Much of the recent euro currency crisis that struck Europe was attributed to the economy of Greece. In 2008, as the Great Recession hit, Greece began having trouble paying its debts to other countries. By 2009, officials admitted they had been understating Greece's debt for years. The result was that lending to Greece was quickly cut off, sending Greece to a state of near-bankruptcy by 2010. George Papandreou, Greece's prime minister, claimed that corruption was at the heart of the crisis. 

Transparency International (TI) is a global civil organization that reports on worldwide corruption. Each year since 1995, TI has published the “Corruption Perception Index,” rating the government corruption of 177 countries on a scale of zero (highly corrupt) to 100 (very clean) through interviews with expert institutions (i.e., these are not the opinions of citizens, though early surveys were based on public opinion). Greece's CPI in 2013 was 40, making it 80th on the list (by 2016, it had improved its score to 46, moving it the 56th on the list). The organization's research indicated that the average bribe to a public service in Greece in 2013 was $1,596. To validate these findings, a random sample of 20 private sector transactions in 2013 was taken and each person was asked to report the amount of any bribe they were offered. The data can also be found in the Excel file titled Bribes.xlsx.

**Chapter 1 Questions**

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For the Bribes data
   
   (a) Construct a frequency distribution and a histogram in Excel.

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.
Citibank 1 In April of 2016, Citibank, which has a major credit card division, rolled out its new marketing campaign, called “Make It Here” (named in honor of the Frank Sinatra song, *New York, New York*) in select markets. The campaign hasn’t been cheap. Citibank’s president of US retail banking said, “What we want is the client who is forward-thinking, aspirational and driven—the ‘Make It Here’ campaign really resonates with those people.” If Citibank can get customers to invest in their dreams and themselves, the hope is that (among other things) its customers will increase their current credit card usage.

To test the effectiveness of this promotion, Citibank instituted the campaign with a stratified group of customers intended to represent a wide range of consumer types. Citibank monitored the monthly credit card balances of these customers before and after the promotion was launched. Citibank hired a French public relations company called Publicis to handle their global media. If the “Make It Here” campaign goes well in its test markets, Citibank will roll out the campaign globally.

The credit card balances of a sample of 25 customers before and after the promotion can be found in the Excel file *Citibank_1.xlsx*.

### Chapter 1 Questions

1. Get to know your data
   
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   
   (a) What are the population(s) and sample(s)?

   (b) What are the parameter(s) and statistic(s)?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
1. For the **Before** column
   
   (a) Construct a frequency distribution and a histogram in Excel.  
   
   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

2. Do you think that the Empirical rule can apply to the data in Before column (credit card balances before promotion)? If so, identify the range within which 95% of the data will fall using Empirical rule. Then check what is the percentage of data contained in that range.

3. Compare the statistics from part 1(c) of the **Before** column with those of the **After** column. What do you notice?
Citibank, which has a major credit card division, has hired an advertising agency to design a promotion to encourage its customers to increase their current credit card usage. To test the effectiveness of this promotion, Citibank instituted the campaign with a small stratified group of customers intended to represent a wide range of consumer types. Citibank monitored the monthly credit card balances of a total of 50 customers, 25 of whom were informed of the promotion and 25 of whom were not aware of it. Assume the population variance for the credit card balances are equal. If the campaign is found to be effective, Citibank will roll it out in other cities. If it cannot be shown effective, Citibank will abandon the promotion and look to a rival advertising agency to improve their chances.

Pretend you work in data analytics for Citibank. They ask you to recommend whether to continue with the current campaign, or switch advertising agencies. The results of the experiment can be found in the Excel file Citibank_2.xlsx.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
1. For the **Aware of Promotion** data.
   (a) Construct a frequency distribution and a histogram in Excel. 
   (b) Describe the shape of the distribution.
   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

2. Compare the statistics you calculated in part 1(c) for the **Aware of Promotion** column with the **Unaware of Promotion** column. What do you notice?
Contributions A recent report by nonprofit fundraising consultant Blackbaud and its partners revealed differences in the charitable habits of four generations of Americans: Baby Boomers, Generation X, Millennials, and those born before 1946. Among other things, they found that Millennials are nearly eight times more likely to have donated through crowdfunding sites like DonorsChoose.org and Kiva.org than have Baby Boomers.

Kiva Microfunds is a non-profit organization matching lenders with borrowers around the world. Students and entrepreneurs from more than 80 countries each post their story on Kiva.org in hopes that others will fund their need, which averages $411.26. Lenders contribute as little as $25 at 0% interest, and when enough lenders contribute to fund a loan, Kiva gives the money to a local microfinancing institution, which uses Kiva lenders’ contributions as capital on the loan. In other words, Kiva’s contributors don’t lend money directly to the borrower; rather, they supply the capital required by the borrower’s local lender to make them the loan. According to Kiva.org, the average contribution for an individual lender on a single loan is $116.74, with most lenders contributing funds on two loans.

The data from Kiva.org show the current contributions for a random sample of 18 Millennials. These data can be found in the Excel file titled Contributions.xlsx.

Chapter 1 Questions

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For the **Contributions** data
   
   (a) Construct a frequency distribution and a histogram in Excel. 
   
   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.
Credit Cards  THE FEDERAL RESERVE SAYS THAT U.S. HOUSEHOLDS HAVE ACCUMULATED CLOSE TO $14 TRILLION IN OUTSTANDING DEBT. That’s enough to buy everything produced in the European Union this year on credit. To get a share of the revenue generated by this borrowing, universities, museums, and zoos ally with major lenders to sponsor an affinity credit card. The sponsor and lender, such as Chase or Bank of America, want to be sure it will be profitable. It costs money to entice customers to sign up, process the application, and manage the account.

Two characteristics determine the initial success of these credit cards. First, a substantial proportion of those getting an offer have to accept. Second, those who accept the offer have to use the card. The big profits lie in the interest earned on outstanding balances. Unless enough customers accept a card and carry a balance, the offering will not generate the profits that justify the start-up costs. How many customers will accept the offer? How large a balance will they carry? Until it answers these questions, a bank cannot make an informed decision on whether to launch a new credit card.

In this case study, we are going to evaluate the launch of a new affinity credit card. The contemplated launch proposes sending pre-approved applications to 100,000 alumni of a large university. Promotional costs would be $300,000 for advertising plus $5 to send each of the 100,000 mailed offers, a total of $800,000. For those who accept the offer, the bank spends another $50 to set up an account. For profits, the bank earns 10% of the average balance. The proportion who will return the application and the average monthly balance that those who accept the card will carry determine whether the card will be profitable.

The credit card issuer sent preapproved applications to a sample of 1,000 alumni. Of these, 140 accepted the offer (14%) and received a card. Now we are interested in the average monthly balance the customers carry. Data is located in the file Credit_cards.xlsx.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For the Profit column
   (a) Construct a frequency distribution and a histogram in Excel. 
      
      TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel. 
      
      TA initial:

2. For the Accept Offer column,
   (a) Construct a bar chart or a pie chart comparing the proportion of those who accepted the offer versus those who didn’t. 
      
      TA initial:

   (b) Describe your finding.
Decey According to the National Retail Federation, inventory shrinkage cost American businesses $44 billion in sales in 2014 and $34 billion in 2015. Inventory shrinkage includes loss of inventory due to shoplifting (estimated to be 38% of total shrinkage), employee theft (34.5%), vendor theft (6.8%), and administrative errors (16.5%). A Pareto analysis quickly reveals that deterring theft would be the most impactful way of lowering shrinkage.

A retail chain is considering installing devices that resemble cameras to deter shoplifting. The devices only look like cameras, saving the expense of wiring and recording video. To test the benefit of this decoy system (i.e. it decreases the amount loss to theft on average), it picked 40 stores, with half to get the decoy and the other half to serve as comparison group (control group). Stores were matched based on typical levels of sales, local market size, and demographics. The comparison lasted for 3 months during the summer. At the end of the period, the retailer used its inventory system to compute the amounts lost to theft in the stores. The results can be found in the Excel file Decoy.xlsx.

Chapter 1 Questions

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population(s) and sample(s)? Specify if you think there are more than one populations or samples.

   (b) What are the parameter(s) and statistic(s)?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data \textbf{Lost to Theft (Decoy)},

(a) Construct a frequency distribution and a histogram in Excel. TA initial:

(b) Describe the shape of the distribution.

(c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel. TA initial:

2. Compare the statistics in part 1. (c) you calculated from data \textbf{Lost to Theft (Decoy)} with those from data \textbf{Lost to Theft (control)}. Describe your findings.
Due to a large grant from Eli Lilly, Indiana Workforce Development has launched a multilevel attempt to increase the number of potential job candidates for many factories in Region 9. One local factory hiring people for tasks on its assembly line gives applicants a test of manual dexterity, in addition to general problem-solving and math skills. This particular test counts how many oddly shaped parts the applicant can install on a model engine in a one-minute period. The results can be found in the Excel file Dexterity.xlsx. Assume that these tested applicants are simple random samples of men and women who represent all applicants for these jobs.

Pretend you work for the HR department of this factory. You know your department requires that your company can show nondiscrimination in case a lawsuit is brought, but it seems the men do not perform as well as the women in the dexterity test (men install less parts than women on average). Your boss, therefore, has asked you to examine the results of this latest pool of applicants and determine whether managers were justified in their decision to hire more women than men.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data
   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population(s) and sample(s)?
   (b) What are the parameter(s) and statistic(s)?
   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data **Parts installed by men**
   
   (a) Construct a frequency distribution and a histogram in Excel. TA initial:  

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel. TA initial:

2. Compare the statistics you calculated in 1. (c) above for data **Parts installed by men** with those for data **Parts installed by women**. Describe your findings.
Direct Mail Advertising  Proponents of direct mail marketing report that over 40% of local retail advertising is still conducted via direct mail. The advertising seems to work best for young adults, and the majority of potential customers consider print marketing to be the most trustworthy type of marketing.

As such, Performance Tires plans to engage in direct mail advertising. It is currently in negotiations to purchase a mailing list of the names of people who bought sports cars within the last three years. The owner of the mailing list claims that sales generated by contacting names on the list will be more than paying for the cost of using the list. (Typically, a company will not sell its list of contacts, but rather provides the mailing services. For example, the owner of the list would handle addressing and mailing catalogs.)

Before it is willing to pay the asking price of $3 per name, the company obtains a sample of 225 names and addresses from the list in order to run a small experiment. It sends a promotional mailing to each of these customers. The data for this exercise show the gross dollar value of the orders produced by this experimental mailing. The company makes a profit of 20% of the gross dollar value of a sale. For example, an order for $100 produces $20 in profit. Should the company agree to the asking price? Data is located in the file Direct_mail.xlsx. (Hint: Think about when you can make a profit from purchasing a mailing list. Compare the average profit you can make from each customer and the price you need per for each name.)

Chapter 1 Questions

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistic?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data **Order Cost**
   
   (a) Construct a frequency distribution and a histogram in Excel. TA initial: 
   
   (b) Describe the shape of the distribution.
   
   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel for the profit the company made from the 225 customers. TA initial:
It’s no secret that Tiger Woods changed the world of golfing, at least not to any of us that were around to see him burst onto the Pro circuit in 1996. For one thing, he increased the amount of money golfers can earn. With his American Express and Nike endorsements, Tiger broke the mold for how much players could earn from equipment deals. Tiger also incited young golfers to work harder and non-golfers to take up the sport. Across the world, the sport gained popularity, thrusting the long-ignored European Tour into notoriety. After he came on the scene, the number of new golf courses built in the US per year more than doubled, from 140 to 307. The influx of new golfers and more serious golfers increased the demand for better equipment. As clubs got better, scores dropped (a good thing). However, the top driving distance has remained relatively flat for the past few years, indicating that the gains players are seeing on the course from better equipment (and better skill) are reaching their peak.

A golf equipment manufacturer took note of this and saw the opportunity to capitalize on an area that still may show room for improvement – golf balls. They developed a new proprietary core element for their golf balls and would like to convince members of a club that its golf balls travel farther than those of a competitor, even for weekend golfers. For the comparison, twelve golfers randomly selected from the club to each drive a ball from this manufacturer and from the rival. The results show the distance traveled, in yards. The results can be found in the Excel file Distance.xlsx.

### Chapter 1 Questions

1. **Get to know your data**
   
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. **Get to know your scientific question**

   (a) What are the population(s) and sample(s)?

   (b) What are the parameter and statistic?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For Data $\text{distance(manuf)}$

   (a) Construct a frequency distribution and a histogram in Excel. 

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

2. Compare the statistics from part 1(c) for the data $\text{Distance (manuf)}$ with those for the data $\text{Distance (rival)}$. What do you notice?
Domestic Airfares  Any firm that wants to compete in today’s market needs to meet the basic requirements of the customers – these are called “order qualifiers” – and stand out in one area, known as their “order winner.” Companies don’t have to be the best at everything; they just need to offer customers some sort of benefit over its competitors (e.g., I frequently buy umbrellas at the IMU on rainy days. I don’t mind that they aren’t the cheapest, because they offer me time utility – having what your customer needs WHEN they need it – and place utility – having what your customer needs WHERE they need it). The airline industry is no different. Every airline needs to cover the basics of safety and timeliness, but they compete on things like fares and amenities. 

The relatively young Spirit Airlines (NASDAQ: SAVE) is competing with price utility as their order winner. With their “Bare Fare,” flyers get only a cramped flight with no snacks and no bags. Everything from water to printing their ticket at the airport (rather than on their own printer at home) will cost their customers extra money. With newer, more fuel efficient planes and all the pennies they save giving stripped down service, they can offer ultra low fares – up to 40% lower than their competitors. But if you have price as your order winner, you have to make sure you’re winning on price!

In November of 2016, Spirit will add more flights to DC/Baltimore. The company needs to know what their competitors are charging on average so they can make sure to beat them by a comfortable margin. The data show the airfares for 24 randomly selected tickets for travel within the United States to BWI Marshall Airport. These data can be found in the Excel file titled Domestic_airfares.xlsx.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population(s) and sample(s)?

   (b) What are the parameter(s) and statistic(s)?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Domestic_airfares
   (a) Construct a frequency distribution and a histogram in Excel. TA initial:

(b) Describe the shape of the distribution.

(c) Compute the sample mean (\( \bar{x} \)), sample variance (\( s^2 \)) and sample standard deviation (\( s \)) in Excel. TA initial:

2. Can Empirical rule apply to the data set? And why?
Golf Scores  If you think golf is a boring game, maybe speed golf is more your style. Although the sport has been around for decades, it has recently become somewhat of a craze. The basic idea is to play a round of golf as fast as you can, without benefit of a cart to drive you from one hole to the next. Instead, players run and hit as quickly as they can. The final score is the combined total of number of strokes plus number of minutes required to complete the course. At the elite level, scores range from the mid-70s to the upper 90s, with times ranging from about 40 minutes to an hour.

Two friends, Brian and John, frequently practice together for the upcoming Speedgolf 2016 World Championships in Illinois. Brian always beats John on time, but John claims Brian’s scores suffer as a result. Brian disagrees.

The Excel file Golf_scores.xlsx contains a random sample of golf scores from the two friends. Assume the population variances for each golfer’s scores are approximately equal.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population(s) and sample(s)?

   (b) What are the parameter(s) and statistic(s)?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data John
   (a) Construct a frequency distribution and a histogram in Excel.

   TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

   TA initial:

2. John claims that Brain’s score suffer as a result, is that true? Describe your finding.
Hawaii Ocean Temperatures  The National Oceanic and Atmospheric Association (NOAA) (https://www.nodc.noaa.gov) is a great resource for secondary data on global climate, including ocean water temperatures around the world. Many snorkelers head to Hawaii hoping to spend some time swimming in the beautiful coral reefs. However, rising water temperatures threaten these rock-like animals. Coral and algae have a symbiotic relationship, but coral expels the algae in warm water. This starves the coral, eventually turning it white. This “coral bleaching” leads to higher mortality rates for the coral; not so great for tourism – or the earth!

A group of environmentalist entrepreneurs are interested in predicting whether coral bleaching will threaten the coast this coming summer. If so, they have plans to sell the State their services in enacting new management techniques for preventing the bleaching that have been suggested by the Pacific Regional Environment Programme (SPREP). In the past, the average water temperature was reported to be 77° Fahrenheit in the winter and 82° F in the summer. Coral bleaching occurs when temperatures rise above 83° F (28.3° Celcius)

The entrepreneurs gathered data from November – February (125 days’ worth). If the average winter water temperatures are warmer than usual, it follows that the summer temperatures may be expected to be warmer, also. The file Hawaii_ocean_temps.xlsx lists daily ocean temperatures for November through the end of February.

Chapter 1 Questions

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistic?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data in excel file **Hawaii_ocean_temps.xlsx**
   
   (a) Construct a frequency distribution and a histogram in Excel.  
   TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.  
   TA initial:

2. If there is a time series dataset, make one timeplot/line chart and describe the trend.  
   TA initial:
Holiday Sales Many retailers count on strong fourth quarter sales to hit their yearly targets. Hence, retailers pay close attention to holiday sales during the Thanksgiving weekend each year, as it serves as a gauge for consumer demand over the remainder of the holiday season. In addition to garnering attention from retailers, the federal government also has an interest in these sales as a measure of economic health and consumer confidence. To monitor these sales, a random sample of shoppers was selected in 2012 and again in 2013 to study whether the average amounts spent on holiday shopping during this weekend differ between the two years. Because of the widespread growth and competition in online shopping, most retailers now offer their products through websites as well as in traditional “brick and mortar” stores. Therefore, the amount of money each shopper spent in both stores and on Web sites was recorded. These data can be found in the Excel file Holiday_sales.xlsx. Past studies show that the population standard deviation for the amount spent during this weekend has not changed much, and is about $170.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data
   
   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population(s) and sample(s)?

   (b) What are the parameter(s) and statistic(s)?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Y2012 (sales in year 2012) in excel file *Holiday_sales.xlsx*

   (a) Construct a frequency distribution and a histogram in Excel. 

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

2. Compare the statistics from part 1(c) for the data in Y2012 column with those of the Y2013 column. What do you notice?
IBM Most people know IBM for its PC business, but IBM’s history began long before the invention of personal computers, and it has continued to be a strong Fortune 500 company since selling its PC business to Lenovo in 2005. IBM began in the 1880s as a merger of three companies, and its early sales were for punch card readers to customers like the railway. Today, IBM’s mission is to be the world leader in high value solutions, and they’re involved in finding solutions for everything from food-born illness outbreaks to the Zika virus. In July of 2016, IBM partnered with ENGIE to help improve the management of cities by looking at patterns, event correlation, anomaly detection and real-time data across all parts of cities. “Whether it’s a traffic jam, a building fire, or a large sporting event, city officials need a reliable solution to adapt to constant changes and challenges.”

However, IBM’s stock performance is not as strong as one might expect from a technology company, which are typically “growth” stocks (meaning the investor takes on extra risk in the hopes of higher-than-market returns).

The year is 2012. A trader is considering “going short” on IBM stock and is looking at data from the past 22 years (from 1990 to 2011). A “short” position involves selling stock you don’t own, which you must then purchase at a later date, hopefully for a lower price. Her plan is to short IBM stock and use the proceeds to invest in the S&P500 index. The average (S&P500) market return over this period is 10%, with an effective monthly rate of 0.66% (that is, 0.0066). Does stock in IBM return less on average than the S&P500? Data is located in the file IBM.xlsx.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population and sample?

   (b) What are the parameter and statistic?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
1. For data Return IBM in excel file IBM.xlsx
   (a) Construct a frequency distribution and a histogram in Excel.
   (b) Describe the shape of the distribution.
   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

2. Do you think that the Empirical rule can apply to the data in Before column (credit card balances before promotion)? If so, identify the range within which 95% of the data will fall using Empirical rule. Then check what is the percentage of data contained in that range.

3. If there is a time series dataset, make one timeplot/line chart and describe the trend.
Millennials (those born 1980–2000) are the largest generation ever, even bigger than the Baby Boomers (born 1946–1964). This means they’re an important focus for marketers. But strategies for marketing to them have to be different, as Millennials tend to have more debt and less money thanks to lower employment levels and higher educational costs. Now Millennials are at an age where they are buying homes for the first time. Lowe’s home improvement chain knows this is the time to reach out to these new homeowners. Marci Grebstein was promoted to Chief Marketing Officer (CMO) in November of 2015, and by March of 2016 she was rolling out a new marketing campaign aimed at Millennials. The campaign, called “Make Your Home Happy,” was developed by BBDO, the world’s most awarded and second largest advertising agency. Using a mix of humor, unicorns, and pink flamingos, BBDO’s executive creative director says the marketing “allows us to inject some humor and personality that resonates with our emerging customers, without alienating our current customers.”

In 2015, the average receipt at Lowe’s was for $67.26. The Excel file labeled Lowes.xlsx lists the receipt total for 350 randomly selected customers for the home improvement chain from 2016. Are Marci and BBDO producing results?

### Chapter 1 Questions

1. Get to know your data
   
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Receipts from the sample

   (a) Construct a frequency distribution and a histogram in Excel. 

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.
Nigeria  Recently, cell phone companies have been heavily investing in the African continent. With only half of its population owning a phone, Nigeria is considered a market that has much potential for growth. However, recent competition has driven prices for cell phones down in the country. It has been reported that the average monthly revenue earned per customer in Nigeria has been falling steadily since 2006 and is currently at $11. In addition, government regulations and penalties are very harsh in Nigeria, with even Nigeria’s top wireless provider (MTN Group, Ltd.) being fined so heavily critics say the loss is “way, way higher than the profits they’re going to make from Nigeria for many years to come.” Many believe the government is unfairly targeting and prosecuting firms in order to cover budget shortfalls.

Vodafone, a U.K. cell phone company, recently entered the Nigerian market through a partnership with Globacom. The joint venture company, called Vodacom, serves private citizens and businesses. The business sales are doing very well, offering VoIP, machine-to-machine connectivity, and enterprise messaging and collaboration. The private citizen sales, however, have been weak, and the potential for high fines for minor regulatory infractions makes staying in this market even less attractive to Vodacom.

Because of its investments in infrastructure and long contracts with suppliers, Vodacom is hesitant to leave the market without strong evidence that the benefits of staying in the market are outweighed by the risks. Vodacom has hired a consulting firm to help make the decision. The senior executives are not willing to share complete sales data with business consultants, but they are willing to provide a sample, given that the sample size needed is small. Unless there is statistical evidence that the average monthly revenue is less than $11, Vodacom will choose to remain in this market; otherwise, the company will shift all of their efforts to business services. The dataset **Nigeria.xlsx** contains the monthly revenue from a random sample of 65 Nigerian mobile phone customers. Based on this sample, what recommendations will you provide Vodacom?

**Chapter 1 Questions**

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?
(b) What are the parameter and statistics?

(c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?

Chapter 2-3 Questions

1. For data **Average monthly revenue of private customers**

   (a) Construct a frequency distribution and a histogram in Excel.  

   TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean (\(\bar{x}\)), sample variance (\(s^2\)) and sample standard deviation (\(s\)) in Excel.

   TA initial:
Despite record-high donations of $114.9 billion in 2014, churches are still feeling the burden of a sluggish economy. According to The Nonprofit Almanac 2012, 45% of households donate an average of $1700 annually to religious organizations. With approximately 350,000 religious congregations in the US, this comes out to about $320k annually for each of them. Of course, some bring in a lot more than that, and some a lot less. And the donations brought in need to pay for mortgages on the property, taxes, maintenance, programming, and all staff salaries. The problem is, the record-high total donations don’t translate into record-high individual donations. In fact, as the number of churchgoers has declined, so has the average donation made by those still attending.

A *Wall Street Journal* article discussed the impact that the recent economic downturn has had on the job market for clergy. Many churches have reduced or frozen the salaries and benefits for their clergy as donations have declined. The article reported that the average salary for an associate pastor with 10 years of experience is $64,000. Suppose that to investigate whether this average has changed recently, the National Association of Church Business Administration conducted a salary survey of pastors. The results can be found in the Excel file *Pastor.xlsx*.

### Chapter 1 Questions

1. Get to know your data
   
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Pastor
   
   (a) Construct a frequency distribution and a histogram in Excel.

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.
United Parcel Service (UPS) was founded by two teenagers in 1907 with a loan of $100. For nearly 100 years, their biggest competitor was the United States Postal Service (USPS), until 1998 when FedEx, founded in 1971 as a cargo airline, expanded into ground delivery. Because both FedEx and UPS offer business-friendly services, such as online scheduling of package pickup, easy online tracking, and reliability of scheduled delivery times, the two dominate the package delivery industry. At the same time, email and e-cards have made the USPS less relevant (though they continue to thrive on business mailing). As a result, the USPS has ramped up its service to businesses, their most valuable service being Priority Mail Flat Rate Box shipping, which allows businesses with small but heavy items to ship their products for much less than the rates of UPS and FedEx. To raise awareness, the USPS developed promotions to advertise its capabilities.

To assess the possible benefit of a promotion, USPS pulled shipping records for a random sample of 50 offices that received the promotion and another random sample of 75 that did not. The offices were not randomly assigned to the two groups; rather, they sampled independently from two populations. The data can be found in the Excel file Promotion.xlsx.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data **USPS offices that were not aware of the promotion**
   
   (a) Construct a frequency distribution and a histogram in Excel. TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$), median in Excel.

   TA initial:

2. Compare the statistics you calculated in 1. (c) above for data **USPS offices that were not aware of the promotion** with those for data **USPS offices that were aware of the promotion**. Describe your findings.
Promotion Response  Thanks to cell phones, the way we communicate with those who live far from us has forever changed. For families with relatives across the ocean, special plans designed to cater to their needs have emerged. For example, Verizon currently offers three international calling plans: the International Value plan, the Call the World Share Plan, and Standard Pay per Minute.  

Verizon launched an advertising program designed to increase the number of minutes of international (long-distance) calls made by its most lucrative plan, the Call the World Share Plan. To get a sense of the benefits of the program, it ran a small test of the promotion. It first selected a sample of 100 customers of the type being targeted by the promotion, those already paying for the Call the World Share Plan. This sample of 100 customers used an average of 185 minutes per month of international long-distance service, at $0.10/minute. The company then included a special flyer in its monthly statement to those customers for the next two billing cycles. After receiving the promotion, these same customers were using 215 minutes per month on average. Did the promotion work?

Let $X_1$ denote the number of minutes used by a customer before the promotion, and let $X_2$ denote the number of minutes after the promotion. Use $\mu_1$ for the mean of $X_1$ (before the promotion) and $\mu_2$ for the mean of $X_2$ (after).

Data is located in the file Promotion_resp.xlsx.

Chapter 1 Questions

1. Get to know your data
   
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   
   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data **Number of minutes used after the promotion**
   
   (a) Construct a frequency distribution and a histogram in Excel. TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}_2$), sample variance ($s_2^2$) and sample standard deviation ($s_2$), median in Excel.

   TA initial:

2. Compare the statistics you calculated in 1. (c) above for data **Number of minutes used after the promotion** with those for data **Number of minutes used before the promotion**. Describe your findings.
Property Taxes  Commercial real estate leases vary, but are based on two basic rent calculation methods: Gross and Net. Gross (or Full Service) leases are all-inclusive, whereas Net leases charge less for the space but don’t pay much or any of the extra costs, such as property taxes, insurance, and maintenance. Net leases are labeled as Single (N), Double (NN), or Triple (NNN) Net. The most common type of lease overall is the NNN lease, wherein the tenant pays a pro-rata (based on their use of space in the building) share of all three “nets”: property taxes, insurance, and common area maintenance charges (CAMS). Because the lessor doesn’t pay the property taxes, raising these taxes is met with far less resistance than raising taxes on residential real estate.

A Midwestern city with a rapidly growing business sector faces an urgent need for more downtown parking. The city’s Planning Commission is recommending the construction of a parking garage. Since the average parking garage costs $8.56 million to build, the mayor is concerned about how the city will raise the money to pay for it. He remembers from his economics class in business school that it’s best to cover the indirect costs of negative externalities (such as a lack of parking) with taxes on those causing the negative externality.

Because the businesses would benefit from the new parking garage, the mayor is considering a tax on commercial real estate property in the city, which would be paid primarily by the businesses leasing property. How much revenue would a 1% tax generate?

The city has 4,500 businesses that lease properties, nearly all as NNN leases; The data are the costs of a random sample of 223 recent leases. Data is located in the file Property_tax.xlsx.

Chapter 1 Questions

1. Get to know your data
   
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Total Lease Cost

   (a) Construct a frequency distribution and a histogram in Excel.

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean (\( \bar{x} \)), sample variance (\( s^2 \)) and sample standard deviation (\( s \)) in Excel.
San Francisco has had a rent crisis for a long time. The problem is not easily solved. On the one hand, affordable housing is nearly impossible to find in the city. This has pushed many families out of the area, with concerns of affordability, space, safety, and education. At the same time, San Francisco is building housing units per capita at a higher rate than any other large city, but housing tends to be purchased by the rich as “pieds a terre” (non-primary residences for those who want a place closer to downtown and can afford them). There is growing resistance in suburban areas to the housing expansions, and some have “closed their doors” to the construction of any new residential units. The result is an urgent need for housing, coupled with a very high percentage of apartments sitting empty most of the time.

To combat the problem, city planners are working with a firm that manages rental properties. Its earnings are proportional to the rents of the properties it manages, and the city is willing to sell the firm a piece of public land in a very good location. In return, the firm will be required to subject their rentals to two important caveats: each unit must be the primary residence of the lessee, and each unit must be rented by a family with at least one child. To cover its costs, the firm needs rents to average more than $1,500 per month. The firm uses an existing dataset of San Francisco rents, but limits their sample to only primary residential units housing families of three or more. Are rents in San Francisco high enough to justify the firm accepting the city planner’s offer? The data can be found in the Excel file Rent_sf.xlsx.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data
   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population and sample?
   (b) What are the parameter and statistics?
   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Rent

   (a) Construct a frequency distribution and a histogram in Excel. TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel. TA initial:
The Great Recession, as it is now called, began in mid-2007, when BNP Paribas, one of the world’s largest banks (headquartered in Paris, France), refused to lend any more money to three hedge funds due to “a complete evaporation of liquidity.” In addition to hedge funds and other fund managers, banks typically lend each other money for maturities as short as overnight or as long as a week, via the Interbank Lending Market at the London Interbank Offered Rate (LIBOR). The LIBOR rate is set daily by the Intercontinental Exchange (ICE) for exchanges of various maturities for the U.S. dollar, Swiss franc, euro, Japanese yen, and the British pound (when one exists). Soon after BNP’s loan termination, banks large and small across the world were not willing to lend money to each other, because so many were heavily invested in the suddenly failing market. This credit crunch rippled through the entire economy, precipitating a global economic crisis that reached its peak in 2008. Ben Bernanke, Chairman of the Federal Reserve at the time, described September and October of 2008 as “the worst financial crisis in global history, including the Great Depression.”

The Great Recession affected everyone, not just those with market investments. Banks stopped lending money to businesses, and at the same time businesses suffered from falling sales as consumers found themselves losing their homes and jobs.

A researcher is interested in determining if there were signs in 2007 of the impending crisis. To investigate, he is looking at a fairly stable sector – women’s clothing. He gathered data that gives the sales volume (in dollars per square foot) for 37 retail outlets specializing in women’s clothing in 2006 and 2007. If sales went up from 2006 to 2007, it would indicate that the crisis was not yet fully understood by consumers and had not yet spread to everyday purchasing habits. If sales went down, it suggests that retailers and consumers were already feeling the effects of the recession. Did sales change by a statistically significant amount from 2006 to 2007? The data can be found in the Excel file Retail_sales.xlsx.

Chapter 1 Questions

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistics?
(c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?

Chapter 2-3 Questions

1. For data Sales of 2007
   (a) Construct a frequency distribution and a histogram in Excel.  
   TA initial:

   (b) Describe the shape of the distribution.

   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.  
   TA initial:

2. What is the sample mean of the Sales of 2006, is it different from the Sales of 2007?
The profit in merging two companies often lies in eliminating redundant staff. The merger of two pharmaceutical companies (call them A and B) allows senior management to eliminate one of their sales forces. Which one should the merged company retain? We can compare data on the performance of the two sales forces. Let’s compare the average level of sales obtained during a recent period by the two sales forces.

Rather than have two independent samples, we have paired samples. Both sales forces market similar products and were organized into 20 comparable geographical districts. For each district, the data give the average dollar sales per representative per day in that district. Because each district has its own mix of population, cities, and cultures, it makes the most sense to directly compare the sales forces in each district. Some districts have higher sales than others because of the makeup of the district. The data can be found in the Excel file `Sales_force.xlsx`.

Chapter 1 Questions

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
1. For data Sales Force A
   
   (a) Construct a frequency distribution and a histogram in Excel. 
       
   (b) Describe the shape of the distribution.
       
   (c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

2. What is the sample mean of Company B? Is it different from the sample mean of Company A? Which company has a stronger sales force?
Coke and Pepsi have both been around for over a century. From the beginning, Coke was always the more popular brand among soda drinkers. While Pepsi was heading straight for its second bankruptcy, Coke was being sold to investors for $25 million – in 1919!

Now both companies have diversified, buying up drink and even food brands to boost revenues. Coke owns Dasani, Odwalla, Minute Maid, Fanta, and more. Pepsi owns Tropicana, Fritos, Quaker, Lays, and more. The result is that, even though Coca Cola continues to dominate Pepsi in terms of cola market share, Pepsi brings in more revenue. Coca Cola also spends about twice as much on advertising. The advertising has created advertisement of its own – Coke has more than twice as many followers on Twitter and more than four times as many Facebook followers than Pepsi. When it comes right down to it, though, what matters is which taste people prefer. Coke has done a great job marketing itself, thanks to celebrity endorsements going all the way back to 1900 and campaigns like the polar bear. But with less diversification than Pepsi, it’s critical that Coke retain its supremacy in the Cola Wars.

Coke would like to perform a blind taste test of its cola vs. Pepsi. To conduct this test, 24 individuals were randomly selected, asked to taste one unmarked cola, and to rate it on a scale of 1-10. The individuals were then asked to do the same for the other unmarked cola. The results are shown in the Excel file Taste_test.xlsx.

**Chapter 1 Questions**

1. Get to know your data
   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question
   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Cola A

(a) Construct a frequency distribution and a histogram in Excel. 

(b) Describe the shape of the distribution.

(c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.

2. What is the sample mean of Cola 2? Is it different from the sample mean of Cola 1? Which Cola is more preferred?
Car dealers have always sold used cars as well as new. When selling used cars, they compete with millions of individual sellers who can easily beat their prices. And let’s face it – car dealers don’t have the best reputation for honesty. In fact, Gallup has a running poll on their website where visitors to the site can share their perception of “Honesty/Ethics in Professions.” Car salesmen are rated as highly ethical by only 8% of those surveyed, tying with telemarketers and members of Congress, and beating out lobbyists by 1% for the lowest spot.

To combat their image problems and create a competitive advantage over private sellers, in the 1990s luxury brands Lexus and Mercedes-Benz began selling what they called “certified pre-owned cars” (CPOs). Today, you can buy any car brand you want as a CPO, and independent certifiers like Carmark have built businesses on independently certifying cars for sale.

The year is 2007. A BMW dealer wants to price their CPOs just right. The certification is coming from the dealership (a “Dealer CPO”) rather than the manufacturer (a “Factory CPO”), so they can’t price them too high. They want to review the average selling price of other BMW dealers’ CPO vehicles, two to five years old, from around the country. These data indicate the prices of 155 used BMW cars. Some have four-wheel drive (the model identified by the xi type) and others two-wheel drive (the model denoted simply by the letter i). The data can be found in the Excel file Used_cars.xlsx.

Chapter 1 Questions

1. Get to know your data

   (a) Identify the types of your data: nominal data, ordinal data or quantitative data

   (b) Identify the types of your data: time series data or cross-sectional data

2. Get to know your scientific question

   (a) What are the population and sample?

   (b) What are the parameter and statistics?

   (c) What is the scientific question? Is this Descriptive Statistics or Inferential Statistics?
Chapter 2-3 Questions

1. For data Used Cars

(a) Construct a frequency distribution and a histogram of the **Price** column in Excel.

(b) Describe the shape of the distribution.

(c) Compute the sample mean ($\bar{x}$), sample variance ($s^2$) and sample standard deviation ($s$) in Excel.