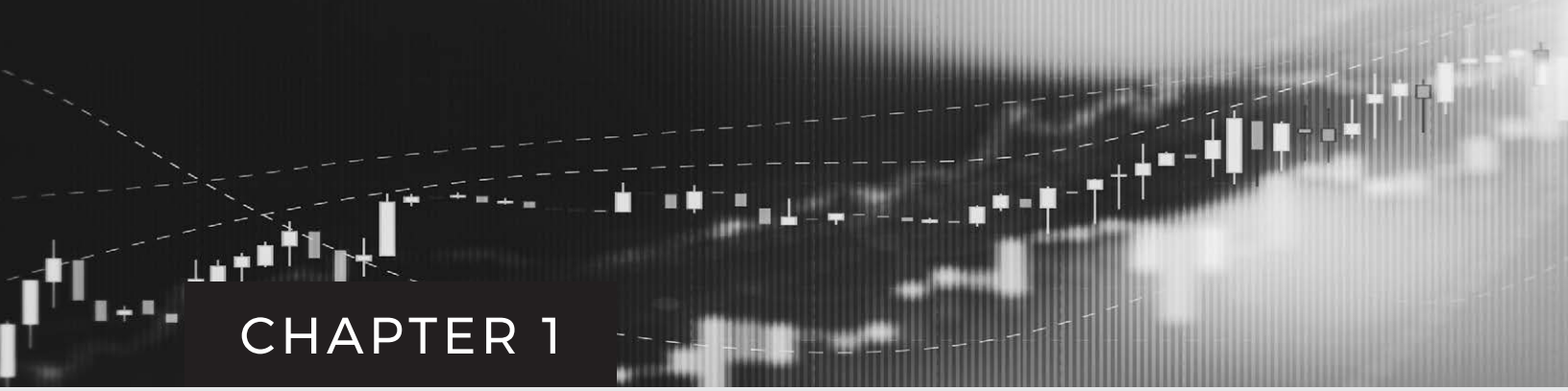

CHAPTER 1

Introduction to Biostatistics

KEY CONCEPTS

- Biostatistics is the branch of statistics concerned with the development and application of methods for collecting, analyzing, and interpreting biological data to assist decision-making.
- Problems in the biological sciences lead to research questions (representing uncertainty to be resolved), which may be addressed using biostatistics.
- Data are information taken from variables.
- A variable is a characteristic that varies from one subject to the next and can take on a specified set of values.
- Quantitative data are information about quantities or numbers.
- Qualitative data are information about the quality, nature, or essence of something.
- Nominal scale refers to data that fall into categories with no logical order or structure.
- Ordinal scale refers to data that fall into categories with an inherent order.
- Interval scale refers to measurements where the difference between the intervals is meaningful, but there is no true definition of zero.
- Ratio scale has the same properties as interval scale, but there is a true zero.
- The four broad areas of biostatistics are descriptive statistics, probability, inference, and statistical techniques.
- Descriptive biostatistics are measures of frequency (e.g., count, percent), central tendency (e.g., mean, median), dispersion (variation) (e.g., standard deviation, range, interquartile range), and position (e.g., percentile, rank) that represent either a sample or a population.
- Probability provides a basis for evaluating the reliability of the conclusions we reach and the inferences we make under uncertainty.
- Inferential statistics draw conclusions about a population based on sample information taken from that population.
- Statistical techniques are analytic approaches that utilize statistical methods to investigate a range of questions.
- Application of statistical software is essential as we address research questions in biostatistics.

This chapter provides an introduction to biostatistics. First, the meaning of the term biostatistics is discussed. Second, the research question and its importance for motivating the application of biostatistics is developed. Third, basic concepts related to data collection, analysis, presentation, and interpretation are covered. Fourth, the four broad areas of biostatistics, upon which this book is organized, are presented: descriptive statistics, probability, inference, and statistical techniques. Fifth, a brief introduction to statistical software and coding is given.

© MicroStockHub/iStock/Getty Images

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

1.1 What Is Biostatistics?

The word statistics is from the German word ***statistik***, which derives from the Latin word ***statisticum*** ("of the state") and the Italian word ***statista*** ("statesman," "politician"). In 1749, it was used to refer to the science dealing with data about the condition of a community or state.[1] Today we refer to statistics as the area of mathematics that involves the collection, analysis, presentation, and interpretation of data. Biostatistics is the science or practice of statistics applied to data that relate to living organisms, which primarily involves human biology, health, and medicine. In other words, it is a contraction of biology and statistics. Biostatistics contributes to a number of fields, including health, medicine, nutrition, genetics, biology, epidemiology, and many more. It is important in the health sciences for a number of reasons, including helping us identify the natural course of a disease; individuals at greatest risk for disease, injury, or death; risk factors for disease, injury, or death; and the value of new drugs, medical procedures, and healthcare interventions.

Biostatistics is the branch of statistics concerned with the development and application of methods for collecting, analyzing, and interpreting biological data to assist decision-making.

Biostatistics finds application in many areas, including clinical trials to evaluate the safety and efficacy of medications or medical devices. In clinical trials, biostatistics plays an important role at all levels: designing the study; determining the sample size; and collecting, analyzing, and interpreting the data. In public health programs, biostatistics is often used to evaluate a program's efficacy, effectiveness, and potential for replication in other areas. In literature reviews and meta-analyses, biostatistics provides the tools for quantifying patterns and trends, thus showing evidence for medical delivery and treatment practices. There are yet several other examples of application areas for biostatistics, many of which will be discussed in this book.

1.2 Questions for Biostatistics

When a problem is identified in the biological sciences, it leads to a research question or questions. A research question reflects the uncertainty that the investigator wishes to resolve by conducting a study. The application of methods in biostatistics allows us to effectively address questions arising in the biological sciences.

Consider the question, "Does physical activity increase your life expectancy?" A study showed that more leisure-time physical activity leads to longer life expectancy.[2] Physical activity equivalent to brisk walking for approximately 75 minutes per week resulted in 1.8 more years in life expectancy compared with no physical activity. Brisk walking for 450+ minutes per week resulted in 4.5 additional years in life expectancy compared with being inactive. Improvements in life expectancy with greater physical activity were found across weight classifications (normal, overweight, and obese). Individuals of normal weight with physical activity equivalent to 150+ minutes per week of brisk walking had 7.2 more years in life expectancy compared to those who were physically inactive and obese. Other questions about physical activity and

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

Table 1.1 Selected Questions Related to COVID-19

What is the frequency and pattern of this disease?
What are the clinical characteristics of this disease?
How do the symptoms vary by age group?
How does population density affect this disease?
How does comorbid health problems affect this disease?
How does nutrition and obesity affect this disease?
How does climate influence the virus?
Who is at the greatest risk?
What is the length of time after exposure before clinical symptoms appear (i.e., incubation period)?
When are infected people most contagious?
How is the disease spread?
How long is an asymptomatic carrier contagious?
What is the probability of reinfection of the disease following recovery?
What is the probability of false negatives and false positives for the test?
How effective is wearing a mask at preventing the spread of the disease?
What is the death-to-case ratio, and how does it vary by selected groups?
What are the long-term health consequences of the disease?
Is there an effective treatment for the disease?
Is there an effective vaccine for the disease?

health have also been addressed in recent years with the application of biostatistics.[3–7]

With the current COVID-19 pandemic, several research questions are being investigated, some of which are listed in **Table 1.1**. As answers are discovered, the disease can be better managed. For example, identifying the duration of time from exposure to first symptoms (incubation period) for COVID-19 tells us about the possible source of the virus, the appropriate length of time for quarantine, and the time a person can spread the virus prior to knowing they have it. Consider the question about the incubation period for COVID-19. One study found that the median incubation period is 5 days (range 1–14 days).[8] Incubation periods for several other infectious diseases are listed elsewhere.[9]

The Centers for Disease Control and Prevention has presented a general list of questions and answers related to COVID-19.[10] In the coming months, data will become increasingly more available to address such questions.

1.3 Data

The heart of biostatistics is data. Data are information and may be thought of as observations or measurements of something of interest.

Data are information obtained through observation, experiment, or measurement of a phenomenon of interest.

Data are either quantitative or qualitative.

Quantitative data are information about quantities or numbers, measured on a numerical scale.

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

Identifying additional years of life associated with increased exercise is an example of quantitative data.

Qualitative data are information about the quality, nature, or essence of something, with information gathered and ordered into larger themes as the researcher works from the specific to the general.

Describing how regular exercise makes people feel is an example of qualitative data.

Qualitative data are also called categorical data when the levels of the variable fit into categories. The levels of categorical data have no logical order. For example, males versus females or disagree, neutral, agree.

Data consist of measurements taken from variables. A variable is a characteristic that varies from one entity to another and can be measured or categorized. A variable can take on a specified set of values. A random variable is any outcome of a variable that occurs by chance, such as for every 1000 COVID-19 patients, how many are hospitalized.

Data measurements are made on different scales: nominal, ordinal, interval, and ratio, listed in **Table 1.2**. The statistical techniques presented and applied in this book depend on the type of data involved.

Interval and ratio data can be divided into one of two types: discrete or continuous. Discrete data have a finite number of measurements based on counts. Continuous data have a theoretically infinite number of measurements; there are no limits in the area between measurements. While discrete data represent things that are counted, continuous data represent things that are measured. Continuous data are more precise and provide more information than discrete data.

1.3.1 Data Collection

Data collection is the systematic approach in which information is gathered and measured in order to obtain an accurate picture of a phenomenon of interest. Data collection involves various methods, depending on whether quantitative or qualitative data are desired. Quantitative data collection methods consist of scales, tests, questionnaires, and surveys, whereas qualitative data collection

Table 1.2 Measurement Scales

Scale	Description	Example
Nominal	Refers to placing data into categories, where there is no logical order or structure	Exposed ("Yes" or "No") Diseased ("Yes" or "No") Sex, race, marital status, educational status
Ordinal	Refers to placing data into categories where the gross order of the categories is informative but the relative positional distances are not quantitatively meaningful	Preference rating (e.g., agree, neutral, disagree) Rank-order scale
Interval	Refers to a measurement where the difference between the intervals is meaningful, but there is no true definition of zero	Temperature, since zero on Fahrenheit or Celsius scales does not mean no temperature Calendar year
Ratio	Has the same properties as interval scale data, but has a clear definition of zero (e.g., height, weight, and blood pressure)	Height, weight, blood pressure Length, duration

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

methods consist of interviews, observations, and documents. In the physical activity and life expectancy study, leisure-time physical activity was associated with life expectancy according to six large cohort studies comprising 654,827 individuals, ages 21–90 years. Physical activity was measured as metabolic equivalent hours per week (Met-h/week). An approximate 10-year follow-up period occurred in the study with the main outcome, variable life expectancy, measured in years.

1.3.2 Data Analysis

Data analysis is the process of evaluating data to discover useful information. While quantitative data can be statistically analyzed, qualitative data are generally used to build concepts, hypotheses, or theories. In the physical activity and life expectancy study, the data were quantitatively assessed using a statistical technique called proportional hazards regression.

1.3.3 Data Presentation

Data presentation is the method in which we summarize, organize, and communicate information. These methods may be textual, tabular, or graphical. Textual methods include ranking data and the stem-and-leaf plot; tabular methods include the frequency distribution table and contingency table; and graphical methods include the bar chart, histogram, frequency polygon, and **pie chart**. In general, graphs and tables are effective ways to communicate data so that they can be easily understood.

1.3.4 Data Interpretation

Once data are collected, they are analyzed and presented in various forms, such as statistics, tables, graphs, diagrams, maps, and so on. At this point, the findings need to be interpreted. Data interpretation is the process of assessing and determining the meaning and importance of the results. Data interpretation involves consideration of questions about the data as they relate to your study questions. The answer to these questions should be organized as results and conclusions. Implications and recommendations may follow.

Data may be interpreted with perspective by comparing the results among groups or with what is typically known. In clinical trials, a control group is critical for assessing new drugs and medical procedures. The control group is also an effective way to control for confounding factors. Controlling for confounding allows us to have confidence that intervention effects are not explained by extrinsic factors. In the physical activity and life expectancy study, meaning and importance of leisure-time physical activity in terms of greater life expectancy was made by comparing life expectancy among people classified into groups according to physical activity and body weight.

1.4 Methods in Biostatistics

Biostatistics has four general method areas: descriptive, probability, inferential, and statistical techniques. They do not stand alone; for example, descriptive biostatistics often requires the use of probability and statistical techniques; inferential biostatistics relies on probability; and statistical techniques often involve inference.

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

1.4.1 Descriptive Biostatistics

Descriptive biostatistics involves methods of organizing, summarizing, and describing biological data. Suppose we want to describe our data so that we can answer who is at greatest risk for disease and where is the disease risk greatest. In recent months, several reports have answered questions about the frequency and pattern of COVID-19. As of July 27, 2020, the observed death-to-case ratio for COVID-19 was 15.2% in the United Kingdom, 14.3% in Italy, 13.9% in France, 10.4% in Spain, 7.2% in Sweden, 6.8% in Ireland, 5.7% in Switzerland, 3.9% in Poland, 3.5% in the United States, and 2.3% in India.[11]

Descriptive biostatistics are measures of frequency (e.g., count, percent, rate), central tendency (e.g., mean, median, mode), dispersion (variation) (e.g., standard deviation, range, interquartile range), and position (e.g., percentile, rank) representing either a sample or a population.

In a study involving 5700 hospitalized COVID-19 patients in New York, the median age was 63 years. The patients were more likely male (60.3% versus 39.7%) and had at least one comorbid condition (88% versus 6.3% with one and 6.1% with none). An estimated 56.6% had hypertension, 41.7% were obese, and 33.8% had diabetes.[12]

However, another study that controlled for underlying conditions and patient characteristics found that hypertension was no longer associated with hospitalization, indicating that underlying factors associated with hypertension (e.g., obesity and diabetes) were explaining the positive association.[13]

1.4.2 Probability

Probability provides a basis for evaluating the reliability of the conclusions we reach and the inferences we make under uncertainty. Biostatistics applies probability in several areas, such as in risk assessment, sampling and evaluating the reliability of an estimator, validity of a diagnostic test, and the chance of survival over a given time period.

Probability is a numerical description of the likelihood of an events occurrence.

A common question in biological health research is whether the probability of an event is related to the levels of a given variable, like exposure status, age, race and Hispanic origin, marital status, treatment, and so on. For example, in the United States, the probability of COVID-19 cases dying from the disease varies considerably by age (**Figure 1.1**).[14]

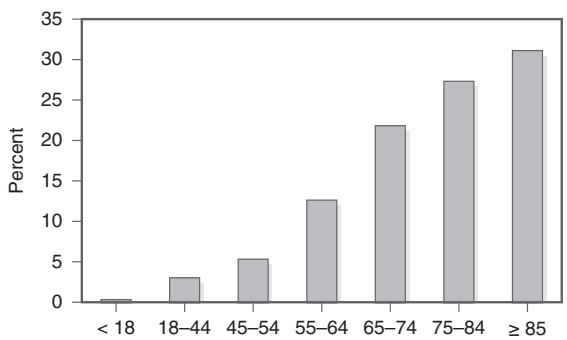

Figure 1.1 Mortality from COVID-19 in the United States by Age, February 12 through May 18, 2020

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

Among hospitalized cases in New York City, the probability of dying is greater among males than females in each age group (**Figure 1.2**).[12] The higher probability of death in males than females is more pronounced in the younger age groups. For example, the risk of death is 3.9 times greater in males compared with females in the age group 20–29 years and 1.4 times greater in the age group ≥ 90 years.

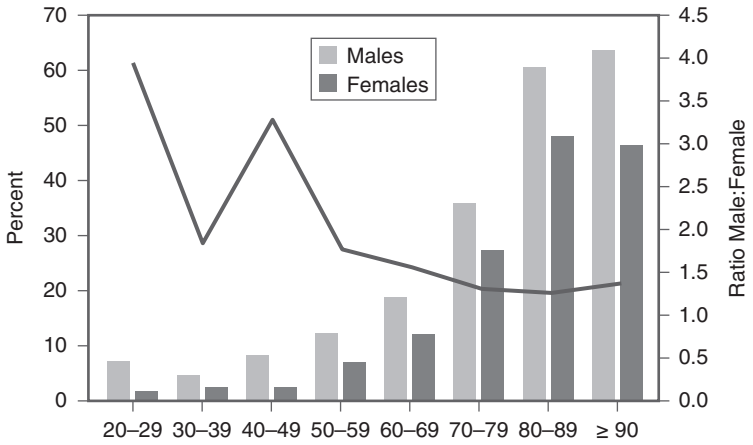

Figure 1.2 Deaths in COVID-19 Patients Hospitalized in the New York City Area by Sex and Age

1.4.3 Inferential Statistics

Inferential statistics involves a process of drawing conclusions about some aspect of a population, typically based on random sampling. For example, according to an observed subset of all customers who shop at a given store, females are more likely than males to wear a mask to prevent the spread of COVID-19.

Inferential statistics is the theory, methods, and practice of drawing conclusions about the parameters of a population based on sample information taken from the population.

The population is all customers.

A **population** is a set or collection of items of interest in a study.

The sample should be a sufficiently large, representative group from the population.

A **sample** is a subset of items that have been selected from the population.

Two methods are commonly used in inferential statistics: estimation and hypothesis testing. We may wish to infer something about a population, based on an estimated value from a sample. The estimated statistic is hopefully close to the true population parameter. Because our estimate contains some uncertainty, an accompanying statement of confidence (confidence interval) should be included.

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

A **statistic** is a measure from a sample that is used in statistics.

A **parameter** is a measure from a population that is often unknown. We are usually not able to calculate its value.

A **confidence interval** is a range of values wherein there is a **level of confidence** that the true parameter value lies within it.

Sampling is the process of selecting participants for the study when the number of people in the population of interest is larger than the required sample size. A number of sampling methods will be considered in this book. We will also address ways to minimize sampling bias where systematic error causes the sample of people in the study to not represent the target population.

An example of estimation, in which an estimate is derived from a sample and a statement of confidence is associated with the estimate, involves the death-to-case ratio for COVID-19. What is the death-to-case ratio for COVID-19? On March 5, 2020, researchers estimated the death-to-case ratio for China, including Hong Kong, Macau, and Taiwan, to be 3.48% (95% confidence interval [CI] 3.35–3.61%).[15]

In some cases, we may wish to infer that an estimated value is significantly different from another value. To assess this difference, we employ the process of hypothesis testing. For example, suppose you want to know whether the death-to-case ratio of COVID-19 is different between the United States and the United Kingdom. Your research hypothesis is that the ratios are different. On July 27, 2020, the number of confirmed cases and deaths in the United States was 4,288,012 and 148,009, respectively.[16] Corresponding values in the United Kingdom were 301,706 and 45,844.[16] The death-to-case ratio was 15.19% in the United Kingdom and 3.45% in the United States. The ratio of these two death-to-case estimates is 4.40 (95% CI = 4.36–4.45). In other words, the death-to-case ratio was 4.40 times greater in the United Kingdom than in the United States. Because the 95% confidence interval does not overlap 1, we conclude that this result is statistically significant at the 0.05 level.

1.4.4 Statistical Techniques

A variety of statistical techniques will be presented in this book. These statistical techniques are often applied to questions in the biological sciences. For example, rate ratios were used to address questions as to whether pregnancy associates with hospitalization and increased risk for intensive care unit admission, receipt of mechanical ventilation, and death among reproductive age women with COVID-19 infection.[17] Rate ratios are commonly used to compare the risk of an outcome in one group with another group. After adjusting for age, race and Hispanic origin, and underlying health conditions, pregnant women were 5.4 (95% CI = 5.1–5.6) times more likely to be hospitalized, 1.5 (95% CI = 1.2–1.8) times more likely to be admitted to the ICU, and 1.7 (95% CI = 1.2–2.4) times more likely to receive mechanical ventilation. Each of these estimates are statistically significant because the confidence intervals do not overlap 1. On the other hand, pregnant women were 0.9 (95% CI = 0.5–1.5) times as likely to die, which is not statistically significant.

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

Statistical techniques are analytic approaches that utilize statistical methods to investigate a range of questions.

News File

During March 27–28, 2020, three University of Texas students showed symptoms and tested positive for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Each of these students had gone to Cabo San Lucas, Mexico for spring break (March 14–19). Symptoms for the students did not manifest until March 22–25. Other student travelers were identified by contact tracing interviews and flight manifests. Of 183 students who traveled to Cabo San Lucas and who were evaluated for SARS-CoV-2, 60 (33%) tested positive. In addition, there was 1 of 13 (8%) household contacts of the travelers and 3 of 35 community contacts of the travelers who tested positive. Of those with positive tests, 20% were asymptomatic, and no one needed hospitalization or died. Testing of all those who were at risk allowed public health officials to identify those who could spread the virus, even asymptomatic people, in an effort to control further COVID-19 outbreaks.[18]

1.5 Statistical Software

Throughout this text, we will give examples of programs and output from computer packages designed to assess statistical data. Common statistical software used to analyze data include Statistical Analysis System (SAS), Statistical Package for the Social Sciences (SPSS), Stata, R, and Microsoft Excel. Although we will concentrate on SAS and Excel, the supplementary content for this class will be programs and output for SPSS, STATA, and R. In this section, we will provide introductions to both SAS and Excel.

SAS has three interactive windows: the Editor, Log, and Output. Each will be briefly described here. SAS programs involve SAS statements. Each SAS statement ends with a semicolon. The semicolon completes the SAS statement. SAS statements are used to create SAS data sets and to run predefined statistical or other routines. A group of SAS statements used to define and create your SAS data set is called a Data Step. A Data Step tells SAS programs about the data. For example,

```
DATA Example;
   INPUT Age Sex $;
DATALINES;
   17 Male
   18 Female
   20 Male
   24 Male
   24 Female
   30 Male
   32 Male
   ;
```

The name of the data set is arbitrary but must begin with a letter. For categorical variables, follow the variable name on the INPUT line with "$".

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

"DATALINES" comes prior to entering the data. Following the data set is a semicolon. It does not matter whether you choose to type words in uppercase or lowercase.

PROC (meaning procedure) begins several predefined routines. Each PROC is followed by a specific option, like PRINT, UNIVARIATE, MEAN, FREQ, REG, and so on. We end the procedure statements with "RUN". For example, to obtain the mean age for males and females separately, we first sort the data according to sex:

```
PROC SORT DATA=Example;
   Title 'Sorted Data by Sex';
   BY Sex;
RUN;
```

Then run PROC MEANS, with the variable that the means are to be computed for preceded by VAR and, because the means are to be computed for males and females, "BY Sex" is added. The MAXDEC option allows you to specify the number of decimal values.

```
PROC MEANS DATA=Example MAXDEC=2;
   TITLE 'Mean Age by Sex';
   VAR Age;
   BY Sex;
RUN;
```

SAS programs are entered into the SAS Editor and then run by clicking "Submit" on the toolbar. When the procedure is executed, it produces a LOG and OUTPUT. The LOG is an annotated copy of the original program, with the data excluded. Any procedure coding errors and information about the data set (e.g., number of observations and variables) will be identified there. The OUTPUT provides the results requested by the PROC statement, which for our example appear as follows:

The MEANS Procedure

Sex=Female

Analysis Variable : Age				
N	Mean	Std Dev	Minimum	Maximum
2	21.00	4.24	18.00	24.00

Sex=Male

Analysis Variable : Age				
N	Mean	Std Dev	Minimum	Maximum
5	24.60	6.39	17.00	32.00

Microsoft Excel features the ability to store, organize, and manipulate data (i.e., make calculations, graphs, and more). It is used to create spreadsheets, which are special documents that allow us to store and organize data in rows (horizontal sets of boxes labeled as 1, 2, 3, etc.) and columns (vertical sets of boxes labeled as A, B, C, etc.). The data can then be read and manipulated. The interaction of each row and column is a cell wherein we enter numbers, text, or formulas.

An Excel document is a workbook. A workbook consists of one or more worksheets. Worksheets are the grid where we store and calculate data. Simple

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

and complex formulas can be calculated in Excel. Excel also offers a variety of charts (e.g., line graph, bar chart, scatter plot) to assess data. To introduce you to Excel, consider the partial table shown here. The data are the same as used in the SAS example. To obtain the statistics in Column B, we entered in cell B8 =COUNT(B2:B6), in cell B9 =AVERAGE(B2:B6), in cell B10 =STDEV.S(B2:B6), in cell B11 =MIN(B2:B6), and in cell B12 =MAX(B2:B6). Statistics for the data in column C are computed similarly.

	A	B	C
1		Male	Female
2		17	18
3		20	24
4		24	
5		30	
6		32	
7			
8	Number	5	2
9	Mean	24.6	21
10	Standard Deviation	6.4	4.2
11	Minimum	17	18
12	Maximum	32	24

Excel also has an add-in feature for performing data analysis. To add this feature, go to the File tab and select Options. Then select Add-ins and choose Analysis ToolPak. In the Manage box, select Excel Add-ins and click Go. Check the Analysis ToolPak and then select OK. Now, in the Excel spreadsheet under the Data tab on the far right of the toolbar, the Data Analysis option will appear.

To use the Data Analysis option, choose the Descriptive Statistics option, identify the input range, check the summary statistics option, and select OK. This will return the summary statistics shown in the previous table, and more.

1.6 Summary

This chapter provides an introduction to the meaning of biostatistics and the role it plays in addressing research questions. In biostatistics, data are the numerical values from variables. A variable is an attribute which describes something and can vary from one entity to another. Variables are quantitative if the values are quantities or numbers, and qualitative if the information is about quality. Data are measured on different scales: nominal, ordinal, interval, and ratio. The scale of measurement determines how the values should be collected, analyzed, and presented.

Four broad areas of biostatistics are descriptive, probability, inference, and statistical techniques. Descriptive statistics are important because they present data in a more manageable, meaningful way that leads to simpler interpretation. Probability is important because it serves as a basis for evaluating the reliability of the conclusions we reach and the inferences we make under uncertainty. Inferential statistics is important because it allows us to draw conclusions about a population based on sample information. Statistical techniques are important because they allow us to explore and answer a range of research questions. Statistical software is an efficient tool for conducting thorough analysis of each of these areas of biostatistics.

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

Exercises

1. List and briefly describe the four broad areas of biostatistics.
2. Classify the following as (A) nominal data, (B) ordinal data, (C) discrete data, or (D) continuous data.

   ___ Integers of counts that differ by fixed amounts, with no intermediate values possible

   ___ Measurable quantities not restricted to taking on integer values

   ___ Ordered categories or classes

   ___ Unordered categories or classes
3. Classify the following as either (A) quantitative or (B) qualitative data.

   ___ Age

   ___ Hometown

   ___ Cholesterol level

   ___ Eye color

   ___ Number of siblings
4. Discuss whether discrete or continuous data provide more precision and information.
5. Descriptive biostatistics involves four types of statistics. List and define these.
6. Define probability and list some of its applications.
7. In what way does inferential statistics involve probability?
8. Classify the following as either (A) a statistic, (B) a parameter, or (C) a confidence interval.

   ___ Gives a range of values for an unknown parameter

   ___ Measure from a sample

   ___ Measure from a population

   ___ A numerical characteristic of a population

   ___ A number that represents a property of a sample
9. The length of stay for a sample of eight patients admitted to the hospital were recorded: 3, 16, 12, 3, 2, 5, 4, 1. Use a statistical software package to compute the number of patients, mean, standard deviation, minimum, and maximum. For the mean and standard deviation, express your answers to the nearest hundredth. For the others, do not round.
10. In this chapter we spoke briefly about confidence intervals. Much more will be said about confidence intervals later in this book. For now, try calculating a confidence interval for the population mean based on data in the previous exercise. Express your answer to the nearest hundredth. (Hint: In SAS use the procedure MEANS and the option CLM. In Excel, go into Data Analysis, choose Descriptive Statistics, identify the input range of your data and select summary statistics and confidence level for the mean. Use the estimated mean and confidence level in the output to derive the confidence interval.)

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

References

1. Van der Zande J. Statistik and history in the German enlightenment. *J Hist Ideas.* 2010;*71*(3):411–432.
2. Moore SC, Patel AV, Matthews CE, et al. Leisure time physical activity of moderate to vigorous intensity and mortality: A large pooled cohort analysis. *PloS Med* 2012;*9*(11):e1001335.
3. US Department of Health and Human Services. *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: CDC; 1996.
4. American Heart Association. Physical activity improves quality of life; 2018. http://www.heart.org/HEARTORG/HealthyLiving/PhysicalActivity/FitnessBasics/Physical-activity-improves-quality-of-life_UCM_307977_Article.jsp#.W2CQd9VKiUk. Accessed July 24, 2020.
5. Holme I, Anderssen SA. [Physical activity, smoking and mortality among men who participated in the Oslo studies of 1972 and 2000]. *Tidsskr Nor Laegeforen.* 2014;*134*(18):1743–1748.
6. US Department of Health and Human Services. *2008 Physical Activity Guidelines for Americans*. Washington, DC: US Department of Health and Human Services; 2008. https://health.gov/paguidelines/pdf/paguide.pdf. Accessed May 3, 2021.
7. US Department of Health and Human Services. *Physical Activity Guidelines Advisory Committee Report*, 2008. Washington, DC: US Department of Health and Human Services; 2008. https://health.gov/sites/default/files/2019-10/CommitteeReport_7.pdf instead? Accessed May 3, 2021.
8. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann Intern Med.* 2020;*172*(9):577–582.
9. Merrill, RM. *Introduction to Epidemiology,* Eighth Edition. Chapter *3*. Burlington, MA: Jones & Bartlett Learning; 2021.
10. Centers for Disease Control and Prevention. Clinical questions about COVID-19: Questions and answers. https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html. Accessed July 24, 2020.
11. Mortality analysis. Johns Hopkins University & Medicine. https://coronavirus.jhu.edu/data/mortality. Accessed July 27, 2020.
12. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA*. 2020;*323*(20):2052–2059.
13. Killerby ME, Link-Gelles R, Haight SC, et al. Characteristics associated with hospitalization among patients with COVID-19—metropolitan Atlanta, Georgia, March–April 2020. *MMWR.* 2020;*69*:790–794.
14. Wortham JM, Lee JT, Athomsons S, et al. Characteristics of persons who died with COVID-19—United States. February 12–May 19, 2020. *MMWR* 2020;*69*(28):923–929.
15. Wilson N, Kvalsvig A, Barnard LT, Baker MG. Case-fatality risk estimates for COVID-19 calculated using a late time for fatality. *Emerg Infect Dis.* 2020;*26*(6):1339–1441.
16. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University & Medicine. https://coronavirus.jhu.edu/map.html. Accessed July 27, 2020.
17. Ellington S, Strid P, Tong VT, et al. Characteristics of women of reproductive age with laboratory-confirmed SARS-CoV-2 infection by pregnancy status—United States, January 22–June 7, 2020. *MMWR.* 2020;*69*:769–775.
18. Lewis M, Sanchez R, Auerbach S, et al. COVID-19 outbreak among college students after a spring break trip to Mexico—Austin, Texas, March 26–April 5, 2020. *MMWR.* 2020;*69*(26):830–835.

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.

© Jones & Bartlett Learning LLC, an Ascend Learning Company. NOT FOR SALE OR DISTRIBUTION.