

# CONTENTS

---

Preface	xi
Acknowledgments	xv
Introduction	xvii
I.1    Statistics in practice	xvii
I.2    Learning statistics	xix
<b>1    Foundations</b>	<b>1</b>
1.1    Identifying and summarizing data	1
1.2    Population distributions	5
1.3    Selecting individuals at random—probability	9
1.4    Random sampling	11
1.4.1    Central limit theorem—normal version	12
1.4.2    Central limit theorem—t-version	14
1.5    Interval estimation	15
1.6    Hypothesis testing	19
1.6.1    The rejection region method	19
1.6.2    The p-value method	21
1.6.3    Hypothesis test errors	24
1.7    Random errors and prediction	25
1.8    Chapter summary	28
Problems	29
	<b>vii</b>

<b>2</b>	<b>Simple linear regression</b>	<b>35</b>
2.1	Probability model for $X$ and $Y$	35
2.2	Least squares criterion	40
2.3	Model evaluation	45
2.3.1	Regression standard error	46
2.3.2	Coefficient of determination— $R^2$	48
2.3.3	Slope parameter	52
2.4	Model assumptions	59
2.4.1	Checking the model assumptions	61
2.4.2	Testing the model assumptions	66
2.5	Model interpretation	66
2.6	Estimation and prediction	68
2.6.1	Confidence interval for the population mean, $E(Y)$	68
2.6.2	Prediction interval for an individual $Y$ -value	69
2.7	Chapter summary	72
2.7.1	Review example	74
	Problems	78
<b>3</b>	<b>Multiple linear regression</b>	<b>83</b>
3.1	Probability model for $(X_1, X_2, \dots)$ and $Y$	83
3.2	Least squares criterion	87
3.3	Model evaluation	92
3.3.1	Regression standard error	92
3.3.2	Coefficient of determination— $R^2$	94
3.3.3	Regression parameters—global usefulness test	101
3.3.4	Regression parameters—nested model test	104
3.3.5	Regression parameters—individual tests	109
3.4	Model assumptions	118
3.4.1	Checking the model assumptions	119
3.4.2	Testing the model assumptions	123
3.5	Model interpretation	124
3.6	Estimation and prediction	126
3.6.1	Confidence interval for the population mean, $E(Y)$	126
3.6.2	Prediction interval for an individual $Y$ -value	127
3.7	Chapter summary	130
	Problems	132
<b>4</b>	<b>Regression model building I</b>	<b>137</b>
4.1	Transformations	138
4.1.1	Natural logarithm transformation for predictors	138
4.1.2	Polynomial transformation for predictors	144

4.1.3	Reciprocal transformation for predictors	147
4.1.4	Natural logarithm transformation for the response	151
4.1.5	Transformations for the response and predictors	155
4.2	Interactions	159
4.3	Qualitative predictors	166
4.3.1	Qualitative predictors with two levels	167
4.3.2	Qualitative predictors with three or more levels	174
4.4	Chapter summary	182
	Problems	184
<b>5</b>	<b>Regression model building II</b>	<b>189</b>
5.1	Influential points	189
5.1.1	Outliers	189
5.1.2	Leverage	194
5.1.3	Cook's distance	196
5.2	Regression pitfalls	199
5.2.1	Nonconstant variance	199
5.2.2	Autocorrelation	202
5.2.3	Multicollinearity	206
5.2.4	Excluding important predictor variables	209
5.2.5	Overfitting	212
5.2.6	Extrapolation	213
5.2.7	Missing data	215
5.2.8	Power and sample size	217
5.3	Model building guidelines	218
5.4	Model selection	221
5.5	Model interpretation using graphics	224
5.6	Chapter summary	231
	Problems	234
<b>6</b>	<b>Case studies</b>	<b>243</b>
6.1	Home prices	243
6.1.1	Data description	243
6.1.2	Exploratory data analysis	245
6.1.3	Regression model building	246
6.1.4	Results and conclusions	247
6.1.5	Further questions	252
6.2	Vehicle fuel efficiency	253
6.2.1	Data description	253
6.2.2	Exploratory data analysis	253
6.2.3	Regression model building	255

6.2.4	Results and conclusions	256
6.2.5	Further questions	261
6.3	Pharmaceutical patches	261
6.3.1	Data description	261
6.3.2	Exploratory data analysis	261
6.3.3	Regression model building	263
6.3.4	Model diagnostics	263
6.3.5	Results and conclusions	264
6.3.6	Further questions	266
<b>7</b>	<b>Extensions</b>	<b>267</b>
7.1	Generalized linear models	268
7.1.1	Logistic regression	268
7.1.2	Poisson regression	273
7.2	Discrete choice models	275
7.3	Multilevel models	278
7.4	Bayesian modeling	281
7.4.1	Frequentist inference	281
7.4.2	Bayesian inference	281
Appendix A:	Computer software help	285
	Problems	287
Appendix B:	Critical values for t-distributions	289
Appendix C:	Notation and formulas	293
C.1	Univariate data	293
C.2	Simple linear regression	294
C.3	Multiple linear regression	295
Appendix D:	Mathematics refresher	297
D.1	The natural logarithm and exponential functions	297
D.2	Rounding and accuracy	298
Appendix E:	Answers for selected problems	299
References		309
Glossary		315
Index		321