Statistical Models For Engineers

Text for STT 363 Wright State University

by

Thaddeus Tarpey

Table of Contents

Preface

This text is for an introductory course covering basic statistical ideas for undergraduate engineering students. Many undergraduate engineer programs require a single course in probability/statistics. However, it is difficult to get to the heart of basic statistical techniques and principals in a single course. Typically introductory texts on engineering statistics spend a great deal of time on basic probability ideas for the first several chapters. In fact, basic probability can easily fill up a standard introductory course. Because engineering students often have only one probability/statistics course, the material needs to be reorganized in order to allow for coverage of statistical methodology.

In this text I give a brief introduction to probability in Chapter 2. However, I also introduce the basics of hypothesis testing in the context of a binomial model in Chapter 2. Typically, hypothesis testing is introduced much later in introductory textbooks and consequently, many engineering students fail to get any exposure to the ideas behind hypothesis testing. Hypothesis testing is actually a natural topic to discuss when introducing the basic ideas of probability theory.

Hypothesis testing ideas are also covered in later chapters dealing with continuous models, regression analysis and analysis of variance. Once students have the basic ideas covered in Chapter 2, the use of hypothesis testing in these later chapters should come naturally.

Continuous probability models are discussed in Chapter 3. First, some basic probability ideas are introduced involving density functions and expectations. As in Chapter 2, once the basic probability ideas are introduced, the focus is shifted to statistical problems of hypothesis testing and parameter estimation. The t distribution is introduced immediately. We do not cover the artificial statistical problems where the variance is assumed to be known and instead move immediately to the t-distribution.

Chapter 4 introduces some basic ideas for multivariate statistics. Some definitions are provided in terms of multiple integration, but multiple integration techniques are not needed for computations. Instead, the focus is again on statistical problems such as the estimation of covariances. This chapter does use some basic matrix algebra ideas. I do not assume the reader has any matrix algebra background. All the necessary matrix algebra background needed for this text is provided in an appendix to Chapter 4. However, students familiar with basic matrix computations will find those skills useful.

Data encountered in practice are often multivariate in nature and some exposure to multivariate ideas will be useful. However, the primary motivation for Chapter 4 is to lay a foundation needed in the regression material. Regression models are defined in terms of two or more parameters. When estimating these parameters, we often end up with correlated estimators. In order to properly design experiments, some basic understanding of these correlations are needed. Therefore, basic ideas from multivariate analysis are needed to completely understand more complex models.

Many introductory textbooks introduce the bivariate normal distribution, but they do so without the aid of matrix notation. As a consequence, it is impractical to introduce natural generalizations of tri-variate and higher dimensional distributions because the density formulas become way too complicated. However, if matrix notation is used, then the density function has the exact same form regardless of dimension. Thus, the multivariate normal distribution is introduced in Chapter 4 using matrix notation.

The matrix approach is also used in Chapter 5 on regression models. Again, authors of introductory texts generally refrain from matrix notation when covering regression models except later in the material when discussing multiple regression. However, since data comes to us in arrays, the matrix approach is natural. The least-squares solution when fitting a regression model is exactly the same for simple linear regression, multiple regression, and polynomial regression when the matrix notation is used. The basic matrix ideas needed to understand these formulas are quite elementary. Therefore, it seems unnatural and overly complicated to introduce regression models without the aid of matrix notation. Introductory texts that do not use matrix notation are littered with lots of complicated and un-intuitive formulas. The matrix approach taken in this text avoids these problems.

Finally, Chapter 6 covers material on comparing means of different groups or populations. The two-sample t-test and ANOVA ideas are covered in this chapter as well as some general discussion of experimental design.

Currently this text does not have any coverage of statistical process control and reliability. This text is designed to cover material in a single class that generally takes two classes to cover. Unfortunately there simply is not enough time to cover additional topics in a single course. Nonetheless, this introductory text does introduce the student to the basic principals of statistical thinking, variability, statistical modeling, hypothesis testing etc. that will make it easier to understand statistical concepts in other settings.

I would like to thank Jolene Splett from the National Institute of Standards in Boulder, Colorado for her help with some of the data sets used in this text and also thank Professor Richard Bethke, Chair of the Mechanical Engineering Department at Wright State University, for helpful comments that motivated the writing of this text.

Thaddeus Tarpey

November 19, 2003