PEARSON NEW INTERNATIONAL EDITION

Statistics for Business
Decision Making and Analysis

Robert Stine Dean Foster

Second Edition

ALWAYS LEARNING

PEARSON

Pearson New International Edition

Statistics for Business
Decision Making and Analysis
Robert Stine Dean Foster
Second Edition

PEARSON®

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsoned.co.uk

© Pearson Education Limited 2014

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a licence permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

PEARSON®

ISBN 10: 1-292-02339-2
ISBN 13: 978-1-292-02339-7

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

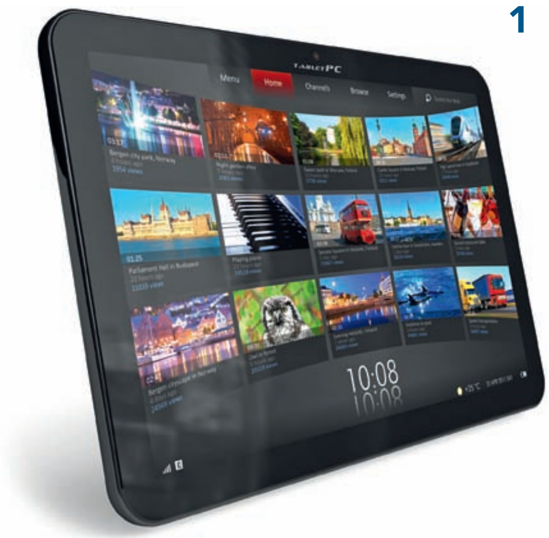
Printed in the United States of America

Table of Contents

Introduction

From Chapter 1 of *Statistics for Business: Decision Making and Analysis*, Second Edition. Robert Stine, Dean Foster.
Copyright © 2014 by Pearson Education, Inc. All rights reserved.

Introduction



1 | WHAT IS STATISTICS?

What do you think statistics is all about?

Statistics answers questions using data, or information about the situation. A **statistic** is a property of data, be it a number such as an average or a graph that displays information. Statistics—the discipline—is the science and art of extracting answers from data. Some of these answers do require putting numbers into formulas, but you can also do every statistical analysis with graphs and tables. It's hard to lie or be fooled when the answer stares you in the face.

Think of statistics as art. An artist must choose the right subject, and a good statistician finds the right picture. Rather than learning to paint, in this course you'll learn how to use statistics to interpret data and answer interesting questions.

Which questions are interesting? The answer is simple: those you care about. Of course, what interests one person may be of no interest to someone else. In this text we apply statistics to a mix of topics, ranging from finance and marketing to personal choices. Most of the questions in this text concern business, but statistics applies more generally. We'll help you appreciate the generality of statistics by solving problems from health and science, too.

statistic A property of data.

Variation

What kinds of questions can be answered with statistics? Let's start with an example. In November 2011, Barnes and Noble debuted its entry into the market for tablet computers called the Nook Tablet™ and joined a growing field of challengers to the market-leading Apple iPad. A question facing Barnes and Noble was, What's the right price for the Nook Tablet?

That's a hard question. To find an answer, you need to know basic economics, particularly the relationship among price, supply, and demand. A little finance and accounting determine the cost of development and production. Then come questions about the customers. Which customers are interested in a tablet computer? How much are they willing to pay? The Nook Tablet is smaller than an iPad with less technology, making it cheaper to produce. It could be sold at a profit for less than the $499 iPad, but how much less? Should it cost more than the competing $199 Kindle Fire?

Suddenly, the initial pricing question branches into several questions, and the answers depend on whom you ask. There's **variation** among customers; customers react differently. One customer might be willing to pay $300 whereas another would pay only $200. Once you recognize these differences among customers, how are you going to set *one* price? Statistics shows how to use your data—what you know about your product and your customers—to set a price that will attract business and earn a profit.

variation Differences among individuals or items; also fluctuations over time.

Here's another interesting question: Why does a shopper choose a particular box of cereal? Modern grocers have become information-rich retailers, tracking every item purchased by each patron. That's why they give out personalized shopping cards; they're paying customers with discounts in return for tracking purchases. Customers keep retailers off balance because they don't buy the same things every time they shop. Did the customer buy that box of cereal because it was conveniently positioned at the end of an aisle, because he or she had a discount coupon, or simply because a six-year-old just saw a commercial while watching *Sponge Bob*? Variation makes the question harder to answer.

Patterns and Models

pattern A systematic, predictable feature in data.

statistical model A breakdown of variation into a predictable pattern and the remaining variation.

Statistics helps you answer questions by providing methods designed to handle variation. These methods filter out the clutter by revealing patterns. A **pattern** in data is a systematic, predictable feature. If customers who receive coupons buy more cereal than customers without coupons, there's a pattern.

Patterns form one part of a **statistical model**. A statistical model describes the variation in data as the combination of a pattern plus a background of remaining, unexplained variation. The pattern in a statistical model describes the variation that we claim to understand. The pattern tells us what we can anticipate in new data and thus goes beyond describing the data we observe. Often, an equation summarizes the pattern in a precise mathematical form. The remaining variation represents the effects of other factors we cannot explain because we lack enough information to do so. For instance, retail sales increase during holiday seasons. Retailers recognize this pattern and prepare by increasing inventories and hiring extra employees. It's impossible, though, for retailers to know exactly which items customers will want and how much they will spend. The pattern does not explain everything.

Good statistical models simplify reality to help us answer questions. Indeed, the word *model* once meant the blueprints, the plans, for a building. Plans answer some questions about the building. How large is the building? Where are the bathrooms? The blueprint isn't the building, but we can learn a lot from this model. A model of an airplane in a wind tunnel provides insights about flight even though it doesn't mimic every detail of flight. Models of data provide answers to questions even though those answers may not be entirely right. A famous statistician, George Box, once said, "All models are wrong, but some are useful."

A simple model that we understand is generally better than a complex model that we do not understand. A challenge in learning statistics is to recognize when a model can be trusted. Models based on physics and engineering often look impressively complex, but don't confuse complexity with being correct. Complex models fail when the science does not mimic reality. For example, NASA used the following elaborate equation to estimate the chance of foam insulation breaking off during take-off and damaging the space shuttle:

$$p = \frac{0.0195(L/d)^{0.45}(d)\rho_F^{0.27}(V - V^*)^{.67}}{S_T^{.25}\rho_T^{.16}}$$

The model represented by this equation failed to anticipate the risk of damage from faulty insulation. Damage from insulation caused the space shuttle *Columbia* to break apart on reentry in 2003.

Models also fail if we mistake random variation for a pattern. People are great at finding patterns. Ancient people looked into the sky and found patterns among the stars. Psychiatrists use the Rorschach ink blot test to probe deep feelings. People even find patterns in clouds, imagining shapes or faces floating in the sky. A key task in statistics is deciding whether the pattern we have discovered is real or something that we've imagined. Finding a pattern allows us to anticipate what is most likely to happen next, to understand the data in order to plan for the future and make better decisions. But if we imagine a pattern when there is none, we become overconfident and make poor decisions.

2 | PREVIEWS

The following two examples preview statistics as theaters preview movies; they show the highlights, with lots of action and explosions, and save the character development for later. These examples introduce a couple of recurring themes and showcase several methods.

Each example begins with a question motivated by a story in the news, and then uses a statistical model to formulate an answer to the question. The first example uses a model to predict the future, and the second uses a model to fill in for an absence of data. These are previews, so we emphasize the results and skip the details.

Predicting Employment

In early November, 2005, national broadcasts announced surprising and disturbing economic results. The big story was not a recession, but rather that the U.S. economy had grown more slowly than expected. The Labor Department reported that only 56,000 jobs had been created in October, 2005, far short of the 100,000 additional jobs expected by Wall Street forecasters.

Financial markets react to surprises. If everyone on Wall Street expects the Labor Department to report large numbers of new jobs, the stock market can tumble if the report announces only modest growth. What should we have expected? What made Wall Street economists expect 100,000 jobs to be created in October? Surely they didn't expect *exactly* 100,000 jobs to be created. Was the modest growth a fluke? These are serious questions. If the shortfall in jobs is the start of a downward trend, it could indicate the start of an economic recession. Businesses need to anticipate where the economy is headed in order to schedule production and supplies.

Was the weather responsible for the modest growth? On August 29, 2005, Hurricane Katrina slammed into Louisiana, devastating the Gulf Coast (see Figure 1). Packing sustained winds of 175 miles per hour, Katrina overwhelmed levees in New Orleans, flooded the city, and wrecked the local economy. Estimates of damages reached $130 billion, the highest ever attributed to a hurricane in the United States, with more than 1,000 deaths. Katrina and the hurricanes that followed during the 2005 season devastated the oil industry concentrated along the Gulf of Mexico and disrupted energy supplies around the country. Did Katrina wipe out the missing jobs?

Let's see if we can build our own forecast. Back in September 2005, how could you forecast employment in October?

We need two things to get started: relevant data and a method for using these data to address the question at hand. Virtually every statistical analysis proceeds in this way. Let's start with data. At a minimum, we need the number employed before October. For example, if the number of jobs had been steady from January through the summer of 2005, our task would be easy; it's easy to forecast something that doesn't change.

FIGURE 1 Hurricane Katrina on August 29, 2005.

The problem is that employment does change. Table 1 shows the number of thousands employed each month since 2003. These are the data behind the story.

TABLE 1 Nonfarm employment in the United States, in thousands on payrolls.

	2003	2004	2005
Jan	130,247	130,372	132,573
Feb	130,125	130,466	132,873
Mar	129,907	130,786	132,995
Apr	129,853	131,123	133,287
May	129,827	131,373	133,413
Jun	129,854	131,479	133,588
Jul	129,857	131,562	133,865
Aug	129,859	131,750	134,013
Sep	129,953	131,880	134,005
Oct	130,076	132,162	134,061
Nov	130,172	132,294	
Dec	130,255	132,449	

Each column gives the monthly counts for a year. The first number in the table represents 130,247,000 jobs on payrolls in January 2003. The following number shows that payrolls in February 2003 fell by 122,000. At the bottom of the third column, we can see that employment increased by 56,000 from September to October 2005, as reported by Reuters. This variation complicates the task of forecasting. We've got to figure out how we expect employment to change next month.

We won't replicate the elaborate models used by Wall Street economists, but we can go a long way toward understanding their models by plotting the data. Plots are among the most important tools of statistics. Once we see the plot, we can decide how to make a forecast.

timeplot A chart of values ordered in time, usually with the values along the *y*-axis and time along the *x*-axis.

The graph in Figure 2 charts employment over time, a common type of display that we'll call a **timeplot**. To keep the vertical axis nicely scaled and avoid showing extraneous digits, we labeled the employment counts in millions rather than thousands.

tip When displaying numerical information, showing fewer digits often produces a better presentation.