variables and loglinear models for count data are explained. A brief theoretical background is given with examples and SAS procedures.

We wish to express our gratitude to everyone who helped us produce this book. We extend our special acknowledgement to Matt Lucy, Duane Keisler, Henry Mesa, Kristi Cammack, Marijan Posavi and Vesna Luzar-Stiffler for their reviews, and Cyndi Jennings, Cinda Hudlow and Dragan Tupajic for their assistance with editing.

Zagreb, Croatia
Columbia, Missouri
March 2004

Miroslav Kaps
William R. Lamberson