**Chapter 1**

**Introduction to Statistics and Business Analytics**

**LEARNING OBJECTIVES**

The primary objective of Chapter 1 is to introduce you to the world of

Statistics and analytics, thereby enabling you to:

1.List quantitative and graphical examples of statistics within a business

Context.

2.Define important statistical terms, including population, sample, and

parameter, as they relate to descriptive and inferential statistics

3.Explain the difference between variables, measurement, and data.

4. Compare the four different levels of data: nominal, ordinal, interval, and ratio

5. Define important business analytics terms including big data, business analytics,

data mining, and data visualization.

6. List the four dimensions of big data and explain the differences between them.

7. Compare and contrast the three categories of business analytics.

**CHAPTER TEACHING STRATEGY**

In chapter 1 it is very important to motivate business students to study statistics by presenting them with many applications of statistics in business. The definition of statistics as a science dealing with the collection, analysis, interpretation, and presentation of numerical data is a very good place to start. Statistics is about dealing with data. Data are found in all areas of business. This is a time to have the students brainstorm on the wide variety of places in business where data are measured and gathered especially in this era of big data and business analytics. It is important to define statistics for students because they bring so many preconceptions of the meaning of the term. For this reason, several perceptions of the word statistics is given in the chapter.

Chapter 1 sets up the paradigm of inferential statistics. The student will understand that while there are many useful applications of descriptive statistics in business, the strength of the application of statistics in the field of business is through inferential statistics. From this notion, we will later introduce probability, sampling, confidence intervals, and hypothesis testing. The process involves taking a sample from the population, computing a statistic on the sample data, and making an inference (decision or conclusion) back to the population from which the sample has been drawn.

In chapter 1, levels of data measurement are emphasized. Too many texts present data to the students with no comment or discussion of how the data were gathered or the level of data measurement. In chapter 7, there is a discussion of sampling techniques. However, in this chapter, four levels of data are discussed. It is important for students to understand that the statistician is often given data to analyze without input as to how it was gathered or the type of measurement. It is incumbent upon statisticians and analysts to ascertain the level of measurement that the data represent so that appropriate techniques can be used in analysis. All techniques presented in this text cannot be appropriately applied to analyze all different levels of data.

In addition, chapter 1 introduces the student to the concepts of big data, business analytics, data mining, and data visualization. Throughout the text reference will be made to business analytics, big data, and data visualization. It is a really nice segue to discuss big data in conjunction with levels of data measurement. In this chapter, the student will be exposed to the four characteristics of big data: variety, velocity, veracity, and volume. From this they will perhaps understand that raw data, as part of data mining, needs to be extracted, transformed or cleaned, loaded into databases such that it can be in a form that is usable and useful for business statisticians and analysts.

In this chapter, the student will be introduced to the three dimensions of business analytics. Currently, most everything in the first business statistics course falls under descriptive business analytics including hypothesis testing and correlation. The idea is that business analysts want to understand the data, its characteristics, its applications, and even its relatedness with other data. A second dimension of business analytics is predictive analytics within which fall regression and forecasting techniques. The third dimension is prescriptive analytics which offer businesses the best options among various alternatives give particular circumstances.

Through the introduction to data visualization as part of descriptive analytics, the student may more fully understand the importance of chapter 2 which presents techniques for visualizing data with chapters and graphs.

# CHAPTER OUTLINE

1.1 Basic Statistical Concepts

1.2 Data Measurement

Nominal Level

Ordinal Level

Interval Level

Ratio Level

Comparison of the Four Levels of Data

1.3 Introduction to Business Analytics

Big Data

Business Analytics

Categories of Business Analytics

Descriptive Analytics

Predictive Analytics

Prescriptive Analytics

Data Mining

Data Visualization

Statistical Analysis Using the Computer: Excel, Minitab, and Tableau

**KEY TERMS**

Big Data Ordinal Level Data

Business Analytics Parameter

Census Parametric Statistics

Data Population

Data Mining Predictive Analytics

Data Visualization Prescriptive Analytics

Descriptive Analytics Ratio Level Data

Descriptive Statistics Sample

Inferential Statistics Statistic

Interval Level Data Statistics

Measurement Variable

Metric Data Variety

Nominal Level Data Velocity

Nonmetric Data Veracity

Nonparametric Statistics Volume

**SOLUTIONS TO PROBLEMS IN CHAPTER 1**

1.1 Examples of data in functional areas:

accounting - cost of goods, salary expense, depreciation, utility costs, taxes, equipment inventory, etc.

finance - World bank bond rates, number of failed savings and loans, measured risk of common stocks, stock dividends, foreign exchange rate, liquidity rates for a single-family, etc.

human resources - salaries, size of engineering staff, years experience, age of employees, years of education, etc.

marketing - number of units sold, dollar sales volume, forecast sales, size of sales force, market share, measurement of consumer motivation, measurement of consumer frustration, measurement of brand preference, attitude measurement, measurement of consumer risk, etc.

information systems - CPU time, size of memory, number of work stations, storage capacity, percent of professionals who are connected to a computer network, dollar assets of company computing, number of “hits” on the Internet, time spent on the Internet per day, percentage of people who use the Internet, retail dollars spent in e-commerce, etc.

production - number of production runs per day, weight of a product; assembly time, number of defects per run, temperature in the plant, amount of inventory, turnaround time, etc.

management - measurement of union participation, measurement of employer support, measurement of tendency to control, number of subordinates reporting to a manager, measurement of leadership style, etc.

1.2 Examples of data in business industries:

manufacturing - size of punched hole, number of rejects, amount of inventory, amount of production, number of production workers, etc.

insurance - number of claims per month, average amount of life insurance per family head, life expectancy, cost of repairs for major auto collision, average medical costs incurred for a single female over 45 years of age, etc.

travel - cost of airfare, number of miles traveled for ground transported vacations, number of nights away from home, size of traveling party, amount spent per day on besides lodging, etc.

retailing - inventory turnover ratio, sales volume, size of sales force, number of competitors within 2 miles of retail outlet, area of store, number of sales people, etc.

communications - cost per minute, number of phones per office, miles of cable per customer headquarters, minutes per day of long distance usage, number of operators, time between calls, etc.

computing - age of company hardware, cost of software, number of CAD/CAM stations, age of computer operators, measure to evaluate competing software packages, size of data base, etc.

agriculture - number of farms per county, farm income, number of acres of corn per farm, wholesale price of a gallon of milk, number of livestock, grain storage capacity, etc.

banking - size of deposit, number of failed banks, amount loaned to foreign banks, number of tellers per drive-in facility, average amount of withdrawal from automatic teller machine, federal reserve discount rate, etc.

healthcare - number of patients per physician per day, average cost of hospital stay, average daily census of hospital, time spent waiting to see a physician, patient satisfaction, number of blood tests done per week.

1.3 Descriptive statistics in recorded music industry -

1) RCA total sales of compact discs this week, number of artists under contract to a company at a given time.

2) total dollars spent on advertising last month to promote an album.

3) number of units produced in a day.

4) number of retail outlets selling the company's products.

Inferential statistics in recorded music industry -

1) measure the amount spent per month on recorded music for a few consumers then use that figure to infer the amount for the population.

2) determination of market share for rap music by randomly selecting a sample of 500 purchasers of recorded music.

3) Determination of top ten single records by sampling the number of requests at a few radio stations.

4) Estimation of the average length of a single recording by taking a sample of records and measuring them.

The difference between descriptive and inferential statistics lies mainly in the usage of the data. These descriptive examples all gather data from every item in the population about which the description is being made. For example, RCA measures the sales on all its compact discs for a week and reports the total.

In each of the inferential statistics examples, a sample of the population is taken and the population value is estimated or inferred from the sample. For example, it may be practically impossible to determine the proportion of buyers who prefer rap music. However, a random sample of buyers can be contacted and interviewed for music preference. The results can be inferred to population market share.

1.4 Descriptive statistics in manufacturing batteries to make better decisions -

1) total number of worker hours per plant per week - help management understand labor costs, work allocation, productivity, etc.

2) company sales volume of batteries in a year - help management decide if the product is profitable, how much to advertise in coming year, compare to costs to determine profitability.

3) total amount of sulfuric acid purchased per month for use in battery production. - can be used by management to study wasted inventory, scrap, etc.

Inferential Statistics in manufacturing batteries to make decisions -

1) take a sample of batteries and test them to determine the average shelf life - use the sample average to reach conclusions about all batteries of this type. Management can then make labeling and advertising claims. They can compare these figures to the shelf- life of competing batteries.

2) Take a sample of battery consumers and determine how many batteries they purchase per year. Infer to the entire population - management can use this information to estimate market potential and penetration.

3) Interview a random sample of production workers to determine attitude towards company management - management can use this survey results to ascertain employee morale and to direct efforts towards creating a more positive working environment which, hopefully, results in greater productivity.

1.5 1) Size of sale ($) per customer in men’s formal wear. Either by taking a sample or using a census, management could compute the average sale in men’s formal wear of a weekly period and compare the number to the same average taken a year ago or a month ago to determine if more is being sold per customer. Other variables might include number of sales per hour, number of people entering the department per day, number of dress shirts sold per day, etc.

2) Number of employees working per day. This variable could indicate the day of the week (certain days have more or less sales), sales activity (how sales are doing overall), or even health of associates. Other variables might include percent of employees absent due to illness, average number of hours worked per week per employee, number of open positions, etc.

3) Inventory turnover rate. How fast are items in the store selling? Other variables might include reorder rate, percent of storage space utilized, number of stockouts per week, etc.

4) Number of customers that enter the store per hour. This figure will vary by day, time of day, and season. Compare figures on this variable from period to period can give some indication of sales trends which can help drive human resource planning, etc. Other variables might include amount of time spent per customer in the store per visit, distance that customers travel to shop in the store, number of referrals that customers make to other people annually, etc.

5) Percentage of people paying with cash. Percentage of people using credit cards. These can be used to expedite pay systems, investigate employee theft, calculate surcharges associated with credit cards, etc. Other variables might include average time per checkout, average wait time in pay line, etc.

1.6 1) Size of bill or tab. This variable is the total amount in dollars spent by a patron per visit to the restaurant. The bill or tab could be for an individual or a group and would include both food and beverages if they are all included in the bill. Of course, the measurement would be in dollars. This information could be very useful for the manager or owner to know the average size of a bill both in projecting out total revenues over a period or as a baseline before a marketing effort to increase sales.

2) Percentage of Capacity Filled. This variable could be measured at various intervals, times, and days of the week. The measurement would be calculated by taking the number of patrons in the restaurant at any one time divided by the total number of seats in the restaurant (capacity). From this, management could make staffing decisions for various times and days of the week. In addition, management could make decisions about when to expand, how much to advertise, and/or when to run specials.

3.) Length of Stay. The measurement is how many minutes people are actually in the restaurant from the time they are assigned a table until they are leaving? From this, management could determine customer turnover rates which have capacity implications. That is, how many times in a day is an average table “turned over”. If people stay longer, do they spend more?

4) Number of Arrivals Per 5-minute intervals. The measurement is how many customers arrive at the front door to be greeted by the maître ‘d in any given five-minute period. This figure will likely vary by day of the week, season of the year, and time of day. Management can use this information for staffing decisions and planning.

1.7 a) ratio

b) ratio

c) ordinal

d) nominal

e) ratio

f) ratio

g) nominal

h) ratio

1.8 a) ordinal

b) ratio

c) nominal

d) ratio

e) interval

f) interval

g) nominal

h) ordinal

1.9 a) The population for this study is the 900 electric contractors who purchased Rathburn wire.

b) The sample is the randomly chosen group of thirty-five contractors.

c) The statistic is the average satisfaction score for the sample of thirty-five contractors.

d) The parameter is the average satisfaction score for all 900 electric contractors in the population.