

**INTRODUCTION TO MODERN MATHEMATICAL
MODELING WITH R**

CONTENTS

Foreword	xi
Glossary	xv

PART I SHORTCUT TO DISCOVERING LAWS OF NATURE AND LINEAR REGRESSION MODELS

1 Dimensional analysis	3
1.1 Dimension and units	3
1.2 Fundamental physics dimensions: $LMT\theta I$ -class	4
1.3 Relationships between magnetic and electric fields	7
1.4 The general principle of dimensional analysis	8
1.4.1 A universal mathematical model for nature	8
1.4.2 Dimensional analysis for a simple pendulum and calculation of the pendulum period	10
1.4.3 The instantaneous speed of a free-fall body of mass m	11
1.5 Shock wave radius of a nuclear explosion	13
Exercises	15
2 Basics of R Programming	19
2.1 Download and install R software package	19
2.2 R Tutorial	20
	v

vi CONTENTS

2.2.1	R as a smart calculator	20
2.2.2	Write a function in R	21
2.2.3	Plot with R	21
2.2.4	Symbolic calculations by R	21
2.2.5	Vectors and matrices	22
2.2.6	Statistics	24
2.3	Online Tutorials	25
2.3.1	Youtube tutorial: for true beginners	25
2.3.2	Youtube tutorial: for some basic statistical summaries	25
2.3.3	Youtube tutorial: Input data by reading a csv file into R	25
	Exercises	26
3	Linear Models by Regression	29
3.1	Introduction to a linear model	29
3.2	Formula derivation and interpretation for the trend and intercept of a linear regression	31
3.3	Evaluate a linear model by regression using R	33
3.4	Research level exploration	39
	Exercises	40
4	Principles of Mathematical Modeling	43
4.1	Principles of mathematical modeling and client report template	43
4.2	Zeroing a rifle	44
4.3	Zero-dimensional Energy Balance Model for Earth's Constant Temperature Climate	48
4.3.1	Earth's energy budget	48
4.3.2	A uniform water-covered Earth	50
4.4	EBM for a uniform Earth with nonlinear albedo feedback	52
4.5	Modeling mortgage payment	55
4.6	Template of a client report	57
4.7	Term Project #1.	58
	Exercises	59
5	Mathematical Modeling by Linear Algebra	61
5.1	Kirchhoff's laws and solution of an electric circuit	61
5.2	Mass balance models for chemical equations	63
5.3	Leontif production model: a balance of the output and input	64
5.4	An SVD representation model for space-time data	66
5.4.1	The fundamental idea of SVD: space-time-energy separation	66
5.4.2	SVD for a 2-Dim spatial domain and 1-Dim temporal domain	70
5.4.3	An SVD algorithm and covariance matrix	72
5.4.4	SVD analysis for El Nino Southern Oscillation data	72

5.4.5	SVD analysis for the tropical Pacific's precipitation data	79
	Exercises	79
6	Mathematical Modeling by Calculus	83
6.1	Optimal production level of oil	83
6.2	Optimal dimensions of food cans	86
6.3	Exponential decrease of pressure in an isothermal layer of atmosphere	88
6.4	Modeling blackbody radiation	89
6.5	Chemical mixture problems in a natural or chemical engineering process	93
	Exercises	94
7	Probabilistic Models	97
7.1	Two-dice probability problems	97
7.2	Optimal design for product test	98
7.3	Buffon's needle problem	98
7.4	Detecting a round cloud by a radar	103
7.5	Markov chains	103
	7.5.1 Example 1	103
	7.5.2 Example 2: Order of fish tanks	107
7.6	Monte Carlo simulations	111
	7.6.1 Use Monte Carlo simulation for numerical integration	111
	7.6.2 Use Monte Carlo simulation to estimate the volume of an n-ball	113
8	Stochastic Models	115
8.1	A nowhere differentiable but everywhere continuous model	115
8.2	Brownian motion	116
8.3	Ito calculus	117
8.4	Fractal dimension and similarity	119
	8.4.1 References	119
	8.4.2 Dimension of Koch curve	119
	8.4.3 Use R to calculate the fractal dimension	120
8.5	Stochastic differential equations	121
8.6	Solving SDE using R	121
9	Visualize Mathematical Models by R	123
9.1	R graphics examples	123
	9.1.1 Plot two different time series on the same plot	123
	9.1.2 Figure setups: margins, fonts, mathematical symbols, and more	124
	9.1.3 Plot two or more panels on the same figure	127
9.2	Contour color maps	129
	9.2.1 Basic principles for an R contour plot	129
	9.2.2 Plot contour color maps for random values on a map	129

9.2.3	Plot contour maps from climate model data in NetCDF files	131
9.3	Visualize regression models using R	134
9.4	Animation of a free fall based on model	134
9.5	Visualize El Niño models and data	134
9.5.1	A sea level pressure model	134
9.5.2	A surface temperature model	134
9.5.3	A precipitation model	134
10	Statistical Models and Hypothesis Tests	135
10.1	Statistical indices from the global temperature data from 1880 to 2015	136
10.2	Commonly used statistical plots	140
10.2.1	Histogram of a set of data	140
10.2.2	Box plot	140
10.2.3	Scatter plot	141
10.2.4	QQ-plot	144
10.3	Probability distributions	145
10.3.1	What is a probability distribution?	145
10.3.2	Normal distribution	148
10.3.3	Student's t-distribution	150
10.4	Estimate and its error	151
10.4.1	Probability of a sample inside a confidence interval	151
10.4.2	Mean of a large sample size: Approximately normal distribution	153
10.4.3	Mean of a small sample size: t-test	160
10.5	Statistical inference of a linear trend	163
10.6	Free online statistics tutorials	165
References		167
	Exercises	168
11	Concept of big data modeling	169
11.1	Big data: books for layman and techies	169
11.2	A guide to propose and review a big data project	170
11.3	The concept of fitting a model to data	170
11.4	Why over fitting is bad?	170
11.5	Good models and decision-analytic thinking	170
11.6	Big data practice	170
11.7	Data visualization	171
11.8	Term Project #3–The Final Project	171
12	Concepts of machine learning	173
12.1	What is machine learning?	174
12.2	K-means clustering	175

12.3	Logistic regression	177
12.4	CART classification	177
12.5	SVM classification	177
References		179
	Exercises	180
13	Artificial Intelligence Models	181
13.1	Introduction	181
13.2	Searching	181
13.3	First-order logic	181
13.4	Planning	181
13.5	Knowledge representation	181
13.6	Uncertainty quantification	181
14	Network models	183
14.1	Introduction	183
14.2	An example of transportation network	185
14.3	An example of communication network	185
14.4	Matching algorithms	185
14.5	Flow maximization	185
14.6	Shortest path between two nodes in a network	185
14.7	Neural network models and their simulations	185
15	Mathematical and Statistical Consulting	187
15.1	How to conduct the first meeting	187
	15.1.1 When to hold the first meeting	187
	15.1.2 What to present at the first meeting as a consultant	187
	15.1.3 What questions to ask	187
15.2	How to write an SOW	187
15.3	Deliver the consulting results	187
15.4	Maintain the conducts	187

FOREWORD

This is primarily a textbook for the first course of undergraduate mathematical modeling, although many of its R codes and modeling methods can be useful research tools in natural sciences, engineering, social sciences, and even applied mathematics. The book includes basic skills of applied mathematics consulting, such as dimensional analysis for exploring possible relationships among the relevant variables, R programs for time series analysis, map plotting, data visualization, basic probability, statistics, linear algebra, calculus, and 5-step method of mathematical modeling principles.

The book is based on the lecture notes I developed for an upper division course “Math 336: Introduction to Mathematical Modeling” at San Diego State University since 2015. The mathematical prerequisite for this course are Calculus I and the first semester of linear algebra. The book includes the following topics: dimensional analysis, R programming, principles of 5-step mathematical modeling, linear regression models, linear algebra models, probability models, calculus models, stochastic models, statistical inference, big data models, machine learning models, artificial intelligence models, network models, R graphics models, and principles of applied mathematics consulting.

Computer programming experience is not required for reading this book. R programming tutorial is described in book and taught in class from beginning, and is the official computer program language for the course. R and R Studio are free for public download and can be installed easily for either PC or Mac.

Mathematical model is a mathematical expression, often a formula or an equation, that describes a phenomenon, such as the free-fall of an object from a height. The distance between of the object and its initial release position is modeled by $(1/2)gt^2$, where g is the gravitational acceleration and t is the time from the release. Science history implies that Galileo Galilei (1564-1642) was the first who invented this formula. He designed a very

smart experiment for this. At that time, it was hard to observe the free fall time t since a body falls down very fast in the free fall environment. He slowed down the free fall by a free roll of a ball on a plate with ticks (see Fig. 0.1). He placed a wire on the plate so that the ball would make a click sound when the ball rolled over the wire. He adjusted the positions of the four wires so that the ball would make click sound in uniform time intervals. He then discovered that the distance after each click sound is

$$(1/2)at^2 [m] \quad (0.1)$$

where $a = g \sin \theta$ and θ is the angle between the plate and the horizontal plane. The four lines' distances from the releasing points are thus

$$0.5a \times 1^2, \quad 0.5a \times 2^2, \quad 0.5a \times 3^2, \quad 0.5a \times 4^2 [m]. \quad (0.2)$$

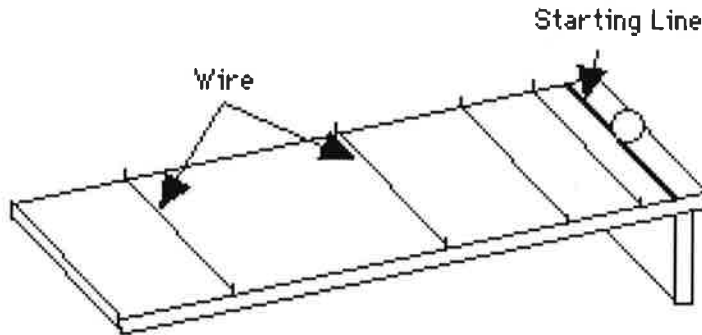


Figure 0.1 Galileo's experiment for a ball falling down on an inclined plate.

The formula $s = (1/2)at^2$ is a mathematical model for the ball rolling down on a plate under gravity. Because of measurement errors, the model is not 100% accurate when compared with the observed data of time and distance. The real world problem is often that when a certain phenomenon is observed, a mathematical model is needed to describe the phenomenon in a quantitative fashion, as accurately as one can. Because observations are necessarily involved in most natural and engineering phenomena, the observational records, called observed data, often used to develop a mathematical model. Linear regression is a commonly used approach to develop a mathematical model. This is an induction approach, deriving a mathematical model using data.

However, some mathematical models can be established from mathematical point of view, whose results are thought to be physically meaningful and to describe the nature. Dimensional analysis is a good approach to develop a mathematical model, such as the problem of an object's free fall. This is a deduction approach, which discover a mathematical model based on mathematical logic and the intrinsic relationships among the variables of the problem. Observational data are still needed to validate the model or to determine one or more critical free parameters of the model.

Both induction and deduction approaches demonstrate the power and beauty of mathematics. This book attempts to show the effectiveness, power, and wide applications of mathematical modeling, using an updated modern approach. The book covers current and future mathematical topics, such as big data, machine learning (ML), networks, artificial

intelligence (AI), mathematical consulting, and R programming and graphics, which are not covered in most of the existing mathematical modeling texts, but these topics are very important in the big data and AI era. They are more frequently encountered by students in their career than calculus or linear algebra alone. The book has another unique characteristics of interdisciplinary approach that uses calculus, linear algebra, statistics, and computing as an integrated tool to solve a practical problem, such as the analysis of spatiotemporal pattern of the El Niño climate phenomenon over the tropical Pacific, rather than treats them as separated and isolated branches of mathematics and statistics. Many existing mathematical modeling books are built on differential equation models, either ordinary differential equation or partial differential equation, and thus involve techniques of solving differential equations, either analytically or numerically. Those books require the background knowledge of Calculus II or III or more advanced mathematics, and are for senior or graduate levels in mathematics physics or engineering. This book is different and does not requires the knowledge of differential equations. It focuses on the current and future needs of mathematical modeling on data and computer programming, such as linear regression, stochastic modeling, and machine learning. This book emphasizes the modeling objectives and results interpretation, although the model development and solutions are also described.

Another feature of this book is to show students how to write short proposals and consulting reports based on mathematical modeling approaches. This helps train students to pursue excellent jobs of mathematical consulting, a career similar to but different from the popular statistical consulting.

By SSPS in San Diego, January 2018