openstax™

Introductory Business

Statistics

2e.

Introductory Business Statistics 2e

SENIOR CONTRIBUTING AUTHORS

ALEXANDER HOLMES, THE UNIVERSITY OF OKLAHOMA

BARBARA ILLOWSKY, DE ANZA COLLEGE

SUSAN DEAN, DE ANZA COLLEGE

openstax™

OpenStax
Rice University
6100 Main Street MS-375
Houston, Texas 77005

To learn more about OpenStax, visit https://openstax.org.
Individual print copies and bulk orders can be purchased through our website.

HARDCOVER BOOK ISBN-13	**978-1-711472-59-1**
B&W PAPERBACK BOOK ISBN-13	**978-1-711472-60-7**
DIGITAL VERSION ISBN-13	**978-1-961584-33-4**
ORIGINAL PUBLICATION YEAR	**2023**

1 2 3 4 5 6 7 8 9 10 CJP 23

OpenStax

OpenStax provides free, peer-reviewed, openly licensed textbooks for introductory college and Advanced Placement® courses and low-cost, personalized courseware that helps students learn. A nonprofit ed tech initiative based at Rice University, we're committed to helping students access the tools they need to complete their courses and meet their educational goals.

Rice University

OpenStax is an initiative of Rice University. As a leading research university with a distinctive commitment to undergraduate education, Rice University aspires to path-breaking research, unsurpassed teaching, and contributions to the betterment of our world. It seeks to fulfill this mission by cultivating a diverse community of learning and discovery that produces leaders across the spectrum of human endeavor.

RICE

Philanthropic Support

OpenStax is grateful for the generous philanthropic partners who advance our mission to improve educational access and learning for everyone. To see the impact of our supporter community and our most updated list of partners, please visit openstax.org/foundation.

Arnold Ventures
Chan Zuckerberg Initiative
Chegg, Inc.
Arthur and Carlyse Ciocca Charitable Foundation
Digital Promise
Ann and John Doerr
Bill & Melinda Gates Foundation
Girard Foundation
Google Inc.
The William and Flora Hewlett Foundation
The Hewlett-Packard Company
Intel Inc.
Rusty and John Jaggers
The Calvin K. Kazanjian Economics Foundation
Charles Koch Foundation
Leon Lowenstein Foundation, Inc.
The Maxfield Foundation
Burt and Deedee McMurtry
Michelson 20MM Foundation
National Science Foundation
The Open Society Foundations
Jumee Yhu and David E. Park III
Brian D. Patterson USA-International Foundation
The Bill and Stephanie Sick Fund
Steven L. Smith & Diana T. Go
Stand Together
Robin and Sandy Stuart Foundation
The Stuart Family Foundation
Tammy and Guillermo Treviño
Valhalla Charitable Foundation
White Star Education Foundation
Schmidt Futures
William Marsh Rice University

Contents

Preface

Welcome to *Introductory Business Statistics 2e*, an OpenStax resource. This textbook was written to increase student access to high-quality learning materials, maintaining highest standards of academic rigor at little to no cost.

About OpenStax

OpenStax is a nonprofit based at Rice University, and it's our mission to improve student access to education. Our first openly licensed college textbook was published in 2012, and our library has since scaled to over 25 books for college and AP® courses used by hundreds of thousands of students. OpenStax Tutor, our low-cost personalized learning tool, is being used in college courses throughout the country. Through our partnerships with philanthropic foundations and our alliance with other educational resource organizations, OpenStax is breaking down the most common barriers to learning and empowering students and instructors to succeed.

About OpenStax resources

Customization

Introductory Business Statistics 2e is licensed under a Creative Commons Attribution 4.0 International (CC BY) license, which means that you can distribute, remix, and build upon the content, as long as you provide attribution to OpenStax and its content contributors.

Because our books are openly licensed, you are free to use the entire book or pick and choose the sections that are most relevant to the needs of your course. Feel free to remix the content by assigning your students certain chapters and sections in your syllabus, in the order that you prefer. You can even provide a direct link in your syllabus to the sections in the web view of your book.

Instructors also have the option of creating a customized version of their OpenStax book. The custom version can be made available to students in low-cost print or digital form through their campus bookstore. Visit the Instructor Resources section of your book page on OpenStax.org for more information.

Art Attribution

In *Introductory Business Statistics 2e*, most art contains attribution to its title, creator or rights holder, host platform, and license within the caption. Because the art is openly licensed, anyone may reuse the art as long as they provide the same attribution to its original source.

To maximize readability and content flow, some art does not include attribution in the text. If you reuse illustrations, graphs, or charts from this text that do not have attribution provided, use the following attribution: Copyright Rice University, OpenStax, under CC BY 4.0 license.

Errata

All OpenStax textbooks undergo a rigorous review process. However, like any professional-grade textbook, errors sometimes occur. Since our books are web based, we can make updates periodically when deemed pedagogically necessary. If you have a correction to suggest, submit it through the link on your book page on OpenStax.org. Subject matter experts review all errata suggestions. OpenStax is committed to remaining transparent about all updates, so you will also find a list of past errata changes on your book page on OpenStax.org.

Format

You can access this textbook for free in web view or PDF through OpenStax.org, and for a low cost in print.

About *Introductory Business Statistics 2e*

Introductory Business Statistics 2e is designed to meet the scope and sequence requirements of the one-semester statistics course for business, economics, and related majors. Core statistical concepts and skills have been augmented with practical business examples, scenarios, and exercises. The result is a meaningful understanding of the discipline which will serve students in their business careers and real-world experiences.

Coverage and scope

Introductory Business Statistics 2e began as a customized version of OpenStax *Introductory Statistics* by Barbara Illowsky and Susan Dean. Statistics faculty at The University of Oklahoma have used the business statistics adaptation for several years, and the author has continually refined it based on student success and faculty feedback.

The book is structured in a similar manner to most traditional statistics textbooks. The most significant topical changes occur in the latter chapters on regression analysis. Discrete probability density functions have been reordered to provide a logical progression from simple counting formulas to more complex continuous distributions. Many additional homework assignments have been added, as well as new, more mathematical examples.

Introductory Business Statistics 2e places a significant emphasis on the development and practical application of formulas so that students have a deeper understanding of their interpretation and application of data. To achieve this unique approach, the author included a wealth of additional material and purposely de-emphasized the use of the scientific calculator. Specific changes regarding formula use include:

- Expanded discussions of the combinatorial formulas, factorials, and sigma notation
- Adjustments to explanations of the acceptance/rejection rule for hypothesis testing, as well as a focus on terminology regarding confidence intervals
- Deep reliance on statistical tables for the process of finding probabilities (which would not be required if probabilities relied on scientific calculators)
- Continual and emphasized links to the Central Limit Theorem throughout the book; *Introductory Business Statistics 2e* consistently links each test statistic back to this fundamental theorem in inferential statistics

Another fundamental focus of the book is the link between statistical inference and the scientific method. Business and economics models are fundamentally grounded in assumed relationships of cause and effect. They are developed to both test hypotheses and to predict from such models. This comes from the belief that statistics is the gatekeeper that allows some theories to remain and others to be cast aside for a new perspective of the world around us. This philosophical view is presented in detail throughout and informs the method of presenting the regression model, in particular.

The correlation and regression chapter includes confidence intervals for predictions, alternative mathematical forms to allow for testing categorical variables, and the presentation of the multiple regression model.

Pedagogical features

- **Examples** are placed strategically throughout the text to show students the step-by-step process of interpreting and solving statistical problems. To keep the text relevant for students, the examples are drawn from a broad spectrum of practical topics; these include examples about college life and learning, health and medicine, retail and business, and sports and entertainment.
- **Practice, Homework, and Bringing It Together** give the students problems at various degrees of difficulty while also including real-world scenarios to engage students.
- **Try It** practice problems immediately follow many examples and give students the opportunity to practice as they read the text.

Changes to the Second Edition

The revision of *Introductory Business Statistics* has been undertaken with the support of many faculty adopters and student users. Most of the edits are focused on currency, clarity, accuracy, and belonging. The author made a number of additions in order to provide more support and practice opportunities for students; these include conceptual explanations, introductory text, new Try It exercises, and more detailed example solution steps. Also, dozens of examples and references have been changed in order to align with contemporary contexts and uses. Accuracy checking and errata resolutions resulted in corrections and clarifications, both in the main text and in the answers/solutions. Finally, the second edition includes several hundred edits related to gender, race, ethnicity, age, academic status, and societal issues, designed to present the most informed and inclusive material.

OpenStax only undertakes revisions when pedagogically necessary, and we work to ensure that the text maintains its organization in order to ease transition for instructors. Many of the above changes have been made within problems and examples. However, in nearly all cases, the problems retain their original numbering, so that instructors and online homework providers can move to the new edition with minimal impact on their assignment structures. (This will not be the case for Try It exercises in many chapters, due to the addition of new exercises.) A detailed transition guide is available on the instructor resource page for this textbook.

Answers to Questions in the Book

Answers to Examples are provided directly under the question. Answers to Try it Questions are provided at the Student Resources page. Answers to an assortment of Practice, Homework, and Bringing it Together questions are available to students in the solutions section at the end of each chapter. All other answers to these questions are provided only to instructors in the Instructor Answer Guide via the Instructor Resources page.

Additional resources

Student and instructor resources

We've compiled additional resources for both students and instructors, including Getting Started Guides, an instructor solution manual, and PowerPoint slides. Instructor resources require a verified instructor account, which you can apply for when you log in or create your account on OpenStax.org. Take advantage of these resources to supplement your

OpenStax book.

Community Hubs

OpenStax partners with the Institute for the Study of Knowledge Management in Education (ISKME) to offer Community Hubs on OER Commons – a platform for instructors to share community-created resources that support OpenStax books, free of charge. Through our Community Hubs, instructors can upload their own materials or download resources to use in their own courses, including additional ancillaries, teaching material, multimedia, and relevant course content. We encourage instructors to join the hubs for the subjects most relevant to your teaching and research as an opportunity both to enrich your courses and to engage with other faculty.

To reach the Community Hubs, visit www.oercommons.org/hubs/OpenStax (https://www.oercommons.org/hubs/OpenStax).

Technology partners

As allies in making high-quality learning materials accessible, our technology partners offer optional low-cost tools that are integrated with OpenStax books. To access the technology options for your text, visit your book page on OpenStax.org.

About the authors

Senior contributing authors

Alexander Holmes, The University of Oklahoma

Barbara Illowsky, DeAnza College

Susan Dean, DeAnza College

Contributing authors

Kevin Hadley, Analyst, Federal Reserve Bank of Kansas City

Reviewers

Birgit Aquilonius, West Valley College
Charles Ashbacher, Upper Iowa University - Cedar Rapids
Abraham Biggs, Broward Community College
Daniel Birmajer, Nazareth College
Roberta Bloom, De Anza College
Bryan Blount, Kentucky Wesleyan College
Ernest Bonat, Portland Community College
Sarah Boslaugh, Kennesaw State University
David Bosworth, Hutchinson Community College
Sheri Boyd, Rollins College
George Bratton, University of Central Arkansas
Franny Chan, Mt. San Antonio College
Jing Chang, College of Saint Mary
Laurel Chiappetta, University of Pittsburgh
Lenore Desilets, De Anza College
Matthew Einsohn, Prescott College
Ann Flanigan, Kapiolani Community College
David French, Tidewater Community College
Mo Geraghty, De Anza College
Larry Green, Lake Tahoe Community College
Michael Greenwich, College of Southern Nevada
Inna Grushko, De Anza College
Valier Hauber, De Anza College
Janice Hector, De Anza College
Jim Helmreich, Marist College
Robert Henderson, Stephen F. Austin State University
Mel Jacobsen, Snow College
Mary Jo Kane, De Anza College
John Kagochi, University of Houston - Victoria
Lynette Kenyon, Collin County Community College
Charles Klein, De Anza College

Alexander Kolovos
Sheldon Lee, Viterbo University
Sara Lenhart, Christopher Newport University
Wendy Lightheart, Lane Community College
Vladimir Logvenenko, De Anza College
Jim Lucas, De Anza College
Suman Majumdar, University of Connecticut
Lisa Markus, De Anza College
Miriam Masullo, SUNY Purchase
Diane Mathios, De Anza College
Robert McDevitt, Germanna Community College
John Migliaccio, Fordham University
Mark Mills, Central College
Cindy Moss, Skyline College
Nydia Nelson, St. Petersburg College
Benjamin Ngwudike, Jackson State University
Jonathan Oaks, Macomb Community College
Carol Olmstead, De Anza College
Barbara A. Osyk, The University of Akron
Adam Pennell, Greensboro College
Kathy Plum, De Anza College
Lisa Rosenberg, Elon University
Sudipta Roy, Kankakee Community College
Javier Rueda, De Anza College
Yvonne Sandoval, Pima Community College
Rupinder Sekhon, De Anza College
Travis Short, St. Petersburg College
Frank Snow, De Anza College
Abdulhamid Sukar, Cameron University
Jeffery Taub, Maine Maritime Academy
Mary Teegarden, San Diego Mesa College
John Thomas, College of Lake County
Philip J. Verrecchia, York College of Pennsylvania
Dennis Walsh, Middle Tennessee State University
Cheryl Wartman, University of Prince Edward Island
Carol Weideman, St. Petersburg College
Kyle S. Wells, Dixie State University
Andrew Wiesner, Pennsylvania State University

1 SAMPLING AND DATA

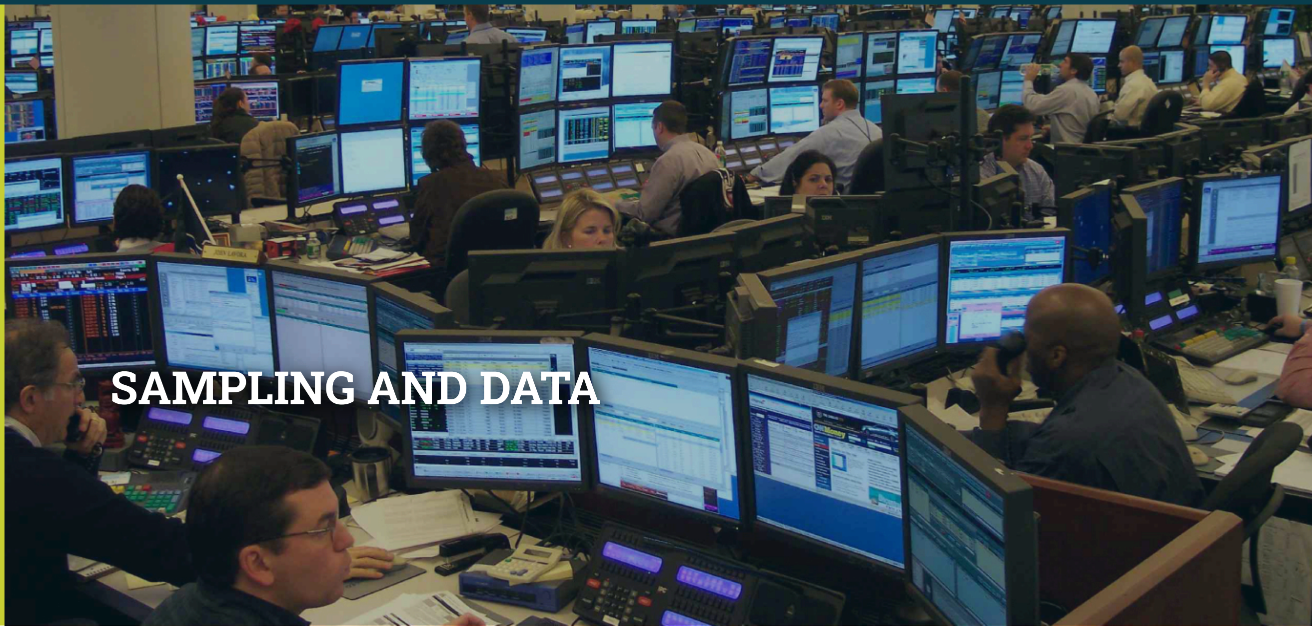

Figure 1.1 We encounter statistics in our daily lives more often than we probably realize and from many different sources, like the news. (credit: modification of work “New office” by Phil Whitehouse/ Flickr, CC BY 2.0)

Introduction

You are probably asking yourself the question, "When and where will I use statistics?" If you read any newspaper, watch television, or use the Internet, you will see statistical information. There are statistics about crime, sports, education, politics, and real estate. Typically, when you read a newspaper article or watch a television news program, you are given sample information. With this information, you may make a decision about the correctness of a statement, claim, or "fact." Statistical methods can help you make the "best educated guess."

Since you will undoubtedly be given statistical information at some point in your life, you need to know some techniques for analyzing the information thoughtfully. Think about buying a house or managing a budget. Think about your chosen profession. The fields of economics, business, psychology, education, biology, law, computer science, police science, and early childhood development require at least one course in statistics.

Included in this chapter are the basic ideas and words of probability and statistics. You will soon understand that statistics and probability work together. You will also learn how data are gathered and what "good" data can be distinguished from "bad."

1.1 Definitions of Statistics, Probability, and Key Terms

The science of **statistics** deals with the collection, analysis, interpretation, and presentation of **data**. We see and use data in our everyday lives.

In this course, you will learn how to organize and summarize data. Organizing and summarizing data is called **descriptive statistics**. Two ways to summarize data are by graphing and by using numbers (for example, finding an average). After you have studied probability and probability distributions, you will use formal methods for drawing conclusions from "good" data. The formal methods are called **inferential statistics**. Statistical inference uses probability to determine how confident we can be that our conclusions are correct.

Effective interpretation of data (inference) is based on good procedures for producing data and thoughtful examination of the data. You will encounter what will seem to be too many mathematical formulas for interpreting data. The goal of statistics is not to perform numerous calculations using the formulas, but to gain an understanding of your data. The calculations can be done using a calculator or a computer. The understanding must come from you. If you can thoroughly grasp the basics of statistics, you can be more confident in the decisions you make in life.

Probability

Probability is a mathematical tool used to study randomness. It deals with the chance (the likelihood) of an event occurring. For example, if you toss a **fair** coin four times, the outcomes may not be two heads and two tails. However, if you toss the same coin 4,000 times, the outcomes will be close to half heads and half tails. The expected theoretical probability of heads in any one toss is $\frac{1}{2}$ or 0.5. Even though the outcomes of a few repetitions are uncertain, there is a regular pattern of outcomes when there are many repetitions. After reading about the English statistician Karl **Pearson** who tossed a coin 24,000 times with a result of 12,012 heads, one of the authors tossed a coin 2,000 times. The results were 996 heads. The fraction $\frac{996}{2000}$ is equal to 0.498 which is very close to 0.5, the expected probability.

The theory of probability began with the study of games of chance such as poker. Predictions take the form of probabilities. To predict the likelihood of an earthquake, of rain, or whether you will get an A in this course, we use probabilities. Doctors use probability to determine the chance of a vaccination causing the disease the vaccination is supposed to prevent. A stockbroker uses probability to determine the rate of return on a client's investments. You might use probability to decide to buy a lottery ticket or not. In your study of statistics, you will use the power of mathematics through probability calculations to analyze and interpret your data.

Key Terms

In statistics, we generally want to study a **population**. You can think of a population as a collection of persons, things, or objects under study. To study the population, we select a **sample**. The idea of **sampling** is to select a portion (or subset) of the larger population and study that portion (the sample) to gain information about the population. Data are the result of sampling from a population.

Because it takes a lot of time and money to examine an entire population, sampling is a very practical technique. If you wished to compute the overall grade point average at your school, it would make sense to select a sample of students who attend the school. The data collected from the sample would be the students' grade point averages. In presidential elections, opinion poll samples of 1,000–2,000 people are taken. The opinion poll is supposed to represent the views of the people in the entire country. Manufacturers of canned carbonated drinks take samples to determine if a 16 ounce can contains 16 ounces of carbonated drink.

From the sample data, we can calculate a statistic. A **statistic** is a number that represents a property of the sample. For example, if we consider one math class to be a sample of the population of all math classes, then the average number of points earned by students in that one math class at the end of the term is an example of a statistic. The statistic is an estimate of a population parameter, in this case the mean. A **parameter** is a numerical characteristic of the whole population that can be estimated by a statistic. Since we considered all math classes to be the population, then the average number of points earned per student over all the math classes is an example of a parameter.

One of the main concerns in the field of statistics is how accurately a statistic estimates a parameter. The accuracy really depends on how well the sample represents the population. The sample must contain the characteristics of the population in order to be a **representative sample**. We are interested in both the sample statistic and the population parameter in inferential statistics. In a later chapter, we will use the sample statistic to test the validity of the established population parameter.

A **variable**, or random variable, usually notated by capital letters such as X and Y, is a characteristic or measurement that can be determined for each member of a population. Variables may be **numerical** or **categorical**. **Numerical variables** take on values with equal units such as weight in pounds and time in hours. **Categorical variables** place the person or thing into a category. If we let X equal the number of points earned by one math student at the end of a term, then X is a numerical variable. If we let Y be a person's party affiliation, then some examples of Y include Republican, Democrat, and Independent. Y is a categorical variable. We could do some math with values of X (calculate the average number of points earned, for example), but it makes no sense to do math with values of Y (calculating an average party affiliation makes no sense).

Data are the actual values of the variable. They may be numbers or they may be words. **Datum** is a single value.

Two words that come up often in statistics are **mean** and **proportion**. If you were to take three exams in your math classes and obtain scores of 86, 75, and 92, you would calculate your mean score by adding the three exam scores and dividing by three (your mean score would be 84.3 to one decimal place). If, in your math class, there are 40 students and 22 are men and 18 are women, then the proportion of men students is $\frac{22}{40}$ and the proportion of women students is $\frac{18}{40}$. Mean and proportion are discussed in more detail in later chapters.

NOTE

The words "**mean**" and "**average**" are often used interchangeably. The substitution of one word for the other is common practice. The technical term is "arithmetic mean," and "average" is technically a center location. However, in practice among non-statisticians, "average" is commonly accepted for "arithmetic mean."

EXAMPLE 1.1

Problem

Determine what the key terms refer to in the following study. We want to know the average (mean) amount of money first year college students spend at ABC College on school supplies that do not include books. We randomly surveyed 100 first year students at the college. Three of those students spent $150, $200, and $225, respectively.

Solution

The **population** is all first year students attending ABC College this term.

The **sample** could be all students enrolled in one section of a beginning statistics course at ABC College (although this sample may not represent the entire population).

The **parameter** is the average (mean) amount of money spent (excluding books) by first year college students at ABC College this term: the population mean.

The **statistic** is the average (mean) amount of money spent (excluding books) by first year college students in the sample.

The **variable** could be the amount of money spent (excluding books) by one first year student. Let X = the amount of money spent (excluding books) by one first year student attending ABC College.

The **data** are the dollar amounts spent by the first year students. Examples of the data are $150, $200, and $225.

TRY IT 1.1

Determine what the key terms refer to in the following study. We want to know the average (mean) amount of money spent on school uniforms each year by families with children at Knoll Academy. We randomly survey 100 families with children in the school. Three of the families spent $65, $75, and $95, respectively.

EXAMPLE 1.2

Problem

Determine what the key terms refer to in the following study.

A study was conducted at a local college to analyze the average cumulative GPA's of students who graduated last year. Fill in the letter of the phrase that best describes each of the items below.

1. Population ____ 2. Statistic ____ 3. Parameter ____ 4. Sample ____ 5. Variable ____ 6. Data ____

a. all students who attended the college last year
b. the cumulative GPA of one student who graduated from the college last year
c. 3.65, 2.80, 1.50, 3.90
d. a group of students who graduated from the college last year, randomly selected
e. the average cumulative GPA of students who graduated from the college last year
f. all students who graduated from the college last year
g. the average cumulative GPA of students in the study who graduated from the college last year

Solution

1. f; 2. g; 3. e; 4. d; 5. b; 6. c

TRY IT 1.2

Determine what key terms refer to in the following study.

A survey of athletes in a university was conducted to study the heights of athletes, in meters. Fill in the letter of the phrase that best describes each of the items below.

1. Population ____
2. Statistics ____
3. Parameter ____
4. Sample ____
5. Variable ____
6. Data _____

a. the average height of athletes in the university
b. the average height of athletes in the survey
c. all athletes in the university
d. all students in the university
e. the height of one athlete
f. a group of athletes randomly selected
g. 1.82, 1.76, 1.69, 1.93

EXAMPLE 1.3

Problem

Determine what the key terms refer to in the following study.

As part of a study designed to test the safety of electric automobiles, the National Transportation Safety Board collected and reviewed data about the effects of an automobile crash on test dummies. Here is the criterion they used:

Speed at which cars crashed	Location of "drive" (i.e. dummies)
35 miles/hour	Front Seat

Table 1.1

Cars with dummies in the front seats were crashed into a wall at a speed of 35 miles per hour. We want to know the proportion of dummies in the driver's seat that would have had head injuries, if they had been actual drivers. We start with a simple random sample of 75 cars.

Solution

The **population** is all cars containing dummies in the front seat.

The **sample** is the 75 cars, selected by a simple random sample.

The **parameter** is the proportion of driver dummies (if they had been real people) who would have suffered head injuries in the population.

The **statistic** is proportion of driver dummies (if they had been real people) who would have suffered head injuries in the sample.

The **variable** X = whether a dummy (if it had been a real person) would have suffered head injuries.

The **data** are either: yes, had head injury, or no, did not.