Chapter 1
Data and Statistics

Case Study: Exploring New Marketing Strategies

Infinity Property and Casualty Corporation (Infinity Auto Insurance) headquartered in Birmingham, Alabama provides personal automobile insurance with a concentration on nonstandard auto insurance. Nonstandard insurance serves individuals unable to obtain coverage through standard insurance companies, which can be due to a driving record with accidents and/or tickets, prior DUI, the driver's age, vehicle type, etc. Infinity Auto Insurance's products include personal automobile insurance for individuals, commercial vehicle insurance for businesses and classic collector insurance for individuals with classic and antique automobiles. Infinity Auto Insurance distributes its products primarily through independent agencies and brokers. Infinity Auto Insurance, a top-performing Infinity brand, provides nonstandard car insurance through more than 12,500 independent agents. Infinity Auto Insurance utilizes Internet-based software applications to provide many of its agents with real-time underwriting, claims and policy information. The Company is licensed to write insurance in all 50 states and the District of Columbia. The company traces its roots back to 1955 doing business as Dixie Insurance company, later known as The Infinity Group. In 1991, Pennsylvania Company (American Premier Underwriters) purchases The Infinity Group. American Premier Underwriters, Inc. and the American Financial Corporation merge and become the American Financial Group (AFG) in 1995.

(continued)

L. Herkenhoff and J. Fogli, *Applied Statistics for Business and Management Using Microsoft Excel*, DOI 10.1007/978-1-4614-8423-3_1,

© Springer Science+Business Media New York 2013

Case Study: (continued)

The company was officially founded in 2002 when AFG transferred all common stock of Infinity and its sister companies to IPCC. In 2003, IPCC was listed on the NASDAQ as a registered public holding.

Food for Thought

Who are some of Infinity Auto Insurance's direct auto insurance competitors?
How large (annual revenue) is the Auto Insurance industry?
Discuss what type of customers Infinity Auto Insurance might target?
If you were a marketing consultant to Infinity Auto Insurance, what market areas would you suggest to target to access the customers identified above?

Possible Answers

Who are some of Infinity Auto Insurance's direct auto insurance competitors?
GEICO, Progressive, Mercury, AIG

How large (annual revenue) is the Auto Insurance industry?
Annual revenue for the Auto Insurance industry ranges in billions of dollars every year.

Discuss what type of customers Infinity Auto Insurance might target?
High risk auto insurance individuals

If you were a marketing consultant to Infinity Auto Insurance, what market areas would you suggest to target to access the customers identified above?
Large metropolitan areas or lower income areas

Key Concepts

Business statistics, Constant, Continuous data, Cross sectional data, Data, Descriptive statistics, Discrete data, Distribution, Frequency, Inferential statistics, Longitudinal data, Population, Probability, Qualitative data, Quantitative data, Random sample, Sample, Scales, Statistics, and Variables.

Discussion

This chapter will include some of the basic concepts and definitions necessary in understanding and applying statistical methods in business. One of the challenges in statistical analysis and subsequent report writing is the confusion over what words and statistical terms really mean. In statistics the mathematical meaning may not match up with the everyday usage of the same term, which can lead to confusion. For example in our daily communications the term "significant" means important i.e. Steve's attendance, as the current CEO, at the award ceremony was very significant to the press. In statistics when a statistic is "significant", it simply means that you are very sure that the statistic is reliable or real and not just due to chance. It doesn't mean the finding is important nor that it has any decision-making utility.

Statistics is often said to be more open to misuse, both deliberately and naively, than any other area in business. Part of that misuse may stem from inconsistency in statistical definitions. The next few pages will attempt to firmly establish the lexicon for moving into the next chapters, but each chapter will also provide content specific definitions. It is in your best interest to take the time to read thought these terms and not assume you have a complete and fully accurate understanding of these key concepts.

Business statistics: Business statistics include the areas of descriptive statistics, probability statistics, and inferential statistics applied to business.

Constant: A value that remains unchanged. For example let **C** be a constant, **C** = number of years of university education required for all entry level accountants in your company and in this case **C** = 4 years.

Cross-sectional data: Cross-sectional data is data that are collected from participants at one point in time (rule of thumb is usually within six or less months). Time is not considered one of the study variables in a cross-sectional research design. In a cross-sectional study, time is assumed to be a random effect that produces only variance, not bias. For example the number of IPOs in the San Francisco area.

Data: Data is specific values of the variables if interest that are collected, analyzed and summarized for presentation and interpretation.

Descriptive statistics: The statistics associated with organizing, summarizing and presenting data in a convenient and informative way. Both graphical and numerical techniques are employed.

Distribution: an arrangement of values of a variable showing their observed or theoretical frequency of occurrence.

Frequency: How often a value occurs.

Inferential statistics: The statistics associated with making an estimate, prediction, or decision from a small group to draw conclusions about a larger group. For example in a random sample we note that 40 % of the voters do not want an increase in business taxes. If the sample has been selected without bias, we can **infer** the population would also vote about 40 % against an increase in business taxes.

Longitudinal data or time series data: Data collected over several time periods. For example the number of IPOs in the San Jose area over the past 5 years. In this example the time period would be defined as 1 year and we are collecting data over 5 of these time periods, or years. Note the term longitudinal has nothing to do with geo-spatial data on the globe.

Population: All the existing individuals, items, or data about which you want to draw a conclusion. For example all of the MBA graduates in the USA, or all the copy machines owned by your company in the past 5 years.

Probability: The analysis associated with the chance of occurrence of one or more possible results of an unpredictable event. When potential students apply to an MBA school, they consider the chance or probability of being accepted.

Qualitative data: Categorical, non-numeric data such as gender, industry type. There is no mathematical relationship between the data values. The data labels can be numeric or non-numeric, and are descriptors of the data. For example the labels may be True/False or could be shown as 1/2. But there is no mathematical relationship between 1 and 2 in this case, they are just labels. In other words 2 is not twice as big as 1.

Quantitative data: Numerical data such as salaries, or sales. There is a mathematical relationship between the data values. The data labels must always be numeric. However quantitative data can be either discrete or continuous. **Continuous** quantitative data arise from a measurement process. Continuous data is information that can be measured on a continuum or scale. Continuous data can have almost any numeric value and can be meaningfully subdivided into finer and finer increments, depending upon the precision of the measurement system. For example **time** may be considered as continuous quantitative data. **Discrete** quantitative data arise from a counting process. An example includes how many text messages you sent this past week.

Random sample: Data selected from a population in a way that ensures each data value has an equal opportunity of being selected.

Sample: A portion of a population selected for analysis.

Scales (measurement): A scale may be defined as any set of items which is progressively arranged according to value or magnitude into which an item can be placed according to its quantification. In other words a scale is a continuous spectrum or series of categories. The purpose of scaling is to represent, usually quantitatively, an item's, a person's or an event's place in the spectrum. The four types of scales are nominal, ordinal, interval and ratio.

- **Nominal scale**
  A scale in which the numbers or letters assigned to objects serve as labels for identification or classification. An example of this scale includes MBA schools.
- **Ordinal scale**
  A scale that arranges objects or alternatives according to their ranking. An example includes the Business Week ratings of MBA schools.

Scale Type	Numerical Operations	Descriptive Statistics
Nominal	Counting	Frequency in each category Percentage in each category Mode
Ordinal	Rank ordering	Median Range Percentile ranking
Interval	Arithmetic operations on intervals between numbers	Mean Standard deviation Variance
Ratio	Arithmetic operations on actual quantities	Geometric mean Coefficient of variation

Fig. 1.1 Descriptive statistics for types of scales

- **Interval scale**
  A scale that not only arranges objects according to their magnitudes, but also distinguishes this ordered arrangement in units or equal intervals. The interval between measurements is a meaningful value. It does not involve a true zero point. Standardized test scores in MBA entrance exams is an example of this scale.
- **Ratio scale**
  A scale having absolute rather than relative quantities and possessing an absolute zero, where there is an absence of a given attribute. The number of years an MBA program has been offered is measured on a ratio scale.

These scales and their associated descriptive statistics are summarized in Fig. 1.1. All statistics that are appropriate for lower-order scales (nominal is the lowest) are appropriate for higher-order scales (ratio is the highest) but not vice versa.

Statistics: Values derived from the data collected from a random sample.

Variables: Characteristics of interest that may change, unlike a constant, within the scope of a given problem. For example the variable $\mathbf{s}$ = executive salary, and the associated values include: $\mathbf{s_1}$ = \$200,000, $\mathbf{s_2}$ = \$580,000, $\mathbf{s_3}$ = \$610,000.

Common Pitfalls

- Be careful of longitudinal data. Make sure the conditions for the data collection process have remained constant. Often environmental factors may change and thus bias the data.
- You cannot convert qualitative data to quantitative data by changing the labels to numeric format.
- Don't show too many decimal places. You cannot make your answer more precise than the data that was used to calculate it.
- Make sure the scales have the correct units of measurement and variable name.

Final Thoughts and Activities

Practice Problems

1. Longitudinal data: Take a look at what has been happening with world record times in the men's mile run since 1900. Extrapolate forward 5 years from today, by simply continuing the same trend of the existing line. Are you surprised?
2. AVERT is an organization that is dedicated to eradicating AIDS in Africa. Explore their website and decide if the statistical data is helpful or is there just too much? Is the data presented in a useful way or is it somewhat misleading? Pay careful attention to the percentage data.
   - ➢ Open www.avert.org
   - ➢ Click on the tab **HIV and AIDS Topics**
   - ➢ Click on **Statistics**
   - ➢ Click on **Africa HIV and AIDS Statistics.** You have two different reports to view on South Africa and Sub-Saharan Africa

Discussion Boards

1. Often political goals can get in the way of accurate reporting of data. Discuss this and any other issues that might be obstacles in getting complete and accurate data on the unrest in Afghanistan.
2. This year's top mutual funds will likely slide in rank next year. Do you agree/disagree? Use statistical arguments to support your position.
3. Statistics can change your life. . . disagree or agree. Provide detailed examples to support your position.
4. Why was *Moneyball* such a popular book and movie?

Group Activity

Productivity Growth, defined as getting more output per hour worked by an employee, has been identified as a key goal in companies within the United States. Often companies provide managers and workers with rewards and special incentives for resulting gains in productivity.

Search the web to find supporting data on one of the positions (pro or con) in addressing the question: Are productivity gains always in the best interest of society?

Parting Thought

It has been proven that the celebration of birthdays is healthy. Statistics show that those people who celebrate the most birthdays are the oldest. . .

Problem Solutions

Problem #1: Longitudinal Data

Answer: (Answers may vary) The world record times have decreased in time since the 1900s. Runners have gotten faster.

The table of data for this problem was retrieved from Wikipedia on March 7, 2012 (http://en.wikipedia.org/wiki/Mile_run_world_record_progression#cite_note-iaaf-4)

Time	Athlete	Nationality	Date	Venue
4:14.4	John Paul Jones	United States	31 May 1913[5]	Allston, Mass.
4:12.6	Norman Taber	United States	16 July 1915[5]	Allston, Mass.
4:10.4	Paavo Nurmi	Finland	23 August 1923[5]	Stockholm
4:09.2	Jules Ladoumègue	France	4 October 1931[5]	Paris
4:07.6	Jack Lovelock	New Zealand	15 July 1933[5]	Princeton, N.J.
4:06.8	Glenn Cunningham	United States	16 June 1934[5]	Princeton, N.J.
4:06.4	Sydney Wooderson	United Kingdom	28 August 1937[5]	Motspur Park
4:06.2	Gunder Hägg	Sweden	1 July 1942[5]	Göteborg
4:06.2	Arne Andersson	Sweden	10 July 1942[5]	Stockholm
4:04.6	Gunder Hägg	Sweden	4 September 1942[5]	Stockholm
4:02.6	Arne Andersson	Sweden	1 July 1943[5]	Göteborg
4:01.6	Arne Andersson	Sweden	18 July 1944[5]	Malmö
4:01.4	Gunder Hägg	Sweden	17 July 1945[5]	Malmö
3:59.4	Roger Bannister	United Kingdom	6 May 1954[5]	Oxford
3:58.0	John Landy	Australia	21 June 1954[5]	Turku
3:57.2	Derek Ibbotson	United Kingdom	19 July 1957[5]	London
3:54.5	Herb Elliott	Australia	6 August 1958[5]	Santry, Dublin
3:54.4	Peter Snell	New Zealand	27 January 1962[5]	Wanganui
3:54.1	Peter Snell	New Zealand	17 November 1964[5]	Auckland
3:53.6	Michel Jazy	France	9 June 1965[5]	Rennes
3:51.3	Jim Ryun	United States	17 July 1966[5]	Berkeley, Cal.
3:51.1	Jim Ryun	United States	23 June 1967[5]	Bakersfield, Cal.
3:51.0	Filbert Bayi	Tanzania	17 May 1975[5]	Kingston
3:49.4	John Walker	New Zealand	12 August 1975[5]	Göteborg
3:49.0	Sebastian Coe	United Kingdom	17 July 1979[5]	Oslo
3:48.8	Steve Ovett	United Kingdom	1 July 1980[5]	Oslo
3:48.53	Sebastian Coe	United Kingdom	19 August 1981[5]	Zürich
3:48.40	Steve Ovett	United Kingdom	26 August 1981[5]	Koblenz
3:47.33	Sebastian Coe	United Kingdom	28 August 1981[5]	Bruxelles
3:46.32	Steve Cram	United Kingdom	27 July 1985[5]	Oslo
3:44.39	Noureddine Morceli	Algeria	5 September 1993[5]	Rieti
3:43.13	Hicham El Guerrouj	Morocco	7 July 1999[5]	Rome

[5] Referenced from "12th IAAF World Championships In Athletics: IAAF Statistics Handbook. Berlin 2009." (PDF). Monte Carlo: IAAF Media & Public Relations Department. 2009. pp. Pages 546, 549–50. Retrieved August 4, 2009.

Excel Output

Let's look at how often the record actually changes and by how much it changes. It seems that the changes as expected are very small. But even so, how fast can humans actually run... there must be some limit. The time between records seems to almost follow a cycle, where several records get set close together and then it takes a while for a new break through time. The last time the record was set was in 1999; based on the pattern, it seems that we are long overdue for a new record and should expect one in the next 5 years, but not by much.

Time	Difference in new record	Date	#years between records
04:14.4		31 May 1913	
04:12.6	00:01.8	16 July 1915	2
04:10.4	00:02.2	23 August 1923	8
04:09.2	00:01.2	4 October 1931	8
04:07.6	00:01.6	15 July 1933	2
04:06.8	00:00.8	16 June 1933	2
04:06.4	00:00.4	28 August 1937	3
04:06.2	00:00.2	1 July 1942	5
04:06.2	00:00.0	10 July 1942	0
04:04.6	00:01.6	4 September 1942	0
04:02.6	00:02.0	1 July 1943	1
04:01.6	00:01.0	18 July 1944	1
04:01.4	00:00.2	17 July 1945	1
03:59.4	00:02.0	6 May 1954	9
03:58.0	00:01.4	21 June 1954	10
03:57.2	00:00.8	19 July 1957	3
03:54.5	00:02.7	6 August 1958	1
03:54.4	00:00.1	27 January 1962	4
03:54.1	00:00.3	17 November 1964	2
03:53.6	00:00.5	9 June 1965	1
03:51.3	00:02.3	17 July 1966	1
03:51.1	00:00.2	23 June 1967	1
03:51.0	00:00.1	17 May 1975	8
03:49.4	00:01.6	12 August 1975	0
03:49.0	00:00.4	17 July 1979	4
03:48.8	00:00.2	1 July 1980	1
03:48.5	00:00.3	19 August 1981	1
03:48.4	00:00.1	26 August 1981	0
03:47.3	00:01.1	28 August 1981	0
03:46.3	00:01.0	27 July 1985	4
03:44.4	00:01.9	5 September 1993	8
03:43.1	00:01.3	7 July 1999	6

Problem #2: AVERT Case Study

Be careful when working with percentages. They can be misleading. It is important to have the actual sample size associated with the percentage. For example 50 % of

100 people is 50 people but 50 % of 2 people is only one person. Also be careful that you are comparing "apples with apples". This organization tends to compare data from countries with different populations and demographics. Also comparing regional data with single country data can be misleading.

The first set of data that is presented in the case study is related to the impact on the health sector in the countries of sub-Saharan Africa. Consider that 40 % of midwives in Zambia were found to be HIV positive. This information, coupled with the decrease in health care workers, may lead you to believe that there is a lack of education for health care workers in how to care for AIDS patients. Perhaps more data on what the African government and AVERT are doing to promote education would be helpful.

The death toll in South African is also a great concern. The average life expectancy in the worst affected sub-Saharan countries has fallen by 20 years. The biggest increase in deaths among 20–49 year old adults accounts for 60 % of all deaths in sub-Saharan Africa. Perhaps an historical snapshot of the life expectancy of both men and women would be statistically helpful in analyzing the historical trend of life expectancy in these regions. Also, the case study refers to the increase in deaths, but does not mention the average (mean) or median age of death. This would provide a clearer picture on the impact that the epidemic is having on the population. Where does that 60 % fall within that age range?

More quantitative data in the area of economics should be provided. Although the case study states that increased ARV coverage by 50 % would reduce the negative effect on the economy by 17 %, AVERT does not mention the specific positive economic outcome as a result of ARV coverage. The study lacks statistical research in this area. As with life expectancy, some historical trend data should be provided as it relates to the impact of AIDS on the economy.

There is some key data that is missing from this case study, which may prevent potential donors from making a final and informed decision on whether to donate funds towards this charity. http://www.avert.org/aids-young-people.htm.

Chapter 2
Introduction to Excel and Basic Charts

Key Concepts

Area charts, Bar and Column charts, Bubble charts, Filtering, Formatting, Line charts, Pie charts, Pivot Tables, Sorting, Radar charts.

Discussion

In this chapter the key concepts are all about commands in Excel. The first step is to load the **Data Analysis Toolpak.** The standard data structure used throughout this book consists of rows of observations and columns of variables. Some of the basic functionality will be introduced to familiarize you in navigating through Excel spreadsheets.

In this chapter the emphasis is on the basic Excel procedures to allow you to successfully analyze and chart you data. These basic procedures will form the framework for future chapters that have more complicated Excel functionalities. The instructions are provided in a detailed step-by-step process so make sure you do not skip steps. One of the most common frustrations is Excel refusing to compute or complete your graph and usually it is because a vital step has been ignored by the user.

There are many versions of Excel for various types of computers under various operating systems. This book has been compiled using Excel 2010 running under the Microsoft® Windows operating system. However the instructions and flow remain very similar regardless of the Excel version you may be using.

These Excel functions and charts will provide a useful companion to your business applications in the workplace or to your business statistics textbook applications in the classroom.

L. Herkenhoff and J. Fogli, *Applied Statistics for Business and Management Using Microsoft Excel*, DOI 10.1007/978-1-4614-8423-3_2,