Preface

Advanced Statistics for Health Research provides a rigorous geometric understanding of models used in the analysis of health data, including linear and nonlinear regression models, and supervised machine learning models. Models drawn from the health literature include: ordinary least squares, two-stage least squares, probits, logits, Cox regressions, duration modeling, quantile regression, and random forest regression. Causal inference techniques from the health literature are presented including randomization, matching, and propensity score matching, differences-in-differences, instrumental variables, regression discontinuity, and fixed effects analysis. Codes for the respective statistical techniques presented are given for STATA, SAS, and R.

The first three chapters introduce statistics in general, and the linear regression model in particular. The basic vocabulary is established. After discussing the data generating properties in the beginning of Chapter 4, the last half of Chapter 4 — and the heart of the book — discusses ways to do statistics badly. If you skip this section, you and your descendants will be cursed for 4.7839 generations (with a standard deviation of 2.0156 generations). Or at least, their research will be cursed.

The liver, spleen, backbone, and lungs of the book are contained in Chapter 5: how to do causal inference by randomization, or by "as if randomized" analysis of data.

All the statistical extremities are contained and appropriately autopsied in the remaining chapters.