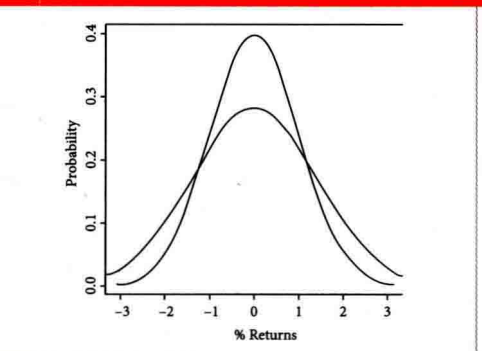
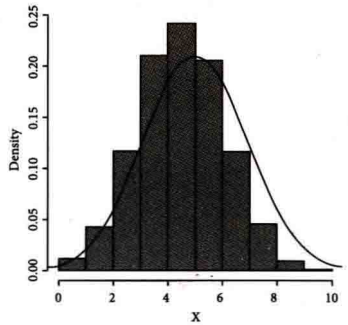
PRINCIPLES OF ECONOMETRICS

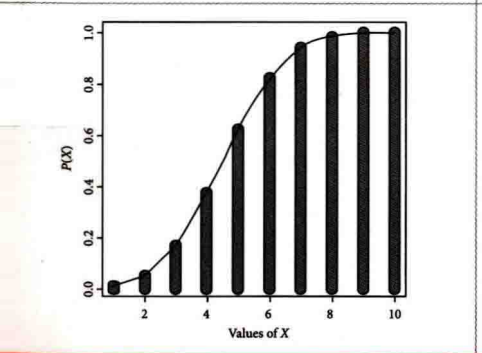
An Introduction (Using R)

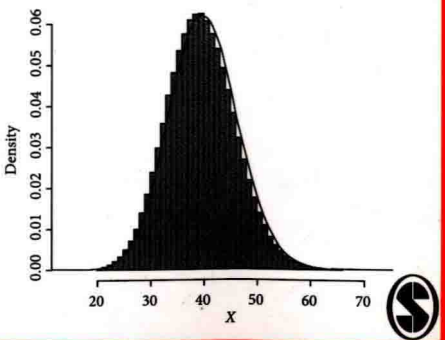
Neeraj R. Hatekar

SAGE TEXTS









Principles of Econometrics

An Introduction (Using R)

NEERAJ R. HATEKAR

SAGE www.sagepublications.com
Los Angeles • London • New Delhi • Singapore • Washington DC

First published in 2010 by

SAGE Publications India Pvt Ltd
B1/I-1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044, India
www.sagepub.in

SAGE Publications Inc
2455 Teller Road
Thousand Oaks, California 91320, USA

SAGE Publications Ltd
1 Oliver's Yard, 55 City Road
London EC1Y 1SP, United Kingdom

SAGE Publications Asia-Pacific Pte Ltd
33 Pekin Street
#02-01 Far East Square
Singapore 048763

Published by Vivek Mehra for SAGE Publications India Pvt Ltd, typeset in 11/13pt Minion Pro by Diligent Typesetter, Delhi and printed at Chaman Enterprises, New Delhi.

Library of Congress Cataloging-in-Publication Data Available

ISBN: 978-81-321-0469-8 (PB)

The SAGE Team: Reema Singhal, Pranab Jyoti Sarma, P.J. Mathew and Umesh Kashyap

Preface

As the name indicates, this text book on the first principles of econometrics is aimed at a student beginning to learn the subject. Let us imagine one such student. She probably has very little or no background in statistics or mathematics. She is keen to learn the subject but is a little bit unsure about what it will hold and whether she will be able to cope with the supposedly 'difficult' subject. I will consider this textbook successful if it can bring for her, some assurance in the initial phase and a genuine passion for the subject as the course unfolds. This textbook emphasizes real-world applications; econometrics is, after all, a way of making sense of the real world. This should enable the student to understand how the subject can help her comprehend the complex reality around her and endow her with a fascination for its methods. This is the main objective of the book.

Today, the teaching and learning of econometrics is inseparable from software. Programme code in R programming language is provided with every chapter. R is a widely popular programming language that is free to use. Students are strongly encouraged to experiment with these programs. If they keep experimenting with the programs, they will surely develop a good understanding of this very flexible and versatile programming language. It will also give the students a substantial professional edge.

The famous Caribbean cricket writer C.L.R. James once famously asked, 'What do they know of cricket who only cricket know?' The same question applies to understanding science. Science is not merely a dispassionate search for truth. It is also a social activity that scientists engage in. The objectives of science, its methods of judging scientific success and failure and its progress are inseparable from

the wider social context within which scientists live and work. In this book, an attempt has been made to provide at least a fleeting introduction to the major dramatis personae. I hope this will provide some context to what otherwise may appear more asocial than it actually is.

Econometrics is a fascinating subject indeed. I have always been deeply fascinated by its method of getting so much information about the world that we inhabit, from apparently dry, dull and boring numbers arrayed in rows and columns with military tedium. My beginning in the subject was rather like the student we have just imagined and my journey has lived up to the promise of the hope that I hold out to her.

Acknowledgements

Many individuals have contributed on the way, though none can be blamed for where I have failed to reach. My first debt is to my Alma Mater, The Department of Economics, University of Mumbai. The jovial, egalitarian and informal atmosphere has ingrained deeply human values among all those who went through its portals. It has created a space for each of us and allowed all to flourish. Teachers like Suhas Joglekar, Medha Deshpande, Abhay Pethe, Ajit Karnik, Ritu Dewan, Romar Correa, late Prof. M.J.M. Rao and colleagues like Avadhoot Nadkarni and Chandrahas Deshpande have shaped me as a student and teacher of econometrics as well as of economics. As a person and a researcher too, I owe much to these rare, warm and gifted individuals. I was particularly lucky to have had Prof. Dilip Nachane as my PhD guide. I cannot think of a way of repaying my debt to him. Friends like K. Sridhar, Geeta Chaddha, Rajesh Save, Mukund Vyas and T.P. Deo in my early student career fanned my passion for econometrics (as well as for several other things equally delightful though somewhat less exalted!). Friends like Biswamohan Pradhan and Rajan Padwal, whom I acquired in my later career keep egging me on, mostly by telling me everyday over innumerable cups of tea that I am the darkest possible blot on the fair face of scholarship.

I would not dare to call my wife Rajni a 'pillar', even in a metaphorical sort of way. She absolutely hates all construction activity. She has generally been a loving, caring wife, as long as she is not exasperated beyond a reasonable limit by my love for large German Shepherds who shed hair copiously on the drawing room sofa or by the occasional harmless tiny snake in my back pack. Our tom cat Mau has brought a fresh twist to life by giving us firsthand experience of living with a feline mafia don. He continues to be true

to his birth and upbringing on the campus of a University by recklessly risking life and limb to satisfy what to lesser humans would be mere idle curiosity.

My greatest debt, however, is to several batches of students, who I have had the privilege to teach in various institutions and at the University of Mumbai as well as those students of the D.G. Ruparel College who sat through a course based on this book during their vacations in an unusually uncomfortable Mumbai summer. If I can call myself a tolerably good teacher, the credit should be entirely due to these long suffering personalities. I would like to record my thanks to each one of them. Some of my students, colleagues and friends have been good enough to give their time to reading various drafts of this book. I have benefited by comments made by Sandhya Krishnan, Krithika Subramanian, Vaidehi Tandel, Nupur Joshi, Ashutosh Sharma, Abodh Kumar, Sumedh Dalwai, Nisha Yadav, Pallavi Rege, Yogendra Sisodia, Wandana Sonalkar, Sneha Thayyil and finally Renita D'Souza as well as Savita Kulkarni, my bubbly pupils and prized project fellows, without whom this book would never have seen the light of the day. I am grateful to all of them, without ascribing to them any responsibility for the errors that might remain.

Contents

List of Tables

List of Figures

Chapter 1

Random Variables

INTRODUCTION

A major objective in econometrics is to be able to make useful statements about parameters that we cannot observe directly. We aim to achieve this by examining data that we can actually observe. For instance, suppose we want to measure the average household income in Mumbai. To find the actual average household income in Mumbai, we will have to go to each household, find its income, add the incomes of all such households and then divide by the number of households. The difficulty of this task will become evident when we account for the fact that there are nearly three million households in Mumbai. We cannot realistically hope to ask each household what their income is. We will have to choose a sample of Mumbai households and infer the average household income for Mumbai on the basis of this sample of households. The number that we will arrive at is not the actual average household income in Mumbai, but an 'estimate' of that income. If we choose a different subset of households as our sample, we will get a different estimate. We can get different estimates of the average income if we choose different samples. We are constrained to make some reasonable statements about the actual average household income in Mumbai, on the basis of such imperfect information. A major part of this book will be concerned with how to make reasonably

good guesses about such unobservable parameters from observable but imperfect bits of information. The concept of random variables is an integral part of the toolkit that we will build in order to achieve this. In this chapter, we will learn about:

1. The idea of a random variable.
2. The probability mass/distribution function of a random variable.
3. The mean and variance of a random variable as well as its skewness and kurtosis.
4. An introduction to some important random variables: Binomial, Poisson and the Normal distribution.

In addition to these statistical ideas, we will also learn about some basic commands in R, which you should try out on your own.

1.1 THE CONCEPT OF A RANDOM VARIABLE

Random variables are central to econometrics. In order to get some idea of what a random variable is, let us take an example. Suppose you and me are playing a game which involves tossing a coin. The coin can land heads up or tails up. A balanced coin is equally likely to land heads up or tails up. Suppose we describe the possible outcomes of our 'experiment' as follows:

$$S = \{\text{Head, Tail}\}$$

The set S is called the 'sample space'. With each of the outcomes, 'head' and 'tail', we can assign a probability. Suppose we assign the probability 1 / 2 to the outcome that the coin lands up heads and 1 / 2 to the outcome that it lands up tails. Then, we can write this as shown in Table 1.1.

Table 1.1
The probability distribution over the outcomes of a toss of an unbiased/balanced coin

Outcome	*Probability of the given outcome*
Head	1 / 2
Tail	1 / 2

A useful way to understand the meaning of the sentence 'probability of a head = 1 / 2' is as follows: Suppose I toss a balanced coin ten times and count the fraction of heads and there were seven heads in this toss, so the fraction of heads = 7 / 10 = 0.7. As against this, let me throw the coin 100 times and then count the fraction of heads. Now, I might expect it to be closer to 0.5, say 0.6. As I keep on tossing the coin more and more times, I should expect the fraction of heads to get closer and closer to 0.5. It is this limiting value of the fraction of heads that we call the 'probability of a head'.

In effect, we have defined the 'probability distribution' of the outcomes of the experiment. Let us reconsider what we have done.

We have specified all the possible outcomes of our 'experiment', and then associated a unique real number between 0 and 1 to each of the outcomes, with the requirement that all such numbers add up to 1. Any such collection of numbers can be regarded as a probability distribution.

For example, suppose our coin is not a balanced coin but the chance that it lands up 'heads' was thrice that of it landing 'tails', then we will have to re-specify our probability distribution as shown in Table 1.2.

Table 1.2
The probability distribution over the outcomes of a toss of a biased/imbalanced coin

Outcome	*Probability of the given outcome*
Head	3 / 4
Tail	1 / 4

Note that the occurrence of a head rules out the occurrence of a tail on the same toss. Such outcomes are called 'mutually exclusive'. The point is that we can assign any of the infinitely many possible numbers to each of our mutually exclusive outcomes, as long as each of the numbers lies between 0 and 1 (both inclusive), each outcome has a unique number associated with it and all such numbers sum to 1.