Contents at a Glance

Introduction ................................................................ 1

Part I: Beginning with Biostatistics Basics ................. 7
Chapter 1: Biostatistics 101.................................................................9
Chapter 2: Overcoming Mathophobia: Reading and Understanding Mathematical Expressions ................................................17
Chapter 3: Getting Statistical: A Short Review of Basic Statistics ........................................31
Chapter 4: Counting on Statistical Software ........................................51
Chapter 5: Conducting Clinical Research ........................................................................61
Chapter 6: Looking at Clinical Trials and Drug Development ........................................77

Part II: Getting Down and Dirty with Data .......... 91
Chapter 7: Getting Your Data into the Computer ..............................................93
Chapter 8: Summarizing and Graphing Your Data.............................................103
Chapter 9: Aiming for Accuracy and Precision ..................................................121
Chapter 10: Having Confidence in Your Results ..................................................133
Chapter 11: Fuzzy In Equals Fuzzy Out: Pushing Imprecision through a Formula............143

Part III: Comparing Groups ....................................... 153
Chapter 12: Comparing Average Values between Groups ......................................155
Chapter 13: Comparing Proportions and Analyzing Cross-Tabulations .................173
Chapter 14: Taking a Closer Look at Fourfold Tables ..........................................189
Chapter 15: Analyzing Incidence and Prevalence Rates in Epidemiologic Data ...........203
Chapter 16: Feeling Noninferior (Or Equivalent) ..................................................211

Part IV: Looking for Relationships with Correlation and Regression .................. 219
Chapter 17: Introducing Correlation and Regression ........................................221
Chapter 18: Getting Straight Talk on Straight-Line Regression ............................233
Chapter 19: More of a Good Thing: Multiple Regression ......................................251
Chapter 20: A Yes-or-No Proposition: Logistic Regression ....................................267
Chapter 21: Other Useful Kinds of Regression .....................................................291
Part V: Analyzing Survival Data ......................... 311
Chapter 22: Summarizing and Graphing Survival Data ........................................... 313
Chapter 23: Comparing Survival Times ................................................................. 331
Chapter 24: Survival Regression ............................................................................. 339

Part VI: The Part of Tens ........................................ 357
Chapter 25: Ten Distributions Worth Knowing ...................................................... 359
Chapter 26: Ten Easy Ways to Estimate How Many Subjects You Need ............ 369

Index ...................................................................................................................... 375
# Table of Contents

## Introduction

- About This Book ................................................................................. 1
- Conventions Used in This Book .......................................................... 2
- What You’re Not to Read .................................................................. 2
- Foolish Assumptions ......................................................................... 3
- How This Book Is Organized ............................................................. 3
  - Part I: Beginning with Biostatistics Basics ....................................... 3
  - Part II: Getting Down and Dirty with Data ...................................... 4
  - Part III: Comparing Groups ............................................................ 4
  - Part IV: Looking for Relationships with Correlation and Regression 4
  - Part V: Analyzing Survival Data ..................................................... 5
  - Part VI: The Part of Tens ................................................................. 5
- Icons Used in This Book ................................................................. 5
- Where to Go from Here ..................................................................... 6

## Part I: Beginning with Biostatistics Basics

### Chapter 1: Biostatistics 101

- Brushing Up on Math and Stats Basics ................................................. 9
- Doing Calculations with the Greatest of Ease ...................................... 10
- Concentrating on Clinical Research .................................................... 11
- Drawing Conclusions from Your Data ............................................... 12
  - Statistical estimation theory ......................................................... 12
  - Statistical decision theory ............................................................ 12
- A Matter of Life and Death: Working with Survival Data .................... 14
- Figuring Out How Many Subjects You Need ....................................... 15
- Getting to Know Statistical Distributions .......................................... 16

### Chapter 2: Overcoming Mathophobia: Reading and Understanding Mathematical Expressions

- Breaking Down the Basics of Mathematical Formulas .................... 18
- Displaying formulas in different ways ................................................. 18
- Checking out the building blocks of formulas .................................... 18
- Focusing on Operations Found in Formulas ...................................... 20
  - Basic mathematical operations ..................................................... 20
  - Powers, roots, and logarithms ....................................................... 22
  - Factorials and absolute values ..................................................... 24
# Biostatistics For Dummies

<table>
<thead>
<tr>
<th>Functions</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple and complicated formulas</td>
<td>25</td>
</tr>
<tr>
<td>Equations</td>
<td>26</td>
</tr>
<tr>
<td>Counting on Collections of Numbers</td>
<td>26</td>
</tr>
<tr>
<td>One-dimensional arrays</td>
<td>26</td>
</tr>
<tr>
<td>Higher-dimensional arrays</td>
<td>27</td>
</tr>
<tr>
<td>Arrays in formulas</td>
<td>28</td>
</tr>
<tr>
<td>Sums and products of the elements of an array</td>
<td>28</td>
</tr>
</tbody>
</table>

## Chapter 3: Getting Statistical: A Short Review of Basic Statistics . . . 31

- Taking a Chance on Probability .......................................................... 31
- Thinking of probability as a number ................................................. 32
- Following a few basic rules ............................................................... 32
- Comparing odds versus probability .................................................. 33
- Some Random Thoughts about Randomness ........................................ 34
- Picking Samples from Populations ..................................................... 35
- Recognizing that sampling isn’t perfect .......................................... 36
- Digging into probability distributions .............................................. 37
- Introducing Statistical Inference ....................................................... 38
- Statistical estimation theory ............................................................. 38
- Statistical decision theory ............................................................... 40
- Homing In on Hypothesis Testing ..................................................... 41
- Getting the language down ............................................................... 41
- Testing for significance .................................................................... 42
- Understanding the meaning of "p value" as the result of a test .......... 43
- Examining Type I and Type II errors ............................................... 43
- Grasping the power of a test ........................................................... 45
- Going Outside the Norm with Nonparametric Statistics ............... 48

## Chapter 4: Counting on Statistical Software .................................. 51

- Desk Job: Personal Computer Software ............................................ 51
- Checking out commercial software .................................................. 52
- Focusing on free software ............................................................... 55
- On the Go: Calculators and Mobile Devices ..................................... 57
  - Scientific and programmable calculators ....................................... 57
  - Mobile devices ............................................................................... 58
- Gone Surfin’: Web-Based Software ................................................. 58
- On Paper: Printed Calculators ......................................................... 59

## Chapter 5: Conducting Clinical Research ...................................... 61

- Designing a Clinical Study ................................................................. 61
- Identifying aims, objectives, hypotheses, and variables .................... 61
- Deciding who will be in the study ..................................................... 63
- Choosing the structure of the study ................................................ 64
- Using randomization ....................................................................... 64
- Selecting the analyses to use .......................................................... 66
# Table of Contents

Defining analytical populations..........................................................67  
Determining how many subjects to enroll........................................67  
Putting together the protocol ............................................................68  
Carrying Out a Clinical Study ............................................................70  
Protecting your subjects..................................................................70  
Collecting and validating data ...........................................................72  
Analyzing Your Data........................................................................73  
Dealing with missing data.................................................................74  
Handling multiplicity.......................................................................74  
Incorporating interim analyses...........................................................76  

**Chapter 6: Looking at Clinical Trials and Drug Development ..........77**  
Not Ready for Human Consumption: Doing Preclinical Studies.........78  
Testing on People during Clinical Trials to Check a Drug’s Safety and Efficacy .................................................................80  
Phase I: Determining the maximum tolerated dose ............................80  
Phase II: Finding out about the drug’s performance ............................82  
Phase III: Proving that the drug works ..............................................84  
Phase IV: Keeping an eye on the marketed drug ..................................85  
Holding Other Kinds of Clinical Trials ..............................................86  
Pharmacokinetics and pharmacodynamics (PK/PD studies) .............86  
Bioequivalence studies .....................................................................88  
Thorough QT studies .....................................................................88  

**Part II: Getting Down and Dirty with Data ......................... 91**  

**Chapter 7: Getting Your Data into the Computer .................93**  
Looking at Levels of Measurement.....................................................94  
Classifying and Recording Different Kinds of Data ............................95  
Dealing with free-text data .................................................................95  
Assigning subject identification (ID) numbers....................................95  
Organizing name and address data ....................................................96  
Collecting categorical data .................................................................96  
Recording numerical data .................................................................98  
Entering date and time data ...............................................................99  
Checking Your Entered Data for Errors .................................................101  
Creating a File that Describes Your Data File ......................................102  

**Chapter 8: Summarizing and Graphing Your Data ..............103**  
Summarizing and Graphing Categorical Data .......................................104  
Summarizing Numerical Data .............................................................106  
Locating the center of your data .......................................................107  
Describing the spread of your data ....................................................110  
Showing the symmetry and shape of the distribution ........................113  
Structuring Numerical Summaries into Descriptive Tables .................114
Graphing Numerical Data .............................................................. 115
  Showing the distribution with histograms .................................. 116
  Summarizing grouped data with bars, boxes, and whiskers ......... 118
  Depicting the relationships between numerical variables with other graphs ........................................... 120

Chapter 9: Aiming for Accuracy and Precision ............................. 121
  Beginning with the Basics of Accuracy and Precision............... 121
  Getting to know sample statistics and population parameters ... 121
  Understanding accuracy and precision in terms of the sampling distribution ............................................. 122
  Thinking of measurement as a kind of sampling ...................... 123
  Expressing errors in terms of accuracy and precision .......... 124
  Improving Accuracy and Precision ............................................. 126
  Enhancing sampling accuracy .................................................. 126
  Getting more accurate measurements ...................................... 126
  Improving sampling precision .................................................. 127
  Increasing the precision of your measurements ...................... 128
  Calculating Standard Errors for Different Sample Statistics ...... 129
  A mean .................................................................................... 129
  A proportion ........................................................................... 130
  Event counts and rates ........................................................... 130
  A regression coefficient .......................................................... 131

Chapter 10: Having Confidence in Your Results ............................ 133
  Feeling Confident about Confidence Interval Basics ................. 133
  Defining confidence intervals ................................................... 134
  Looking at confidence levels .................................................... 134
  Taking sides with confidence intervals .................................... 135
  Calculating Confidence Intervals .............................................. 136
  Before you begin: Formulas for confidence limits in large samples .............................................................. 136
  The confidence interval around a mean .................................. 137
  The confidence interval around a proportion ......................... 139
  The confidence interval around an event count or rate .......... 140
  The confidence interval around a regression coefficient ......... 141
  Relating Confidence Intervals and Significance Testing .......... 141

Chapter 11: Fuzzy In Equals Fuzzy Out: Pushing Imprecision through a Formula ......................................................... 143
  Understanding the Concept of Error Propagation .................... 144
  Using Simple Error Propagation Formulas for Simple Expressions ..... 146
  Adding or subtracting a constant doesn’t change the SE ......... 146
  Multiplying (or dividing) by a constant multiplies (or divides) the SE by the same amount ......................... 147
  For sums and differences: Add the squares of SEs together ....... 147
For averages: The square root law takes over ......................... 148
For products and ratios: Squares of relative SEs
    are added together .................................................. 148
For powers and roots: Multiply the relative SE by the power ..... 149
Handling More Complicated Expressions ............................. 149
Using the simple rules consecutively .................................. 150
Checking out an online calculator ...................................... 150
Simulating error propagation — easy, accurate, and versatile ... 152

Part III: Comparing Groups ............................................. 153

Chapter 12: Comparing Average Values between Groups .......... 155
Knowing That Different Situations Need Different Tests ............ 155
Comparing the mean of a group of numbers
    to a hypothesized value ............................................. 156
Comparing two groups of numbers ..................................... 156
Comparing three or more groups of numbers ....................... 157
Analyzing data grouped on several different variables .......... 158
Adjusting for a “nuisance variable” when
    comparing numbers ................................................... 158
Comparing sets of matched numbers .................................. 159
Comparing within-group changes between groups ............... 160
Trying the Tests Used for Comparing Averages .................... 161
Surveying Student t tests ................................................. 161
Assessing the ANOVA .................................................... 164
Running Student t tests and ANOVAs from summary data ...... 168
Running nonparametric tests .......................................... 169
Estimating the Sample Size You Need for Comparing Averages . 169
    Simple formulas ...................................................... 169
    Software and web pages ............................................. 170
    A sample-size nomogram .......................................... 170

Chapter 13: Comparing Proportions and Analyzing
Cross-Tabulations ......................................................... 173
Examining Two Variables with the Pearson Chi-Square Test ....... 174
Understanding how the chi-square test works ....................... 175
Pointing out the pros and cons of the chi-square test .......... 180
Modifying the chi-square test: The Yates continuity correction .. 181
Focusing on the Fisher Exact Test .................................... 181
Understanding how the Fisher Exact test works ................... 181
Noting the pros and cons of the Fisher Exact test .............. 182
Calculating Power and Sample Size for Chi-Square
    and Fisher Exact Tests ............................................. 183
Analyzing Ordinal Categorical Data with the Kendall Test ...... 185
Studying Stratified Data with the Mantel-Haenszel Chi-Square Test ... 187
Chapter 14: Taking a Closer Look at Fourfold Tables ............... 189
  Focusing on the Fundamentals of Fourfold Tables ..................... 190
  Choosing the Right Sampling Strategy ................................... 191
  Producing Fourfold Tables in a Variety of Situations ................. 192
    Describing the association between two binary variables .......... 193
    Assessing risk factors ..................................................... 194
    Evaluating diagnostic procedures ...................................... 197
    Investigating treatments .................................................. 199
  Looking at inter- and intra-rater reliability ............................ 201

Chapter 15: Analyzing Incidence and Prevalence Rates
in Epidemiologic Data ......................................................... 203
  Understanding Incidence and Prevalence ................................ 203
    Prevalence: The fraction of a population
      with a particular condition .............................................. 204
    Incidence: Counting new cases .......................................... 204
    Understanding how incidence and prevalence are related ......... 205
  Analyzing Incidence Rates .................................................. 205
    Expressing the precision of an incidence rate ...................... 205
    Comparing incidences with the rate ratio .......................... 206
    Calculating confidence intervals for a rate ratio ................. 207
    Comparing two event rates .............................................. 207
    Comparing two event counts with identical exposure ............ 209
  Estimating the Required Sample Size .................................. 209

Chapter 16: Feeling Noninferior (Or Equivalent) ...................... 211
  Understanding the Absence of an Effect ................................ 212
    Defining the effect size: How different are the groups? .......... 212
    Defining an important effect size: How close is close enough? ... 213
    Recognizing effects: Can you spot a difference
      if there really is one? .................................................... 213
  Proving Equivalence and Noninferiority ................................ 214
    Using significance tests ................................................. 214
    Using confidence intervals ............................................. 215
    Some precautions about noninferiority testing .................... 217

Part IV: Looking for Relationships
with Correlation and Regression ........................................... 219

Chapter 17: Introducing Correlation and Regression ................. 221
  Correlation: How Strongly Are Two Variables Associated? ........ 222
    Lining up the Pearson correlation coefficient ...................... 222
    Analyzing correlation coefficients .................................... 223
Regression: What Equation Connects the Variables? ............................. 227
Understanding the purpose of regression analysis.................................. 227
Talking about terminology and mathematical notation ....................... 228
Classifying different kinds of regression............................................. 229

Chapter 18: Getting Straight Talk on Straight-Line Regression ........ 233
Knowing When to Use Straight-Line Regression .................................. 234
Understanding the Basics of Straight-Line Regression ........................... 235
Running a Straight-Line Regression ...................................................... 236
  Taking a few basic steps ................................................................... 237
  Walking through an example............................................................ 237
Interpreting the Output of Straight-Line Regression ............................. 239
  Seeing what you told the program to do ........................................... 240
  Looking at residuals ..................................................................... 241
  Making your way through the regression table ............................... 243
  Wrapping up with measures of goodness-of-fit .............................. 247
  Scientific fortune-telling with the prediction formula ................... 248
Recognizing What Can Go Wrong with Straight-Line Regression .......... 249
Figuring Out the Sample Size You Need ............................................. 249

Chapter 19: More of a Good Thing: Multiple Regression .............. 251
Understanding the Basics of Multiple Regression ............................... 251
  Defining a few important terms....................................................... 252
  Knowing when to use multiple regression ..................................... 253
  Being aware of how the calculations work ..................................... 253
Running Multiple Regression Software ............................................. 254
  Preparing categorical variables ...................................................... 254
  Recoding categorical variables as numerical ................................. 255
  Creating scatter plots before you jump into your multiple regression ...................................................... 256
  Taking a few steps with your software .......................................... 258
Interpreting the Output of a Multiple Regression ............................... 258
  Examining typical output from most programs .............................. 259
  Checking out optional output available from some programs .... 260
  Deciding whether your data is suitable for regression analysis ... 261
  Determining how well the model fits the data ................................ 262
Watching Out for Special Situations that Arise in
  Multiple Regression ........................................................................ 263
  Synergy and anti-synergy ............................................................. 263
  Collinearity and the mystery of the disappearing significance .. 263
Figuring How Many Subjects You Need ............................................. 265
Chapter 20: A Yes-or-No Proposition: Logistic Regression . . . . . . . .267
Using Logistic Regression ................................................................. 267
Understanding the Basics of Logistic Regression .............................. 268
Gathering and graphing your data ..................................................... 268
Fitting a function with an S shape to your data .............................. 270
Handling multiple predictors in your logistic model ...................... 274
Running a Logistic Regression with Software .................................. 274
Interpreting the Output of Logistic Regression ................................. 275
   Seeing summary information about the variables ......................... 276
   Assessing the adequacy of the model ........................................... 276
   Checking out the table of regression coefficients ....................... 278
   Predicting probabilities with the fitted logistic formula ............... 278
   Making yes or no predictions ...................................................... 280
Heads Up: Knowing What Can Go Wrong with Logistic Regression ... 285
   Don’t fit a logistic function to nonlogistic data .......................... 285
   Watch out for collinearity and disappearing significance ............ 285
   Check for inadvertent reverse-coding of the outcome variable ... 286
   Don’t misinterpret odds ratios for numerical predictors .......... 286
   Don’t misinterpret odds ratios for categorical predictors .......... 286
   Beware the complete separation problem ................................ 287
Figuring Out the Sample Size You Need for Logistic Regression .... 288

Chapter 21: Other Useful Kinds of Regression  . . . . . . . . . . . . . . . . . . . .291
Analyzing Counts and Rates with Poisson Regression .................... 291
   Introducing the generalized linear model ................................ 292
   Running a Poisson regression .................................................... 293
   Interpreting the Poisson regression output ................................. 295
   Discovering other things that Poisson regression can do .......... 296
Anything Goes with Nonlinear Regression ..................................... 298
   Distinguishing nonlinear regression from other kinds ............... 299
   Checking out an example from drug research ............................ 300
   Running a nonlinear regression ................................................ 302
   Interpreting the output ............................................................ 304
   Using equivalent functions to fit the parameters
      you really want ..................................................................... 305
Smoothing Nonparametric Data with LOWESS .............................. 306
   Running LOWESS .................................................................. 307
   Adjusting the amount of smoothing ........................................ 309
Part V: Analyzing Survival Data ................................................. 311

Chapter 22: Summarizing and Graphing Survival Data ............... 313
Understanding the Basics of Survival Data ................................... 314
Knowing that survival times are intervals ................................... 314
Recognizing that survival times aren’t normally distributed ....... 314
Considering censoring ............................................................... 315
Looking at the Life-Table Method ................................................. 318
Making a life table ...................................................................... 319
Interpreting a life table ............................................................... 323
Graphing hazard rates and survival probabilities
  from a life table ...................................................................... 324
Digging Deeper with the Kaplan-Meier Method ............................ 324
Heeding a Few Guidelines for Life Tables and
  the Kaplan-Meier Method ....................................................... 326
Recording survival times the right way ..................................... 327
Recording censoring information correctly ................................ 327
Interpreting those strange-looking survival curves .................. 328
Doing Even More with Survival Data ......................................... 329

Chapter 23: Comparing Survival Times ..................................... 331
Comparing Survival between Two Groups with the Log-Rank Test ..... 332
  Understanding what the log-rank test is doing ......................... 333
  Running the log-rank test on software ..................................... 333
  Looking at the calculations ..................................................... 334
  Assessing the assumptions .................................................... 336
Considering More Complicated Comparisons ............................ 337
Coming Up with the Sample Size Needed for Survival Comparisons ... 337

Chapter 24: Survival Regression .............................................. 339
Knowing When to Use Survival Regression .................................. 339
Explaining the Concepts behind Survival Regression ...................... 340
  The steps of Cox PH regression .............................................. 341
  Hazard ratios ......................................................................... 345
Running a Survival Regression .................................................. 346
Interpreting the Output of a Survival Regression ........................... 347
  Testing the validity of the assumptions ................................... 349
  Checking out the table of regression coefficients ....................... 350
Homing in on hazard ratios and their confidence intervals .......... 350
Assessing goodness-of-fit and predictive ability of the model ...... 351
Focusing on baseline survival and hazard functions .................... 352
How Long Have I Got, Doc? Constructing Prognosis Curves ..........353
Running the proportional-hazards regression .........................353
Finding h .................................................................................354
Estimating the Required Sample Size for a Survival Regression ....356

Part VI: The Part of Tens ........................................... 357

Chapter 25: Ten Distributions Worth Knowing ................. 359
The Uniform Distribution ......................................................360
The Normal Distribution .......................................................360
The Log-Normal Distribution ..............................................361
The Binomial Distribution .....................................................362
The Poisson Distribution ......................................................362
The Exponential Distribution ...............................................363
The Weibull Distribution .......................................................364
The Student t Distribution .....................................................364
The Chi-Square Distribution .................................................366
The Fisher F Distribution .....................................................367

Chapter 26: Ten Easy Ways to Estimate How Many 
Subjects You Need .............................................................369
Comparing Means between Two Groups ................................370
Comparing Means among Three, Four, or Five Groups ........370
Comparing Paired Values .....................................................370
Comparing Proportions between Two Groups .......................371
Testing for a Significant Correlation .....................................371
Comparing Survival between Two Groups .............................371
Scaling from 80 Percent to Some Other Power ......................372
Scaling from 0.05 to Some Other Alpha Level .......................373
Making Adjustments for Unequal Group Sizes ......................373
Allowing for Attrition .........................................................374

Index .................................................................................. 375